Optimization of Electricity Systems Under Uncertainty

By

Tomás Valencia Zuluaga

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Shmuel S Oren, Chair
Professor Ilan Adler
Professor Duncan Callaway

Summer 2024

Optimization of Electricity Systems Under Uncertainty

Abstract

Optimization of Electricity Systems Under Uncertainty

By

Tomás Valencia Zuluaga

Doctor of Philosophy in Engineering - Industrial Engineering and Operations Research

University of California, Berkeley

Professor Shmuel S Oren, Chair

In this dissertation, we study a market design problem and two investment planning problems: three different issues at the intersection of optimization under uncertainty and applications in electricity systems. Addressing the uncertainty associated with renewable energy and decentralized resources through computational optimization often leads to large-scale problems that are difficult to solve numerically. We propose models, solution methods and computational implementations in which we address this issue to approach some of the challenges faced by modern electricity systems.

We start with a market design problem in retail aggregation markets. Adopting a linear cooperative game model, we propose a mechanism to distribute collective savings in a community residential customers with local generation (so-called *prosumers*). We exploit the strong duality of Linear Programs to prove that stable imputations in a desired class, the uniform price core, are guaranteed to exist and can be found efficiently, overcoming scalability issues of the methods previously existing in the literature. We then show that looking for fair distributions in that class may come at the expense of computational tractability for large problem instances. To overcome this challenge, we propose a sampling methodology which we prove returns imputations in an acceptable relaxation of the uniform price core.

Next, we consider an investment planning problem from the perspective of a wind power producer: sizing a colocated battery to mitigate the revenue uncertainty caused by the intermittency in wind power production. We propose a Markovian model whose structure we exploit to develop a reduction technique that allows formulating an infinite horizon, average profit dynamic program as a low-dimensional numerical problem that can be solved with off-the-shelf software. The model obtained is a simple, stylized model that lends itself well for high-level sensitivity analyses of interest for project developers.

Finally, we adopt a central planner's perspective and consider a Capacity Expansion Planning (CEP) problem for climate resilience. In this part of our work, we address the need

for a software tool that can solve a two-stage stochastic CEP model on a realistic instance of the California power system, incorporating weather data from downscaled climate projection models. To successfully solve the resulting large-scale Mixed Integer Linear Program utilizing High Performance Computing resources, the problem is decomposed by scenarios by way of the Progressive Hedging Algorithm. We first show results of a successful software implementation on an instance with over 8,000 buses and 360 representative weather days, using a simplified network flow model instead of a more accurate representation of power flow. We then address this simplification by considering different models of the DC power flow formulation, including $b\theta$ and Power Transfer Distribution Factors formulations that incorporate transmission losses. We assess the impact of these improved models on time performance and solution quality.

Per a la Sara, el millor de tots els resultats obtinguts en aquest doctorat.

Contents

List of Figures

List of Tables

List of Acronyms

Acronyms used, indexed by chapter number.

BESS Battery Energy Storage System. 1–3

CARRM California Regionally Refined Model. 4

CATS California Test System. 4–6

CCA Community Choice Aggregation. 1, 2, 6

CDF Cumulative Distribution Function. 5

CEP Capacity Expansion Planning. 0, 1, 4–6

CTMC Continuous-Time Markov Chain. 3

E3SM Energy Exascale Earth System Model. 4

GCM Global Circulation Model. 4

GEP Generation Expansion Planning. 4

HPC High Performance Computing. 1, 4, 5

LLNL Lawrence Livermore National Laboratory. 1, 4–6

LMP Locational Marginal Price. 4

LP Linear Program. 0–2, 6

MILP Mixed Integer Linear Program. 1, 4, 5

MIP Mixed Integer Program. 1, 4–6

MMFQ Markov-Modulated Fluid Queue. 3

MPI Message Passing Interface. 4, 6

OPF Optimal Power Flow. 4, 5

PHA Progressive Hedging Algorithm. 1, 4–6

PTDF Power Transfer Distribution Factors. 0, 1, 5

PV Photovoltaic. 2

REC Renewable Energy Certificates. 4

RES Renewable Energy Sources. 3

RPS Renewable Portfolio Standards. 0, 4

TEP Transmission Expansion Planning. 1, 4, 5

WPP Wind Power Producer. 1, 3

Acknowledgments

I am extremely grateful to all those who supported me throughout this journey. While I can only name a few people here, I am grateful to everyone whose contribution made completing this PhD possible.

I want to first thank my advisor, Professor Shmuel Oren, for his unwavering support and encouragement throughout my PhD. Professor Oren has always shown interest not just in my academic progress, but in my overall growth and well-being, always eager to provide guidance and advice, usually about research and academics, but sometimes about arguably more important topics: what to include in a great trip to Alaska. There are occasions when research should take a backseat and our personal lives need to come first, in times of unforeseen difficulties, for example, but also to take time for celebrating. He has helped me see that life does not need to be put on pause to be a good researcher.

I am also immensely grateful to my collaborators at Lawrence Livermore National Lab, especially my mentor Amelia Musselman, whose guidance and constant encouragement have shaped my path as a researcher. I met Amy during during a remote internship in the Summer of 2020, in which we started a collaboration that brought me back for two more summers and an upcoming postdoc. During my second internship at LLNL, I met Jean-Paul Watson, who has been a great source of knowledge, ideas and support. JP and Amy have been key in making the last stretch of my PhD a delight. I am incredibly grateful to both, and am very excited to continue working with such a fantastic team. I also want to thank Georgios Patsakis, who first brought me to LLNL, and Ignacio Aravena, my first mentor at the lab. I learned a lot from Ignacio and hope to continue learning from him.

I would like to thank the members of my PhD dissertation committee, Duncan Callaway and Ilan Adler, for their time and helpful comments, which have contributed to improving this work. Professor Adler was also a great mentor to me as a Graduate Student Instructor; I truly enjoyed that academic experience and am very grateful for it.

The journey to complete this dissertation has been about more than just academics.

I came to Berkeley to get a degree, but had the incredible fortune of finding so much more than that: a friend, lover, and now, wife. Sara, you are the cornerstone of my support system and would definitely not be here without you. I am forever grateful for your love, support and sacrifice, and cannot wait to spend the rest of our lives together. *T'estimo!*

Having a longtime friend already in the Bay Area made moving here much easier. I want to thank Sebastián Rojas Neva for his unconditional friendship and being my home away from home when I first moved to Berkeley. I also want to thank Isaac Meza, from whom I learned so much during his time in Berkeley, and who quickly became my tutor, roommate, and in a word, friend.

Sometimes I needed to take a break from the PhD and recharge. My lifelong LAdC friends (David, Juanita, Andrés and Christian) were always there to make sure nothing, not even a PhD, is ever taken too seriously.

The pandemic made the start of my PhD a specially a challenging time. Nicolás Gamba and Liz Medrano helped keep me sane during the lockdown months. That time was also a lot more fun because of Lucía González and Elena Bonnelly, who introduced me to Sara. I am forever grateful to all.

Finding a Latin-American community made Berkeley feel like home. I want to thank Gabriel Matute, and the *Cafecito* group (especially Federico Mora, María Díaz de León and Carlos Ng) for welcoming me into that family.

Diego Rodríguez, at Gers, a friend and mentor more than a boss, and Sergio Rivera, my master's advisor at Universidad Nacional de Colombia, gave me the push I needed to pursue a doctorate. I am grateful to both for convincing me to take this leap.

I want to acknowledge the Colombian Fulbright Commission and the Colombian Ministry of Science and Technology for financially supporting me throughout my PhD. I would also like to acknowledge the Power Systems Engineering Research Center (PSERC) and Lawrence Livermore National Laboratory for funding different parts of this research.

I have become the researcher I am today because I was a curious child whose curiosity was always encouraged and nurtured. Every achievement I have obtained has only been possible because of the unconditional support and love I have received from parents my entire life. *¡Muchas gracias, los quiero muchísimo!*

Chapter 1

Introduction

1.1 Optimization in modern electricity systems

From charging a phone to launching a space rocket, electricity is essential in modern societies. Often called the largest machine in the world, the power grid is the backbone of our modern standard of life, as well as a key enabler of ongoing technological advancements. Although the early development of the power system occurred before the advent of computational optimization, the US power grid started taking its modern shape through the regional interconnections of the 1950s and 1960s, in part enabled by the simultaneous development of computational optimization.

Over the years, advancements in optimization tools and complexification of power systems have occurred through a symbiotic relationship. For example, the Stochastic Dual Dynamic Programming Method, a now widely used tool in multi-stage stochastic optimization, was first developed for scheduling hydro dam reservoir operation in Brazil's power system [83]. In the opposite direction, improvements in large-scale linear optimization software, and later on mixed integer linear optimization, have been fundamental for routinely solving the unit commitment and economic dispatching problems that modern electricity markets rely on for pricing and scheduling.

We could classify the applications for optimization in power systems into three categories of problems: investment (long-term) planning, operation (short-term) planning, and market design. These categories are not independent, nor mutually exclusive. Take, for instance, the project of installing a battery on-site at a wind power plant, the setting of the problem we address in Chapter 3. The investment planning question can be addressed by formulating an optimization problem to find the best size of battery to purchase to maximize profit. The optimal size depends of course on the revenue or savings the battery can generate during its lifetime, which depends on how the battery is operated. So in order to answer the question, a second optimization problem must be posed to determine the operation, which in turn depends on the rules of the markets to which the battery has access. To complete the circle, a market designer will try ensuring that the market rules provide incentives so

that investment and operation decisions are made in a socially optimal manner. In this dissertation, we concern ourselves with optimization problems of market design and long-term planning, but by necessity, also dip into modeling optimal operation.

The work presented in this dissertation lies at the intersection between electricity systems and optimization techniques. At the time of writing, as we approach the end of the first quarter of the 21st century, the two main challenges driving new development in that intersection are the increasing penetration of renewable, intermittent sources in the generation mix, and the new role played by decentralized resources, increasingly active in both generation and demand. In this thesis, we confront those challenges through a wide range of contributions, including modeling and theoretical results as well as software development and implementation of algorithms.

Sometimes, the opportunity for model innovation and theoretical contributions appears by bridging gaps between the problem and the state of the art in other fields of Operations Research (OR). For example, in Chapter 3, we propose a novel Markovian model for battery sizing in a wind power plant that can be reduced to a numerically optimizable low-dimensional problem by utilizing results and algorithms from the fluid queue literature. Similarly, in Chapter 2, we obtain a model that provably addresses the challenge of scalability in peer-to-peer electricity market design by identifying structural similarities with linear production games, an old and well studied subject in OR. A further challenge of scalability arises when we modify the mechanism to obtain fairer outcomes. Computational tractability can be recovered again by drawing from the statistical learning literature in a problem with different application but similar mathematical structure.

In other cases, innovation is driven by necessity. Recent events like the Texas winter storm in February 2021, or increasingly common prolonged heat domes in the western U.S. have made apparent the need for incorporating the effect of a changing climate into the planning process of power grid infrastructure planning. Climate projections of increasingly high resolution and long-ranging scope are becoming available, but incorporating them into the power system Capacity Expansion Planning (CEP) process in a computationally tractable manner is not a trivial task. In Chapters 4 and 5, as part of a project led by Lawrence Livermore National Laboratory (LLNL), we propose a parallel computing implementation of an existing scenario decomposition method for a CEP model using the Progressive Hedging Algorithm (PHA), to take advantage of increasingly powerful High Performance Computing (HPC) platforms.

We next briefly go over some background that we consider useful for the entire dissertation, before providing an overview of the outline of this document and the contributions of this thesis.

1.2 Theoretical background

Providing exhaustive background on the optimization techniques used in this thesis is out of the scope of this dissertation. In this section, we introduce some key concepts that are used

throughout the document, and provide useful references for further reading. As is common in the optimization literature, we adopt here the language of cost minimization, which is the setting of all chapters except Chapter 3.

Stochastic optimization

In Chapters 4 and 5, and Section 2.7 of Chapter 2, we model the problems of decision making under uncertainty as two-stage stochastic optimization problems. In this setting, a decision must be made in the present, called first stage, at a time when some of the future data, and hence the costs associated with that decision, are still uncertain, but have some degree of structure that is known. The second stage occurs after the uncertainty is revealed. Additional decisions, called second-stage decisions, can then be made. Second-stage decisions may encompass multiple time periods, but it is assumed that no more uncertainty persists at that point. In this thesis, we consider only risk-neutral decision makers, i.e. decision makers who in the face of uncertainty opt for minimizing the expected future cost[1].

Following the notation and exposition of [9], a generic linear two-stage stochastic optimization problem can be written as in (1.1). Uncertainty is modeled through random experiments with possible outcomes Ω. An individual outcome is denoted ω. In stochastic optimization, an outcome is more commonly called a *scenario*. x is the vector of first-stage variables, which must be decided upon before the uncertainty is cleared. y is the vector of second-stage variables, which can be chosen after the outcome is known. ξ is a random vector that encompasses all the scenario-dependent second-stage data. Note that the problem contains deterministic constraints on first-stage variables (1.1b) and scenario-dependent constraints coupling first-stage and second-stage variables (1.1c).

$$\min_{x} \left\{ c \cdot x + \mathbb{E}_{\xi} \left[\min_{y} q(\omega) \cdot y \right] \right\} \tag{1.1a}$$

$$\text{s.t.} \quad Ax = b \tag{1.1b}$$

$$T(\omega)x + Wy(\omega) = h(\omega) \tag{1.1c}$$

$$x \geq 0, y(\omega) \geq 0 \tag{1.1d}$$

In all the two-stage stochastic models we consider, we model random variables as discrete, which allows writing the two-stage linear stochastic optimization problem as in (1.2).

$$\min_{x,y} c \cdot x + \sum_{\omega \in \Omega} p(\omega) q(\omega) \cdot y(\omega) \tag{1.2a}$$

$$\text{s.t.} \quad Ax = b \tag{1.2b}$$

$$T(\omega)x + Wy(\omega) = h(\omega) \ \forall \omega \in \Omega \tag{1.2c}$$

$$x \geq 0, y(\omega) \geq 0 \ \forall \omega \in \Omega \tag{1.2d}$$

[1]Some of the models proposed in this thesis could be easily extended to handle risk-aversion. A useful textbook for this approach is [94].

Note that if Ω is finite, problem (1.2) can be fully written and is a conventional Linear Program (LP), which can be solved by passing it to a commercial LP solver. This approach is called solving the problem in *Extensive Form*.

If some of the first-stage variables are integer or binary, as is the case in Chapters 4 and 5, (1.2) is a Mixed Integer Linear Program (MILP). Note moreover that it has a block-angular structure, where the vector x contains all the variables that couple all constraints. If x is fixed, the resulting problem can be decomposed into a series of smaller, independent problems. Hence, problem (1.2) can be solved in Extensive Form by passing it directly to a MILP solver, or by exploiting the block-angular structure to perform some decomposition technique, which is the avenue pursued in Chapter 4. Decomposition techniques in stochastic and deterministic MILPs is a well-studied topic; for textbook references, see e.g. [9, 21].

Stochastic infinite-horizon dynamic programming

We find it useful to briefly highlight here the difference between the two-stage stochastic model presented above and the approach taken to model uncertainty in Chapter 3, an infinite-horizon, average cost problem with an initial decision, expressed in generic form in (1.3). We omit here details about formulating dynamic programs of this type. A useful textbook reference is [5].

$$\min_{\substack{x\in\mathcal{X}\\ \mu(\cdot)}} c\cdot x + \lim_{T\to\infty} \mathbb{E}_{\xi_t}\left[\frac{1}{T}\int_0^T f(y_t,\xi_t;x)dt \,\middle|\, y_0\right] \tag{1.3a}$$

$$\dot{y_t} = g\left(y_t, \mu(y_t), \xi_t; x\right) \tag{1.3b}$$

Here, x is a first-stage decision vector with known cost that will affect the dynamics of the system, as well as the costs incurred, $\mu(\cdot)$ is a stationary policy for determining the action taken at each time, and the stochastic process ξ_t affects both the dynamics of the system and the costs incurred. If the optimal policy $\mu(\cdot)$ can be expressed as a function of x, and ξ_t is an ergodic stochastic process taking values in a discrete state space, then the limit in (1.3a) can be expressed with help of a limiting distribution ψ over a discrete space $\mathcal{S}$ of states, obtaining (1.4).

$$\min_{x\in\mathcal{X}} c\cdot x + \sum_{s\in\mathcal{S}} \left(\psi_s(x)\cdot f(s;x)\right) \tag{1.4}$$

At first glance, the resemblance between (1.4) and (1.2a) might lead us to believe that (1.4) is also a two-stage stochastic program; we have indeed a first-stage variable with known cost and a second-stage probability-weighted sum over states that might be thought of as scenarios. Note, however that the probabilities in (1.2) are part of the problem data, while in (1.4) they are a function of the first-stage variable x. This fact, moreover, makes problem (1.4) a non-linear optimization problem, unlike the linear problems presented before. And

finally, the reduction of (1.3) to the two-stage-looking expression in (1.4) was a result of exploiting some of the assumptions in the model, but the uncertainty was not assumed to be completely revealed after one stage. On the contrary, the stochastic process ξ_t spans over all $t \geq 0$. So despite the similar final appearance, both the representation of uncertainty and the approach taken to optimize under it in Chapter 3 are quite different from that of other chapters.

Optimality gaps in MILPs

The stochastic optimization problems of Chapters 4 and 5 are in fact stochastic MILPs. The solution approach implemented in that work relies on commercial MILP solvers for the solution of the subproblems posed, which employ enhanced branch-and-bounding algorithms. A good textbook reference for Mixed Integer Program (MIP) and branch and bounding methods is [73]. We omit here a full description of branch and bounding, but do consider helpful introducing the concept of MIP gap, which plays an important role in assessing the performance of the methods proposed in Chapters 4 and 5.

$$
\begin{aligned}
(P): \quad & z = \min c \cdot x + d \cdot y && \text{(1.5a)}\\
\text{s.t.} \quad & Ax + My = b && \text{(1.5b)}\\
& x \geq 0, y \geq 0 && \text{(1.5c)}\\
& x \in \mathbb{Z}, y \in \mathbb{R} && \text{(1.5d)}
\end{aligned}
$$

$$
\begin{aligned}
(LP_j): \quad & z^{LP_j} = \min c \cdot x + d \cdot y && \text{(1.6a)}\\
\text{s.t.} \quad & Ax + My = b && \text{(1.6b)}\\
& g_i x + h_i y = e_i, \ \forall i \in \mathcal{C}_j && \text{(1.6c)}\\
& x \geq 0, y \geq 0 && \text{(1.6d)}\\
& x \in \mathbb{R}, y \in \mathbb{R} && \text{(1.6e)}
\end{aligned}
$$

Consider the generic MILP given in (1.5), and consider its linear relaxation (1.6), to which we have added a set of valid cuts $\mathcal{C}_j$, initially empty, i.e. $\mathcal{C}_0 = \emptyset$. We say the cut $g_i x + h_i y = e_i$ is valid if we can guarantee that it is satisfied by the optimal solutions (x^*, y^*) of (P). Thus, if all cuts in $\mathcal{C}_j$ are valid, (x^*, y^*) is guaranteed to be in the feasible set of (LP_j) and thus $z^{LP_j} \leq z$ for all j. Let $(\hat{x}, \hat{y})$ be an integer-feasible candidate solution, i.e. a point in the feasible set of (P). In particular, consider that we keep track of candidate feasible solutions found and let $(\hat{x}^j, \hat{y}^j)$ be the best feasible solution of (P) that the algorithm has found up to iteration j; let $\hat{z}^j$ be its objective value. Then, the following two-sided inequality holds:

$$z^{LP_j} \leq z \leq \hat{z}^j$$

The quantity $\frac{\hat{z}^j - z^{LP_j}}{|\hat{z}^j|}$ is called the MIP gap[2] obtained by the algorithm by iteration j. In the branch and bound method implemented by commercial solvers, a series of heuristic and cut generation routines are used to obtain a sequence of better integer-feasible solutions $(\hat{x}^j, \hat{y}^j)$ and tighter cuts $\mathcal{C}_j$ that lead to a smaller MIP gap. When the gap drops below a predefined threshold ε, the algorithm stops and the feasible solution $(\hat{x}^j, \hat{y}^j)$ is returned. Although branch and bounding algorithms are guaranteed to converge to 0 gap in a finite number of iterations, for large-scale problems this can take unreasonably long. This issue underscores the importance of achieving an acceptable gap soon enough in the execution of the algorithm, as discussed in Chapters 4 and 5.

1.3 Outline of this dissertation

In this dissertation, we study three independent problems that all fall in the intersection of optimization under uncertainty and at least one of the power system challenges mentioned in Section 1.1, i.e. problems of long-term planning and market design under a landscape of increased uncertainty driven by intermittent energy sources and decentralized participants. Chapter 2 is about addressing the challenge of scale posed by decentralization in a market design problem. In Chapter 3 we study an investment planning problem under wind and price uncertainty from the perspective of a private owner. In Chapters 4 and 5, we adopt the perspective of a central planner and do a computational implementation of an investment planning problem under climate and weather uncertainty. Chapters 3 and 2 are self-contained. Chapters 4 and 5 are conceived as a self-contained sequential unit.

Chapter 2

In this chapter, we consider a problem of stable and fair market design in Community Choice Aggregation (CCA) settings. We adopt an existing cooperative game model from the literature to distribute the collective savings resulting from aggregation of demand in a group of prosumers at the retail level, and make several major theoretical contributions to improve it and complement it. First, we characterize a set of allocations of interest, which we call *uniform price stable allocations*, and prove that they are guaranteed to exist. Moreover, we propose a computationally scalable methodology, based on solving a pair of primal, dual LPs, to find a uniform price stable allocation. Next, we illustrate that uniform price allocations may not be unique and may not be equally preferred by a social planner. We propose an optimization framework for selecting the most-preferred uniform price allocation. Finally, we show that although a computationally efficient and scalable methodolgy for exactly solving the optimization problem proposed is not available at the moment, satisfactory and computationally tractable solutions can be obtained via sampling by bridging existing results from the statistical learning literature. Most of the chapter is dedicated to studying the

[2] Note that a special case, omitted here, needs to be considered if the denominator is 0.

deterministic case, but a stochastic version that takes into account the uncertainty induced by renewable sources and demand is also discussed.

The results of this chapter are under consideration for publication at the time of publishing of this thesis. A part of the results of this chapter were presented in [106].

Chapter 3

In this chapter, we consider the case of a Wind Power Producer (WPP) who wants to install a Battery Energy Storage System (BESS) on site to mitigate the uncertainty in the production of electric power by firming its output and thereby increasing revenue. The WPP is interested in finding the optimal size to maximize their long-term average profit. We propose a modeling framework in which the wind power output of the plant and the market prices of eletricity are modeled as a joint Markovian stochastic process. The ergodicity of the process is exploited to reduce the expression of the infinite-horizon average profit to a low-dimensional function that can be optimized numerically. Special attention is paid to the case where market prices are constant, i.e. the WPP sells its power through forward contracts. A full solution of the model is derived in that case, and a numerical implementation is presented to illustrate how the model lends itself to performing sensitivity analyses.

The contributions of this chapter are: a stylized Markovian stochastic model for battery sizing of colocated storage in WPPs; a solution method for the aforementioned model in the case with constant prices via a bridge between a recent algorithm for numerically computing limiting distributions in multiregime fluid queues; and initial results in characterizing the optimal control policy for BESSs for the case with variable prices, potentially providing a path towards a usable implementation for that case as well. A substantial part of the results of this chapter were published in [107].

Chapters 4 and 5

In these chapters, we consider a long-term investment planning problem: developing parallel computing software for stochastic joint generation, transmission and storage CEP. The work presented in these chapters was performed in collaboration with LLNL, as part of a project aiming to bridge the gap between climate projection models and infrastructure planning.

In Chapter 4, we present the CEP that we adopt and extend, and the implementation of the PHA proposed to solve the large-scale MILP that results from it. Both the model and the solution method through scenario decomposition presented in this chapter had been proposed in the literature before. The contributions of our work are in the software implementation, which makes use of *Pyomo*, a Python modeling language for optimization, and *mpi-sppy*, a Python extension for solving stochastic optimization problems in HPC clusters. This implementation allows handling a stochastic nodal CEP larger than what had been solved in the literature before.

The model used in Chapter 4 uses a simplified network flow representation of power flow. In Chapter 5, we consider other linear representations of power flow and assess their

impact on the execution time and solution quality obtained. In particular, we test two common representations of DC power flow: the $b\theta$ formulation and the Power Transfer Distribution Factors (PTDF) formulation. The conventional versions of these representations neglect transmission losses; versions of the model that take these losses into account are proposed and some statistically justifiable heuristics are introduced to maintain the computational tractability of the model. The work in this chapter contains contributions on both modeling and software implementation. On the modeling part, the CEP model with PTDF formulation, which had been proposed in the literature for Transmission Expansion Planning (TEP), was extended to also include generation and storage. We also propose a methodology to include transmission losses under this formulation, which had not been proposed in the literature before. On the implementation part, a systematic test on 365 different loading and generation conditions, corresponding to a year of realistic data in a 500-bus testcase system, was performed to evaluate the time and quality performance of the different models, obtaining very promising results for the proposed PTDF model, which dominates the more common $b\theta$ model in the lossless case, and offers an interesting tradeoff in the lossy case.

The software developed as part of this work will be released as open software after the end of the ongoing project, which has not occurred at the time of submission of this dissertation. A substantial part of the results of Chapter 4 have been presented in [105] and will be published in [104].

Chapter 6

Finally, in Chapter 6 we discuss avenues for future research that spawn from the work performed during the completion of this dissertation.

Chapter 2

Stable and Fair Uniform Price Allocations of Community Choice Aggregation Gains in Retail Electricity Markets [1]

Abstract

The advent of differentiated prices for consuming (buying) and injecting (selling) electricity at the residential level promotes the emergence of local peer-to-peer electricity markets for prosumers, which can deliver savings to participants as long as an attractive cost sharing mechanism can be designed. Building on cooperative game theory models that have been proposed in the literature, we define the uniform price core, a class of desirable distribution of savings in this context, and prove constructively that it is not empty. We propose the shadow price imputation, a computationally efficient stable uniform price imputation, which we show to be equivalent to a dual imputation in the sense of cooperative linear production games. In the second part of the chapter, we compare the shadow price imputation to other imputations in the uniform price core through the lens of fairness. To overcome the challenge of tractability for larger numbers of participants, we extend an existing sampling methodology and apply it to optimization problems devised to obtain a fair imputation. The long-term incentive implications for different stakeholders, as well as a generalization to a stochastic case where the uncertainty in renewable production and local demand is considered, are also discussed. We present theoretical results and numerical experiments and examples to illustrate our approach.

[1] A substantial part of the text that constitutes this chapter was submitted for publication under the name "Uniform Price Allocations of Community Choice Aggregation Gains in Retail Electricity Markets", authored by Tomas Valencia Zuluaga and Shmuel S. Oren, and was under review at the time of publishing of this thesis. A conference paper version of the sampling methodology presented in Section 2.5 was presented and published in [106].

2.1 Introduction

After years of incentive measures like *Net Metering*, and even mandatory installation of rooftop Photovoltaic (PV) and local Battery Energy Storage Systems (BESSs) in new buildings in certain jurisdictions ([81]), the so-called prosumer, a consumer of electricity with local means of production, is expected to become nearly ubiquitous in modern distribution systems. With the advent of the end of net metering subsidy programs, the use of different retail prices for consuming (buying) and injecting (selling) electricity is likely to become more common. In this way, public utilities may reconcile their non-convex cost structure, driven by significant fixed costs, with a need to maintain linear (volumetric) prices, as shown by [63]. In this context, prosumers find that trading among each other can offer savings compared to the conventional scheme where all transactions are settled with the public utility, and so forming local markets, usually called peer-to-peer markets or local aggregation markets becomes naturally attractive. One of the main challenges when setting up such a market is designing a mechanism to distribute the collective savings in such a way that participants are incentivized to remain in the market.

In this chapter, we approach this problem through a cooperative game theoretic model, where we formally define some desired characteristics of a distribution mechanism: uniform prices, computational efficiency and a certain notion of fairness.

We propose an efficient mechanism that satisfies the two former desiderata, and analyze a framework to obtain more fair distributions in a computationally efficient way. We also investigate through numerical examples if more fair distributions exist as the number of participants grows. We close with a discussion of long-term incentive implications of the scheme proposed here.

2.1.1 Related research

There are numerous examples of cooperative game models applied to peer-to-peer electricity markets in the literature, which differ in the market structure (centralized or decentralized), the representation of the grid and flows (copper-plate, linear, non-linear), among numerous other model details. Recent, comprehensive reviews are available in [16] and [59].

Our approach is closest to the models of [38] and [54], which analyze the obtention of a profit distribution mechanism using a centralized, linear model with several time periods linked through storage devcies, like ours. The former authors find that the Shapley value is often not in the core of the game and can fail to incentivize participation. The nucleolus is proposed as a stable, albeit computationally challenging, alternative. The latter authors, whose formulation and notation we adopt in this chapter, propose overcoming the computational challenges of the nucleolus computation with a linear optimization problem to obtain the least core allocation, although the least core Linear Program (LP) proposed still requires an exponential number of constraints in its description. They also restrict allocations to the ones that can be described as uniform prices, but do not formalize it or analyze it like we do here.

In our model, the value of a coalition is obtained from solving an LP. In this chapter, we exploit LP strong duality to obtain an imputation that is guaranteed to be in the core of the game. This is the same approach followed for example by [19] in the context of inventory centralization problems. All these problems are special cases of the *linear production games* proposed by [78], who shows that such games are totally balanced and that the dual imputation, sometimes called in the literature *Owen imputation*, is guaranteed to be in the core. [78] also shows that although the core and the set of dual imputations are not the same for general linear games, they become the same in the limit as the number of players grows to infinity through replication.

Fairness in the context of cooperative games can have numerous interpretations and has been the subject of extensive study (for a textbook reference, see e.g. [69]). A classic notion of fairness in this context is that of Shapley, who proposes through an axiomatic approach that a fair allocation should reflect each individual's contribution to other potential allocations. It is in this light that we analyze unfairness in Example 2.4.1: essential participants may be allocated zero profits by a dual imputation, which are thus unfair in this regard. This phenomenon was identified by [82] for general linear production games, where they characterize a class of more fair allocations, but find that their existence in the core cannot be guaranteed.

This analysis of unfairness in our context is related to finding a uniform price that is intermediate between the utility selling and buying prices and thus distribute savings between buyers and sellers. This is addressed by [102] by proposing a *mid-market rule (MMR)*, which is somewhat similar to our price-control mechanism introduced in Section 2.4. However, their model is a single-period model without storage, and the MMR is not in the core of the game, as shown in the numerical tests of [54].

Our approach of finding the best core allocation via optimization can be interpreted as the social planner's problem of finding a social optimum, with our objective value being a *social welfare function*, in the sense of classic Welfare Economics as covered e.g. in [60]. We do not expand on this angle of analysis, but point the interested reader to [29] for a related approach of finding an *egalitarian allocation* in cooperative games.

Describing the core in general games requires a number of constraints exponential in the number of players. This is a known shortcoming of finding stable imputations in general settings. In the statistical learning literature, [3] propose two probabilistic relaxations of the core and show that imputations therein can be learned from a number of samples that is polynomial in the number of players. This is adapted by [106] to the same setting studied in our work to find the least core allocation in a relaxation of the core of the game for p2p markets with larger numbers of participants. In Section 2.5, we apply that approach to optimization problems with general objective functions.

2.1.2 Contributions of this work

The contributions of this work can be summarized as follows:

- In the context of P2P electricity markets, we formalize a class of imputations that is desirable in a cost sharing mechanism: imputations that can be described through uniform prices. Although numerous models in the literature use this class of imputations, we have not found examples where the difference between this class and the regular game core is formally analyzed.

- We prove that core of the game and the uniform price core are not the same, which justifies the formalization of the uniform price core. We prove that the uniform price core is not empty for the peer-to-peer electricity market.

- We give a methodology to obtain a uniform price imputation in the core of the game, which we call the shadow price imputation. This methodology only requires solving a primal, dual LP pair of moderate size, and is thus computationally efficient for p2p markets with larger numbers of participants. We show that this imputation is a dual, or Owen imputation, when our model is interpreted as a linear production game in the sense of [78].

- To our knowledge, the relation of this structure of p2p market with the aforementioned linear production games has not been established in the literature before. Besides its academic interest, this bridge is of value because it allows applying the asymptotic result that dual imputations are the unique core imputations for large enough markets, for which we find some evidence in numerical tests.

- We show that shadow price imputations may be unfair in a Shapley sense and propose a framework for obtaining better uniform price imputations through optimization. The proposed framework has scalability challenges. To overcome them, a sampling technique is proposed based on statistical learning results that are shown to be applicable to this setting.

2.1.3 Structure of this chapter

In Section 2.2, we present the setting of the peer-to-peer electricity market and describe in detail the cooperative game model adopted. We also present the preliminary results about cooperative game theory necessary for our approach. In Section 2.3, we introduce the class of uniform price imputations and show that the set of such imputations in the core is not empty through the introduction of shadow price imputations. In Section 2.4, we show how shadow price imputations, despite their desirable characteristics, might be unfair and propose an optimization framework to find the most desirable imputation. In Section 2.5, we present a sampling methodology to overcome the challenge of scalability in implementing that optimization framework in markets with larger numbers of prosumers, and present numerical tests that validate the methodology. In Section 2.6, we discuss some considerations about long-term responses of participants and the public utility which are not captured in our model. Section 2.7 discusses the generalization of this model to a stochastic

case that considers uncertainty in the local production and electricity demand. Finally, conclusions are presented in Section 2.8.

2.2 Setting and preliminaries

In this section, we present the setting for our peer-to-peer electricity market, and introduce the cooperative game chosen to model it. Other than some minor exposition and notation choices, the model is identical to that of [54].

After presenting the model, we define the class of *uniform price imputations* and the *uniform price core*, a related refinement of the core, and illustrate how they differ through a small example. Finally, we prove that the uniform price core is not empty and provide a methodology to efficiently find an imputation therein, which constitutes one of the main contributions of this work.

2.2.1 Nomenclature

Sets
N: Set of all market participants, i.e. prosumers. Indexed by i.
T: Set of time periods considered in horizon, indexed by t.
Variables
Primal
$s^c_{i,t}$: energy (in kWh) charged into i's BESS during period t.
$s^d_{i,t}$: energy (in kWh) drawn from i's BESS during period t.
$e_{i,t}$: energy (in kWh) in i's BESS at the end of period t.
$\ell_{i,t}$: net consumption of i during period t.
z_t: net collective consumption of all prosumers during period t.
w_t: net collective generation of all prosumers during period t.
Dual
π_t: dual variable of energy balance constraint at period t.
$\alpha_{i,t}$: dual variable of i's BESS charge limit constraint in period t.
$\beta_{i,t}$: dual variable of i's BESS discharge limit constraint in period t.
$\underline{\gamma}_{i,t}, \overline{\gamma}_{i,t}$: dual variable of i's BESS lower and upper energy storage limit constraint in period t.
δ_i: dual variable of i's BESS zero net usage constraint.
Parameters
p_t: utility price for consuming (buying) electricity during period t (in \$/kWh).
h_t: utility price for injecting (selling) electricity during period t (in \$/kWh).
$d_{i,t}$: Prosumer i's gross demand (in kWh) during period t.
$g_{i,t}$: Prosumer i's gross PV production (in kWh) during period t.
$\underline{s}_i, \overline{s}_i$: i's discharge and charge limit during one period.
$\underline{E}_i, \overline{E}_i$: i's lower and upper energy storage level limit.

η_i^c, η_i^d: prosumer i's BESS charge and discharge efficiency.
e_i^0: prosumer i's BESS initial state of charge.

Notation

Unless otherwise specified, the vector is noted by omitting the corresponding index, e.g. $\pi = [\pi_t]_{t \in T}$.

2.2.2 Model description

We consider a set of prosumers N who have some electricity consumption needs, and some of which have local generation (solar PV) and/or a BESS. The utility company has a pair of predetermined import and export prices p_t, h_t for each period t in the day, with $p_t > h_t$. During period t, any customer may purchase any amount of electricity from the utility company at a price p_t \$/kWh, and sell any amount of electricity for a price h_t \$/kWh. A subset $S \subseteq N$ of prosumers may get together and form a coalition; a coalition is treated as a single customer by the utility company, i.e. the coalition is charged for the combined net consumption of all members, and remunerated for the combined net injection.

Members of a coalition can coordinate the usage of their BESSs to shift their consumption in order to minimize the collective electricity bill of the coalition. We assume that utility prices, as well as the generation and demand profile for each customers is known for the entire horizon (day) when the coordination plan is determined. The coordination problem for a coalition $S \subseteq N$ is denoted (P_S) and given by (2.1).

$$
\begin{aligned}
&(P_S) : C(S) = \\
&\min_{\substack{s^c, s^d \\ z, w}} \sum_{t \in T} p_t z_t - h_t w_t && && (2.1a)\\
\text{s.t } & z_t - w_t = \sum_{i \in S} \ell_{i,t} && \forall t \in T && [\pi_t] && (2.1b)\\
& s^c_{i,t} \leq \overline{s_i} && \forall i \in S, t \in T && [\alpha_{i,t}] && (2.1c)\\
& s^d_{i,t} \leq \overline{s_i} && \forall i \in S, t \in T && [\beta_{i,t}] && (2.1d)\\
& \underline{E_i} \leq e^0_i + e_{i,t} \leq \overline{E_i} && \forall i \in S, t \in T && [\underline{\gamma_{i,t}}, \overline{\gamma_{i,t}}] && (2.1e)\\
& \sum_{t \in T} \left(\eta^c_i s^c_{i,t} - s^d_{i,t} \right) = 0 && \forall i \in S && [\delta_i] && (2.1f)\\
& \ell_{i,t} = d_{i,t} + s^c_{i,t} - \eta^d_i s^d_{i,t} - g_{i,t} && \forall i \in S, t \in T && && \\
& e_{i,t} = \sum_{\tau=0}^{t} \left(\eta^c_i s^c_{i,\tau} - s^d_{i,\tau} \right) && \forall i \in S, t \in T && && \\
& s^c_{i,t}, s^d_{i,t}, z_t, w_t \geq 0 && && && (2.1g)
\end{aligned}
$$

Note $\ell_{i,t}$ and $e_{i,t}$ are just shorthands for customer i's net consumption and BESS level during period t respectively, and can be substituted directly into (2.1b) and (2.1e), so they are not assigned a dual variable. Constraints (2.1c) and (2.1d) enforce the charging and discharging limits of the batteries, while (2.1e) ensures that battery levels remain within limits. To avoid end-of-horizon effects, (2.1f) makes sure that the battery ends the day with the same level as it started. Constraint (2.1b) computes the combined net load of the coalition during each time period t and assigns its positive and negative parts to auxiliary variables z_t and w_t, which are used in (2.1a) to compute the total cost (revenue) of energy purchased from (sold to) the utility. It is not hard to verify that because $h_t < p_t$, any optimal solution will satisfy $z_t \cdot w_t = 0$, which justifies the interpretation $z_t = \left(\sum_{i \in S} \ell_{i,t} \right)^+$, $w_t = \left(\sum_{i \in S} \ell_{i,t} \right)^- \ \forall\, t \in T$.

We consider a cooperative game model G defined by the set of prosumers N and the characteristic function $V : S \in 2^N \to \mathbb{R}^+$ given by (2.2).

$$V(S) = \sum_{i \in S} C(\{i\}) - C(S) \tag{2.2}$$

The characteristic function V can be interpreted as the surplus that can be achieved by a coalition comparing their collective bill to the situation where each customer deals with the utility individually. We next show that getting together is attractive for prosumers.

2.2.3 Other preliminaries

Next, we go over some preliminaries of cooperative game theory and results from related research that is relevant to our approach.

Definition 2.2.1. We say a function $f : S \subseteq N \to \mathbb{R}$ is superadditive if for all disjoint sets $S, T \subseteq N$, we have $f(S \cup T) \geq f(S) + f(T)$.

Theorem 2.2.1 (Theorem 1 of [54])**.** *The value function V is superadditive.*

Proof. To make this thesis self-contained and to account for differences in model description, a proof is provided in Section 2.A. □

Theorem 2.2.1 implies that for any partition $\mathcal{P}$ of the set N, $V(N) \geq \sum_{S \in \mathcal{P}} V(S)$, and thus, it is socially optimal for all customers to join the same coalition, which is called the *grand coalition*. The next natural question is if all participants can be incentivized to do the socially optimal action of joining the grand coalition.

Definition 2.2.2. An *allocation* or *imputation* is a vector $x \in \mathbb{R}^{|N|+} : \sum_{i \in N} x_i = V(N)$.

The value of the grand coalition $V(N)$ is the total surplus that the entire set of prosumers can achieve by coordinating. An allocation is one possible distribution of this surplus among the participants.

Definition 2.2.3. We say that a coalition S *blocks* or *is blocking* for allocation x if $\sum_{i \in S} x_i < V(S)$

If a coalition S is blocking for allocation x, the prosumers in S could abandon the grand coalition, coordinate only among themselves and find an alternative allocation in which they would all receive no less surplus than they are receiving under allocation x, with some receiving strictly more. Therefore, the existence of blocking coalitions provokes the collapse of the grand coalition. This introduces the concept of stability: an imputation x is said to be *stable* if there exists no coalition that is blocking for x. The set of all stable coalitions is called the *core* of game G. We denote the core of G with $\mathcal{C}$.

The *Shapley allocation* is an important concept from the cooperative game theory literature; it represents the allocation that best remunerates each participant's contribution to the group, and is the most fair allocation in that sense. It is unique, denoted by ϕ and can be computed from (2.3). Note that the expression in (2.3) requires enumerating all combinations of coalitions in N, which becomes computationally intractable for larger games. Moreover, the Shapley imputation is only guaranteed to be in the core of the game for convex games.

$$\phi_i = \frac{1}{N} \sum_{S \subseteq N \setminus \{i\}} \frac{(V(S \cup \{i\}) - V(S))}{\binom{n-1}{n-|S|-1}}, \; i \in N \tag{2.3}$$

Theorem 2 of [54] shows that this game is balanced, and thus has a nonempty core. However, the game is not convex, so the Shapley imputation is not guaranteed to be stable. [38] find in their numerical tests several such instances. We provide a small example of an unstable Shapley imputation in an instance of our game in Example 2.4.2.

We are not interested in just any stable imputation; in the next section, we define a particular class of imputations on which we will focus for the remainder of the chapter.

2.3 Uniform price imputations

In this section, we present the first of our main contributions. We formalize the set of imputations that we are interested in obtaining, prove that this set is not empty, and provide an efficient method to get an imputation therein.

First, it will be useful in our context to interpret allocations not just as distributions of surplus, but also as proposals for splitting the collective bill, so we introduce one more definition.

Definition 2.3.1. A bill splitting is a vector $b \in \mathbb{R}^{|N|} : \sum_{i \in N} b_i = C(N)$ with $b_i \leq C(\{i\})$ for all $i \in N$.

We interpret b_i as the amount charged to customer i's by the local electricity market, i.e. their local electricity bill. If negative, $|b_i|$ is the amount paid to customer i. It is easy to verify that there is a 1:1 correspondence between bill splittings and imputations given by $b_i = C(\{i\}) - x_i$ for each $i \in N$. Similarly, a coalition S is blocking for x if and only if $\sum_{i \in S} b_i > C(S)$. For a given imputation, we refer to its corresponding bill splitting as its *bill-splitting form*. We denote the core in bill-splitting form $\mathcal{B}$. The core can be expressed in regular and bill-splitting form respectively as:

$$\mathcal{C} = \left\{ x \in \mathbb{R}^{|N|+} : \sum_{i \in N} x_i = V(N), \sum_{i \in S} x_i \geq V(S) \ \forall S \subseteq N \right\} \tag{2.4}$$

$$\mathcal{B} = \left\{ b \in \mathbb{R}^{|N|} : \sum_{i \in N} b_i = C(N), \sum_{i \in S} b_i \leq C(S) \ \forall S \subseteq N \right\}. \tag{2.5}$$

2.3.1 The uniform price core

To facilitate the implementation of the cost sharing mechanism, we want imputations to be describable through uniform prices. In other words, we want to define local selling and buying prices of electricity, such that each participant's bill can be expressed as the result of a sequence of buying and selling transactions, performed at prices that are common to all market participants.

Table 2.1: Values of the collective bill for all coalitions in Example 2.3.1.

S	$C(S)$	S	$C(S)$
$\{A\}$	-2	$\{A,B\}$	-1
$\{B\}$	5	$\{A,C\}$	5
$\{C\}$	15	$\{B,C\}$	20
$\{A,B,C\}$	10		

Definition 2.3.2. A *uniform price imputation* or *uniform price allocation* in the context of game G is an imputation x (or b, in bill-splitting form) for which there exist prices $\lambda^b, \lambda^s \in \mathbb{R}^{|T|+}$ such that $h_t \leq \lambda^b_t, \lambda^s_t \leq p_t$ for each $t \in T$, and:

$$b_i = \sum_{t \in T} \left(\lambda^b_t \left(\ell^*_{i,t}\right)^+ - \lambda^s_t \left(\ell^*_{i,t}\right)^- \right) \ \forall i \in N \tag{2.6}$$

, where $\ell^*_{i,t}$ is the value of $\ell_{i,t}$ in an optimal solution of (P_N).

We denote the set of uniform price allocations with $\mathcal{U}$, and its bill-splitting form with $\mathcal{V}$. Its intersection with the core is called the *uniform price core*. We have found no examples in the P2P electricity market literature where the concepts of uniform price allocations and uniform price core are formalized as we have done here, although they are implicitly defined by [54]. We believe this formalization is not a moot point, because the *core* and the *uniform price core* are not the same, as shown in Example 2.3.1.

Example 2.3.1 ($\mathcal{U} \cap \mathcal{C} \neq \mathcal{C}$)**.** Consider a simple case with three participants $N = \{A, B, C\}$, and one time period. We drop the time subscript for this example. The utility prices are p=5\$/kWh, h=1\$/kWh. Participant A has a net generation of 2kWh. B has a net consumption of 1kWh and C a net consumption of 3kWh. There are no BESSs in this example. Table 2.1 gives the collective bill $C(S)$ for each coalition, i.e. the optimal value of (P_S) for each $S \subseteq N$.

Substituting the values in Table 2.1 into (2.5) and working through some algebra we get the characterization of the core in bill-splitting form:

$$\mathcal{B} = \{(b_A, b_B, b_C) = (\sigma, 5, 5 - \sigma), \sigma \in [-10, -6]\}.$$

Now let us characterize the uniform price core for this example. With λ^s the local price for selling and λ^b the local price for buying, uniform price imputations in bill-splitting form are

$$\begin{aligned} \mathcal{V} = \big\{ & (b_A, b_B, b_C) = (-2\lambda^s, \lambda^b, 3\lambda^b), \\ & 2\lambda^b - \lambda^s = 5, 1 \leq \lambda^s, \lambda^b \leq 5 \big\}. \end{aligned}$$

, which implies $b_C = 3b_B$, so that the only uniform price allocation in the core is obtained making $\sigma = -10$, and is $\mathcal{B} \cap \mathcal{V} = \{(b_A, b_B, b_C) = (-10, 5, 15)\}$. In other words, allocations

in bill-splitting form $(b_A, b_B, b_C) = (\sigma, 5, 5-\sigma), \sigma \in (-10, -6]$ are all in the core, but cannot be expressed through uniform prices.

2.3.2 Shadow price imputations

We are now ready to introduce shadow price imputations. These imputations are related to (D_S), the dual of (P_S). In fact, we will show that shadow price imputations are *Owen imputations* for our linear game. A description of (D_S) is given in (2.7).

$$(D_S): W(S) =$$

$$\max_{\substack{\pi,\alpha,\beta \\ \underline{\gamma},\overline{\gamma},\delta}} \sum_{t\in T, i\in S} \pi_t \left(d_{i,t} - g_{i,t}\right) + \alpha_{i,t}\overline{s_i} + \beta_{i,t}\overline{s_i} + \underline{\gamma_{i,t}}\left(\underline{E_i} - e_i^0\right) + \overline{\gamma_{i,t}}\left(\overline{E_i} - e_i^0\right) \tag{2.7a}$$

$$\text{s.t. } -\pi_t + \alpha_{i,t} + \sum_{\tau=t}^{|T|-1} \eta_i^c \underline{\gamma_{i,\tau}} + \eta_i^c \overline{\gamma_{i,t}} + \eta_i^c \delta_i \le 0 \ \forall i \in S, t \in T \qquad [s_{i,t}^c] \tag{2.7b}$$

$$\eta_i^d \pi_t + \beta_{i,t} - \sum_{\tau=t}^{|T|-1} \left(\underline{\gamma_{i,t}} + \overline{\gamma_{i,t}}\right) - \delta_i \le 0 \ \forall i \in S, t \in T \qquad [s_{i,t}^d] \tag{2.7c}$$

$$\pi_t \le p_t \ \forall t \in T \qquad [z_t] \tag{2.7d}$$

$$-\pi_t \le -h_t \ \forall t \in T \qquad [w_t] \tag{2.7e}$$

$$\alpha_{i,t}, \beta_{i,t}, \overline{\gamma_{i,t}} \le 0, \underline{\gamma_{i,t}} \ge 0 \tag{2.7f}$$

The first main result of this work is given by the following theorem.

Theorem 2.3.1. *Let* $\left(z^*, w^*, s^{c*}, s^{d*}\right)$ *and* $(\pi^*, \alpha^*, \beta^*, \underline{\gamma}^*, \overline{\gamma}^*, \delta^*)$ *be an optimal solution pair to the primal-dual pair* $(P_N), (D_N)$*. Then the imputation given in bill-splitting form for each* $i \in N$ *by (2.8) is in the uniform price core of the game.*

$$\theta_i = \sum_{t\in T} \pi_t^* \ell_{i,t}^* \tag{2.8}$$

, where $\ell_{i,t}^*$*, as usual, is a shorthand for* $d_{i,t} - g_{i,t} + s_{i,t}^{c*} - \eta_i^d s_{i,t}^{d*}$*.*

The crux of our proof is in the following proposition.

Proposition 2.3.1. *Let* $(z^*, w^*, s^{c*}, s^{d*})$*,* $(\pi^*, \alpha^*, \beta^*, \underline{\gamma}^*, \overline{\gamma}^*, \delta^*)$ *be a pair of optimal primal-dual solutions to the primal-dual pair* $(P_N), (D_N)$*. Then, for each* $i \in N$*:*

$$\begin{aligned}\sum_{t\in T} \pi_t^* \ell_{i,t}^* = \sum_{t\in T} \big[&\pi_t^* \left(d_{i,t} - g_{i,t}\right) \\ &+ \alpha_{i,t}^* \overline{s_i} + \beta_{i,t}^* \overline{s_i} \\ &+ \underline{\gamma}_{i,t}^* \left(\underline{E_i} - e_i^0\right) + \overline{\gamma}_{i,t}^* \left(\overline{E_i} - e_i^0\right)\Big]\end{aligned} \tag{2.9}$$

Proof. See Appendix 2.A.2. □

Proof of Theorem 2.3.1. We first check that θ is indeed a uniform price allocation in bill-splitting form. θ satisfies the form of Definition 2.3.2 with $\lambda^b = \lambda^s = \pi^*$, with $h_t \leq \pi_t^* \leq p_t$ by (2.7d) and (2.7e), so we need only verify that it is an allocation, i.e. that $\sum_{i\in N} \theta_i = C(N)$. Applying Proposition 2.3.1 and summing over $i \in N$, we get

$$\begin{aligned}\sum_{i\in N} \theta_i &= \sum_{i\in N, t\in T} \pi_t^* \ell_{i,t}^* \\ &= W(N) \\ &= C(N)\end{aligned}$$

, where the second equality is true because summing over $i \in N$, the right-hand side of (2.9) becomes (2.7a), the optimal value of (D_N), and the last equality is true by strong duality of $(P_N), (D_N)$.

Next, we need to check that no coalition blocks allocation θ, i.e. we need to verify that $\sum_{i\in S} \theta_i \leq C(S)$ for all $S \subseteq N$. Consider any $S \subseteq N$. Then applying Proposition 2.3.1 and summing over $i \in S$, we have:

$$\sum_{i\in S} \theta_i = \sum_{i\in S, t\in T} \left[\pi_t^* \left(d_{i,t} - g_{i,t}\right) + \alpha_{i,t}^* \overline{s_i} + \beta_{i,t}^* \overline{s_i} + \underline{\gamma}_{i,t}^* \left(\underline{E_i} - e_i^0\right) + \overline{\gamma}_{i,t}^* \left(\overline{E_i} - e_i^0\right)\right]. \tag{2.10}$$

Observe that the right-hand side of (2.10) is the objective value of (D_S) corresponding to the solution $\left(\pi^*, \left(\alpha^*, \beta^*, \underline{\gamma}^*, \overline{\gamma}^*, \delta^*\right)_{i\in S}\right)$, which it is easy to check is feasible for (D_S). Then:

$$\sum_{i\in S} \theta_i \leq W(S) = C(S)$$

, where in the last equality we applied strong duality to the pair $(P_S), (D_S)$.

□

Remark 2.3.1. Proposition 2.3.1 in the proof of Theorem 2.3.1 shows that shadow price allocations are just a different way of expressing a dual (or Owen) imputation of game G. As mentioned before, dual imputations are shown by [78] to be in the core for games with a *linear production* structure. Therefore, a different way to prove this result would be to verify that game G has the structure of a linear production game, and prove that the shadow price imputation corresponds to the Owen imputation, by e.g. invoking Proposition 2.3.1. Since the latter already includes most of the work of the whole proof, we decided to make this proof self-contained.

Theorem 2.3.1 gives us a uniform price allocation that is guaranteed to be in the core. Moreover, it is a uniform price allocation that has equal selling and buying prices for each time period. This implies the two corollaries below.

Corollary 2.3.1 (Non-emptiness of the uniform price core)**.** *The uniform price core of game G is not empty.*

Corollary 2.3.2 (Non-emptiness of the uniform price core with identical prices)**.** *The set of allocations in the uniform price core of game G with equal selling and buying prices is not empty.*

To close this section, we discuss some important consequences of Theorem 2.3.1. First, in [54], optimization problems are formulated where the feasible set is restricted to the set of uniform price imputations. Because the existence of core imputations was proven, it is assumed by the authors that stable uniform price imputations also exist, but there was no guarantee that the least-core optimization problem is feasible. Corollary 2.3.1 guarantees the existence of such imputations, and thus the soundness of that method.

In fact, Theorem 2.3.1 does more than just that: it also makes the method somewhat obsolete. The methodology in [54] requires solving a linear problem with an exponential number of constraints. [106] propose a sampling methodology to address this issue by working with a relaxation of the core. Both return an imputation in the uniform price core (or a relaxation thereof), without any preference among different imputations in that set. The shadow price imputation is guaranteed to be in the uniform price core and only requires solving a pair of primal-dual linear problems of small size, so it outperforms these two methods.

Now that we have a guarantee that the uniform price core is not empty, it makes sense to ask ourselves which of the imputations therein is the best one. We delve into that topic, through the lens of fairness, in Section 2.4.

2.3.3 Asymptotic uniqueness of shadow price imputations in the core

[78] also shows that although the core and the set of dual solutions are not the same (cf. Example 2.3.1), they do asymptotically become the same as the number of players grows to infinity through replication. In particular, all core allocations of the limit game are dual allocations. For our application, this has the important implication that as the number of participants grows to infinity, shadow price imputations are the only core imputations. This suggests that for markets with very large numbers of participants, no core allocation other than the shadow price imputation may exist, so the search for more fair allocations could be a pointless endeavor, unless core membership is sacrificed.

Two important clarifications are necessary here: first, Owen's result only shows that in the limit, all core imputations are dual imputations. However, because optimal solutions to (D_N) may not be unique, it is not true in general that, in the limit, the core is a singleton set. Second, this is only an asymptotic result, so numeric tests are necessary to see if this behavior is perceivable for markets with a meaningful number of participants.

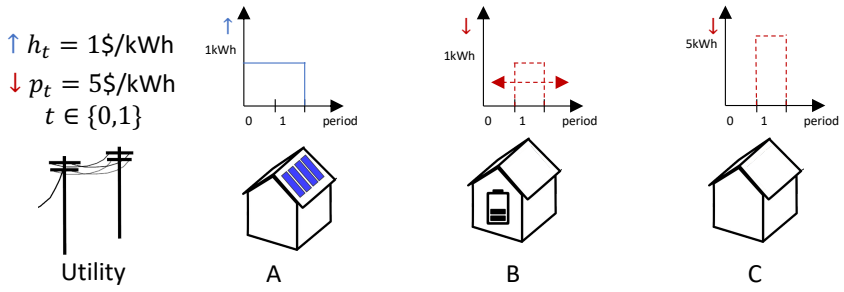

Figure 2.1: Schematic representation of Example 2.4.1. Case with three participants $N = \{A, B, C\}$, and two time periods $T = \{0, 1\}$. Participant A is a net producer of 1kWh each period; C is a net, inflexible consumer of 5kWh in the second period and 0kWh in the first period; and B has a lossless BESS, so B is a flexible consumer: B needs to consume 1kWh, but it can be in either period.

2.4 Fairness of stable uniform price allocations

2.4.1 Motivation through some examples

At the end of Section 2.3, we introduced the shadow price imputation, an imputation that satisfies two of our stated desiderata: it is in the uniform price core and can be computed efficiently. In this section, by taking the perspective of fairness, we analyze whether other uniform price imputations might be preferred over the shadow price imputation. We first motivate this section with a couple of examples.

First, we introduce some notation: we denote $\mathcal{U}^\dagger$ (and $\mathcal{V}^\dagger$ in bill-splitting form) a restriction of the set of uniform price allocations to allocations where the buying price and the selling price are equal to each other at each time period.

Example 2.4.1 (Unfairness of the shadow price imputation)**.** Consider a simple case with three participants $N = \{A, B, C\}$, and two time periods $T = \{0, 1\}$. Participant A is a net producer of 1kWh each period; C is a net, inflexible consumer of 5kWh in the second period and 0kWh in the first period; and B has an ideal BESS (no losses), so B is a flexible consumer: B needs to consume 1kWh, but it can be in either period. The situation of Example 2.4.1 is summarized in Figure 2.1.

It is not hard to verify that minimal collective cost is achieved by having B shift their consumption to the first period. Table 2.2 gives the collective bill $C(S)$ for each coalition.

We will first restrict ourselves to uniform price imputations with equal selling price and buying price, i.e. $\lambda_t^b = \lambda_t^s = \lambda_t, t = 0, 1$. In bill-splitting form, these imputations are $(b_A, b_B, b_C) = (-\lambda_0 - \lambda_1, \lambda_0, 5\lambda_1)$, with $b_A + b_B + b_C = C(N) = 20$. This implies $\lambda_1 = 5$, so that the set of uniform price imputations with same selling price and buying price is given, in bill-splitting and regular imputation form respectively, by:

Table 2.2: Values of the collective bill for all coalitions in Example 2.4.1.

S	$C(S)$	S	$C(S)$
$\{A\}$	-2	$\{A, B\}$	-1
$\{B\}$	5	$\{A, C\}$	19
$\{C\}$	25	$\{B, C\}$	30
$\{A, B, C\}$	20		

$$\mathcal{V}^\dagger = \{(b_A, b_B, b_C) = (-5 - \lambda_0, \lambda_0, 25)\,, \lambda_0 \geq 0\}$$
$$\mathcal{U}^\dagger = \{(x_A, x_B, x_C) = (3 + \lambda_0, 5 - \lambda_0, 0)\,, \lambda_0 \geq 0\}$$

Note that participant C is being allocated \$0 savings by all these imputations, which include the shadow price imputation. However, C makes a positive contribution to the grand coalition: $V(A, B, C) = \$8$, $V(A, B) = \$4$. In this case, the shadow price imputation would be unfair, as it is not remunerating participant C for their contribution. Such unfairness could fail to motivate C to join the local market, resulting in an overall loss for the whole community.

As discussed in section 2.1, the best allocation in the sense of remunerating average contribution is the Shapley value, which we can compute from 2.3 and the values in Table 2.2 as $(\phi_A, \phi_B, \phi_C) = (4, 2, 2)$. By relaxing the constraint of equal selling and buying prices $\lambda_t^b = \lambda_t^s$, the set of achievable imputations becomes, in bill-splitting form, $\mathcal{V} = \left\{(b_A, b_B, b_C) = (-\lambda_0^s - \lambda_1^s, \lambda_0^b, 5\lambda_1^b), -\lambda_0^s - \lambda_1^s + \lambda_0^b + 5\lambda_1^b = 20\right\}$. The Shapley allocation can thus be achieved with $\lambda_1^b = 3, \lambda_2^b = 4.6, \lambda_1^s = \lambda_2^s = 3$.

Example 2.4.1 gives a situation where the shadow price allocation is unfair, in the sense that it allocates zero savings to a participant that provides a positive contribution to the grand coalition. In this example, allowing selling and buying prices to be different allows the implementation of a preferable allocation, namely the Shapley imputation. However, the Shapley value is not necessarily a suitable target imputation, since it may not be in the core of game G, as shown in Example 2.4.2.

Example 2.4.2 ($\phi \notin \mathcal{C}, \phi \notin \mathcal{U}$)**.** Consider again the situation of Example 2.3.1. It was established there that the core imputations in bill-splitting form are given by

$$\mathcal{B} = \{(b_A, b_B, b_C) = (\sigma, 5, 5 - \sigma)\,, \sigma \in [-10, -6]\}\,.$$

We can compute the Shapley value in this example from 2.3 and Table 2.1 as $(\phi_A, \phi_B, \phi_C) = (14/3, 2/3, 8/3)$, which it is straightforward to check is not in $\mathcal{C}$ because it is blocked by $S = \{A, C\}$. Moreover, the set of uniform price allocations is

$$\mathcal{V} = \left\{(b_A, b_B, b_C) = \left(-2\lambda^s, \lambda^b, 3\lambda^b\right), -\lambda^s + 2\lambda^b = 5\right\},$$

which implies $b_C = 3b_B$, so the Shapley imputation is not implementable through uniform prices either.

[38] propose the *nucleolus* as a desirable alternative to the Shapley value, since it is guaranteed to be in the core of the game. As the Shapley value, it quickly becomes computationally intractable for larger games, so it is a challenging target. Moreover, in our case, it may not be describable through uniform prices, as shown in Example 2.4.3.

Example 2.4.3 (Nucleolus not in $\mathcal{U}$)**.** Consider again the situation of Example 2.3.1. Now, we target the nucleolus allocation, which can be computed to be $(x_A, x_B, x_C) = (6, 0, 2)$, or, in bill-splitting form, $(-8, 5, 13)$. Since the set of uniform price allocations has the form $(b_A, b_B, b_C) = \left(-2\lambda^s, \lambda^b, 3\lambda^b\right)$, the nucleolus is not implementable through uniform prices either.

Although admittedly toy-sized, the previous examples are aimed at illustrating the following three ideas: there are other stable uniform price allocations that could be preferred over the shadow price imputation; preferable mechanisms could be achieved by allowing buying and selling prices to be different (Example 2.4.1); and the Shapley value and the nucleolus, conventional cost-sharing mechanisms from the cooperative game theory literature, besides being computationally intractable, may be unsuitable target allocations in our case. These three observations motivate the framework we propose next.

2.4.2 Finding the most fair allocation through optimization

In Section 2.3.2, we showed that the uniform price core is not empty and that one imputation therein, namely the shadow price imputation, can be efficiently computed. However, this allocation is not necessarily the most desirable one in $\mathcal{C} \cap \mathcal{U}$, as illustrated in Example 2.4.1. Here, we propose explicitly finding a most-desirable imputation in the uniform price core via an optimization problem:

$$\min_{x \in \mathcal{C} \cap \mathcal{U}} f(x),$$

, where f is an objective function that represents a preference over imputations. We can interpret $f(x)$ as a metric of the unfairness (or, more generally, the undesirability) of imputation x. f can also be thought of as a *welfare function*, as discussed in Section 2.1. In fact, we allow the preference over uniform price imputations to also be affected by the associated prices λ^b, λ^s, so the optimization problem can be written, with imputations expressed in bill-splitting form, as

$$(A): \quad \min_{b \in \mathcal{B} \cap \mathcal{V}} f(b, \lambda^b, \lambda^s).$$

Mechanisms of this form admit a version where selling and buying prices are restricted to be the same, which we denote with the superscript †:

$$(A^\dagger): \quad \min_{b \in \mathcal{B} \cap \mathcal{V}^\dagger} f(b, \lambda).$$

We next illustrate our framework by introducing three possible options for the function f and discuss some of their characteristics.

Egalitarian mechanism In the egalitarian mechanism, the imputation sought is that which minimizes the dispersion in the allocations to participants. This is achieved by making the objective function in (A) equal to f_e defined below:

$$f_e(x) = \sum_{i \in N} \left| x_i - (1/N) \sum_{j \in N} x_j \right|,$$

i.e. the standard deviation of the allocations with respect to the L^1 norm.

Minmax mechanism In the *minmax* mechanism, we define f_m to measure the maximal difference in allocation of savings between prosumers, so that the mechanism finds the stable uniform price allocation that minimizes that difference.

$$f_m(x) = \max_{i \in N} x_i - \min_{i \in N} x_i.$$

Note that the ideal allocation with respect to both functions (i.e. the one with minimal value) would be one where all participants receive equal savings. So if the *equal division* allocation is feasible, both mechanisms are identical. However, the equal division allocation is not guaranteed to be in the uniform price core (cf. Example 2.3.1). If that is the case, these two mechanisms differ. An optimal allocation in the egalitarian mechanism is an allocation in $\mathcal{C} \cap \mathcal{U}$ that is closest to the equal division allocation, in L^1 distance (i.e. the L^1 projection onto $\mathcal{C} \cap \mathcal{U}$). In the minmax mechanism, on the other hand, only the imputations of the two most unequal participants matter.

These two definitions of fairness, although intuitive, have an important drawback in our context: users would have an incentive to misrepresent themselves as multiple agents and thereby multiply their allocations. For the scope of this thesis, we assume that it is verified that no agent misrepresents themselves.

Price-control mechanism The last mechanism we consider is specific to our context and is a *price-control* mechanism. A way of describing the unfairness in Example 2.4.1 is that no surplus is allocated to the inflexible, net consumer C and most surplus is allocated to the net producer A. This is likely to happen when electricity is priced at the marginal cost of procurement, as the shadow price allocation does. An intuitive compromise to make sure that both consumers and producers perceive benefits is to settle all local transactions at an intermediate price between the utility's export and import prices.

In this vein, one can define an ideal imputation as that induced by having the local selling and buying prices be equal to each other, and equal to the average between the utility's import and export prices, for each time period $\overline{p_t} = (1/2)(p_t + h_t)$. Since those prices are

not guaranteed to support a stable, uniform price allocation, one can use its projection onto the uniform price core. With an L^1 projection, this results in an objective function f_p as follows,

$$f_p(b, \lambda^b, \lambda^s) = \sum_{t \in T} \left| \lambda_t^b - \overline{p_t} \right| + \sum_{t \in T} \left| \lambda_t^s - \overline{p_t} \right| .$$

2.4.3 Preliminary discussion

The three mechanisms presented before are simple, intuitive mechanisms to obtain the most-preferred allocation among all stable uniform price allocations according to three different, but reasonable preferences. Here, we discuss some limitations of the objective functions chosen, and possible improvements that are out of the scope of this thesis, but could be achieved with the framework presented here. We discuss this at this stage to highlight the relevance and versatility of this framework.

Projected Shapley value In the examples presented as motivation for these mechanisms, we looked at fairness through a Shapley lens, i.e. the unfairness was understood as a failure of allocations to reflect the average marginal contribution of each participant to the grand coalition. However, the mechanisms proposed here do not compute or incorporate contributions when defining the allocation, and are instead based on simple, albeit reasonable, heuristics.

A better mechanism could be aimed at finding the uniform price stable allocation that is closest to the Shapley value, i.e. the projection of the Shapley allocation onto $\mathcal{C} \cap \mathcal{U}$. For smaller cases, the Shapley value is easily computable, so this is an easily implementable mechanism. However, for larger markets, this would become an intractable approach. Applying machine learning results to try to learn the projection of Shapley onto the uniform price core through sampling is an interesting avenue of research, but is considered out of the scope of this thesis.

Weighted egalitarian An immediate extension of the egalitarian mechanism presented here is a weighted version, with different weights given to each participant. Such weights could be updated with some frequency based on historical data and be intended to reflect the historical marginal contribution of each participant, in the vein of learning the Shapley projection from data.

Fairness could also be seen through a social lens. In this case, these weights could be fixed values dependent on some structural data, like household income or household size. So this framework is also compatible with a *social* interpretation of fairness.

L^p projections In all the mechanisms presented here, we use L^1 for all projections and measures of dispersion of allocations. We have done this for simplicity, to maintain the linearity of all the optimization problems of type (A). However, note that the the non-emptiness of $\mathcal{C} \cap \mathcal{U}$ is independent of the structure of function f, so non-linear objective

functions, including L^p projections with $p > 1$ could be used instead. Investigating the effect of such choices on the resulting imputations is an interesting problem, but is considered out of the scope of this thesis.

Adding a regularization term The uniform prices resulting from these mechanisms may oscillate frequently between the utility sell and buy prices. This could be considered undesirable, and could be corrected through the inclusion of a regularization term of the form $\rho(\lambda_t - \lambda_{t-1})^2$. As mentioned in the previous paragraph, non-linearities are not an issue for this framework, so regularization could be added to any of these mechanisms.

Restricting prices to be equal Since problems of form (A) are a relaxation of problems of form $(A^\dagger)$, their objective functions will be no worse than that of the restricted counterparts. Moreover, Example 2.4.1 shows a case where allowing different prices results in a more desirable imputation that was unreachable before. Thus, it seems that mechanisms with same prices are weakly dominated by mechanisms that allow different prices and there should be no reason to consider the restricted version. We give here two reasons why we consider them.

First, note that the weakly dominance is true *with respect to each problem's objective function*, but not across metrics. So it is possible that a mechanism with prices restricted to being equal outperforms others with respect to an external benchmark. Secondly, a mechanism with the same price for selling and buying could be considered easier to communicate to participants, so it could be preferred by market designers if its performance is comparable to that of the mechanism with different prices.

Scalability All of these mechanisms share one issue: the description of the uniform price core requires an exponential number of constraints. This is not a problem for markets of moderate size, like the ones used for the numerical tests that are presented next, but leads to an intractable mechanism as the number of participants increases. The issue of scalability is addressed in Section 2.5.

2.4.4 Numerical tests

To test the proposed framework and observe how the different proposed objective functions might lead to different outcomes, numerical tests are conducted on two test cases. The test cases are constructed by taking a set of buildings from the Resstock database of [111], which contains simulated consumption and PV production data for buildings across the U.S. for an entire year at 15-minute intervals. This dataset does not contain storage information, so it is completed by assigning BESSs of two sizes to certain prosumers, with BESS data adapted from [54]. Two instances are created: one with $n = 4$ prosumers and one with $n = 8$ prosumers. The time horizon is a day, so for each case, 365 instances of problem (A) for each of the three definitions of the objective function f are solved. For each mechanism, we solve

the version with and without the restriction of equal selling and buying prices, i.e. (A) and $(A^\dagger)$. Note that to describe each of these optimization problems, we need to solve (P_S) for each $S \subseteq N$, but this only needs to be done once, since the optimal schedule is the same for all (A) and $(A^\dagger)$. The resulting imputations are compared to the shadow imputation, which only requires solving (P_N), (D_N) once. Additionally, these two test cases are sufficiently small that the Shapley value can be computed for each of the 365 instances using (2.3), so that the distance to the Shapley value can be used as benchmark for comparing results.

Extensive details about the test cases and results obtained are presented in Appendix 2.C.2. Due to space constraints, we present here only Fig. 2.2, where we use the distance to the Shapley imputation as benchmark. Three observations are noteworthy here. First, all three proposed mechanisms serve their purpose: imputations about 30% more fair on average are obtained with this framework for the smaller test case, compared to the shadow imputation. The improvement decreases to about 18% for the larger test case. Second, no mechanism seems to clearly outperform the others with respect to this benchmark. The same is true for allowing prices to be different: there seems to be no significant difference between the two versions of each mechanism, and no version is consistently better than the other. Lastly, the differences between the mechanisms proposed are significantly reduced for the larger case, both with respect to each other, and with respect to the shadow imputation. This is in line with the asymptotic result of Section 2.3.3, and suggests that for moderate-size markets, the size of the core is already sufficiently small, that implementing the shadow price imputation does not constitute a significant sacrifice of achievable fairness. Identifying if this tendency continues as the number of participants increases requires a special handling of the exponential number of constraints in the description of $\mathcal{C} \cap \mathcal{U}$, which we address in Section 2.5.

2.5 Addressing scalability through learning

Solving any of the optimization problems of type (A) or $(A^\dagger)$ introduced in Section 2.4 requires enumerating an exponential number of constraints (cf. (2.5)). If an efficient algorithm can be devised to test whether a candidate imputation is in the uniform price core or not, then the number of constraints is not an issue, because a constraint-generation strategy can be easily implemented to only include relevant constraints in the representation of (A).

We do not have a proof at hand to show that this uniform price membership test is *hard* in a formal sense. However, in the absence of an efficient separation algorithm, we propose a sampling approach based on statistical learning to obtain satisfactory, scalable versions of the fair mechanisms proposed in Section 2.4.

2.5.1 Relaxations of the uniform price core

[3] introduce the concept of *probably stable core* and *probably, approximately stable core*. We adapt those concepts to the setting of peer-to-peer markets and formalize a definition for

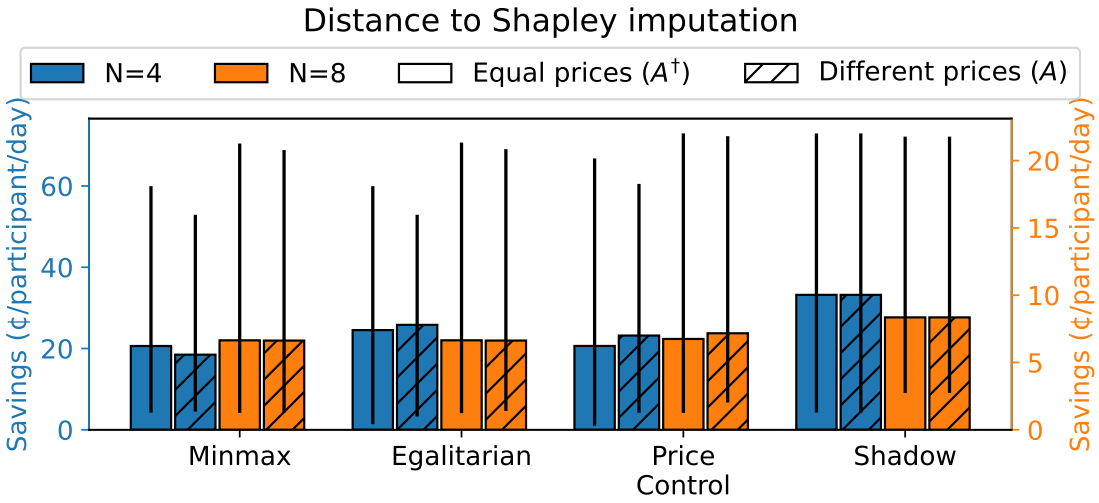

Figure 2.2: Euclidean distance between each the resulting allocation and the Shapley imputation for each mechanism. To meaningfully compare distances of different dimensions, we divide by the number of participants, so the distance plotted here is $(1/|N|)\left(\sum_{i\in N}(\phi_i - x_i)^2\right)^{1/2}$, where x is the imputation returned by each mechanism and ϕ is the Shapley imputation. Main bars represent the average across the 365 daily instances of the test case. Error bars show the min and max distances across the same dataset. Average savings achieved by the aggregation market are 54.56¢/participant/day for $n = 8$ and 99.99¢/participant/day for $n = 4$.

our uniform price core by considering the intersection with $\mathcal{V}$, as follows.

Definition 2.5.1. Given $\delta > 0$, we say a uniform price imputation in bill-splitting form $b \in \mathcal{V}$ for game G is in the *δ-probably stable uniform price core* of game G if for any distribution $\mathcal{D}$ on coalitions, we have

$$\mathrm{P}_{S\sim\mathcal{D}}\left[\sum_{i\in S} b_i \leq C(S)\right] \geq 1-\delta$$

Definition 2.5.2. Given $\delta, \epsilon > 0$, we say a uniform price imputation in bill-splitting form $b \in \mathcal{V}$ for game G is in the *(ϵ,δ)-probably, approximately stable uniform price core* of game G if for any distribution $\mathcal{D}$ on coalitions, we have

$$\mathrm{P}_{S\sim\mathcal{D}}\left[(1-\epsilon)\sum_{i\in S} b_i \leq C(S)\right] \geq 1-\delta$$

If a proposed bill-splitting b is in the (ϵ,δ)-probably, approximately stable uniform price core, then for any subset S of prosumers drawn from a distribution $\mathcal{D}$, the probability that the coalition S would be better off by forming their own P2P market and could profit by a fraction of at least ϵ of their collective bill dictated by b is less than δ.

In other words, a mechanism that gives imputations in the (δ, ϵ)-probably, approximately stable core with δ and ϵ sufficiently small is a satisfactory mechanism: participants are unlikely to find a combination of prosumers with which it would be profitable to deflect from the grand coalition. They could find one, but it is unlikely that the collective profit compared to the proposed mechanism would be sufficiently large to motivate a deflection.

2.5.2 A sampling method to obtain satisfactory imputations

[3] show for general cooperative games that allocations in the two aforementioned relaxations of the core can be learned efficiently as the number of players in the game increases by sampling only a subset of all the possible coalitions. A conceivable strategy could therefore be to first find a general allocation that minimizes the desired objective function, and then find a set of uniform prices that supports the obtained allocation. We discard this strategy for two reasons: first, it would exclude functions f that depend on the local prices themselves; secondly, not all allocations in the core are guaranteed to be supported by uniform prices (cf. Example 2.3.1).

We show next that allocations supported by prices can be learned directly via sampling by solving an optimization problem of the form $(A^\star)$,

$$
\begin{aligned}
(A^\star): \quad & \min_{b,\lambda^b,\lambda^s} f(b, \lambda^b, \lambda^s) && \text{(2.11a)}\\
& \text{s.t.} \sum_{i\in N} b_i = C(N) && \text{(2.11b)}\\
& b_i \le C(\{i\}) \ \forall i \in N && \text{(2.11c)}\\
& \sum_{i\in N} b_i \le C(S) \ \forall S \in \mathcal{S} && \text{(2.11d)}\\
& b_i = \sum_{t\in T} \lambda^b_t \left(\ell^*_{i,t}\right)^+ - \lambda^s_t \left(\ell^*_{i,t}\right)^- \ \forall i \in N && \text{(2.11e)}\\
& h_t \le \lambda^s_t, \lambda^b_t \le p_t \ \forall t \in T && \text{(2.11f)}
\end{aligned}
$$

, which is a relaxation of (A) where constraint (2.11d) is enforced just for the coalitions in a sample set $\mathcal{S}$ instead of for all $S \subseteq N$.

Theorem 2.5.1. *Assuming the uniform price core is not empty, solving optimization problem $(A^\star)$ with $|\mathcal{S}| = m$ samples of coalitions results in an imputation in the δ-probably stable uniform price core with probability $1-\Delta$. The number of samples necessary m has complexity $O\left(\frac{|N|+\log(1/\Delta)}{\delta^2}\right)$*

Proof. See Section 2.A.3. □

Theorem 2.5.2. *Assuming the uniform price core is not empty, solving optimization problem $(A^\star)$ with $|\mathcal{S}| = m$ samples of coalitions results in an imputation in the (ϵ, δ)-probably*

approximately stable uniform price core with probability $1-\Delta$. *The number of samples necessary* m *has complexity* $O\left(\nu^2 \frac{\log|N|+\log(1/\Delta)}{\epsilon^2\delta^2}\right)$, *where* $\nu = \left(\max_S C(S)\right) / \left(\min_{S\neq\emptyset} |C(S)|\right)$ *is the spread of function* C.

Proof. See Section 2.A.4. □

Theorems 2.5.1 and 2.5.2 are important because they guarantee, asymptotically, that with a sufficiently large, but tractable, number of samples m, an allocation can be obtained such that it will be unlikely that a participant can find a set of other prosumers with whom it could be profitable to form a new peer-to-peer market. Moreover, if they find such a set, it is unlikely that the profit they could reap by deflecting is significant. Since these are only asymptotic guarantees, they need to be validated with numerical experiments.

Note $(A^\star)$ is a relaxation of (A). Theorems 2.5.1 and 2.5.2 give an asymptotic, probabilistic guarantee that, although the imputation obtained with $(A^\star)$ may not be strictly feasible for (A), it can be arbitrarily close to the feasible set, in a probabilistic sense. However, they do not give any bounds on the quality of the approximation in terms of objective value. We want to argue that this is not a weakness of the approach.

One could be tempted to give some structure to the metric functions f chosen (e.g. Lipschitz continuity) and obtain similar bounding results for the approximation quality. This could be of value if the approach was driven by a tradeoff between stability and fairness, i.e. if we were adopting $(A^\star)$ instead of (A) because we are willing to sacrifice some guarantee of stability to achieve a more fair distribution of savings. In fact, our approximation is driven by computational intractability. The solution obtained from $(A^\star)$ is sufficiently close to the feasible set of (A) to prevent the collapse of the local market. Moreover, the solution obtained is guaranteed to be at least as fair as the optimal solution of (A). Therefore, this methodology results in a satisfactory mechanism, provided the number of samples necessary is not too large.

2.5.3 Numerical tests

We do a numerical implementation to test the results presented above. First, since the results of Theorems 2.5.1 and 2.5.2 are only asymptotic, we want to verify that they can be observed with a reasonably small number of samples. Secondly, Section 2.3.3 suggests that for large enough markets, all core imputations may be shadow price imputations, so it may not be possible to find stable allocations significantly more fair than the shadow price imputation. We want to check if this behavior can be observed in numerical tests.

The tests are performed on cases with 10, 12, 15, 20 and 50 participants, constructed in the same fashion of Section 2.4.4, for one arbitrary day in the dataset. For each case and objective function, problem $(A^\star)$ is solved on a random sample $\mathcal{S}$ of coalitions, with increasing values of $|\mathcal{S}|$. The allocation returned is then tested against all coalitions in a verification set. For cases with $n \leq 15$, the verification set is 2^N. For the others, 20,000 coalitions are sampled and this set is used for verification. For simplicity, coalitions were

sampled uniformly, so that probabilities correspond to just proportion of coalitions that were found to be blocking.

Fig. 2.4 shows a validation of the sampling method for the minmax mechanism. For space constraints, we omit the results for the other mechanisms, which are similar and can be consulted in the appendix. In short, the validation of Theorems 2.5.1 and 2.5.2 is successful: the introduction of more samples in $\mathcal{S}$ consistently reduces the probability of finding blocking coalitions. By slightly relaxing the stability requirement that defines when a coalition is blocking, ($\epsilon = 0.005$), that probability drops much more quickly: with as little as 200 samples, the number of blocking coalitions found is close to 0 for all test cases. The bottom plot in Fig. 2.4 provides perhaps a more direct interpretation of this result: as the number of samples in $\mathcal{S}$ is increased, the expected shortfall for the remaining blocking coalitions decreases very quickly, and therewith the potential motivation for leaving the p2p market.

For these larger cases, computing the Shapley value is too much of a computational burden; however, the shadow imputation can be efficiently found for all these instances from solving the dual solution to (D_N). We compute the distance of the imputation obtained with each of the three mechanisms to the shadow imputation. Results are shown in Fig. 2.3. The trend observed in Fig. 2.2, and suggested by the asymptotic result of [78], mentioned in Section 2.3.3, indeed continues. As the number of participants increases, the imputations obtained with all the mechanisms are closer to the shadow imputation.

Further research in this topic on the rate of convergence, or otherwise put, the rate at which the size of the core shrinks, would be welcome. The results presented here provide a strong argument in favor of using the shadow price imputation as a mechanism in local, p2p markets.

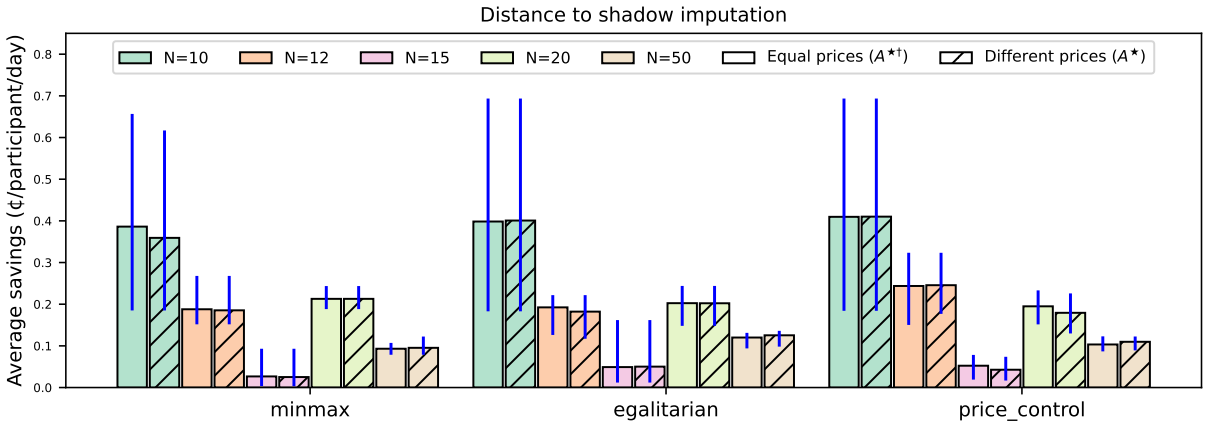

Figure 2.3: Euclidean distance[a] of resulting allocation from shadow imputation for each mechanism. Main bars represent the average distance across the 10 repetitions. Error bars show the min and max. For the day tested, the savings achieved by the aggregation market are 68.69 ¢/participant/day, so the difference allocations obtained are all within 0.5% of each other.

[a] To meaningfully compare distances of different dimensions, we divide by the number of participants, so the distance is $(1/|N|)\left(\sum_{i\in N}(\theta_i - x_i)^2\right)^{1/2}$, where x is the imputation returned by each mechanism and θ is the shadow imputation.

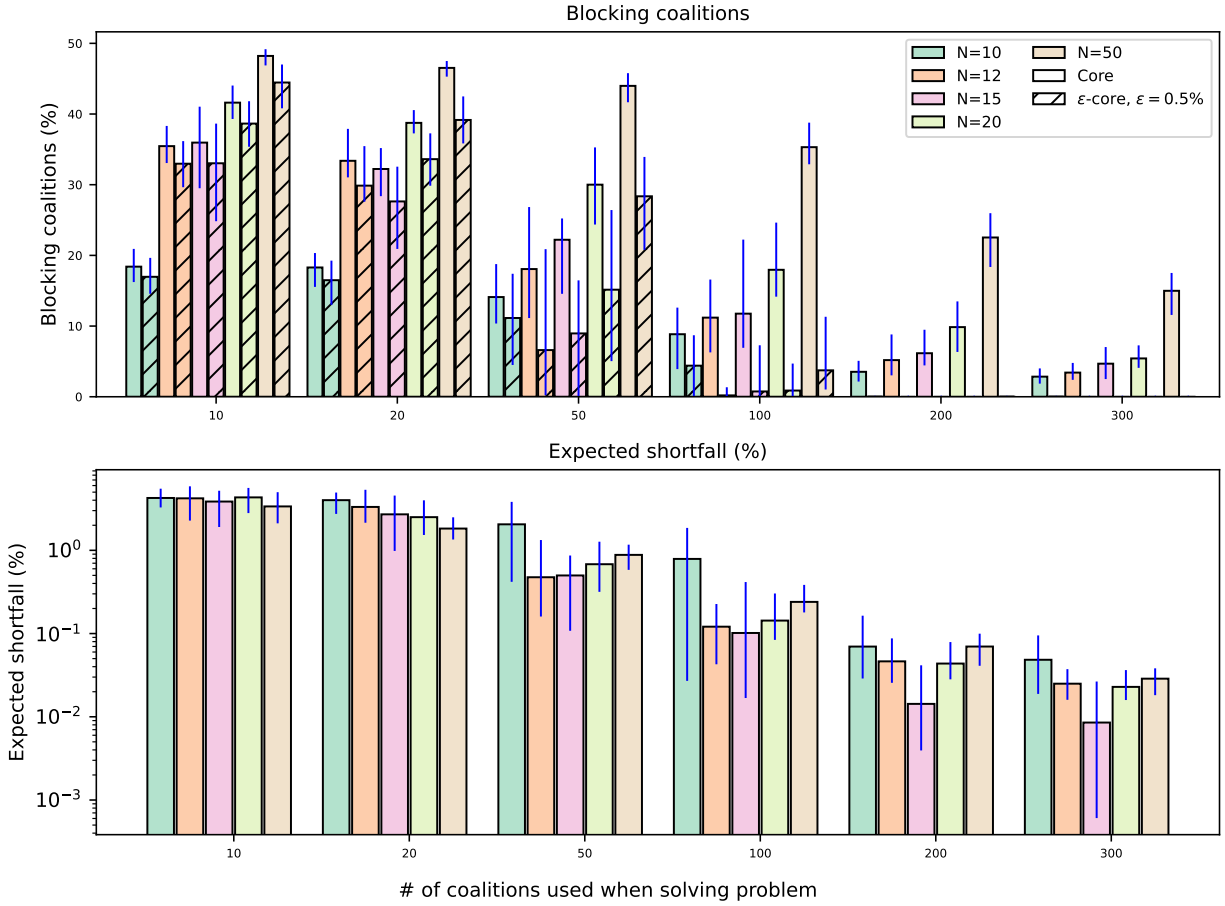

Figure 2.4: Results of validation tests of sampling methods for the minmax mechanism, with different test case sizes. *Top*: Proportion of coalitions that block the allocation proposed by the mechanism by at least a proportion ϵ of the collective bill, for $\epsilon = 0$ and $\epsilon = 0.005$. *Bottom*: Expected conditional shortfall[b] expressed as a proportion of collective bill in log scale. Similar results are obtained with the price-control and egalitarian mechanisms (see Appendix 2.C.2). Main bars represent the average over 10 repetitions. Error bars show the whole range of results obtained.

[b] Let Λ be the collection of blocking coalitions, i.e. $\Lambda = \left\{S \subseteq N : \sum_{i \in S} b_i > C(S)\right\}$, then the conditional expected shortfall shown is $(1/|\Lambda|) \sum_{S \in \Lambda} \left(\sum_{i \in S} b_i - C(S)\right) / \left(\sum_{i \in S} b_i\right)$.

2.6 Long-term participant and utility response

None of the modeling and analysis presented before consider the long-term response of the different stakeholders. In response to the distribution mechanism proposed, prosumers may choose to invest in local generation and/or storage to get a larger piece of the collective savings. Similarly, the utility may try to change its tariff structure in response to the revenue reduction that such an aggregation scheme would impose on them. Although a complete incorporation of these questions into our model is out of the scope of this work, we discuss some important points about the long-term considerations, and how the results presented here remain relevant in that context.

2.6.1 Long-term investments and participants response

In Examples 2.3.1 and 2.4.1, producers get a significantly higher share of the collective savings than pure consumers. This is also the dynamic we observe in the numeric tests (see Appendix 2.C.2). Under these conditions, in the long term, net consumers would have an incentive to invest in solar panels and/or battery storage to reap a larger fraction of the collective savings. However, that might be socially detrimental. In the extreme case, if all participants become net producers they would have no choice but to sell their electricity to the public utility, and there would be no possibility of achieving savings by aggregating demand.

An interesting avenue of future research for this work would thus be analyzing and characterizing long-term equilibrium conditions when individual investment decisions are considered. Alternatively, the investment decisions could be added to the cooperative game model. If investments can be modeled as linear variables, the mathematical results of this work could be easily extended to that case. This would correspond to a community where some generation and/or storage equipment is collectively owned, and the associated collective benefits need to be distributed among members.

2.6.2 Regulatory constraints and utility response

The analysis and concepts presented in this chapter are predicated on the assumption that the regulatory framework and the distribution system infrastructure enable the aggregation and peer-to-peer market that our mechanism supports. This is however not the case in most jurisdictions. In California, for example, separate properties may not, save limited exceptions, share one electricity meter, and allowing the distribution of generation credits corresponding to collectively owned solar systems among multiple customers is the subject of recent regulation and ongoing debate [20, 64]. Injections of power by prosumers may induce reverse energy flows in distribution systems, which usually have a radial topology. While coping with this reality has been the subject of research and development for well over a decade, most distribution systems maintain significant restrictions to bidirectional flow, posing limitations to peer-to-peer mechanisms like the ones proposed here. Changes

to both the regulatory framework and the distribution system infrastructure would thus be necessary, before an implementation of any of the mechanisms proposed here can occur.

For these changes to happen, it is necessary to understand how public utility companies might react to an eventual sprout of local community aggregation schemes. As we have shown, significant savings by customers can be achieved by aggregating demand, exploiting the spread between the public utility's buying and selling prices. Those savings come directly at the utility's expense, who would presumably react to such a reduction of revenue. We list below a few conceivable responses, and analyze how our proposed mechanism could be affected by them.

Adding a fixed charge Perhaps the most likely response by the utility would be to replace its volumetric-only pricing structure with a two-part tariff with a fixed charge per billing period and a per-kWh charge. In the context of natural monopolies like electricity retail, such schemes have some desirable characteristics, like allowing electricity to be priced at a value closer to its marginal cost, thus enabling more efficient market outcomes. Additionally, if the fixed value charged to customers is made a function of their wealth or income, as proposed in [12], the mechanism would have the aggregated value of contributing to social equity.

Adding a capacity charge A different, but related option would be to introduce a non-volumetric part of the tariff, but make it a charge per-kW of installed capacity. This could serve as a long-term incentive to reduce oversizing of installations and peak-shaving at the distribution level, both of which would increase long-term efficiency. Both of these options respond not only to the challenge of emerging Community Choice Aggregations (CCAs), but to the reduction of revenue driven by the installation of local generation by individual residential customers, even in the absence of aggregation schemes.

Charging transaction costs Another conceivable response from the utility company could be to charge prosumers a usage fee for using the utility's grid to trade energy among members of the CCA. In this scenario, volumetric rates would be maintained, but a usage fee per-kWh would be charged to prosumers who inject power into the utility's grid. Provided that the sum of the usage fee and the utility's injection price is less than the utility's selling price, the incentive for aggregation would still exist, although attainable collective savings would be smaller.

The three alternatives considered above guarantee that the utility recoups at least part of their fixed costs, even if the revenue from energy sales decreases. Under such a scenario, at least part of the reason for using distinct prices for energy consumed and energy injected would disappear, so the spread between these prices (h_t and p_t in our model) could be reduced. As long as the spread is positive, the results that we present in this chapter remain valid: there is an incentive for community aggregation and the mechanisms proposed here are

implementable methodologies to distribute the savings, possibly with a measure of fairness in mind.

It is even conceivable that, as proposed in the transactive energy system paradigm, the utility may become a pure infrastructure provider and bow out of the energy retail business, remaining only as a provider of last resort [4]. In such a scenario, the utility would have an incentive to make sure that agents trade chiefly among themselves. A high spread between buying and selling price would be a way to achieve that, making the results of this chapter relevant for that framework as well.

A thorough analysis of the dynamics of spread reduction and the formation of CCAs is out of the scope of this work, but is relevant and necessary research that should be taken into account by the regulator when determining the just and reasonable transaction fee or fixed charge that the utility may charge its customers.

2.7 Addressing the stochastic case

Before concluding, we discuss how the methodology presented here can be extended to handle the stochastic case, i.e. the case where solar PV production and electricity consumption are not known at the time when the scheduling problem (2.1) is solved. We consider a thorough analysis and development of the extended stochastic model out of the scope of this thesis and will only briefly outline how a stochastic model could be posed, and which challenges need to be overcome in that avenue of research.

Consider the case where the aforementioned uncertainty is represented through a discrete set of scenarios Ω. The scenarios could thus be added to a new dimension in the linear problem (2.1) along with their corresponding probabilities, so that (2.1) becomes a stochastic program. Under the assumption of relatively complete recourse, the validity of the strong duality results that we leverage in this work continue to hold [9]. It is thus also possible to solve the dual problem and obtain dual imputations which will be guaranteed to be in the core of the game. One important remark here is that imputations would now correspond to the result of transactions that are settled at uniform prices that are determined *ex-ante*.

The result above has an important implication regarding the strength of the mechanism. In particular, the concept of stability and core would be somewhat weaker than in the deterministic case. We will consider a uniform price imputation to be stable if no coalition can obtain a better outcome *in expectation* (i.e. ex-ante) Blocking ex-post coalitions, however, may exist. Stronger notions of stability have been proposed, but finding stable mechanisms is more challenging in those cases. See [103] for a deeper analysis of this issue in general linear stochastic games.

In the case where the net demand of participants is uncertain, the community demand aggregation can be thought of to a certain extent as an inventory centralization problem. In that vein, the methodology described before is analogous to that proposed in [19] in the context of stochastic inventory centralization games, which served as inspiration for this proposed extension.

2.8 Conclusions

In this chapter, we have adopted a game-theoretic approach to analyze the problem of fair cost sharing in peer-to-peer electricity markets where aggregation of prosumers is socially optimal. We have formally described a set of desired distributions of savings in this setting: uniform price allocations, and showed that the set of stable uniform price allocations form a strict subset of the core of the game. Observing that the model considered is a linear production game, we have shown that imputations stemming from the dual problem, i.e. pricing electricity at the shadow price of the energy balance constraint, results in an imputation that satisfies all these desiderata and is computationally efficient for larger numbers of participants.

We then showed that in spite of its good qualities, the shadow price imputation could be considered unfair, and proposed a framework to obtain the most preferred allocation through optimization. A sampling approach is utilized to overcome the challenge of tractability posed by solving these optimization problems, supported by theoretical, asymptotic results from the statistical learning literature. Numerical tests validate the approach proposed.

By noting that the results of [78] apply to the setting considered here, we conclude that for sufficiently large markets, shadow price imputations are the only stable imputations in the market. Our numerical tests suggest that this trend begins to show for markets in the order of tens of participants. On the one hand, this offers strong support for using the shadow price imputation in real local markets of a certain size with a centralized aggregator, since no other stable alternative may exist. But, as we have demonstrated, this may result in net consumers failing to have an incentive to join the local community, so on the other hand this may be considered a reason to limit the size of such aggregations, so that there may be room in the uniform price core for more fair solutions, which can be found via optimization as we propose here.

Finally, we discussed two important considerations that were out of the scope of the work presented here. The first was long-term incentives and response by market participants and by the public utility. Although a thorough development of that analysis would be necessary and welcome before an implementation of the scheme proposed here is possible, we argue that our approach would still be relevant once those implications are considered. The second extension discussed is how the mechanism can be turned into a stochastic model to address the uncertainty associated with PV production and electricity demand. We outline that our results can be extended to handle that case as well.

Appendix

2.A Proof of theorems

This section contains the theorem proofs that were not included in the main text.

2.A.1 Proof of Theorem 2.2.1

Theorem 2.2.1 is identical to Theorem 1 of [54]. We provide a proof for completeness, to account for differences in notation and representation of the model.

Proof. Consider $S, T \subseteq N$, such that $S \cap T = \emptyset$. We will use q as a shorthand for (z, w, s^c, s^d). Let q^* be an optimal solution of $(P)_S$ and q^{**} an optimal solution of $(P)_T$. We construct $\hat{q}$ as follows: for $i \in S \cup T$, $\hat{s_i}^c = s_i^{c*}$ for $i \in S$, $\hat{s_i}^c = s_i^{c**}$ for $i \in T$. $\hat{s_i}^d$ is defined in the same manner. z and w are defined by $\hat{z}_t = \max\{0, z_t^* - w_t^* + z_t^{**} - w_t^{**}\}$, $\hat{w}_t = \max\{0, -(z_t^* - w_t^* + z_t^{**} - w_t^{**})\}$ for $t \in T$.

It is straightforward to verify that $\hat{q}$ is feasible for $(P)_{S \cup T}$. Let u be the objective value of $(P)_{S \cup T}$ corresponding to the feasible solution $\hat{q}$. We have therefore $C(S \cup T) \leq u$. We want to prove that $u \leq C(S) + C(T)$.

We introduce an auxilliary variable $v_t, t \in T$ to decompose $u - (C(S) + C(T))$, as follows.

$$u - (C(S) + C(T)) = \sum_{t \in T} v_t$$

$$v_t = p_t\left(\hat{z}_t - (z_t^* + z_t^{**})\right) - h_t\left(\hat{w}_t - (w_t^* + w_t^{**})\right)$$

We claim that for all t, $v_t \leq 0$. Recall that because $h_t < p_t$, optimal solutions to $(P)_S$ and $(P)_T$ satisfy $z_t^* w_t^* = 0$ and $z_t^{**} w_t^{**} = 0$ for all $t \in T$. We have thus four cases:

- $w_t^* = w_t^{**} = 0 \Rightarrow \hat{w}_t = 0, \hat{z}_t = z_t^* + z_t^{**} \Rightarrow v_t = 0$
- $z_t^* = z_t^{**} = 0 \Rightarrow \hat{z}_t = 0, \hat{w}_t = w_t^* + w_t^{**} \Rightarrow v_t = 0$
- $w_t^* = 0$, $z_t^{**} = 0$. Two cases:
  - If $z_t^* \geq w_t^{**}$, $\hat{w}_t = 0$, $\hat{z}_t = z_t^* - w_t^{**} \Rightarrow v_t = w_t^{**}(h_t - p_t) \leq 0$
  - If $z_t^* < w_t^{**}$, $\hat{z}_t = 0$, $\hat{w}_t = w_t^{**} - z_t^* \Rightarrow v_t = z_t^*(h_t - p_t) < 0$

- Same as *case 3*, swapping the roles of z_t^*, w_t^* and $z_t^{**}, w_t^{**} \Rightarrow v_t \leq 0$

Therefore, $v_t \leq 0$ for all $t \in T$. Summing over t, we get $u \leq C(S) + C(T)$, and thus $C(S \cup T) \leq u \leq C(S) + C(T)$. Finally, we have:

$$\begin{aligned}
& C(S \cup T) \leq C(S) + C(T) \\
\Leftrightarrow\ & -C(S \cup T) \geq -C(S) - C(T) \\
\Leftrightarrow\ & \sum_{i \in S \cup T} C(\{i\}) - C(S \cup T) \\
& \geq \sum_{i \in S} C(\{i\}) - C(S) + \sum_{i \in T} C(\{i\}) - C(T) \\
\Leftrightarrow\ & V(S \cup T) \geq V(S) + V(T)
\end{aligned}$$

□

2.A.2 Proof of Proposition 2.3.1

Proof. We denote $(LR - P)_S(\pi)$ the problem obtained by doing a Lagrangian relaxation of constraint (2.1b) in (P_S) and assigning it multiplier π. The objective function in $(LR-P)_S(\pi)$ is thus:

$$\begin{aligned}
& \sum_{t \in T} \left[p_t z_t - h_t w_t + \pi_t \sum_{i \in S} \left[d_{i,t} - g_{i,t} + \eta_i^d s_{i,t}^d - s_{i,t}^c \right) - (z_t - w_t) \right] \Bigg] \\
= & \sum_{i \in S, t \in T} \pi_t (d_{i,t} - g_{i,t}) + \sum_{t \in T} \left[(p_t - \pi_t) z_t - (h_t - \pi_t) w_t + \pi_t \sum_{i \in S} (\eta_i^d s_{i,t}^d - s_{i,t}^c) \right]
\end{aligned}$$

With constraint (2.1b) relaxed, there are now no constraints coupling z, w with s^c, s^d. Similarly, there are no constraints coupling s_i^c, s_i^d for different values of $i \in S$. Thus, $(LR - P)(\pi)$ can be separated into subproblems $(LR-P)_i(\pi)$ as described below. Let $(LR-D)_i(\pi)$ be the dual of $(LR - P)_i(\pi)$.

$$
\begin{aligned}
(LR-P)_S(\pi): & \sum_{i\in S, t\in T} \pi_t(d_{i,t}-g_{i,t}) && + \min_{z,w} \sum_{t\in T} [(p_t-\pi_t)z_t - (h_t-\pi_t)w_t] \\
& && + \sum_{i\in S} (LR-P)_i(\pi) \\
(LR-P)_i(\pi): & \min_{s_i^c, s_i^d} \sum_{t\in T} \pi_t(\eta_i^d s_{i,t}^d && - s_{i,t}^c) \\
& \text{s.t.(2.1c)} - \text{(2.1f) for just } i && \\
(LR-D)_i(\pi): & \max_{\substack{\alpha_i,\beta_i,\delta_i \\ \overline{\gamma_i},\underline{\gamma_i}}} \sum_{t\in T} \overline{s_i}\alpha_{i,t} + \overline{s_i}\beta_{i,t} && + \overline{\gamma}_{i,t}\left(\overline{E_i} - e_i^0\right) + \underline{\gamma}_{i,t}\left(\underline{E_i} - e_i^0\right) \\
& \text{s.t.(2.7b)} - \text{(2.7c) for just } i &&
\end{aligned}
$$

By strong duality, $(z^*, w^*, s^{c*}, s^{d*})$ is also an optimal solution of $(LR-P)_N(\pi^*)$, so $\left(s_i^{c*}, s_i^{d*}\right)$ is an optimal solution to $(LR-P)_i(\pi^*)$ for each $i \in N$.

We claim that $(\alpha^*, \beta^*, \underline{\gamma}^*, \overline{\gamma}^*, \delta^*)_i$ is also optimal for $(LR-D)_i(\pi^*)$ for each $i \in N$. We prove this by contradiction and use q as shorthand for $(\alpha, \beta, \underline{\gamma}, \overline{\gamma}, \delta)$. It is easy to check that q_i^* is feasible for $(LR-D)_i(\pi^*)$. So if q_i^* is not optimal for $(LR-D)_j(\pi^*)$ for some $j \in S$, there exists q_j^{**} with higher objective value than q_j^*. But since the feasible sets are uncoupled, the solution $\left(\pi^*, (q_i^*)_{i\in S\setminus\{j\}}, q_j^{**}\right)$ is feasible for (D_N) and would have higher objective value than (π^*, q^*), which violates the optimality of (π^*, q^*) for $(D)_N$.

We thus have that $(s^{c*}, s^{d*}, (\alpha^*, \beta^*, \underline{\gamma}^*, \overline{\gamma}^*, \delta^*)_i$ is an optimal primal-dual solution for pair $(LR-P)_i(\pi^*), (LR-D)_i(\pi^*)$. Applying again strong duality to this pair, we obtain:

$$
\begin{aligned}
\sum_{t\in T} \pi_t^* \left(\eta_i^d s_{i,t}^{d*} - s_{i,t}^{c*}\right) = & \sum_{t\in T} \overline{s_i}\alpha_{i,t}^* + \overline{s_i}^*\beta_{i,t}^* \\
& + \overline{\gamma}_{i,t}^*\left(\overline{E_i} - e_i^0\right) \\
& + \underline{\gamma}_{i,t}^*\left(\underline{E_i} - e_i^0\right)
\end{aligned}
$$

, which adding $\sum_{t\in T} \pi_t^* (d_{i,t} - g_{i,t})$ to both sides yields (2.9). □

2.A.3 Proof of Theorem 2.5.1

Proof. Theorem 2.5.1 is a replica of Theorem 3 of [3], substituting the probably stable core with the probably stable uniform price core we have defined. We follow the steps of their proof, i.e. we find a class of functions that contains the uniform price core, verify that its VC-dimension is low, and use that to find an asymptotic bound on the number of samples necessary to find points in the uniform price core.

Proposition 2.A.1. *Let $n = |N|$, $u \in \mathbb{R}^{n+1}$, and $\ell \in \mathbb{R}^{n|T|}$. Let $\mathcal{H}'$ be the class of functions that define uniform price allocations in bill-splitting form, i.e.*

$$\mathcal{H}' := \Bigg\{ u \to \operatorname{sign}\left(\sum_{i=1}^{n} b_i u_i - u_{n+1}\right) b \in \mathbb{R}^n, \lambda^b, \lambda^s \in \mathbb{R}^{|T|}, \\ \sum_i b_i = C(N), b_i = \sum_{t \in T} \left(\lambda_t^b \left(\ell_{i,t}\right)^+ - \lambda_t^s \left(\ell_{i,t}\right)^-\right) \ \forall i \in N \Bigg\}.$$

Then, $\mathcal{H}'$ has VC-dimension at most $n+1$.

Proof. Following the argument of Corollary 1 of [3]:

$$\mathcal{H}' \subseteq \left\{ u \to \operatorname{sign}\left(\sum_{i=1}^{n} b_i u_i - u_{n+1}\right) : b \in \mathbb{R}^n \right\} \subseteq \mathcal{H} := \left\{ u \to \operatorname{sign}(w^\top u) : w \in \mathbb{R}^{n+1} \right\}$$

By Theorem 2 of [3], $\mathcal{H}$ has VC-dimension $n+1$. If $\mathcal{H}' \subseteq \mathcal{H}$, then $\operatorname{VCdim}(\mathcal{H}') \leq \operatorname{VCdim}(\mathcal{H})$. □

To conclude, we follow the last step of the proof of Theorem 3 of [3] as follows. Consider $q^* = (b^*, \lambda^{b*}, \lambda^{s*})$, an optimal solution to $(A^\star)$, which is guaranteed to exist because the uniform price core is non-empty by assumption. We will show q^* is in the δ probably stable core. Let $h(u) = \operatorname{sign}\left(\sum_{i=1}^n b_i^* u_i^S - u_{n+1}^S\right)$, where $u_i^S = \mathbf{1}_{i \in S}$ and $u_{n+1}^S = C(S)$. Let $f(u) = -1$ for all u. Since (2.11d) is satisfied for all samples in $\mathcal{S}$, $f(u) = h(u)$ on all the samples, and thus, $(1/m)\sum_{i=1}^m \mathbf{1}_{h(u) \neq f(u)} = 0$.

$$\begin{aligned}
&\mathrm{P}_{S \sim \mathcal{D}}\left[\sum_{i \in S} b_i \leq C(S)\right] \\
&= 1 - \mathrm{P}_{u^S : S \sim \mathcal{D}}\left[\operatorname{sign}\left(\sum_{i=1}^{n} b_i u_i^S - u_{n+1}^S\right) \neq -1\right] \\
&= 1 - \mathrm{P}_{u^S : S \sim \mathcal{D}}\left[h(u^S) \neq f(u^S)\right] \\
&= 1 - \left|\mathrm{P}_{u^S : S \sim \mathcal{D}}\left[h(u^S) \neq f(u^S)\right] - (1/m)\sum_{i=1}^{m} \mathbf{1}_{h(u^S) \neq f(u^S)}\right| \\
&\geq 1 - \delta
\end{aligned}$$

, where the last step is true by Theorem 1 of [3], with $\mathcal{O}\left(\left(n + \log(1/\Delta)\right)/\delta^2\right)$ samples because the VC-dimension of our hypothesis class $\mathcal{H}'$ is at most $n+1$ by Proposition 2.A.1. □

2.A.4 Proof of Theorem 2.5.2

The proof is a replica of the proof of Theorem 5 of [3], substituting the probably, approximately stable core with the probably, approximately stable uniform price core. It is easy to check that allocations in the uniform price core satisfy all the conditions of the theorem: they have bounded ℓ_1 norm because they are in the core (Lemma 2 of their proof), and candidate solutions to $(A^\star)$ can be found efficiently because it is a convex problem. It only remains to verify that $\min_{S\subset N, S\neq\emptyset} C(S) > 0$, where $C(S)$ is the characteristic function of the game in cost-minimization form.

The assumption is not true for game G, as we may have $C(S) \leq 0$. We show here that we can construct a game G' that satisfies this condition and has a 1:1 correspondence with game G.

Let $K \geq 0$ be a constant such that $C(S) > -K\,|S|$ for all $S \subseteq N$. Then, we can define G' as the game with set of players N and value function $V'(S) = \sum_{i\in S} C'(\{i\}) - C'(S)$, with $C'(S) = C(S) + |S|\,K$. Note that $V'(S) = V(S)$ for all $S \subseteq N$, $S \neq \emptyset$, so the correspondence between G and G' is clear, i.e. C' is a cost-minimization version of V, and clearly, $C'(S) > 0$ for all $S \subseteq N$. Therefore, we only need to prove that such a K exists.

Claim 2.A.1. *$K = |T|\,(\max_{t\in T} p_t)\,(\max_{i\in N, t\in T} g_{i,t} + \max_{i\in N} \overline{s}_i)$ satisfies $C(S) > -K\,|S|$ $\forall S \subseteq N$.*

Proof. Let $p^\star = \max_{t\in T} p_t$, $g^\star = \max_{i\in N, t\in T} g_{i,t}$, $\overline{s}^\star = \max_{i\in N} \overline{s}_i$

Then, for any nonempty $S \subseteq N$, we have:

$$\begin{aligned}
C(S) &= \sum_{t\in T} p_t z_t^* - h_t w_t^* \\
&= \sum_{t\in T} p_t \left(\sum_{i\in S} \ell_{i,t}^*\right)^+ - h_t \left(\sum_{i\in S} \ell_{i,t}^*\right)^- \\
&\geq \sum_{t\in T} -h_t \left(\sum_{i\in S} \ell_{i,t}^*\right)^- \\
&> -\sum_{t\in T} p_t \left(\sum_{i\in S} \ell_{i,t}^*\right)^- \\
&\geq -\sum_{t\in T} p_t \sum_{i\in S} \left(g_{i,t} + \eta_i^d s_{i,t}^{d*}\right) \\
&\geq -\sum_{t\in T} p_t\,|S|\,(g^\star + s^\star) \\
&\geq -|T|\,p^\star\,|S|\,(g^\star + s^\star) \\
&= -|S|\,K
\end{aligned}$$

□

Table 2.B.1: Battery energy storage systems and price data used in test cases, adapted from [54].

Storage Type	$\overline{E_i}$	$\underline{E_i}$	$\overline{s_i}$	η_i^c	η_i^d
Small	7kWh	0.7 kWh	11.2kW	0.95	0.95
Big	14kWh	1.4 kWh	22.4kW	0.95	0.95
Time of day			p_t	h_t	
Midnight - 7AM			7 ¢/kWh	4.03 ¢/kWh	
7AM - Midnight			14.71 ¢/kWh	4.03 ¢/kWh	

Table 2.B.2: Composition of the constructed test cases. Number of prosumers having PV and/or ESS for each test case used. In parenthesis, name assigned to participant in plot results when applicable.

Number of prosumers with	Total number of prosumers $N =$						
	4	8	10	12	15	20	50
No ESS, No PV	1 (D)	3 (A,D,H)	4	2	1	5	6
No ESS, PV	1 (A)	1 (C)	2	2	4	5	14
Small ESS, No PV	1 (C)	0	0	1	3	1	3
Small ESS, PV	0	2 (E,F)	2	4	3	3	12
Big ESS, No PV	0	1 (G)	0	1	1	2	6
Big ESS, PV	1 (B)	1 (B)	2	2	3	4	9

2.B Description of test cases for numerical implementations

All test cases are constructed by taking generation and consumption timeseries for residential buildings from the Resstock database [111]. We pick the census microdata area in California with the largest number of buildings with PV installations in the dataset, and filter the data so that only single-family houses remain. The test cases are built by arbitrarily picking buildings with and without solar photovoltaic. The data does not include energy storage systems (ESS), so battery ESS are arbitrarily assigned to some of the buildings. The characteristics of the ESS and the utility prices are taken equal to those reported in [54] and are summarized in Table 2.B.1.

Table 2.B.2 describes the types of agents included in each of the test cases considered. For the numerical tests of Section 4, all 365 days in the dataset were used. For the validation of the sampling method, one day in the data set was selected randomly.

The implementation was done in python using the *pyomo* modelling package and solved with *Gurobi*. All tests are run on a laptop with 2.7GHz Intel Core i5 and 8GB of RAM.

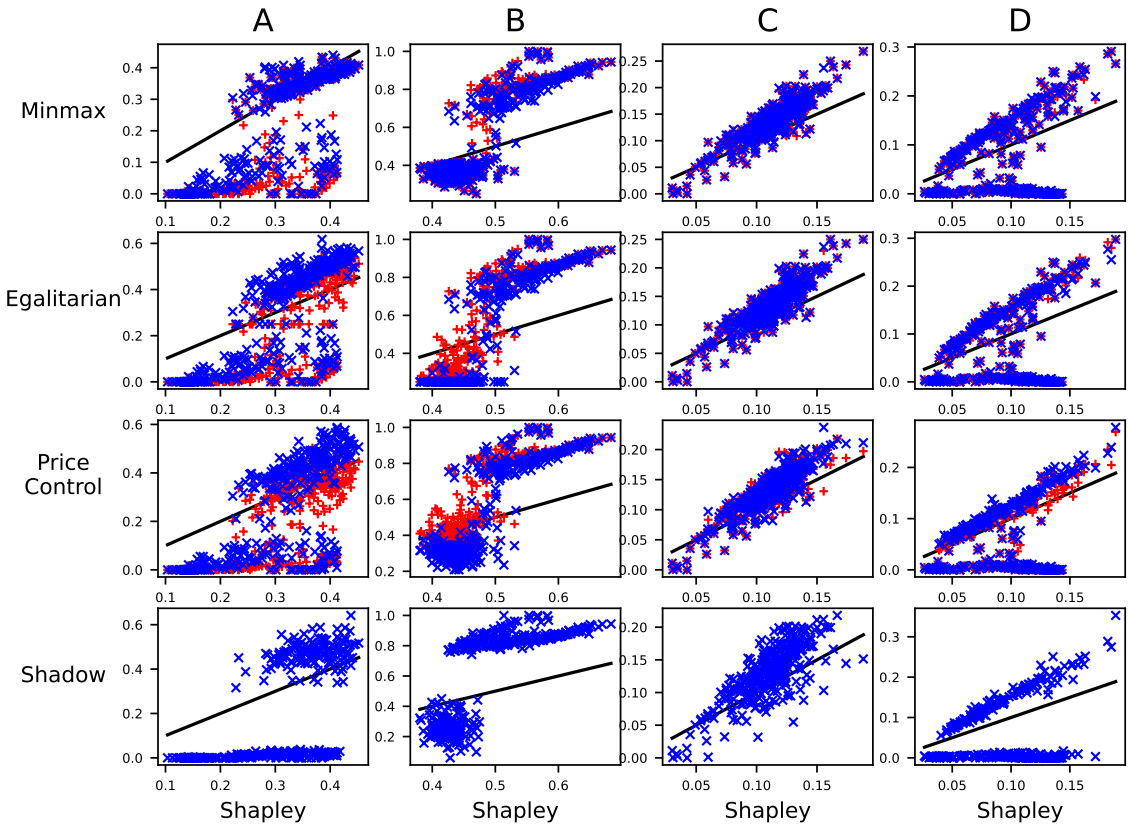

Figure 2.C.1: Scatter plot of proportion of savings allocated to each participant vs. corresponding Shapley allocation, for the four mechanisms considered. Each row corresponds to a mechanism, each column to a participant. Each point corresponds to a day in the year. The solid black line represents the $y = x$ line, so points above and under the line are overallocated and underallocated respectively, with respect to the Shapley allocation. $\times$ for allocations allowing for different selling and buying prices. $+$ for allocations restricted to having equal selling and buying prices. The shadow imputation has identical selling and buying prices.

2.C Extended results of numerical implementations

2.C.1 Optimization framework for most fair uniform price allocation

Figures 2.C.1 and 2.C.2 show detailed results of the numeric tests performed on the test case with $n = 4$ participants. Figure 2.C.3 show the results for the case with $n = 8$ participants.

Figure 2.C.4 shows the annual savings per participant under the shadow imputation

Figure 2.C.2: Proportion of savings allocated to each participant for the four mechanisms considered, case with $n = 4$ participants. Error bars show the range of allocations obtained during the 365 days. Main bars show the average across the year. Projected Shapley is the uniform price allocation in the core that is closest to the Shapley value in ℓ_1 norm.

for different market sizes. The results show that while it is true that the range of savings achieved by different participants vary significantly, all participants achieve positive savings, with a majority of the savings in the range of $100's.

2.C.2 Validation of sampling methodology

The results of the validation tests for the sampling methodology for the other mechanisms are presented in Figures 2.C.5, 2.C.6 and 2.C.7.

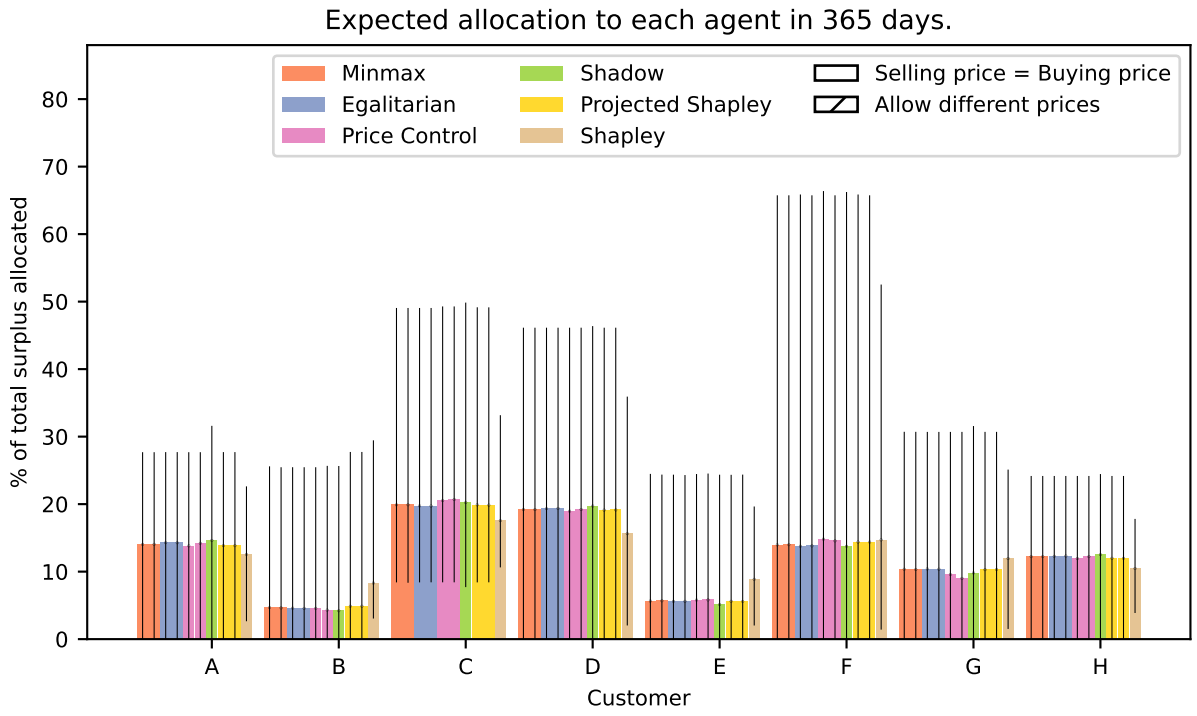

Figure 2.C.3: Proportion of savings allocated to each participant for the four mechanisms considered, case with $n = 8$ participants. Error bars show the range of allocations obtained during the 365 days. Main bars show the average across the year. Projected Shapley is the uniform price allocation in the core that is closest to the Shapley value in ℓ_1 norm.

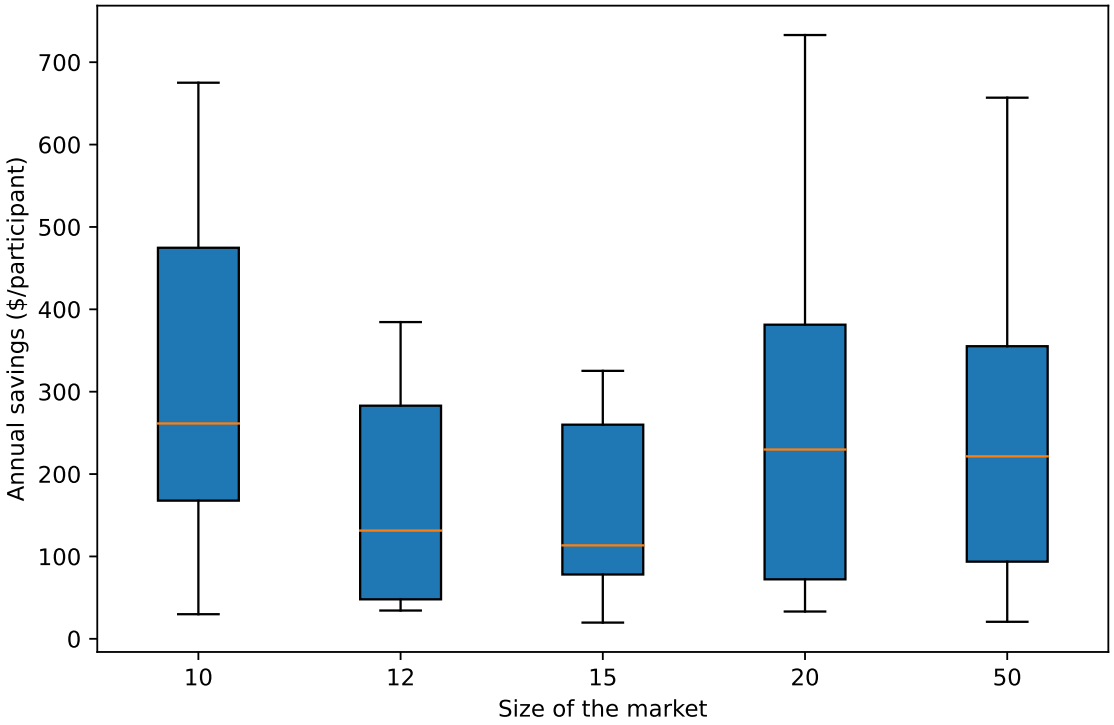

Figure 2.C.4: Boxplot showing the distribution of savings per participant under the shadow price imputation for different market sizes, over the number of participants in each market. The box extends from the first quartile (Q1) to the third quartile (Q3) of the data, with a line at the median. Whiskers cover the whole range of data.

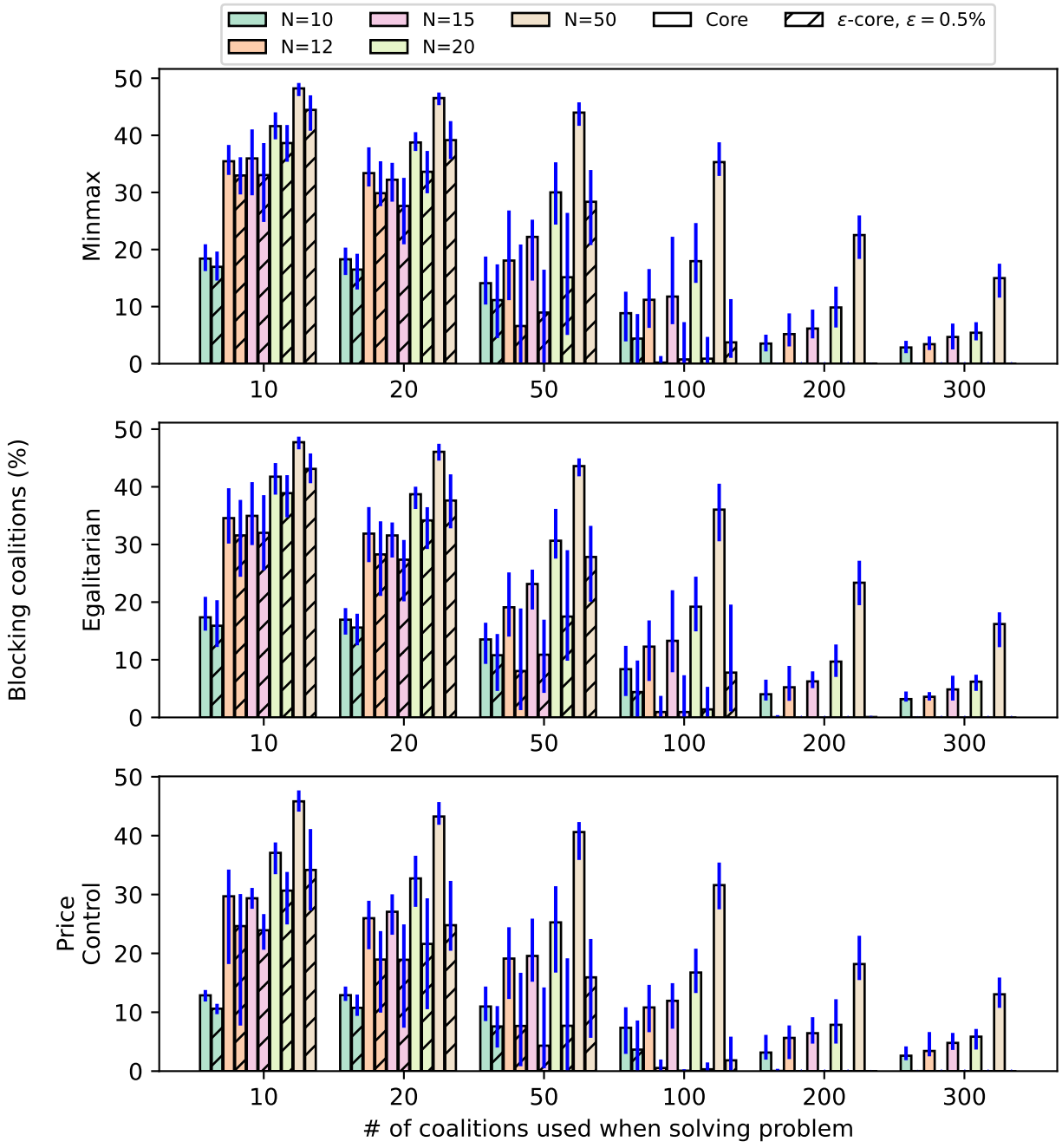

Figure 2.C.5: Blocking coalitions found for obtained allocations, for different market and sample sizes, for all mechanisms. Error bars show the range of results obtained over 10 repetitions. Main bars show the average. The proportion of coalitions found to be blocking by at least a proportion ϵ of the collective bill, for $\epsilon = 0$ and $\epsilon = 0.005$ are shown.

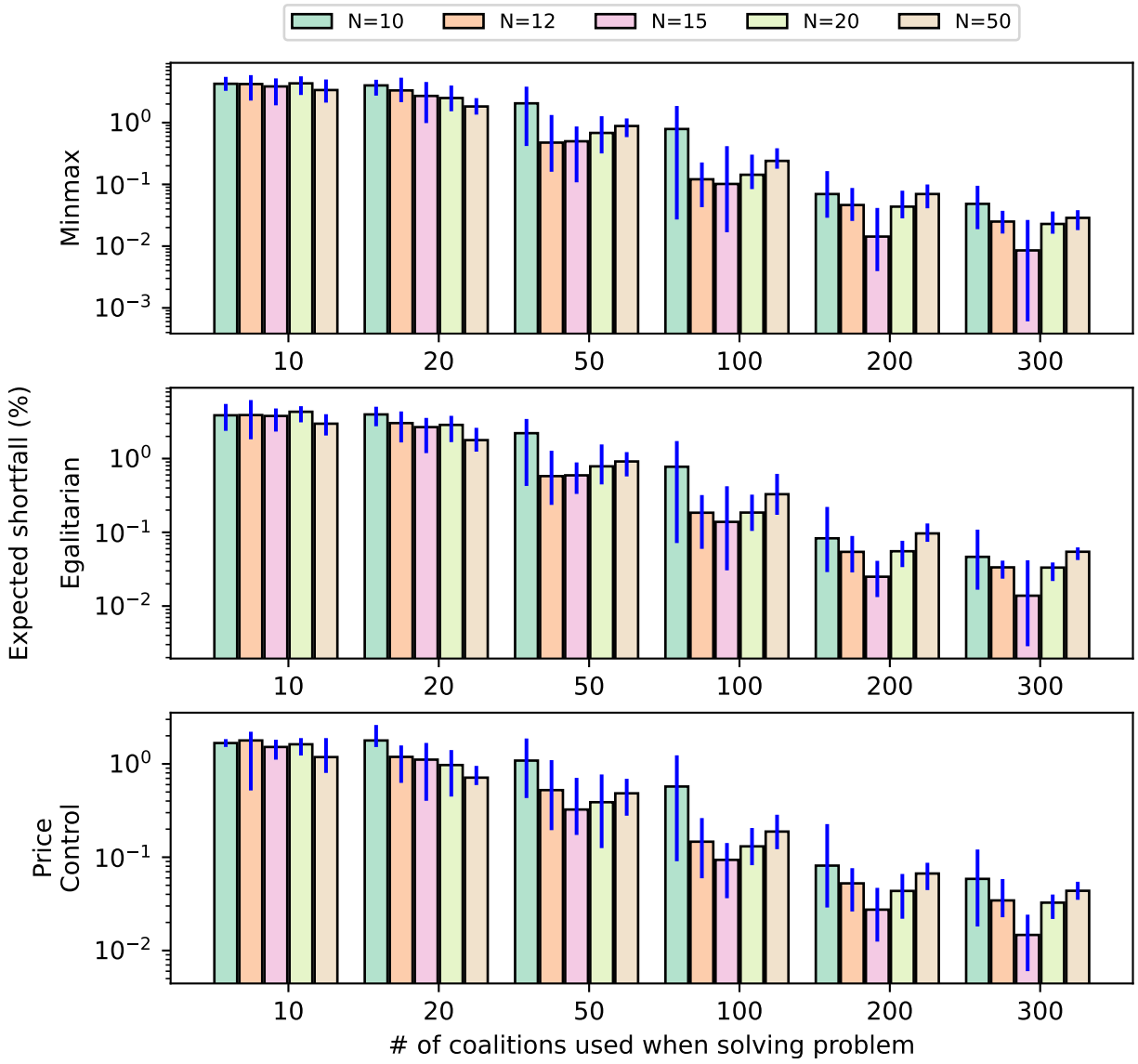

Figure 2.C.6: Expected shortfall for blocking coalitions, for different market and sample sizes, for all mechanisms. Expected conditional shortfall[a] expressed as a proportion of collective bill in log scale. Main bars represent the average over 10 repetitions. Error bars show the whole range of results obtained.

[a] Let Λ be the collection of blocking coalitions, i.e. $\Lambda = \left\{S \subseteq N : \sum_{i \in S} b_i > C(S)\right\}$, then the conditional expected shortfall shown is $(1/|\Lambda|) \sum_{S \in \Lambda} \left(\sum_{i \in S} b_i - C(S)\right) / \left(\sum_{i \in S} b_i\right)$

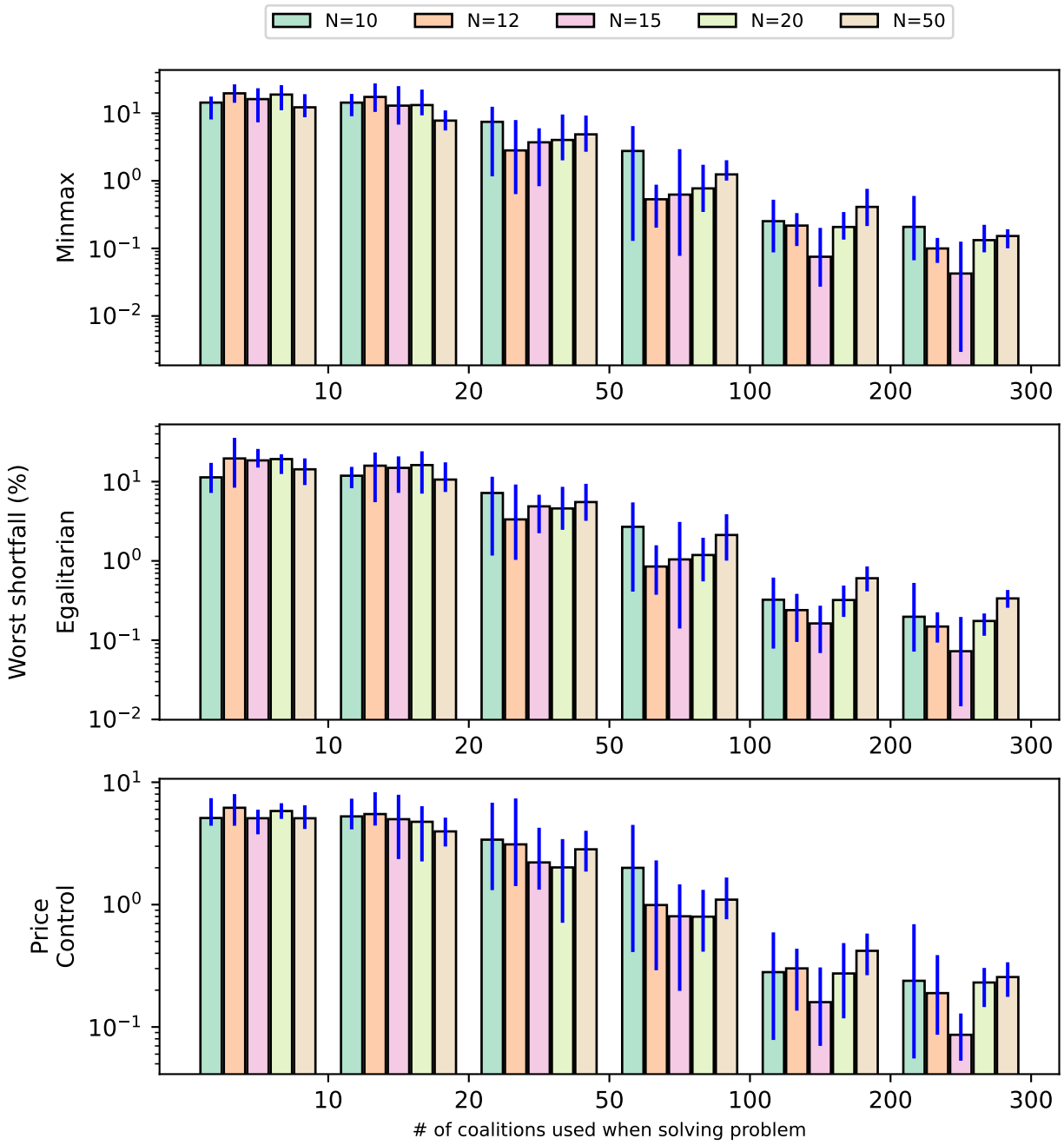

Figure 2.C.7: Worst shortfall for blocking coalitions, for different market and sample sizes, for all mechanisms. Shortfall expressed as a proportion of collective bill in log scale. Error bars show the range of allocations obtained during the 365 days. Main bars show the average across the year.

Chapter 3

A Markovian model for optimal sizing of colocated storage for wind power production [1]

Abstract

We propose a high-level stochastic steady-state model to analyze the value of co-located energy storage systems for wind power producers that participate in an electricity market through forward contracts and use storage to unlock access to capacity payments. In particular, we try to find optimal storage and contract sizing, as well as stationary operating policies for profit maximization in the long-run. We propose a stylized model calibrated to actual wind power production that allows us to obtain limiting distributions of battery storage levels, assess the value of storage size and perform a sensitivity analysis on key parameters such as contract prices, capacity payments and storage efficiency. We develop the case with contracts of constant price, outline how this model can be extended to a variable-price setting and discuss potential challenges in that avenue.

Acknowledgments

We are grateful to Professor Nail Akar of Bilkent University in Ankara, Turkey, for making us aware of the improved method of [47] and for generously sharing his code implementation of it with us, which we have used for obtaining the results presented here.

[1] A substantial part of the text that constitutes this chapter was published in *IEEE Transactions in Energy Markets, Policy and Regulation* under the name "Data-Driven Sizing of Co-located Storage for Uncertain Renewable Energy", authored by Tomas Valencia Zuluaga and Shmuel S. Oren [107]. An early version of that manuscript was presented as a conference paper at the 2022 IREP Conference in Banff, Canada. The conference paper was published in [108]. This work was supported by the Power Systems Engineering Research Center (PSERC).

3.1 Introduction

The impending consequences of climate change have driven massive proliferation of Renewable Energy Sources (RES) around the world. However, a key obstacle to large-scale integration of RES in power systems is the short-term uncertainty and variability of their output. This poses both technical challenges for the reliable operation of the power system and financial challenges for investors in RES, since it is hard to guarantee a reliable income flow without RES incentives through policies such as feed-in tariffs and mandatory contracting for retailers. Such policies however are not sustainable and are being challenged due to their cost to consumers; many systems are now moving toward imposing scheduling requirements and forward commitments on RES forcing them to compete on a level playing field with other resources. The addition of Battery Energy Storage Systems (BESSs) to renewable power plants can help mitigate their uncertainty and thus can help towards solving both the technical and financial issues. While it is intuitive to understand how storage can mitigate uncertainty, sizing and managing BESSs is not obvious.

Optimizing contracting and operation for intermittent power plants has been a very active research topic in both theory and application. In [48], a theoretical analysis is performed for the case of a wind farm that uses storage to optimize its bidding strategy on the day-ahead market in order to minimize imbalance penalties. In this, and other works that study the behavior of Wind Power Producers (WPPs) in day-ahead markets [10, 27], it is observed that the optimal bidding strategy for WPPs takes the form of an optimal fractile, as in the solution to the well-known Newsvendor problem. This highlights the fact that the bidding problem of a WPP is a *reverse* newsvendor problem, where the uncertainty lies not on the demand but on the supply side.

There are also numerous references that address the issue of optimal sizing and optimal management of co-located energy storage. In [39], the infinite-horizon average cost of electricity purchases is minimized by finding an optimal storage management policy and optimal storage size for a power plant that serves a local demand and purchases any shortfall from the grid in presence of dynamic pricing. The authors prove the optimality of a dual threshold policy , reminiscent of optimal (s, S) policies for inventory control. In [88], the optimal size of a BESS is found for a grid-connected photovoltaic system that can purchase and sell energy from the grid under time-of-use pricing, and the convexity of profit in storage size is shown. In [58], the value of co-located storage is analyzed for the case of the UK market as subsidies are phased out by comparing different contracting schemes with help of a model using stochastic differential equations describing wind power production and prices.

The application of high-scale, steady state models, such as fluid queues, to energy storage remains relatively rare. Fluid queue models provide a method to characterize steady-state behavior of a fluid container whose intake/outlet operates stochastically under certain conditions, without need for Monte Carlo simulations, which is relevant for long-term decisions such as sizing said container, as we do in this work. Limiting distributions for fluid queue models have been most commonly used in high-speed communication networks, but also for manufacturing [65], and energy systems, with early applications in hydro dam management

models [34]. In an approach qualitatively similar to ours, in [77], a two-state Markov-Modulated Fluid Queue (MMFQ) is used for optimizing the size and management policy of the US national strategic petroleum reserve.

There are also more recent applications of fluid queue models to renewable power plants with BESSs. In [18], a model-predictive control algorithm is proposed for optimally sizing and managing storage, and the MMFQ framework is used to analyze the reliability performance of the method. In [28], the authors consider a case in which there is no control over the charge/discharge rate (akin to a balancing policy), and use MMFQs to find an asymptotic relation between the battery size and the loss of load probability for a given grid configuration, so that the problem of finding the optimal size of storage to hit a target loss-of-load probability can be solved. Unlike in these cases, here we use the MMFQ framework in the objective function of our optimization model directly.

While short term operation of co-located storage can be optimized using dynamic programming techniques and accounting for current information on state of charge, wind forecast and prices, the optimal sizing of the co-located storage is based on long run average behavior of the production/storage system under an optimized stationary policy. For this purpose we employ a method based on the spectral analysis for characterizing limiting distributions of n-dimensional MMFQs described in [65, 52]. This is a numerically challenging problem for which algorithms with proven numerical stability have been developed [92, 93], but with the limitation of having a single state with negative drift, too restrictive for our setting. In [1], an algorithm is proposed for calculating limiting distributions of MMFQs without solving any eigenvalue problems, and was improved and extended to also handle multi-regime fluid queues in [47], which we use in our approach. The method employed there makes no assumptions regarding the structure of the fluid queue, which is a key feature for general application models like ours.

In this chapter, we use limiting distributions directly to obtain simple expressions of long-run profits of the system, suitable for use in a sensitivity analysis. In particular, we are interested in investigating the effect of some key parameters such as available contract prices on the financial outcome of the project, the optimal storage size and the optimal contract size of a wind power plant trying to maximize its profit. This simple model would be of interest to a wind power plant operator and project developer for early-stage project feasibility analysis. It could also be of interest for policy designers, to evaluate the financial viability of storage projects for wind power plants with and without special incentives.

There are several examples of modeling wind speed and wind power output as a Markov chain in the literature, with applications in simulation of wind data series [57, 99, 114] but also in long-run analysis [28, 18]. It has been found that performing a max likelihood parameter estimation can result in a very good approximation of the limiting distribution, but that a key metric to obtain a more accurate model is the autocorrelation [13]. In [13], in the context of co-located storage sizing for robust operation of microgrids, it is found that Markovian models with an autocorrelation that poorly reflects that of the original data series can lead to underestimation of necessary storage by as much as 50%. Autocorrelation performance can be improved by increasing the order of the Markov chain, but this increases

the size of the state space exponentially, so that only chains of second or third order are of practical relevance. In [79], a rolling-average method is proposed to obtain higher autocorrelation performance in the lower range (0-6h) without increasing the size of the state space. More recently, authors have proposed non-homogeneous Markov chains [91, 114] to more accurately reproduce the autocorrelation of real data series, with much better results replicating the daily behavior of wind (i.e. the autocorrelation around integer multiples of 24h). These results can also be leveraged to capture seasonal changes of longer duration in the wind distribution. The expected range of optimal storage sizes should also inform the decision of how much autocorrelation needs to be captured by the model. If these sizes lie in the 0-6 hour range, capturing daily autocorrelation is less crucial than if storage is expected to be in the 24h+ range.

We could find no references in the literature of modelling wind power production and electricity prices as a joint Markov chain as our general framework proposes, but the discussion above provides what we deem sufficient justification for our modelling of wind power output as such a process in the constant prices case that is addressed in this chapter, as supported by the empirical results shown in Section 3.4.1. We do not claim that a Markovian assumption is appropriate to find the actual day-to-day operating policy of a wind power plant with co-located storage. As mentioned above, dynamic programming approaches (possibly with reinforcement learning) that utilize updated status and forecasts are best-suited for those purposes (see e.g. [18]). However, for investment planning, a simpler, higher-level representation is not only appropriate but necessary to get a tractable and thus useful model. This is not in essence different from other commonly performed simplifications, just as unrealistic but widely accepted, like using linear regressions to represent complex socio-economic dynamics in the context of capacity expansion. We leave a more thorough discussion of the variable-price case for future work in which that version of this model is developed.

The remainder of this chapter is organized as follows. Section 3.2 describes the setting for our problem. In Section 3.3, we present in detail our optimization model, as well as the solution algorithm. Results are presented and discussed in Section 3.4. Section 3.5 presents an exploration of the path to obtain a full solution in the case of long-term contracts with variable prices (supply functions). Finally, we conclude and discuss possible next steps for the development of this model in Section 3.6.

3.2 Problem setting

We consider the case of a WPP that participates in the wholesale electricity market and has access to a long-term forward market. The WPP is also evaluating the construction of a collocated BESS, which will be used to mitigate imbalance costs. In our setting, the wind farm operator needs to answer two main questions: how much storage should be installed and how much energy should be sold in a long-term contract. To answer these questions, it is also relevant to determine a stationary management (charge/discharge) policy for the BESS. Further aspects of our setting are described in more detail in the paragraphs below.

We consider that the WPP is a price-taker in the long-term forward market, where it has access to three contracts:

1. A long-term fixed quantity contract, where it can choose how much to sell for a fixed price in every period of the wholesale electricity market.

2. A pay-as-demanded forward contract, which can be used to cover any shortfall in generation with respect to the commitment in contract 1.

3. A pay-as-produced forward contract, which can be used to sell any excess generation with respect to the commitment in contract 1.

In order to avoid arbitrage opportunities, which cannot exist in efficient markets in the steady-state case considered here, contract 2 must have a higher price than contract 1 and contract 3 must have a lower price than contract 1.

In this setting, the WPP commits to delivering any quantity up to its capacity before the actual output, over which he has no control, is known. An imbalance penalty is paid for the difference between the energy commitment and the actual energy delivered. The case studied here is the special case where the commitment is the same for all periods, the imbalance penalty for shortfall is non-negative and constant, and the imbalance penalty for surplus is non-positive (i.e. it is a reward, not a penalty) and constant.

To mitigate the imbalance cost, the wind power producer can charge/discharge the BESS, so that the net output of the plant, i.e. that perceived by the market is the combined production of the wind farm and the BESS. Having a co-located BESS will have the effect of firming the WPP's energy output, thus giving it access to the capacity market. We also include this source of income in our model, by considering that the WPP is a price taker in the capacity market and is remunerated for the capacity committed in contract 1, derated by the long-run probability of not honoring that contract.

The scope of our work is a high-level analysis that could be of interest for early-stage project developers or policy makers. We are interested in looking at infinite-horizon average profits by considering the steady-state behavior of the model proposed and performing sensitivity analyses on a number of key parameters. To be consistent with this approach, we do not consider the possibility of using storage for arbitrage, since planned use of capacity for opportunistic arbitrage would not be profitable in the long run if the market is efficient.

Admittedly, this is a limited representation of real electricity markets, where producers can and do participate simultaneously in different ways: short-term arbitrage, ancillary services, and a sequence of medium-term forward contracts to name a few. However, in the long-run efficient equilibrium, a long-term contract price will adjust so that participation through the scheme modeled here will be equivalent to the other forms available to the WPP. In essence we invoke here a no arbitrage principle, which should hold in a long run equilibrium, implying financial equivalence of any market participation mode of a fully utilized production/storage facility. This equivalence enables us to choose the most convenient market participation mode for the purpose of our analysis even if in reality the resulting

production/storage facility will participate in the market differently. Hence, there are two direct interpretations of the assumptions used in our model: the forward contract is assumed to always renew at maturity with identical conditions, or, alternatively, it is a long-term forward contract with a price that makes the contract equivalent to a sequence of contracts with shorter maturities or any other form of efficient market participation (that does not leave money on the table).

This raises the question of whether the results obtained by the model presented here are applicable to actual market participants who may not have a long-term contract but rather sequence of forward contracts with shorter maturity and possibly changing prices and quantities. The possible changes in future forward prices can be accounted for through sensitivity analyses on the price of the contract, that our method allows. We are also assuming that efficient utilization of a production/storage facility under a long term contract involves capturing of capacity payments, if available, which is consistent with commitment not changing significantly across subsequent forward contracts. For WPPs that participate through a more varied combination of market mechanisms (arbitrage, balancing, etc.), it is more of an open question whether this model will be directly applicable, although this model will still be of interest.

Our approach suggests the assumption of existence of a long-run equilibrium in the forward market, which depends on many factors that are not considered in our model, such as long-run demand behavior and a stabilization of the cost of new capacity (which in the case of storage, for instance, is actually expected to continue declining for some years [11]). We do not make any such strong assumptions. While such an equilibrium may be far from being reached, this model is a high-level analysis for which this coarse approximation of reality is sufficient.

Finally, we do not specify any particular battery technology. We include in our analysis the issues of charge/discharge conversion efficiency and energy dissipation. However, we do not include the degradation of the battery because of usage and aging in our model. It has been found that this can be an important characteristic to take into account in sizing studies [58], so this could be an interesting feature to add to future versions of our model. In our closing remarks in Section 3.6.1 we discuss how considering this issue could affect our model and outline how this feature could be included in an extension, but otherwise consider this topic out of the scope of this thesis.

3.3 Mathematical Model

3.3.1 Description of the model

To emphasize the flexibility of our model and lay the groundwork for a future extension, we describe first the model of a forward contract that allows variable prices, and then focus on the special case addressed in this chapter.

The energy market

We consider a market in which the WPP is a price taker, so it commits to produce an amount q_t (in MWh) during market period t at a price p_t (in \$/MWh), which is known at the time the commitment is made. During period t, the actual wind power output of the farm is w_t (in MWh), while r_t is the amount energy injected into the BESS ($r_t < 0$ if the energy is extracted), so that the net output of the plant is $w_t - r_t$. The imbalance is $y_t = q_t - (w_t - r_t)$ and the imbalance penalty charged to the WPP is $\Xi_t = \Xi(y_t, p_t)$. We assume $\Xi(\cdot)$ to be a known, deterministic, time-invariant function of the imbalance and the energy price. The profit Π_t at period t is thus given by (3.1).

$$\Pi_t = p_t q_t - \Xi(y_t, p_t) \tag{3.1}$$

For the case with constant prices, p_t is constant and $\Xi(\cdot)$ is defined by (3.2). Note that the negative sign in front of κ' implies that the WPP is not penalized for excess injection, but, on the contrary, sells it on the forward market. To avoid arbitrage opportunities, with $\kappa', \kappa \geq 0$, we must have $\kappa' < 1$ and $\kappa > 1$.

$$\Xi(y_t, p_t) = \begin{cases} \Xi^+(y_t, p_t) = -\kappa' p_t y_t & \text{if } y_t \geq 0 \\ \Xi^-(y_t, p_t) = \kappa p_t y_t & \text{if } y_t < 0 \end{cases}. \tag{3.2}$$

The capacity market

The WPP is also assumed to be a price taker in the capacity market. We define here capacity payments for the case with constant prices, i.e. $q_t = q$, and note that this definition would need to be extended to handle variable prices. The WPP is remunerated periodically (for simplicity, we assume hourly) for the capacity committed for sale in forward contracts at a price p^{cap} (in \$/(MW·h)). This committed capacity is derated by the historic probability of not honoring the contract, which in our model corresponds to the steady-state probability of being in shortfall. The capacity payment Π^{cap} is thus given by (3.3)

$$\Pi^{cap} = p^{cap} \cdot q \left(1 - \lim_{t \to \infty} \mathbb{P}\left(y_t < 0\right)\right) \tag{3.3}$$

The BESS

Let ρ_c and ρ_d be the conversion efficiencies of charge and discharge respectively. Thus, the round-trip efficiency is $\rho = \rho_c \rho_d$. The quantity r_t is measured from the exterior of the BESS, so the energy effectively injected to the battery is $\rho_c r_t$ (for $r_t > 0$) and the energy effectively extracted from the battery is $-r_t/\rho_d$ (for $r_t < 0$).

The capacity (size) of the battery is b (in MWh) and its power inversion capacity, i.e. the maximum rate at which energy can be charged to or discharged from the battery, is g (in MW). For simplicity, we assume the limit is the same for charging and discharging, but the method can easily be extended to consider the asymmetric case. In order to determine an

optimal size, we need to model the cost of installing and operating the battery. We break it down into a cost for storage capacity and a cost for inversion capacity. Both are expressed as linear, amortized costs and are denoted by c_s in \$/(MWh·h) and c_p in \$/(MW·h) respectively. By doing this, we assume that the battery is replaced at the end of its lifetime with the same capacity and at the same cost.

An important factor in BESSs is energy dissipation, i.e. the proportion of energy stored in the BESS that is spontaneously lost without any charging or discharging performed. This is usually expressed as a fraction η of stored energy per unit time, which is the representation adopted in our model, with some limitations described in section 3.3.2.

Objective

In the previous paragraphs, we have referred to t as a period for ease of exposition, given its similarity with standard electricity markets. However, we propose here a continuous-time model, so t actually, and in all instances in the remainder of this document, refers to an instant, and, consequently, the quantities q_t, w_t, r_t represent power levels at time t (in MW). Note that the price p_t is indeed in \$/MWh, so that our profit Π_t is an instantaneous profit rate at time t, in \$/h.

We are interested in the long-run average profit Π, as defined in (3.4). The expectation is taken with respect to the stochastic process of interest here, (w_t, p_t), as described next.

$$\begin{aligned}\Pi &= \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T (\Pi_t + \Pi^{cap})\, dt\right] - c_s b - c_p g \\ &= \Pi^{cap} + \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T \Pi_t dt\right] - c_s b - c_p g \end{aligned} \tag{3.4}$$

Sources of uncertainty

We consider two sources of uncertainty in our model: wind power output and energy prices. This is done by considering (w_t, p_t) as a joint continuous-time stochastic process. In particular, we model it as a Continuous-Time Markov Chain (CTMC) with a discrete state space $\mathcal{S} = \mathcal{W} \times \mathcal{P}$, with $\mathcal{W}$ and $\mathcal{P}$ being the discrete state spaces of w_t and p_t respectively. This stylized model allows for a convenient formulation of limiting distributions, as described in the next subsection.

Optimization model

Putting the previous pieces together, we are interested in solving the infinite-horizon average profit optimization problem in (3.5).

$$\begin{aligned} \max_{q_t, r_t, b, g} \quad & \Pi \\ \text{s.t.} \quad & b, g \geq 0, \\ & q_t \in [0, W] \; \forall t, \\ & q_t, r_t, b \in \mathbb{R} \; \forall t \end{aligned} \tag{3.5}$$

W is the plant capacity (in MW). Furthermore, we intend here to model an electricity market, so we will restrict our analysis to stationary policies where the bid is a function of the price, i.e. $q_t = q(p_t)$. We can also write this as $q_t = q^s$ when $p_t = p^s$, for $s \in \mathcal{S}$.

Finally, we will also be interested in stationary charge-discharge policies $r_t = r(w_t, q_t, p_t)$, which we write $r_t = r(w_t, q_t, p_t) = r^s$ when $(w_t, q_t, p_t) = (w^s, q^s, p^s)$. We must note, however, that this definition must be overriden if the storage is empty or full as summarized below.

$$\begin{aligned} r^s > 0 \;\&\; \text{storage full} \;&\Rightarrow r_t = 0 \\ r^s < 0 \;\&\; \text{storage empty} \;&\Rightarrow r_t = 0 \end{aligned}$$

In this chapter, we focus on cases where strategic storage of wind-generated energy is not attractive because prices and imbalance penalties are constant in time. For this case, it is known that the optimal (cost-minimizing) policy is a balancing policy, i.e. $r^s = w^s - q^s$ [39, 10]. It is not hard to see that this is still true in the presence of capacity payments, since any deviation will only increase the amount by which the committed capacity is derated and hence reduce capacity payments. Coupling this policy with the finite power inversion capacity gives (3.6).

$$r^s = \max\left(\min\left(w^s - q^s, g\right), -g\right) \tag{3.6}$$

3.3.2 Steady-state analysis

By the ergodicity of CTMCs, and since we are restricting our analysis to stationary policies, we can express the long-run average profit of (3.4) in terms of limiting distributions as in (3.7). In words, the long-run average profit is the sum over all states of the income minus the imbalance penalty, for which there are two cases: if storage is available, the policy can be followed; if storage is not available (empty or full), the policy must be overriden.

$$\begin{aligned}\Pi =&\Pi^{cap} + \sum_{s\in\mathcal{S}} \Big(p^s q^s \pi^s - \psi^s \Xi^s (q^s - w^s) \\ &- (\pi^s - \psi^s)\Xi^s(q^s + r^s - w^s)\Big) - c_s b - c_p g \\ =&\Pi^{cap} + \sum_{s\in\mathcal{S}} \Big(p^s q^s \pi^s - \psi^s \kappa p^s (q^s - w^s)^+ \\ &+ \psi^s \kappa' p^s (w^s - q^s)^+ - (\pi^s - \psi^s)\kappa p^s (q^s + r^s - w^s)^+ \\ &+ (\pi^s - \psi^s)\kappa' p^s (w^s - q^s - r^s)^+\Big) - c_s b - c_p g \end{aligned} \tag{3.7}$$

, where r^s is given by (3.6), $\pi = (\pi^s)_{s\in\mathcal{S}}$ is the limiting distribution of the CTMC (w_t, p_t) and $\boldsymbol{\psi} = (\psi^s)_{s\in\mathcal{S}}$ is the long-run probability of storage unavailability (empty or full). The second equality is obtained by plugging in the definition of the imbalance in (3.2).

We present first the case with uncapacitated power conversion ($c_p = 0$ so that g can be made as large as necessary to make $r^s = w^s - q^s$); the capacitated case is presented at the end of this section. In this uncapacitated case, since the optimal charge/discharge policy is balancing the output, $q^s + r^s - w^s = 0$ whenever storage is available, so that we can simplify (3.7) and obtain (3.8).

$$\begin{aligned}\Pi =&\Pi^{cap} + \sum_{s\in\mathcal{S}} \Big(p^s q^s \pi^s - \psi^s \kappa p^s (q^s - w^s)^+ \\ &+ \psi^s \kappa' p^s (w^s - q^s)^+\Big) - c_s b \end{aligned} \tag{3.8}$$

$\boldsymbol{\pi}$ can be easily determined from the generator of the CTMC (w_t, p_t). The long-run probability of unavailable storage $\boldsymbol{\psi}$ can be determined from some results of fluid queue theory, as shown next.

Limiting distribution

Our model corresponds to the model of a Markov-modulated fluid queue with finite buffer. A characterization of the long-run distribution of this process can be obtained through spectral analysis, which we overview next. This is based on the presentation in [52], with more details of the proofs available in [65]. For ease of exposition, in the following overview we omit the efficiency factor ρ.

Define $\mathbf{r} = [r^s]_{s\in\mathcal{S}}$, a vector with the discrete values taken by r_t, $\mathbf{D} = \text{diag}(\mathbf{r})$, a diagonal matrix with $\mathbf{r}$ in its diagonal. In the field of fluid queues, $\mathbf{r}$ is called the drift vector. We assume for now that $r^s \neq 0 \forall s \in \mathcal{S}$. The special case with $r^s = 0$ is considered at the end. Let $\mathbf{Q}$ be the infinitessimal generator matrix of the CTMC and $\mathbf{F}$ the limiting distribution of the level state of the battery, i.e.: $F(x,s) = \lim_{t\to\infty} \mathbb{P}\Big(X_t \leq x, (w_t, p_t) = s\Big)$, $\mathbf{F}(x) = [F(x,s)]_{s\in\mathcal{S}}$.

Then, it can be shown [52] that $\mathbf{F}$ satisfies the differential equation

$$\frac{d\mathbf{F}}{dx}\mathbf{D} = \mathbf{FQ} \tag{3.9}$$

with boundary conditions

$$F(0,s) = 0 \text{ if } r^s > 0$$
$$F(b,s) = \pi^s \text{ if } r^s < 0$$

A spectral solution to these equations can be obtained introducing generalized eigenvalues λ and eigenvectors $\mathbf{u}$, so that $\lambda\mathbf{uD} = \mathbf{uQ}$. The general solution to (3.9) takes then the form

$$\mathbf{F}(x) = \sum_{i=1}^{|\mathcal{S}|} a_i \exp(\lambda_i x)\mathbf{u}^i$$

, where the values of coefficients a_i can be found by solving a linear system from the boundary conditions. For brevity in later use, we denote $\psi_0^s = F(0,s)$, $\psi_f^s = F(b,s)$ and $\psi^s = \psi_0^s + \psi_f^s$.

In this manner, the long-run probability of unavailable storage can be calculated as a function of $\mathbf{r}$ and b. Note, however, that the method requires solving a generalized eigenvalue problem, which has three implications of importance for our work. First, we cannot obtain a closed-form expression of $\boldsymbol{\psi}_0$ or $\boldsymbol{\psi}_f$ in terms of b and $\mathbf{r}$, so that numerical calculating approaches are necessary. Secondly, this function is not convex in general, which makes our optimization problem possibly non-convex as well. Finally, the linear system posed by the boundary conditions can be very ill-conditioned because of the presence of both very large and very small eigenvalues, which is a major challenge for the method. We address this in more detail in the description of the algorithm.

Cases with zero drift If $r^s = 0$ for some state $s \in \mathcal{S}$, then $F(x,s)$ can be expressed as a linear combination of $F(x,z)$ for states $z : r^z \neq 0$. Thus, for these cases, the method described above is performed on a reduced system that includes only states $\{z : r^z \neq 0\}$. Then those values are used to find the distribution for the null states. Details are omitted here, but can be found in the appendix of [93] or in [47].

Dissipation

We next address the issue of modelling dissipation. As mentioned before, it is standard to consider that energy dissipates from the battery at a rate that is a fix multiple of the current storage level. This would mean that an additional term should be included in the drift vector: $r_t = r^s - \eta x_t$, where x_t is the current storage level in the battery (in MWh), and η is the proportion of energy stored lost per unit time (here, per hour). The addition of this term makes the drift dependent on the storage level, so that the spectral analysis performed earlier would no longer be valid.

A way to handle this within the framework of fluid queues is through multi-regime fluid queues. The key idea here is that the battery capacity is divided in bins (discretized), with drifts allowed to variate across different bins, but being level-independent within each bin. Introducing some notation, this can be expressed as in (3.10). Then, a system of ordinary differential equations like in (3.9) can be formulated for each bin, with additional boundary conditions for inter-bin boundaries.

$$\begin{aligned} r_t &= r(w_t, q_t, p_t, x_t) \\ &= r^{s,j} \text{ for } (w_t, q_t, p_t) = (w^s, q^s, p^s),\ b_{j-1} \leq x_t < b_j \end{aligned} \tag{3.10}$$

, where the BESS is divided into J bins with boundaries $0 = b_0 < b_1 < \cdots < b_{J-1} < b_J = b$. Note that this significantly increases the dimension of the eigenvalue problem, so that having a numerically stable method to find the limiting distribution becomes paramount for the success of this method. For brevity, we don't describe this modification of the approach in further detail and point the interested reader to [52] or [47].

The introduction of $\eta > 0$ could affect the optimality of the balancing policy. Indeed, for sufficiently large η, it could be more profitable to sell surplus energy immediately than to store it and have a large proportion of that lost to dissipation. To simplify our analysis, we make the following restriction: we make a sensitivity analysis on η with $\kappa' = 0$ and a sensitivity analysis on κ' with $\eta = 0$, so that the balance policy is optimal for all these situations.

3.3.3 Units

In the previous paragraphs, we defined all quantities in their appropriate physical units. However, it is more convenient and illustrative for the purposes of this work to express them in per unit of power plant capacity and storage capacity, by introducing:

$$\begin{array}{ccc} \mathbf{w} = \tilde{\mathbf{w}} W & q = \tilde{q} W & b = \tilde{b} W \\ g = \tilde{g} W & \mathbf{r} = \tilde{\mathbf{r}} b & c_s = \tilde{c}_s p^{max} \\ p = \tilde{p} p^{max} & p^{cap} = \tilde{p}^{cap} p^{max} & c_p = \tilde{c}_p p^{max} \end{array}$$

, with $p^{max} = \max \mathcal{P}$, so that $\tilde{\mathbf{w}}, \tilde{q}, \tilde{p}, \tilde{g} \in [0, 1]$, $\tilde{b}$ is in hours of storage of full plant capacity, and $\tilde{\mathbf{r}}$ is in units of [p.u. of b]/h, which is interpreted as the number of times that the total storage would be charged starting from an empty state in one hour at full plant capacity. Equation (3.8) becomes:

$$\frac{\Pi}{p^{max} W} = \frac{\Pi^{cap}}{p^{max} W} + \sum_{s \in \mathcal{S}} \left(\tilde{p}^s \tilde{q}^s \pi^s - \psi^s \kappa \tilde{p}^s (\tilde{q}^s - \tilde{w}^s)^+ + \psi^s \kappa' \tilde{p}^s (\tilde{w}^s - \tilde{q}^s)^+ \right) - \tilde{c}_s \tilde{b} \tag{3.11}$$

It is not hard to check that this change of units does not affect the spectral decomposition and hence the values of $\boldsymbol{\psi}$. In this chapter we are interested in the case where the price is constant, i.e. $|\mathcal{P}| = 1$, $p^s = p^{max}\ \forall s \in \mathcal{S}$ and hence $\tilde{p}^s = 1$ for all s. Finally, for readability and ease of notation, the tildes will be omitted in the remainder of this chapter, but we will always refer to values in per unit.

With these unit changes and simplifications, the optimization model (3.5) for the uncapacitated case with constant prices is written as in (3.12).

$$\begin{aligned} \max_{q,\,b} \quad & \sum_{s\in\mathcal{S}} \left(q\pi^s - \psi^s \kappa (q - w^s)^+ + \psi^s \kappa' (w^s - q)^+ \right) \\ & + p^{cap} q \left(1 - \sum_{s\in\mathcal{S}} \psi_0^s \right) - c_s b \\ \text{s.t.} \quad & b \geq 0, \\ & q \in [0,1] \end{aligned} \tag{3.12}$$

3.3.4 Solution algorithm

There are two challenges in finding the optimal solution to problem (3.12). The first difficulty lies in evaluating the objective function. The spectral method described in Section 3.3.2 requires solving a linear system with coefficients that stem from both very large and very small exponential terms, which makes the system very ill-conditioned and the method numerically unstable.

This was overcome thanks to the algorithm proposed by [47], which uses a stable matrix decomposition to obtain the limiting distribution of fluid queues without explicitly finding eigenvalues or relying on exponentially growing terms and results on enhanced numerical stability. This method allows multi-regime fluid queues like the one in our model when considering non-zero dissipation and does not require any special structure in generator or drift.

The second difficulty is that we have a non-linear, possibly non-convex objective function, so that using a gradient descent algorithm does not provide a guarantee of global optimality. For the scenario with constant prices, our search space only has a dimension of 2; it is thus reasonable to start exploring the search space by means of a grid, to then use the best candidate as starting point for a finite-difference descent algorithm, which is the algorithm that was implemented in our tests.

3.3.5 Capacitated power conversion

We finally address the case with capacitated power conversion. In the capacitated case with symmetric capacities, $r^s = \max\left(\min\left(w^s - q^s, g\right), -g\right)$, so $q^s + r^s - w^s$ is not always 0 and (3.8) does not hold in general. We must therefore use the expression for Π given by (3.7). The computation of $\boldsymbol{\psi}$ is only affected by this change to the extent that the entries in $\mathbf{r}$

change, but the method itself is still valid, so the only consequence of this capacitation is having two more terms in the objective function and one extra decision variable, as well as changing the expression of the capacity payments, which would be now given by (3.13). We can, however, show with help of Theorem 3.3.1 that these changes do not make the optimization significantly more difficult.

$$\Pi^{cap} = p^{cap} q \sum_{s \in \mathcal{S}: w^s - r^s - q \geq 0} (\pi^s - \psi_0^s) \tag{3.13}$$

Theorem 3.3.1. *Let (q^*, b^*, g^*) be an optimal solution to optimization problem (3.5). If $c_p > 0$, then $g^* = \left|w^{s^*} - q^{s^*}\right|$ for some $s^* \in \mathcal{S}$ or $g^* < \min_s |w^s - q^s|$.*

Proof. If $g^* \leq \min_s |w^s - q^s|$, the statement clearly holds, so we only need to prove that if $g^* > \min_s |w^s - q^s|$, then $g^* = \left|w^{s^*} - q^{s^*}\right|$ for some $s^* \in \mathcal{S}$. We prove by contradiction.

Suppose $g^* > \min_s |w^s - q^s|$ but the statement does not hold. To create our contradiction, let $\tilde{s} \in \operatorname{argmax}_s \{|w^s - q^s| : |w^s - q^s| < g^*\}$ and let $\tilde{g} = \left|w^{\tilde{s}} - q^{\tilde{s}}\right|$. We want to show that the point $(q^*, b^*, \tilde{g})$ achieves a higher profit than (q^*, b^*, g^*), which would contradict its optimality. First, note that by construction of $\tilde{g}$, for any $s \in \mathcal{S}$, $\max\left(\min\left(w^s - q^s, \tilde{g}\right), -\tilde{g}\right) = \max\left(\min\left(w^s - q^s, g^*\right), -g^*\right)$. This implies that the charge/discharge vectors in both cases are identical: $r(q^*, b^*, \tilde{g}) = r(q^*, b^*, g^*)$. Since $\boldsymbol{\psi}$ is a function of b and the drift vector, $\boldsymbol{\psi}(b^*, \tilde{r}) = \boldsymbol{\psi}(b^*, r^*)$. Thus, the terms inside the summation in (3.7) are identical for $\Pi(q^*, b^*, g^*)$ and $\Pi(q^*, b^*, \tilde{g})$. And hence, $\Pi(q^*, b^*, g^*) - \Pi(q^*, b^*, \tilde{g}) = c_p\left(\tilde{g} - g^*\right) < 0$, which concludes the proof. □

The takeaway from Theorem 3.3.1 is that we do not need to consider the entire feasible set spanned by g when optimizing (3.5), as only a handful candidate solutions need to be tried. We can thus define $\overline{\Pi}$ as in (3.14), and optimize for $\overline{\Pi}$ over $\mathbf{q}$ and b only.

$$\overline{\Pi}(\mathbf{q}, b) = \max_{s \in \mathcal{S}} \{(3.6) : g = |w^s - q^s|\} \tag{3.14}$$

For the case with constant prices, this takes the form of (3.15).

$$\begin{aligned} \max_{q,\, b} \quad & \overline{\Pi}(q, b) \\ \text{s.t.} \quad & b \geq 0, \\ & q \in [0, 1] \end{aligned} \tag{3.15}$$

To close this section, we make two comments about this method. First, each evaluation of $\overline{\Pi}$ requires $|\mathcal{S}|$ calls to the routine that calculates $\boldsymbol{\psi}$. Thanks to the algorithm in [47], this routine is very efficient, so this calls are not very costly computationally. They can also be easily parallelized if several cores are available. Moreover, for the case with constant prices, $\mathcal{S}$ is the discretized state space of wind power outputs, which we expect would have a cardinality in the order of tens in most applications [79], keeping the computational cost of each evaluation of $\overline{\Pi}$ manageable. Finally, our tests suggest that when sorting $|w^s - q^s|$

in ascending order, expression (3.7) is concave in g. If this can be verified, it can further reduce the number of calls necessary to evaluate $\overline{\Pi}$, which could be of value in extensions to variable prices, where $|\mathcal{S}|$ may be larger.

And secondly, we note that the method outlined above only finds an optimal solution to (3.5) if the condition in Theorem 3.3.1 holds, i.e. if $g^* \geq \min_s |w^s - q^s|$. We make two observations to argue, omitting some mathematical details, that this would not be an issue in practical applications. Suppose the condition is not true. Then, the optimal solution has $g^* < \min_s |w^s - q^s|$, which implies $r^s = -g$ if $w^s - q^s < 0$ and $r^s = g$ if $w^s - q^s > 0$. Note also that g^* is upper-bounded by the mesh size of $\mathcal{W}$, so it is a relatively small value. This in turn implies that $q^s - w^s$ and $q^s + r^s - w^s$ are close to each other, which can also be said of the penalties $\Xi(q^s - w^s)$ and $\Xi(q^s + r^s - w^s)$. The summation term in (3.7) for this point would therefore be similar to that of the no-storage case. Since the case with storage must also take into account the cost terms, this can only be possible if both b^* and g^* are quite small. In other words, our method may fail to find the optimal solution in a case where the optimal solution would be to build a storage of positive, but very small size in both energy capacity and power conversion capacity. Such a case can be constructed, but would not be of practical interest.

3.4 Results and discussion

3.4.1 Model data estimation

Wind model

To obtain the values of the generator matrix that defines the CTMC, the methodology outlined in [79] is followed. The methodology consists in first passing the wind power output through an averaging window of one hour, then discretizing the output with $N = 15$ levels, and finally performing a max-likelihood estimation on the resulting data sequence, i.e. finding the transition probabilities from counting transitions in the sequence. The number of states chosen is identified in [79] as a level of resolution that captures well the autocorrelation of the wind power series used as data source without the need for introducing a second-order Markov model. As in [18], the wind data available in [76] was used.

Reference parameters

We perform sensitivity analyses on the other model parameters. Reference values are given in Table 3.4.1. We comment on the choice of the reference cost of storage in the following paragraphs.

Table 3.4.1: Reference values for sensitivity analyses

Parameter	Value	Parameter	Value	Parameter	Value
κ	1.35	c_s	0.005	η	0
κ'	0	c_p	5E-4	ρ_d	0.95
p^{cap}	0			ρ_c	0.95

Empirical validation

To provide an empirical reference for comparison with our model, we use the same data series to solve the ex-post sizing and contracting optimization problem (3.16).

$$\begin{aligned} \max_{q,b,g} \quad & h(q,b,g) \\ \text{s.t.} \quad & b,g \geq 0, \\ & q \in [0,1] \end{aligned} \tag{3.16}$$

, where $h(q,b,g)$ is a function that computes the ex-post average profit over the time available in the data series for contract quantity q and storage size b. The value of function g can be computed easily using the fact that the the balancing policy is known to be optimal. A detailed description of the algorithm used to compute $h(q,b,g)$ is provided as an appendix.

The results of the model (3.12) and (3.16) are shown in Figure 3.4.1 and discussed below. In terms of model validation, it is worth mentioning that there is a good level of agreement between the curves given by the model and the empirical ex-post best. An interesting continuation of this work would be to perform a similar analysis using a non-homogeneous Markov chain to model the wind power output as in [91, 114].

3.4.2 Value of storage

As a first step before doing sensitivity analyses, we are interested in observing the value of storage for the WPP in the setting described. To do this, we fix b and g at different values and solve (3.12) and (3.16) for q only to find the optimal commitment and corresponding profit. The profit is compared to the profit that would be obtained from a feed-in-tariff contract with the same price. For the case without capacity payments, this profit is an upper-bound reference for comparison.

We can make four important observations from the results in Figure 3.4.1. First, without capacity payments, as expected, even with unlimited storage, recovering feed-in-tariff profits is not possible, which is explained by efficiency losses in storage discharge. If there is access to capacity payments, the additional profits are higher, but the overall shape of the value curve remains unchanged. Secondly, we note that the optimal bidding function does not have the form of an optimal fractile of wind power production here. Observe that since the support $\mathcal{W}$ is discrete, such a function would be step-shaped, while our optimal curve is not.

Third, in line with the results of other authors under different structural assumptions [39, 10], the value of storage in the reference case studied has diminishing marginal profit. This allows identifying two quantities of interest. First, the marginal profit at $b = 0$ gives the critical amortized cost of storage: if storage costs more than this quantity, it is not worth having co-located storage. The second quantity is the optimal size; for an amortized cost of storage c_s, the optimal size of storage is the quantity for which the marginal profit is equal to c_s. These results justify our choice of a descent algorithm to find the optimal storage size. These observations also seem to hold for the value of power conversion.

Finally, as anticipated, expected imbalances are reduced as storage sizes grows without limit. It is worth observing, however, that positive imbalances increase for the lowest range of storage sizes. This can be explained by the fact that in our model for the given reference values, the optimal contracting strategy in absence of storage is to commit the entire capacity of the plant, which eliminates positive imbalances. As storage drives the optimal commitment down, positive imbalances appear. With enough storage, these imbalances are then brought down too.

For the reference case shown in Fig. 3.4.1, the critical cost is found at $c_o = 0.0193$, which for a reference price of energy of 60\$/MWh, corresponds to an amortized cost of 10.14\$/kWh-yr. Note that this is a whole order of magnitude below the reference cost of 100\$/kWh-yr for 2020 [66], meaning that under this setting and the reference values used, installing co-located storage is not profitable. Access to capacity payments does not seem to significantly change this conclusion (Fig. 3.4.1, bottom-right). Even taking into account that the cost of storage is declining rapidly, a tenfold reduction is beyond what we should expect to see in the near future. So according to these results, a different setting would need to be in place for co-located storage to be attractive.

In fact, in our setting, the WPP is not sufficiently exposed to uncertainty to justify paying for expensive storage. Indeed, the WPP is already covered against prices higher than κp by the pay-as-demanded forward contract, so that further risk hedging via storage is only modestly attractive. In Fig. 3.4.1 (bottom-right), we see that as the coverage cap κ is increased and the WPP is exposed to higher penalties, the critical cost c_o at which it is willing to invest in storage also rises. However, the pay-as-demanded forward contract would need to be available at a premium of several times the WPP's normal selling price to get anywhere near the actual current critical cost.

Figure 3.4.2 depicts the value of storage considering non-zero energy dissipation η. Two observations are worth highlighting here. First, considering a non-zero value of losses to dissipation has an important effect of reducing the value of storage. For a reference value of $\eta = 0.2\%/h$, this reduction can be as high as 50%. This shows the importance of considering this effect when evaluating collocated storage. Results suggest that the overall shape of the value curve remains unchanged, i.e. we have decreasing marginal value of storage, with an upper bound being met asymptotically for very large values of storage. This asymptotic behavior can only be observed when the refinement of the discretization of storage is increased.

The second set of observations has to do with the limitations of the model used here

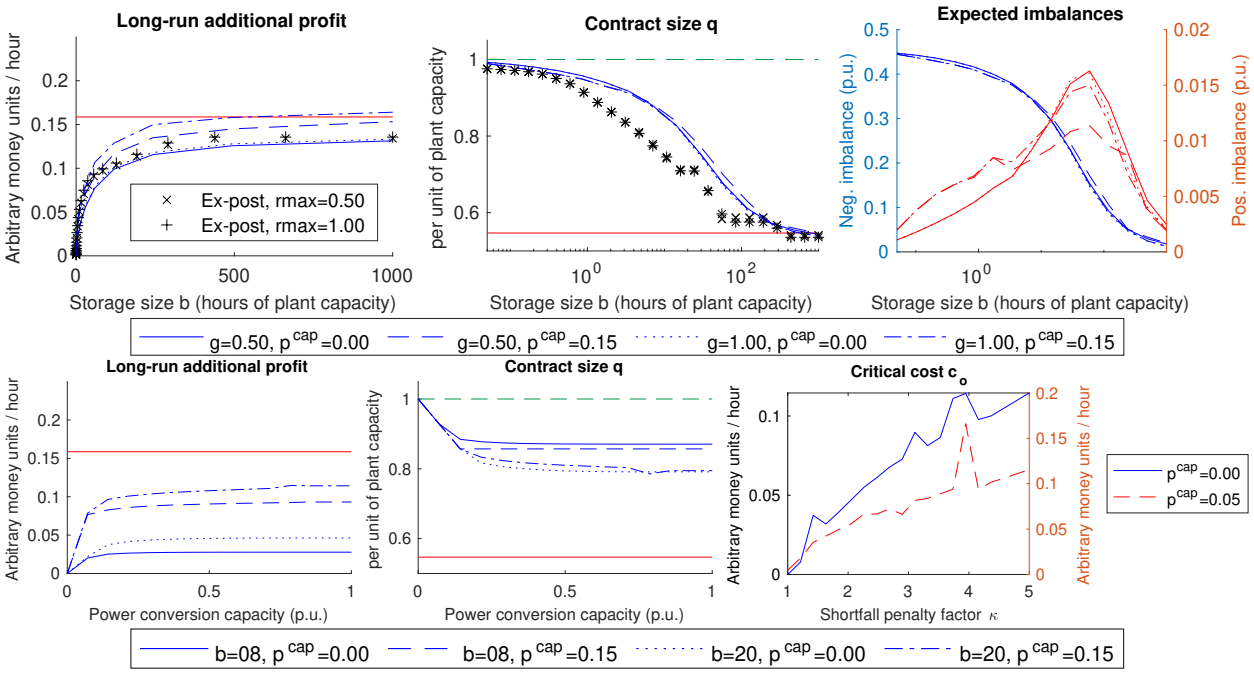

Figure 3.4.1: Plots of value of storage obtained solving the contract size problem for fixed storage capacity. *Top-left*: Long-run average profit in excess of no-storage profit for different storage sizes. The feed-in-tariff benchmark (red) is equal to the average output of the power plant (because price is normalized to 1). *Top-center*: Optimal energy commitment for different storage sizes compared to no-storage and average plant output benchmarks. *Top-right*: Expected positive and negative imbalance for different storage sizes. *Bottom-left*: Long-run average profit in excess of no-storage profit for different power inversion capacities. *Bottom-left*: Long-run average profit in excess of no-storage profit for different power inversion capacities. Feed-in-tariff benchmark in red. *Bottom-center*: Optimal energy commitment for different power inversion capacities compared to no-storage and average plant output benchmarks. *Bottom-right*: Critical storage cost for different values of shortfall penalty factor and capacity payment. Critical storage cost is the slope of the profit curve vs. storage size at $b = 0$.

and the effect of the granularity of the discretization of storage levels when considering dissipation. For very large values of storage, the model assumption of considering dissipation level-independent within each bin is too coarse and leads to artificial reductions in the value of storage. As expected, this effect is mitigated by increasing the number of bins in the discretization (at the cost of more computational burden).

For values of storage in the order of less than ten hours, it seems like a level-independent model is sufficiently accurate. For storage in the order of 10 to 100 hours, it seems necessary to include a multi-regime formulation with at least 10 to 20 bins to accurately model the value of storage. This implies a state space with size in the hundreds of states, which highlights the importance of having a numerically stable algorithm in this state range.

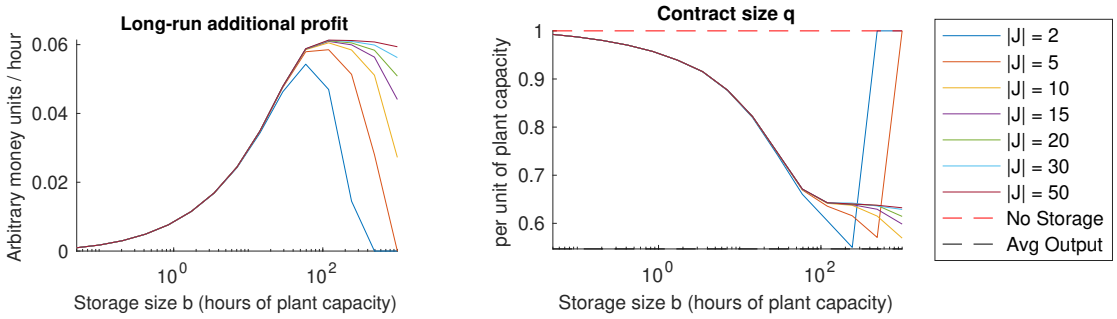

Figure 3.4.2: Value of storage and optimal contract size for $\eta = 0.2\%/\text{h}$ with multi-regime fluid queue model, for different number of regimes $|J|$. *Left*: Long-run average profit in excess of no-storage profit for different storage sizes. *Right*: Optimal commitment for different storage sizes compared to no-storage and average plant output benchmarks.

3.4.3 Sensitivity analyses

The main scope of the model proposed here is performing sensitivity analyses at a high level, such as during a project feasibility study or for policy evaluation. In this section, we show the results of some of these analyses. Despite the previous discussion regarding current costs, sensitivity analyses to changes in some key parameters are of interest, since they provide understanding regarding the interplay between some key quantities.

We are interested in observing the behavior of the model as key parameters are modified. The current cost of storage, $c_s \approx 0.2$ is above the critical cost for the reference case, so the optimal size would be 0, and the behavior of the results to changes in other parameters would be hidden. To avoid this, we take a reference storage cost of $c_s = 0.005$. Although unrealistically low, it allows observing the behavior of the optimal size as some parameters of interest change. Sensitivity curves are shown in Figure 3.6.2 and are commented in the following paragraphs.

Forward contract prices

The sensitivity to changes in shortfall penalty is aligned with intuition: larger storage becomes more attractive as the penalty for energy shortfall grows, i.e. as the exposure to high penalties increases. On the other hand, the change with respect to the price of the pay-as-generated forward contract shows a more interesting dynamic. As κ' is raised and the value of surplus energy increases, it becomes less undesirable to have excess energy, so that the optimal commitment q decreases. This leaves room for increasing the optimal storage size, as more energy is available for accumulation. After a certain point, however, the price of the pay-as-generated contract is so close to the price of the long-term fixed contract that the latter becomes less attractive, so that the commitment drops. With a very low commitment in the long-term contract, the motivation for investing in storage also disappears progressively, leading to a smaller optimal storage size.

Capacity payments

Access to capacity payments provides an important incentive for storage deployment, for both storage and powe¡r inversion capacities.

Efficiency and dissipation

The behavior when the round-trip efficiency and dissipation vary is in line with our expectation. As the system becomes more inefficient, storage becomes less attractive. For discharge efficiency parameter ρ_d, the relation between optimal size and efficiency appears to be close to linear. The sensitivity to the charge efficiency ρ_c is very similar and is omitted. Again, the curves highlight the importance of including dissipation in our model, with optimal storage capacity being most sensitive to this parameter for values near 0.

3.5 Extension to variable prices

The model introduced in 3.3 allows for a general Markovian process where both prices and wind power production are stochastic. However, throughout this chapter, we have focused on the case where prices are constant. A complete incorporation of the varable-price case into the model is out of the scope of this thesis. However, in this section, we discuss some challenges to our solution approach when this assumption is relaxed, briefly overview how they could be overcome, and present some preliminary results in that avenue. Although the work presented in this section is in a relatively early stage of maturity, we include here as it may prove helpful to a reader interested in extending our work to handle variable prices.

3.5.1 Challenges

There are two main challenges to extending the methodology presented in this chapter to a case with variable prices. The first one is one of validity: while assuming that the stochastic process of wind power output has a Markovian behavior may be an acceptable approximation for a high-level analysis, as we presented before, it is less clear that this assumption can be justified for the joint process of wind power output and prices. Some validation work based on historical data should be performed to guarantee the soundness of that approach. Assuming the validity of that assumption, a technical challenge remains: the balancing policy is no longer optimal in general in the variable-price case. We overview next how this challenge could be overcome leveraging known literature results.

3.5.2 Potential solution strategy

Consider the extension of the optimization problem of (3.5) to the case with variable prices:

$$\begin{array}{ll} \max\limits_{q,r,b,g} & \Pi \\ \text{s.t.} & b, g \geq 0 \end{array} \tag{3.17}$$

, where b and g are scalars, as in the case with constant prices, but q and r are now general functions of the state of the system. In a case with variable prices, a conceivable q would take the form of a supply function, so we would have $q = q(p_t)$, and if we restrict ourselves to stationary policies under a discrete state space, q is a vector with same size as the state space of the stochastic process p_t. To reduce the search space for r, we leverage the results of [39], where it is shown that the optimal charging policy in the discrete time case, when surplus energy is curtailed at no cost and no penalty (i.e. $\kappa' = 0$ in our model), is a double-threshold policy.

Double-threshold policy The double-threshold policy is described in equation form in (3.18) and graphically for the iid[2] case in Fig. 3.5.1. In words, the charging policy is battery-level dependent. If the level x_t is below a threshold h^-, then the optimal policy dictates that the battery should be charged back to h^-, regardless of the imbalance status. If the level is above a threshold h^+, the battery should be used for balancing, i.e. store any surplus generation and compensate any shortfall in generation by discharging from the battery as necessary, going no further than h^+. If the level is between both thresholds, any excess should be stored, but no power should be extracted from the battery. The values of the thresholds h^- and h^+ are functions of the state of the system, i.e. functions of the price p_t and the imbalance $q_t - w_t$.

$$r_t = \begin{cases} h^- - x_t & \text{if } x_t < h^- \\ (w_t - q_t)^+ & \text{if } h^- \leq x_t \leq h^+ \\ w_t - q_t & \text{if } x_t > h^+ \end{cases} \tag{3.18}$$

For simplicity, the version given in (3.18) is a simplified version where we assume perfect round-trip efficiency of the charge/discharge cycle, and ignore ramping limitations on the charge and discharge and energy dissipation. The results of [39] consider these cases as well. We note here three additional results of [39] that are of relevance for our work. First, if the round-trip efficiency of the charge/discharge cycle is perfect, $h^- = h^+$. Next, for states where p_t takes its maximum value (recall we assume a finite discrete state space), $h^+ = h^- = 0$. And finally, for the case where the stochastic process $(q_t - w_t, p_t)$ is iid, the thresholds depend only on the price p_t, not on the imbalance.

Conjecturing that the optimal policy in the continuous time case also has a double-threshold structure, optimization problem (3.17) takes the form of (3.19).

[2] iid=independent and identically distributed

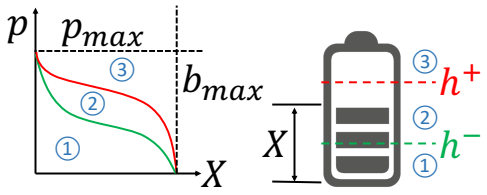

Figure 3.5.1: Graphical description of double-threshold policy for the case where thresholds only depend on the price process (not the imbalance). ①: stock up to h^-, ②: Store excess only, ③: balancing policy. Observe the thresholds h^- and h^+ depend on the price p_t.

$$\begin{array}{ll} \max & \Pi \\ q, h^+, h^-, b, g & \\ \text{s.t.} & b, g \geq 0 \end{array} \tag{3.19}$$

, where h^+ and h^- are functions of the state of the system, so they can be represented by vectors with same size as the state space. Moreover, if the optimal thresholds depended only on prices, the size of the variables h^+ and h^- would only be that of the space state of p_t, which considerably reduces the search space.

3.5.3 Preliminary results

We put this conjecture to the test, i.e., whether the optimal charge/discharge policy has a double-threshold structure, and if so, whether the thresholds depend on prices only or also on the imbalance, by conducting an experiment on a small test instance, which we describe in this section.

Consider the continuous-time setting described in this chapter, where the commitment $q_t = q$ is constant and known, and the processes w_t and p_t are each an independent CTMC with discrete, finite state-space $\mathcal{W}$ and $\mathcal{P}$ respectively, so that the joint stochastic process (w_t, p_t) is a CTMC with discrete finite state space $\mathcal{S} = \mathcal{W} \times \mathcal{P}$. We next consider a discrete-time problem by sampling from (w_t, p_t) periodically at intervals of length Δt.

From standard results in Markov chains, we know that if (w_t, p_t) has generator $\mathbf{Q}$, this sampling procedure results in a discrete-time stochastic process (w_{t_k}, p_{t_k}) with transition matrix $\mathbf{P}^{\Delta t} = \exp(\Delta t \mathbf{Q})$. If we assume that the wind power output and price remain constant during the period $[t_k, t_{k+1})$, the resulting problem falls exactly in the setting of [39]. Therefore, we know that the optimal policy for that case has a double-threshold form. If, as conjectured, the optimal policy in continuous time has a double-threshold form as well, we expect those thresholds to converge as $\Delta t \to 0$. We next make this idea more precise.

Let $\xi_t = (w_t, p_t)$, let $\xi_k^{\Delta t} = (w_{t_k}, p_{t_k})$, $k \in \mathbb{N}$, be the discrete-time stochastic process resulting from sampling from ξ_t at periodic intervals Δt. Let $h^{\Delta t}(w, p)$, for $(w, p) \in \mathcal{S}$ be the function that returns the optimal threshold for each state of the discrete system with sampling period Δt. Recall that since we are considering perfect round-trip efficiency, both

thresholds coincide, i.e. $h^+ = h^- = h$. We expect to obtain $\lim_{\Delta t \to 0} h^{\Delta t}(\cdot) = h^*(\cdot)$ for some h^*.

Next, we present the method used to obtain the optimal thresholds for each discretized problem. In [39], conventional dynamic programming techniques (LP formulation) are used to obtain the optimal values for the numerical examples solved, which involves discretizing the space of storage levels also as space $\mathcal{B}$. In this context, an action is how much energy to charge or discharge into (from) the BESS. Therefore, if the storage level is discretized, this automatically discretizes the action space as well. In fact, we can interpret the system dynamics in the two following equivalent forms.

- An action u_{t_k} is the energy (in MWh) that is injected into the BESS (which we interpret as extraction if < 0), so that the BESS level at the next time period is $x_{t_{k+1}} = x_{t_k} + u_{t_k}$
- An action μ_{t_k} is the BESS level at which the BESS will be at the next time period: $x_{t_{k+1}} = \mu_{t_k}$.

It is clear that these two definitions of an action are related by $\mu_{t_k} = x_{t_k} + u_{t_k}$. The second definition makes evident that discretizing the state space of x_{t_k} also discretizes the action state space. We will use both definitions to describe the policy iteration algorithm.

One important consideration in our discretization as Δt becomes smaller is maintaining precision. Let Δb be the granularity of the discretization of the storage capacity (in MWh). If we maintained this granularity constant, the energy imbalance $(w - q)\Delta t$ would become negligible compared to Δb as $\Delta t \to 0$, so our analysis would become meaningless, as results would trivially converge due to lack of precision. We thus need to adjust the granularity of the storage discretization. To avoid noise introduced by rounding, we choose Δb such that $(w - q)\Delta t$ is a multiple of Δb for all $w \in \mathcal{W}$. This significantly augments the size of our problem as $\Delta t \to 0$ and makes it harder to solve, but is a necessary step.

3.5.4 Policy iteration algorithm

We omit here many details and assume the reader is familiar with optimality in infinite-horizon average cost problems. For a deeper presentation of this topic, see [5]. To describe the policy iteration algorithm used, consider first the optimality equation (3.20) of the infinite-horizon average cost problem, with $v(\cdot)$ the relative cost function and λ the optimal average cost. This optimality equation is satisfied because our process satisfies the *Weak Accessibility* assumption of [5], as verified in the proof of Theorem 8 of [39].

$$v(x_{t_k}, w_{t_k}, p_{t_k}) + \lambda = \min_{r_t} \Big\{ p_{t_k} \big(q - (w_{t_k} - r_{t_k})\big)^+ + \mathbb{E}\left[v\big(x_{t_k} + r_{t_k}, w_{t_{k+1}}, p_{t_{k+1}}\big) \middle| w_{t_k}, p_{t_k} \right] \Big\} \tag{3.20}$$

Since our state space is finite, we can express this equation introducing a vector V to represent the relative cost function, a vector g_r to represent the cost incurred at the current step and a matrix $\mathbf{P}_r$ to represent the the transition matrix under policy r. Formally,

$$V = [v(x,w,p)]_{(x,w,p)\in\mathcal{B}\times\mathcal{S}}$$
$$g_r = [p(q-w+r(x,w,p))^+]_{(x,w,p)\in\mathcal{B}\times\mathcal{S}}$$
$$\mathbf{P}_r\big((x_1,s_1),(x_2,s_2)\big) = \mathbb{P}\Big((x_{t_{k+1}},s_{t_{k+1}}) = (x_2,s_2)\Big|(x_{t_k},s_{t_k}) = (x_1,s_1), r_{t_k} = r(x_1,s_1)\Big)$$

, so that the policy iteration algorithm can be implemented as in Algorithm 1. In words, we start from a balancing policy: $r(x,w,p) = w - q\ \forall (x,w,p) \in \mathcal{B}\times\mathcal{S}$. In the policy evaluation step, we construct the matrix $\mathbf{P}_r$ (which can be defined as a sparse matrix to enhance performance) and solve a linear system to update the values of V and λ. In the policy improvement step, for each $(x,w,p) \in \mathcal{B}\times\mathcal{S}$, the action that minimizes the right-hand-side in (3.20) is found by enumeration of the entire action space.

In the general form shown in Algorithm 1, the entire action space is searched looking for the optimal policy. As Δt gets smaller and $N_b = |\mathcal{B}|$ becomes larger, this enumeration of the action space becomes prohibitively time-consuming.

An alternative, in order to make possible testing smaller values of Δt is thus to replace the policy improvement step with one in which only threshold policies are considered. This is shown in Algorithm 2. Note that for this class of policies, once the optimal action for the first storage level $x = 0$ is determined for some (w,p), the action for all other storage levels x corresponding to the same (w,p) is determined as well: we have $\mu(x,w,p) = \min\{b_{max}, \max\{x+w-q, \mu(0,w,p)\}\}$. This allows saving one loop execution of length N_b, which is significant since N_b is the size of the largest space in our problem.

The size of the search space can be further reduced if thresholds depend only on prices, as shown in Algorithm 3. In this case, once the optimal action has been determined for $w = 0, x = 0$ for one value of p, it is just copied over for all (x,w) for the same value of p as $\mu(x,w,p) = \min\{b_{max}, \max\{x+0-q, \mu(0,0,p)\}\}$. Note that if thresholds that depend on prices only are optimal, the values of w and x chosen to perform the policy improvement step will not matter.

3.5.5 Numerical tests

The algorithms presented earlier were implemented in Matlab to run tests on a small system as a first approach to test our conjectures. Historical wind power output and price data were obtained from [76] and used to get a generator $\mathbf{Q}$ that represents (w_t, p_t), following the method described in [79]. To obtain a model of manageable size, the discretization of the state space was done with $N = |\mathcal{W}| = 3$, $N_p = |\mathcal{P}| = 2$. Separate generator matrices were first obtained for each process, and then combined into a larger generator matrix $\mathbf{Q}$ making sure to maintain independence.

All quantities were converted to per unit (p.u.) with respect to the wind power plant capacity. Storage capacity and level is expressed in hours of storage (at full plant capacity).

/* We start with balancing policy */
for $x \in \mathcal{B}$ **do**
  for $w \in \mathcal{W}$ **do**
    for $p \in \mathcal{P}$ **do**
      $u^0(x, w, p) \leftarrow w - q$
      $\mu^0(x, w, p) \leftarrow \min\{b_{max}, \max\{0, x + w - q\}\}$
    end
  end
end
$\ell \leftarrow 0$
repeat
  // Policy evaluation
  $\mathbf{P}_r \leftarrow \mathbf{0}_{NN_kN_b \times NN_kN_b}$ **for** $x \in \mathcal{B}$ **do**
    for $w \in \mathcal{W}$ **do**
      for $p \in \mathcal{P}$ **do**
        $\mathbf{P}_r\big((x, w, p), (\mu^\ell(x, w, p), :, :)\big) \leftarrow \mathbf{P}^{\Delta t}$
        $g_r\big((x, w, p)\big) \leftarrow p(q - w + u^\ell(x, w, p))^+$
      end
    end
  end
  /* Solve linear system to obtain V and λ */
  $\mathbf{A} \leftarrow [\mathbf{P}_r, -\mathbf{1}; 1, \mathbf{0}]$
  $b \leftarrow [-g_r; 0]$
  $\chi \leftarrow A^{-1} \cdot b$
  $V \leftarrow \chi(1 : \text{end} - 1)$
  $\lambda \leftarrow \chi(\text{end})$
  // Policy improvement
  for $x \in \mathcal{B}$ **do**
    for $w \in \mathcal{W}$ **do**
      for $p \in \mathcal{P}$ **do**
        $u^{\ell+1}(x, w, p) \leftarrow$
          $\arg\min_y \left\{ p(q - w + y)^+ \sum_{w' \in \mathcal{W}, p' \in \mathcal{P}} \mathbf{P}^{\Delta t}(w', p') \cdot V(x + y, w', p') \right\}$
        $\mu^{\ell+1}(x, w, p) \leftarrow u^{\ell+1}(x, w, p) + x$
      end
    end
  end
  $\ell \leftarrow \ell + 1$
until $\mu^{\ell+1} == \mu^\ell$ *or iteration limit*

Algorithm 1: General policy iteration algorithm used

```
// Policy improvement, threshold policies only
for w ∈ W do
    for p ∈ P do
        u^{ℓ+1}(0, w, p) ← arg min_y { p(q−w+y)^+ Σ_{w'∈W, p'∈P} P^{Δt}(w', p') · V(0+y, w', p') }
        μ^{ℓ+1}(0, w, p) ← u^{ℓ+1}(0, w, p) + x
        for x ∈ B do
            μ^{ℓ+1}(x, w, p) ← min{b_max, max{x + w − q, μ^{ℓ+1}(0, w, p)}}
            u^{ℓ+1}(x, w, p) ← μ^{ℓ+1}(x, w, p) − x
        end
    end
end
```

Algorithm 2: Policy improvement step searching only the space of threshold policies.

```
// Policy improvement, price-dependent threshold policies only
for p ∈ P do
    w ← 0
    u^{ℓ+1}(0, w, p) ← arg min_y { p(q − w + y)^+ Σ_{w'∈W, p'∈P} P^{Δt}(w', p') · V(0 + y, w', p') }
    μ^{ℓ+1}(0, w, p) ← u^{ℓ+1}(0, w, p) + x
    for x ∈ B do
        for w ∈ W do
            μ^{ℓ+1}(x, w, p) ← min{b_max, max{x + w − q, μ^{ℓ+1}(0, 0, p)}}
            u^{ℓ+1}(x, w, p) ← μ^{ℓ+1}(x, w, p) − x
        end
    end
end
```

Algorithm 3: Policy improvement step searching only the space of price-dependent threshold policies.

Prices are also expressed in per unit (normalized) with respect to the maximum price. The most important parameters are summarized in Table 3.5.1.

For the test system defined, the discrete-time problems were solved successively for decreasing values of Δt. The results are discussed next.

3.5.6 Results

As expected from the results in [39], even when searching the space of all possible policies, the optimal policy found has always a double-threshold form. Moreover, as predicted, we obtain $h^- = h^+$ and $h = 0$ for the states with maximum price (w, p), $p = p_{max} = 1$. The results of the thresholds obtained for system states with price $p = 0.208$, for the different

Parameter	Value
$N = \|\mathcal{W}\|$	3
$\mathcal{W}$	{0,0.5,1}
$N_p = \|\mathcal{P}\|$	2
$\mathcal{P}$	{0.22,1}
Δw	0.1 p.u.
b_{max}	2 hours (at 1 p.u. capacity)
q	0.4 p.u.

Table 3.5.1: Numerical test parameters

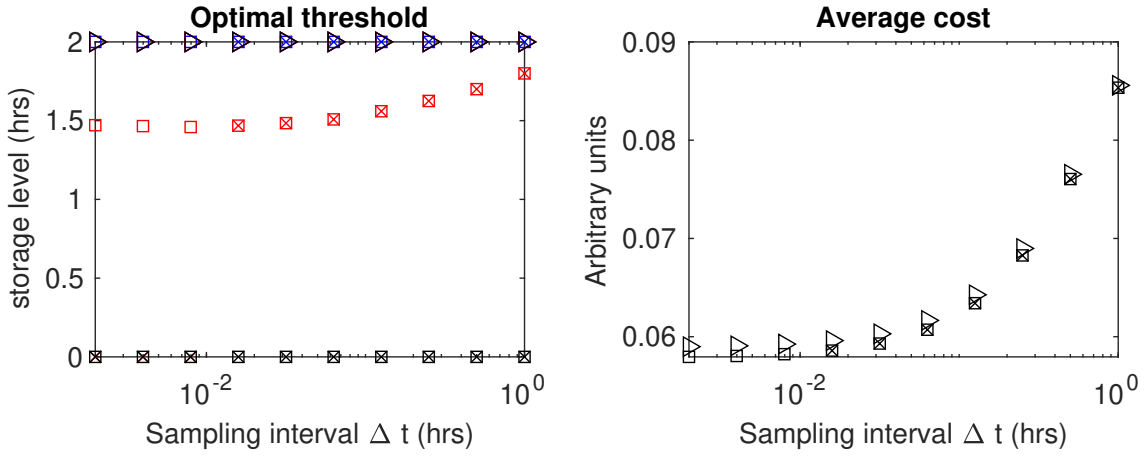

Figure 3.5.2: Optimal thresholds obtained for discrete-time model with decreasing sampling interval Δt. At the top, optimal threshold $h^+ = h^- = h(w,p)$ for each value of Δt. Colors represent different values of w: $w = 0$ in blue, $w = 0.5$ in red, $w = 1$ in black. All values correspond to $p = 0.22$. It was found $h(w,1) = 0 \forall w$ and is not included in the plot.
At the bottom, optimal average cost obtained, in arbitrary money units. For both plots, $\times$ represent values obtained when searching the entire space of policies (this was not possible within the time execution limit defined for the three smallest values of Δt tested). $\square$ represents values obtained when restricted to threshold policies, $\triangleright$ represents values obtained when restricted to price-dependent threshold policies.

values of Δt tested, are shown in Figure 3.5.2.

There are several insights to extract from these plots. First, contrary to what we hoped for, and although the price and wind power output processes are independent from each other, we find that the optimal thresholds depend on the value of the wind power output, i.e. they are different for different values of $w \in \mathcal{W}$. This implies,that restricting the space of policies to policies where the threshold depends only on the price excludes the optimal policy.

This is confirmed when looking at the optimal average cost, which is higher for the latter class of policies. It is worth noting, however, that the cost increases by less than 2% for all values of Δt tested. This suggests that although suboptimal, price-dependent policies might provide an acceptable heuristic.

Next, observe that it was not possible to run the policy iteration algorithm to completion when searching the entire space of policies for the three smallest values of Δt. However, the optimal solution found when restricting the search to threshold policies is the same as the one found when searching the entire policy space for all values of Δt for which both were computed, which suggests the optimal solutions found for the smallest values of Δt would also be globally optimal. However, note that we do not have a guarantee of this at this point, as it is pointed out in [39] that convexity of the problem could be lost when applying this restriction, and hence convergence to local optima is possible.

Finally, and most importantly, there seems to be a convergence of optimal thresholds as Δt is decreased. This is consistent with the conjecture of the optimal policy for the continuous-time case also having a double-threshold form. This is of course only one particular case and is hence far from being a proof, but is encouraging for pursuing this research avenue. We might be tempted to also suggest that the discretization and policy iteration path taken here could be a way to approximately compute the optimal thresholds of the continuous-time case. Although in principle this is correct, for a problem of a more realistic size, it seems unlikely that the discrete-time problem could be solved quickly enough for this approach to be appealing.

3.6 Conclusion and future research

In sum, we are proposing a model to perform a high-level steady-state analysis of the value of co-located storage for a wind power producer that participates in the electricity market through long-term forward contracts. In particular, we assess the optimal size of storage and optimal quantity to sell in forward contracts under different values of key parameters regarding contract prices and storage efficiency. We find that in a setting such as the one considered in this chapter the producer is not exposed to enough uncertainty in income for storage to be attractive at current prices. This is likely to be different with variable prices, which is the natural extension of our model.

In this avenue, we regard the result of the model with non-zero dissipation as very important because of its implications for the extension to a variable-price setting, which

are twofold. First, these results demonstrate the ability of the method to handle larger dimensions, which in our case appear as we refine the discretization of the storage levels. In the tests performed, we introduced as many as 750 states, which the algorithm of [47] could handle without issues. This is encouraging since introducing variable prices increases the dimension of the system in a similar way. Second, it is proved in [39] that the optimal charge/discharge policy when prices are variable is a dual-threshold policy, a policy in which the charge/discharge decision is level-dependent, which lends itself to be easily handled through the formulation proposed here. These two facts in conjunction allow regarding an extension of this method to variable prices as promising from a technical point of view. In [47], they find the algorithm to be numerically stable with a state size of more than 2000. Combining the factors discussed above, it would not be hard to reach a model of this size, which is why it is key for these cases to verify that the algorithm remains numerically stable as the size grows and to reduce as much as possible the number of states introduced, e.g. by using non-uniform bin sizes (with lower resolution at lower levels) and only the number of bins necessary to accurately depict storage sizes in a reasonable range using engineering criteria.

We close this work with a discussion of how we envision addressing the problem of battery aging and battery replacement through an extension of the model presented in this thesis.

3.6.1 Battery replacement

All the models covered here assume that battery capacity does not decay during its lifetime and that when it is replaced, it can be done at the same cost and size as the initial decision. All these assumptions can be lifted by considering an additional decision variable in the problem, namely the time of replacement of the battery. Figure 3.6.1 depicts the pattern that is obtained from considering a linear capacity decay with periodic replacement. As can be seen there, in order to get a system with an average storage capacity over its lifetime equal to a certain desired value (e.g. the constant value for which our current model designs), some oversizing of the initial installation is necessary, which implies some additional cost. There is a tradeoff between the magnitude of this oversizing and the replacement period T. More frequent replacement reduces the magnitude of oversizing necessary to maintain some average capacity, and hence the total installation cost, but increases the amortized cost, as it reduces the amortization period.

A first simple approximation to incorporate the effect of this phenomenon into our model could be an ad hoc post-processing stage where after finding the optimal battery size, a replacement period is chosen arbitrarily, and the battery oversize is selected to maintain the average size that was designed. The cost implications of this oversizing can then be computed and fed back to adjust the selected battery size. A more sophisticated approach, which could balance better this tradeoff, would be to include T as a decision variable in an extended version of our model.

Finally, we note that this same framework can be utilized to address one more relevant feature in the context of battery storage investment, which is that of falling costs due to

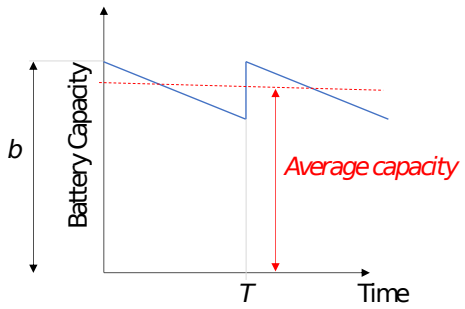

Figure 3.6.1: Pattern of battery capacity decay with linear decay and periodic replacement T. The initial capacity b must be oversized in order to obtain some desired average capacity.

technological advancements. In this case, the cost of replacing the battery at the end of the replacement period would be a decreasing function of time. An additional tradeoff would therefore appear, as a longer replacement period would require a larger initial oversizing, and would decrease the amortized cost as mentioned before, but would in addition allow for a cheaper installation cost for the replacement battery. It is conceivable that the abundant literature in machine replacement problem, which handles a case with many similarities to this, would be of great value to formulate this model extension.

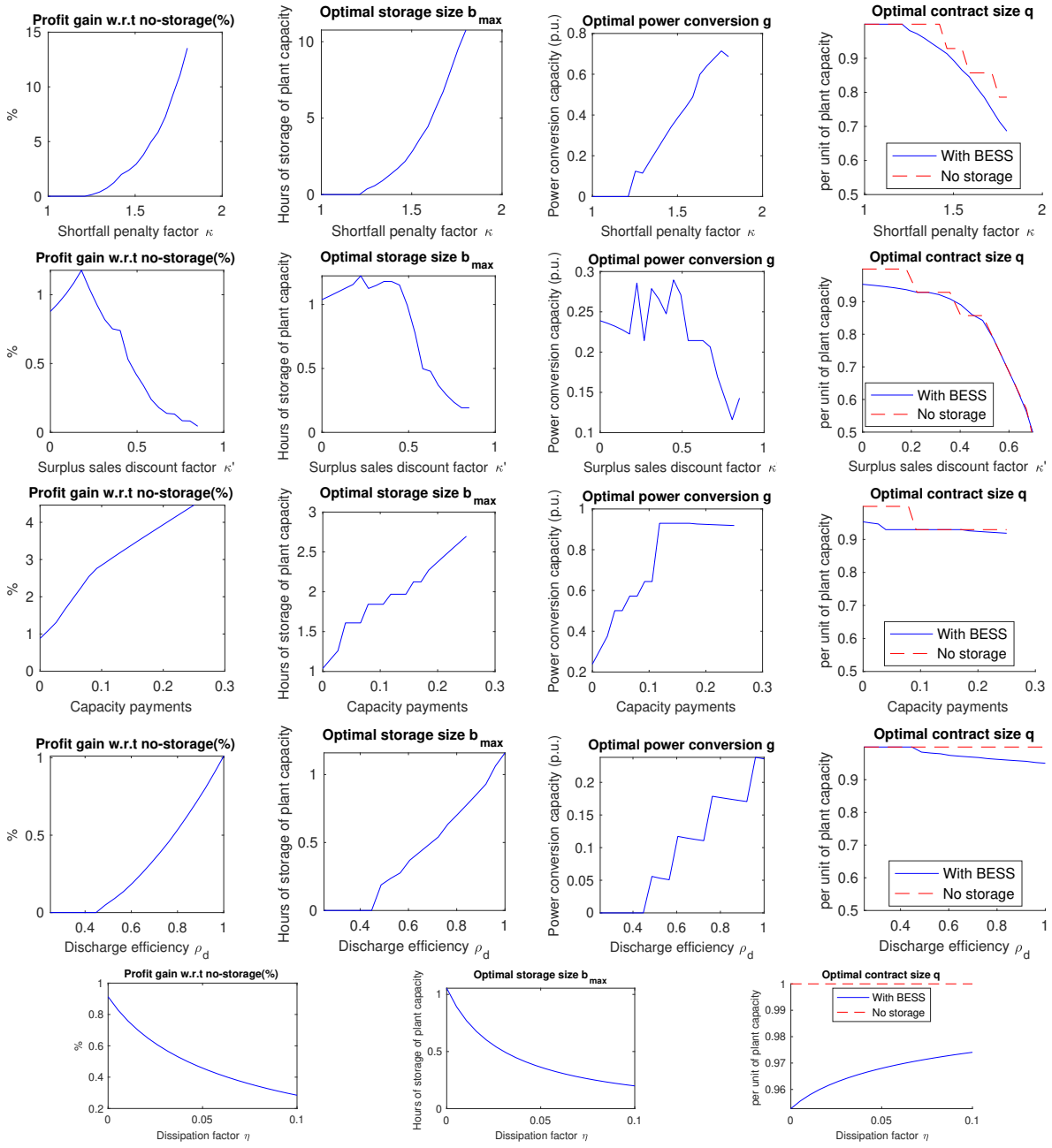

Figure 3.6.2: Sensitivity of optimal quantities (by column) to some model parameters (by row). **Columns**: *Left*: Profit gain with respect to no-storage. *Center-left*: Optimal storage size. *Center-right*: Optimal power inversion size. *Right*: Optimal contract size. **Rows**: Sensitivities with respect to changes in... *Top row*: Shortfall penalty factor κ. *Second row*: Discount factor κ'. *Third row*: Capacity payments p^{cap}. *Fourth row*: Discharge efficiency ρ_d. *Bottom row*: Dissipation losses factor η (in p.u. of storage lost per hour). These tests were run on the uncapacitated model and thus do not include the optimal power inversion output.

Appendix

3.A Pseudocode of ex-post computation

Input: $q, b, g, \mathbf{w}, \kappa, \kappa', \rho, \Delta$
$\mathbf{w}$ // Wind power output data series
Δ // Sampling period of wind power output data series
$T \leftarrow \texttt{length}(\mathbf{w})$ $x_0 \leftarrow 0$ // Initial state of charge
x, x_{prev} // Current, previous state of charge
$\Pi \leftarrow q \times T \times \Delta$ // Average profit
for $t = 1$ **to** T **do**
  $r \leftarrow \min\left(\max\left(w_t - q, -g\right), g\right)$ // Capacitated BESS injection
  $\texttt{empty} \leftarrow x_{prev} + \Delta r < 0$ // Determine if storage is empty
  $\texttt{full} \leftarrow x_{prev} + \rho\Delta r > b$ // Determine if storage is full
  $\texttt{excess} \leftarrow \Delta\left(w_t - q - r\right)^+$
  $\texttt{shortfall} \leftarrow \Delta\left(q + r - w_t\right)^+$
  if full **then**
    $\texttt{excess} \leftarrow \texttt{excess} + \Delta r - \left(b - x_{prev}\right)/\rho$
  end
  if empty **then**
    $\texttt{shortfall} \leftarrow \texttt{shortfall} - \Delta r - x_{prev}$
  end
  $\Pi \leftarrow \Pi + \kappa'\texttt{excess} - \kappa\texttt{shortfall}$
  if $r < 0$ **then**
    $x \leftarrow \max\left(\min\left(x_{prev} + \Delta r, b\right), 0\right)$
  end
  else
    $x \leftarrow \max\left(\min\left(x_{prev} + \rho\Delta r, b\right), 0\right)$
  end
  $x_{prev} \leftarrow x$
end
$\Pi \leftarrow \frac{\Pi}{T\Delta}$
return Π

Algorithm 4: *Ex-post* function to obtain average profit for given contract size, storage size and wind power output data

Chapter 4

Parallel Computing for Large-Scale Stochastic Capacity Expansion Planning [1]

Abstract

We implement a nodal stochastic generation and transmission Capacity Expansion Planning (CEP) planning model that incorporates the output from high-resolution climate projection models through load and generation availability scenarios. We implement our model in Pyomo and perform computational studies on a realistically-sized test case of the California electric grid in a high performance computing environment. We propose model reformulations and algorithm tuning to efficiently solve this large problem using a variant of the Progressive Hedging Algorithm. We utilize the parallelization capabilities and overall versatility of *mpi-sppy*, exploiting its hub-and-spoke architecture to concurrently obtain inner and outer bounds on an optimal expansion plan. Our results show that instances with 360 representative days on a system with over 8,000 buses can be solved to within 5% of optimality in under 4 hours of wall clock time. These are encouraging results on the path towards solving a large-scale power system expansion planning problem across a wide range of climate-informed operational scenarios.

[1] The results, as well as parts of the text in this chapter were presented at the 23rd Power Systems Computation Conference (PSCC 2024) [105], and will be published in the associated journal publication [104] under the name "Parallel computing for power system climate resiliency: solving a large-scale stochastic capacity expansion problem with mpi-sppy", authored by Tomas Valencia Zuluaga, Amelia Musselman, Jean-Paul Watson and Shmuel Oren. See Acknowledgements at the beginning of this chapter for a brief description of the individual contributions of the author of this thesis to the work presented in this chapter. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and was supported by the LLNL-LDRD Program under Project 22-SI-008, and by the Advanced Grid Modeling Program of the Office of Electricity of the U.S. Department of Energy. I would also like to thank Gurobi for providing the academic license used to run all tests.

Acknowledgements of individual contributions

The work presented in chapters 4 and 5 is the result of a collaboration with Lawrence Livermore National Laboratory (LLNL), during two summer internships at LLNL and subsequent LLNL-funded graduate student research appointments at UC Berkeley. As is natural in such collaborations, the efforts of different people contributed to the resulting work. I do not intend to claim that the totality of the models and algorithms presented in this chapter are my own. This section clarifies individual contributions of the LLNL team members to the work presented here.

The model proposed is based on the model by Go, Muñoz, and Watson [35]; the modifications proposed in this chapter are the result of a joint effort with Amelia Musselman. Complementing the California Test System (CATS) with data for our capacity expansion model was an effort led by Amelia Musselman. Details about this process can be found in a forthcoming paper [72].

Minda Monteagudo procured the climate projection data and the electricity load forecast data used as baseline to obtain the weather-adjusted load timeseries. The model to estimate weather impacts on electricity demand is also her work [67]. Matthew Signorotti developed the software tool that produces timeseries of potential generation availability for wind and solar power from the output of downscaled projected climate data [95], and provided the generation availability timeseries used in all tests. The model reformulations for minimizing spurious transmission losses (Section 4.3.6), as well as the runtime-dependent Mixed Integer Program (MIP) gap heuristic feature (Section 4.4.2) were proposed by the author of this thesis. The software implementation of these features and the overall methodology in mpi-sppy (Section 4.4) was mainly undertaken by the author of this thesis, as was preparing and running all the computational tests. Jean-Paul Watson was the project lead; he and Amelia Musselman contributed numerous useful suggestions during weekly conversations.

4.1 Introduction

Widespread penetration of renewable, intermittent, and decentralized generation resources is rapidly transforming the power grid and increasing its sensitivity to weather. Maintaining a resilient power grid as this transition unfolds will require strategic infrastructure investments. The recent *Net-Zero America* study [45] estimates that achieving carbon neutrality by 2050 will require quadrupling previous average annual build rates for renewable generation and increasing the total installed transmission capacity by a factor between 3 and 5. Incorporating climate projections into expansion planning tools can improve their relevance to decision makers, but the results are only meaningful if the uncertainty associated with these projections is also considered. Stochastic optimization is well-suited for this purpose, but often leads to computationally challenging optimization problems, for which high-performance computation can be a very valuable asset.

Developing implementable expansion plans for an efficient and resilient power system

requires both sufficiently high geographical resolution to accurately represent power system components and a sufficiently large representative set of scenarios to encompass all relevant potential weather impacts. The curse of dimensionality usually forces planners to face a tradeoff between these two dimensions.

The first computational challenge arises due to the size of the power system model. It is common to use very coarse geographic resolution models, representing entire states or nations as a single node in the power network. Obtaining actionable decisions from these unrealistic representations of the power grid requires iteratively solving increasingly high-resolution optimization models for each subregion of the grid. Solving power system planning models in this way not only leads to sub-optimal solutions but also underestimates the variability of wind and solar resources and load by aggregating these resources across a large area. To address the need for a realistic power system model suitable for research studies, the CATS testcase was recently developed [101]. This system, which we adopt and extend for the tests presented in this chapter, includes over 8,000 buses for the state of California. For comparison, a model often used for expansion plans in the literature has 240 buses for the entire U.S. Western Interconnection [115].

The second computational challenge is in adequately capturing the uncertainty, which we address in this work using stochastic optimization. Even without consideration of climate-dependent uncertainties, stochastic optimization has been shown to have significant economic benefits for power grid capacity expansion planning when compared to deterministic optimization and heuristic scenario planning [35, 70]. Climate uncertainty underscores these benefits. For example, in [68], a multi-level model is proposed to optimize a convex combination of expected and worst-case cost, as a balance between cost and security in the Brazilian power system, but the long-term uncertainty is reduced to just three states regarding occurrence of El Niño. A rapidly changing climate requires climate model projections, rather than just historical data or stylized low-dimensional models, to be fed into capacity expansion models. However, despite recent advances in global circulation model downscaling algorithms to obtain multi-decadal, high-resolution weather projections, the uncertainty associated with these projections is substantial, and so can be the potentially adverse effect of picking a single deterministic timeseries to inform investment plans.

Despite this necessity, the usage of stochastic optimization in actual planning processes is still modest, with computational limitations being an often-cited reason for preferring deterministic models. Muñoz et al. [71, 70] and Go, Muñoz, and Watson [35] lay the ground-work for using stochastic programming for capacity expansion planning, addressing the aforementioned tradeoff through model simplifications. In [71], a scalable stochastic expansion model is proposed, using decomposition by scenarios, with tests conducted on a universe of 8,760 one-hour scenarios that is reduced to 100-500 scenarios after clustering, on the geographically simplified 240-bus WECC test case. However, interconnection between time periods are neglected, which does not allow storage or other inter-temporal resources to be represented. Inter-temporal storage decisions are included in [35], but only a small, 24-bus, test-case with 5 scenarios is solved.

In the examples above, the scenarios of the stochastic optimization problem are obtained

by constructing a probability space from historical data, then reducing that probability space to a sample space of representative days or hours through statistical clustering techniques. A similar approach can be taken by substituting future weather projections for historical data, which is what we do in this work. We propose thus a climate-dependent stochastic generation, transmission, and storage capacity expansion model adapted from the model proposed in [35] and solve this model for a realistically-sized test case of California, leveraging *mpi-sppy* and High Performance Computing (HPC) resources at LLNL.

Structure of this chapter

In the work presented in this chapter, we solve a stochastic, nodal capacity expansion planning problem for the aforementioned high spatial-resolution, realistic test system of the California electricity grid [101], with a large number of scenarios derived from the output of a downscaled climate model. In Section 4.2, we go over some preliminaries about CEP, parallel computing for stochastic optimization, the *mpi-sppy* tool, and climate models. In Section 4.3, we present the mathematical model proposed. In Section 4.4, we present the solution approach followed: implementing the Progressive Hedging Algorithm (PHA) to utilize scenario decomposition in a HPC cluster. We present some generalities of the algorithm and focus on the key features of our implementation. Finally, results are presented and discussed in Section 4.5.

4.2 Preliminaries

In this section, we briefly review some necessary preliminaries about CEP models, the *mpi-sppy* tool for using parallel computing for stochastic optimization, and climate models.

4.2.1 Capacity expansion planning

Power system infrastructure projects have long (multi-year) construction times, require very large investments, often of public funds, and affect the cost and reliability of the electricity service for all consumers, which is essential in modern societies. Accordingly, transmission and generation expansion planning is a very well-studied problem, for which copious literature can be found. Models proposed for CEP differ by their decision-making structure, the scope of the expansion plan, the level of detail of the model and how uncertainty is addressed, among other aspects. We attempt here a very brief overview to provide the relevant context for the model that is proposed in this chapter. Useful reviews for the interested reader can be found in [22, 36, 41, 70, 71], among many other references.

Decision-making structure The electrical power system developed as a vertically integrated natural monopoly, so all decisions, including both long-term investment decisions and everyday operational decisions, were made under the framework of a cost-minimizing,

centralized decision maker. Accordingly, most models use a central planner perspective that minimizes total cost (and/or maximizes social welfare). After the liberalization of electricity markets in many jurisdictions starting in the late 1980's, the generation and retail portions of the business were opened to competition, but the transmission and distribution activities remained a regulated, natural monopoly[2]. As a consequence, generation and transmission expansion decisions were no longer made by the same agents. Decentralized, multi-level equilibrium, leader-follower models have been proposed to better capture that reality [84, 90].

Even for regions where the expansion plan is no longer centrally mandated, some central oversight of private long-term plans is required [26, 97], so a model adopting a central planner's perspective continues to be a relevant tool in CEP.

Scope of the expansion plan Driven by computational challenges, most conventional CEP models address the Generation Expansion Planning (GEP) problem and the Transmission Expansion Planning (TEP) problem separately. As MIP tools and computational capabilities improve over the years, co-optimization of generation, transmission, and storage resources have become more common [36], which can prove to be very valuable as penetration of intermittent renewable resources increases [35].

Temporal and geographic resolution Power system investments are very costly investments that can take years to build and decades to pay off. CEP models must therefore span a long horizon, often of several decades. It quickly becomes computationally intractable to maintain a high level of detail with such a large temporal scope; a tradeoff ensues affecting many model features, including geographic and temporal resolution, formulation of power flow, and the timeline of decisions.
Zonal representations are common, especially in GEP, but at such a high level of aggregation, accurate estimates of transmission congestion, which is key in determining the need for load shedding and hence for assessing climate resiliency, may be lost. Instances of capacity expansion models that maintain the full detail of the transmission network, and thus have comparable sizes to the one of this work, exist in the literature, but are solved via deterministic scenario planning [86], or, if considering uncertainty, for a significantly reduced number of scenarios [62].

Temporal resolution is also sacrificed: rather than trying to model every time period of the planning horizon, only a handful of snapshots, usually called representative hours, are kept. That comes at the cost of dropping inter-temporal constraints like unit commitment, ramping limitations, and storage dynamics.

Power flow representation The conventional representation of power flow in transmission grids is the so-called AC power flow. Although very accurate, this formulation is nonlin-

[2] For more about this history, see e.g. [49, 96, 98].

ear and nonconvex[3], so its usage for long-term planning and large-scale system is limited. A common, tractable alternative is the so-called DC power flow representation, in which representation of voltage variations, reactive power flow and losses are ignored. An even simpler representation is the network flow model, sometimes called transportation or pipeline model, where full control over the flow of electricity in a grid is assumed. This is the representation of power flow adopted in this chapter. The DC power flow, very common in small and medium-sized systems, can become challenging at larger scales. We look into these aspects with more detail in Chapter 5.

Multistage vs. two-stage CEP is a continuous process in which decisions are made by different stakeholders at different points in time, affecting subsequent decisions being made. A multistage formulation allowing for investment decisions at several points in time can therefore better capture the decision process, but will make a stochastic formulation significantly more challenging [2]. In two-stage formulations, it is assumed that all investment decisions are made at the same time, and that all projects are constructed during some time previous to the target horizon, which typically lies several years in the future. Although less representative of the true decision process, this compromise simplifies the computational challenge while still being informative for decision makers.

Uncertainty representation Models also differ by how they address the uncertainty associated with such a large temporal scope. There are numerous sources of uncertainty in CEP models: investment and operational costs, future demand, price and availability of fuel for electricity generation, availability of intermittent energy sources for hydro, wind and solar power, public policy at different jurisdictional levels, and technological advances to just name the main ones. Including all of these would lead to an intractable model, so it is common to formulate deterministic optimization problems and address these issues via sensitivity analyses of a handful of parameters.
Alternatively, the uncertainty can be directly incorporated into the model, usually via a discretization (or sampling) of the probability space. Appropriate levels of risk aversion can then be achieved by considering a corresponding risk measure, ranging from risk-neutral stochastic optimization where expected cost is minimized, to risk-adjustable conditional Value-at-risk formulations, all the way to robust optimization formulations, or even combinations thereof [68].

4.2.2 Parallel computing for stochastic optimization

Two-stage stochastic optimization problems are classic examples of optimization problems with a block-angular structure, i.e. optimization problems where only a handful of variables (first-stage, investment variables) couple otherwise independent variables (second-stage, operation variables). Thus, if first-stage variables are fixed to a certain value, the stochastic

[3]Although numerous convex relaxations and approximations exist. See e.g. [100].

optimization problem can be decomposed into much smaller operational subproblems that can be solved independently. Problems with block-angular structure are thus very good candidates for iterative solution algorithms with the broad structure of Algorithm 5. Two algorithms that implement that structure are the L-shaped method [9] and the Progressive Hedging Algorithm. [9, 21, 94].

repeat
  Fix the first-stage variables to some value.
  Solve all second-stage subproblems for that fixed value.
  Update the value of the first-stage variable, possibly by solving a first-stage master problem, based on the subproblems results.
until *some termination criterion is met*

Algorithm 5: General structure of a decomposition algorithm for problems with block-angular structure

Since subproblems are independent once the first-stage variable is fixed, they can be solved simultaneously if computational resources are available. This is what makes these problems very well suited for implementation in HPC clusters, where hundreds of computation cores are available for parallel execution. Setting up a code to run in parallel in an HPC cluster is not a trivial task, so having a tool that facilitates it is very helpful.

mpi-sppy

mpi-sppy [50] is a recently developed extension of the *Pyomo* [14, 40] optimization package in Python, specifically designed to solve stochastic optimization problems in HPC environments using the Message Passing Interface (MPI). We highlight here the two main features that make this tool so valuable for this work: the simplified implementation of the PHA in an HPC cluster, and its *hub-and-spokes* architecture. Note that we describe these features here in terms of the PHA because that is the method chosen in this work, but mpi-sppy does not use that algorithm exclusively: the L-shaped method is also implemented, and the framework is compatible with other scenario-decomposition-based algorithms.

Simplified implementation Any parallel program requires some degree of communication between the different processes that run simultaneously, in order to assign tasks to each process and collect the results of the distributed work. MPI is a very commonly used standard for this purpose. For an algorithm following the structure of Algorithm 5, such communication steps include, for instance, assigning subproblems to individual processes, communicating updated values of first-stage variables to all processes and collecting second-stage results from them. mpi-sppy uses the Python implementation *mpi4py*[25] of MPI to perform these communication tasks, so that the user does not need more than a few lines of code to setup the model.

Moreover, mpi-sppy already has an implementation of the Progressive Hedging Algorithm, including numerous common heuristics and features, e.g. fixing variables. Some of the mpi-sppy features that were used and extended in this work are mentioned in Section 4.4.2.

Hub-and-spokes architecture[4] In the mpi-sppy architecture, processes[5] are sorted into different functional groups, called *cylinders*. The main cylinder is the *hub*, which solves the conventional PHA, with subproblem solutions being parallelized across the processes that belong to it. A known shortcoming of the PHA is that before convergence, there is no guarantee regarding the quality of the solution, i.e. no lower or upper bounds on the optimal cost of the CEP. Furthermore, for MIPs, interim solutions are most likely non-integer and hence unfeasible, so it does not provide a candidate solution in case of early termination of the algorithm.

In mpi-sppy, optimality bounds are obtained by the other groups of processes, called *spokes*. Each spoke runs an algorithm that can be decomposed by scenario subproblems and parallelized across the spoke's processes. They can get information from the hub to form and update subproblems, and return information to the hub from which an outer or inner bound[6] of the stochastic optimization problem can be obtained. The architecture is depicted in Fig. 4.2.1.

An example of an outer bound spoke is running a variant of PHA without quadratic terms, i.e. the Lagrangian subproblem. This provides valid outer bounds of the subproblem optimal cost, and under certain conditions, may follow an independent multiplier update rule [33]. An example of an inner bound spoke is successively evaluating the first-stage solution of each subproblem on all other subproblems to obtain a fully implementable (feasible) solution. Note that inner bounds also provide a solution candidate in case of early termination of the algorithm. Several spokes of each type may be run concurrently, so that the best bound obtained so far can be used by the hub. Outer and inner bounds are used by the hub to obtain optimality gaps and decide on an early termination of the overall algorithm.

4.2.3 Climate models

Among the numerous sources of uncertainty in CEP, we focus in this work on climate and weather uncertainty. In particular, our model takes into account how climate and weather may affect the availability of generation resources and electricity demand during the planning horizon, which affects the need for generation and transmission expansion. The paragraphs that follow are a summary of [24], a great introductory-level text about climate projection data for models of energy systems. The reader is encouraged to consult that text and the references therein for more details.

[4]This paragraph assumes some familiarity with the PHA. The unfamiliar reader may prefer to come back to this paragraph after reading Section 4.4.1.

[5]Note that each process may, and in our tests, does, have access to several CPU cores.

[6]For minimization problems, outer bounds are lower bounds and inner bounds are upper bounds.

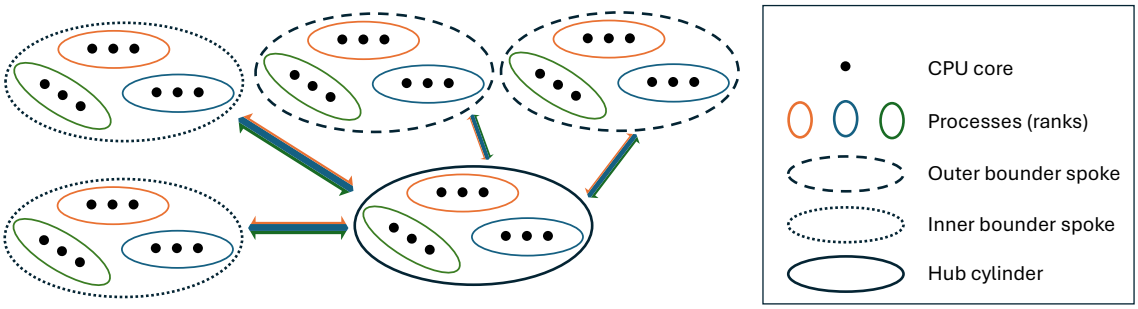

Figure 4.2.1: Diagram of mpi-sppy's hub-and-spokes architecture. Each process (*rank* in MPI nomenclature) corresponds to a scenario subproblem. Typically, information going from the hub to the spokes includes PHA weights for each scenario and the solution average across scenarios. Spokes typically send an upper or lower bound back to the hub. Note that each process may be assigned several CPU cores to exploit lower-level parallelization in certain tasks, e.g. linear algebra operations and branch & bound exploration by the optimization solver.

Craig et al. [24] enumerate five requirements for climate data for CEP models. The data should:

1. have sufficient spatial and temporal resolution,
2. be synchronous (maintain relevant correlation between inputs),
3. be convenient to process,
4. be computationally manageable, and
5. accurately represent relevant meteorological data.

Traditional CEP models have used historical climate data, which satisfy these requirements, and restrict their analysis to geographically and temporally reduced scopes. However, as human-made climate change increases the frequency and intensity of events that deviate from historical behavior, the probability of such an approach leading to suboptimal investment plans increases, and the traditional approach becomes less sensible. In this work, the output of climate projection models is used instead.

Climate projection models for energy systems can be seen as a serial sequence, or pipeline, of processes, depicted in Fig. 4.2.2, based on the methodology of [95]. The process starts with Global Circulation Models (GCMs), physics-based models of the thermodynamics of atmospheric circulation at coarse spatial resolution, with a scope of several decades for the entire planet. A multitude of parameters and modeling choices affect the outcome of GCMs, so great effort is spent in tuning and validating models against available past data [44]. Different GCMs may have different spatial and temporal scopes, and define different metrics for validating the model, which leads to different degrees of validity in CEP applications.

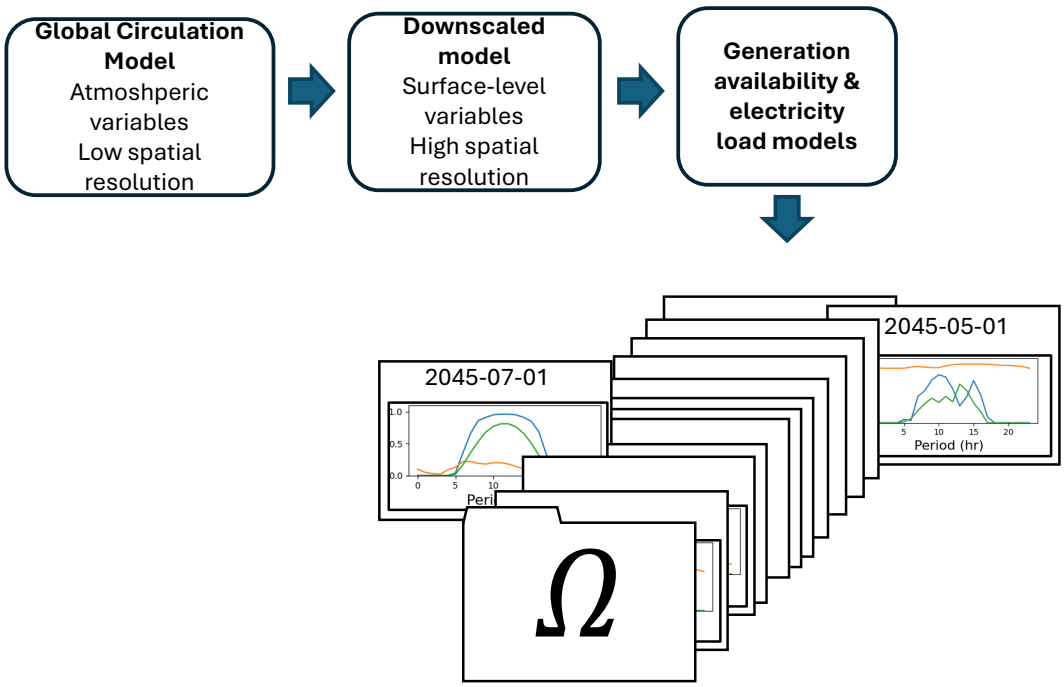

Figure 4.2.2: Data pipeline for climate and weather inputs for the CEP model. The output of GCMs, low-resolution atmoshpheric variables is input into a downscaling model, which produces surface-level variables with high spatial resolution, like surface temperature, humidity and wind speed. These variables are the input of specific models that produce hourly timeseries for the availability of each generation technology and the electricity load at each location in the network.

The next step is downscaling, a process by which the coarse atmospheric data output of GCMs is converted to surface-level quantities relevant for modeling energy systems. Different methodologies, including statistics-based and dynamics-based approaches, exist for performing this process, and, once again, can have significantly different results depending on the final-use in mind for the data. In our tests, we use the output of the California Regionally Refined Model (CARRM) [116].

Finally, surface-level climate and weather timeseries are fed to appropriate models to obtain impacts on the availability of energy resources (e.g. output of solar and wind power plants, inflow into hydropower dams, transmission line load-carrying capability, etc.) and forecasts of electricity demand (e.g. impact on heating and air-conditioning demands).

The climate science and energy system modeling communities are becoming increasingly aware of the gap between the climate data output by climate projection models and the input needed by models of energy systems[7]. Accordingly, models and methods have been

[7]The project at LLNL under which this work was performed is evidence of this awareness.

and continue to be developed which specifically aim to be relevant and valid for use in the planning of energy infrastructure, e.g. the Department of Energy's Energy Exascale Earth System Model (E3SM) [53], which we use in our tests.

Nonetheless, it remains true that including climate data from just one year of one parameterization of one model may question the validity of the investment plan obtained from the CEP model. CEP models that aim to enhance climate resiliency with some credibility should therefore be capable of handling data from multiple sources simultaneously to account for the associated uncertainty. This leads to large-scale CEP models that pose computational challenges. This is the problem we address in the work presented in this chapter.

A note about the term 'scenario' Among the numerous exogenous parameters that affect the outcome of GCMs, one to which a lot of attention is paid, because of its significant impact, is the assumed behavior of the biosphere, and in particular, human activity. To account for this unpredictability, GCM datasets are published for different parameterizations of the model. Most importantly among them are Representative Concentration Pathways [61], different assumptions agreed upon by the climate science community, about the rate at which humans will continue to emit CO2 into the atmosphere over the next decades.

These different parameterizations are referred to in the climate science field as "*scenarios*". In the context of stochastic programming, *scenarios* are possible realizations of stochastic processes. In particular, for two-stage stochastic optimization with a discrete probability space, like we have here, a scenario is each representative operational horizon. Certain parameters and variables in the model are indexed by scenarios, and may have scenario-dependent values. In this thesis, we always use the term *scenario* in the stochastic optimization sense.

4.3 Mathematical model for CEP

Our capacity expansion model is an adaptation of the model proposed in [35]. We give first a high-level description of the model before providing a full model description. At the end of the section, we discuss briefly about our model features that differ from the reference.

4.3.1 General description

Our capacity expansion model is a two-stage stochastic program that co-optimizes generation, transmission and storage investments in the first stage and solves a multi-period Optimal Power Flow (OPF) problem in the second stage. Second-stage suproblems, i.e. representative days, differ from each other in the hourly nodal demand and generation availability.

Renewable generation and storage investments are modeled as continuous (installed capacity), thermal generation investments as integer (number of installed units), and transmission investments as binary (build or no-build for candidates) variables. All investment costs

are assumed to be linear. Storage investments consist of two decisions[8] : energy storage capacity (in MWh, which can be interpreted as storage duration) and instantaneous power charging/discharging capacity (in MW). These two decisions are made independently and have linear costs.

In this chapter, the second-stage OPF is modeled with a transportation relaxation of power flow incorporating transmission losses. Improved linear representations of power flow are considered in Chapter 5. Ramping and start-up constraints are disregarded. Storage levels are assumed to be cyclic (level at the end of the last period must equal the initial storage level).

4.3.2 Nomenclature

Sets
$\mathcal{B}$: Set of all buses (nodes) in the network. Indexed by b (or i, m, n where noted).
$\mathcal{G}$: Set of generation types, indexed by g.
$\mathcal{G}^{\mathcal{R}}$: Set of renewable generation types.
$\mathcal{G}^{\mathbb{Z}}$: Set of generation types modeled with integer variables (turbine-generator set).
$\mathcal{G}^{\mathbb{R}}$: Set of generation types modeled with continuous variables (inverter-based).
$\mathcal{S}$: Set of storage types, indexed by s.
$\mathcal{L}$: Set of transmission branches, i.e. transmission lines and transformers. Indexed by ℓ.
$\mathcal{L}^{\dagger}$: Set of existing transmission branches.
$\mathcal{L}^{\star}$: Set of candidate transmission branches.
$\mathcal{L}^{\circ}$: Set of transmission branches modeled as lossless.
$\mathcal{L}^{\bullet}$: Set of transmission branches modeled as lossy.
$\mathcal{T}$: Set of periods in a representative horizon. Indexed by t.
Ω: Set of scenarios in the uncertainty set of the stochastic optimization problem. Indexed by ω.

$\mathcal{G}^{\mathbb{Z}}$ and $\mathcal{G}^{\mathbb{R}}$ constitute a partition of $\mathcal{G}$, i.e. $\mathcal{G}^{\mathbb{Z}} \cap \mathcal{G}^{\mathbb{R}} = \emptyset$ and $\mathcal{G}^{\mathbb{Z}} \cup \mathcal{G}^{\mathbb{R}} = \mathcal{G}$. Similarly, $\mathcal{L}^{\dagger}$ and $\mathcal{L}^{\star}$ constitute a partition of $\mathcal{L}$. $\mathcal{L}^{\circ}$ and $\mathcal{L}^{\bullet}$ constitute a possibly different partition of $\mathcal{L}$.
Index maps
$o(\ell)$: Origin bus (also called 'from' bus) of branch ℓ
$d(\ell)$: Destination bus (also called 'to' bus) of branch ℓ

Parameters
Existing resources
$X^{\mathcal{G}}_{b,g}$: existing generation of type g at bus b, in MW.

[8] An alternative formulation for which data is easier to procure is to have only storage power conversion capacity as decision variable, and use fixed ratios of storage duration (i.e. energy capacity) to power conversion capacity. Different durations can then be included as different storage types. A version of the model with this alternative formulation is currently under development at LLNL, and will be published after the conclusion of this thesis.

$X^{\mathcal{S}}_{b,s}$: existing storage capacity of type s at bus b, in MWh.
$X^{\mathcal{S}\text{-PC}}_{b,s}$: existing storage power conversion capacity of type s at bus b, in MW.
Fixed costs
$C^{\mathcal{G}\text{-cap}}_{\ell}$: Capital cost for generator type g, amortized over its lifetime, in \$/(MWy)
$C^{\mathcal{S}\text{-cap}}_{g}$: Capital cost of storage for storage type s, amortized over its lifetime, in \$/(MWhy)
$C^{\mathcal{S}\text{-PC-cap}}_{s}$: Capital cost of power conversion for storage type s, amortized over its lifetime, in \$/(MWy)
$C^{\mathcal{L}\text{-cap}}_{s}$: Capital cost for branch candidate ℓ, amortized over its lifetime, in \$/y
$C^{\mathcal{G}\text{-FOM}}_{g}$: Fixed operation and maintenance cost for generation type g, in \$/(MWy).
$C^{\mathcal{S}\text{-FOM}}_{s}$: Fixed operation and maintenance cost for storage type s, in \$/(MWhy).
$C^{\mathcal{S}\text{-PC-FOM}}_{s}$: Fixed power conversion operation and maintenance cost for storage type s, in \$/(MWy).
Variable costs
$C^{\mathcal{G}\text{-fuel}}_{g}$: Fuel cost, or equivalent per-MWh operating cost, for generation type g, in \$/(MWh).
$C^{\mathcal{G}\text{-VOM}}_{g}$: Variable operation and maintenance cost for generation type g, in \$/(MWh).
$C^{\mathcal{S}\text{-dch}}_{s}$: Variable discharge cost for storage type s, in \$/(MWh).
$C^{\mathcal{S}\text{-VOM}}_{s}$: Variable operation and maintenance cost for storage type s, in \$/(MWh).
C^{sh}: Cost of load shedding, in \$/MWh.
Operational parameters
$P^{\mathcal{G}}_{g}$: capacity per unit of generator type g, in MW.
$D_{b,t,\omega}$: demand at bus b during period t of scenario ω.
$\alpha_{b,g,t,\omega}$: fraction of generation capacity of type g that is available at bus b during period t of scenario ω.
$\eta^{\mathcal{S}\text{-ch}}$: power conversion efficiency for storage type s when charging.
$\eta^{\mathcal{S}\text{-dch}}$: power conversion efficiency for storage type s when discharging.
$\eta^{\mathcal{L}}_{\ell}$: efficiency of branch ℓ.
χ_{ℓ}: reactance of branch ℓ, in p.u.
r_{ℓ}: resistance of branch ℓ, in p.u.
F_{ℓ}: transmission capacity of branch ℓ in MW.
Construction and planning parameters
$K^{\mathcal{B}}_{b}$: maximum buildable capacity across generation types at bus b, in MW.
$K^{\mathcal{G}}_{g}$: maximum buildable capacity of generation type g across all buses, in MW.
$K^{\mathcal{BG}}_{b,g}$: maximum buildable capacity of generation type g at bus b, in MW.
$K^{\mathcal{S}}_{s}$: maximum buildable capacity of storage type s across all buses, in MWh.
$K^{\mathcal{BS}}_{b,s}$: maximum buildable capacity of storage type s at bus b, in MWh.
ρ^{RPS}: renewable portfolio standard, i.e. fraction of generated energy that must be from qualified renewable generation types.
λ^{RPS}: penalty for RPS non-compliance, in \$/MWh .
Other model parameters
T^{rep}: number of times that the representative time horizon repeats in a year.
τ: length of each period t in a representative time horizon, in h.

π_ω: probability assigned to scenario ω in the stochastic optimization problem.

Decision variables

Investment variables

$x^{\mathcal{G}}_{b,g}$: new generation of type g at bus b, expressed as number of generation units.

$x^{\mathcal{S}}_{b,s}$: new storage capacity of type s at bus b, in MWh.

$x^{\mathcal{S}\text{-PC}}_{b,s}$: new storage power conversion capacity of type s at bus b, in MW.

$x^{\mathcal{L}}_{\ell}$: binary variable indicating whether candidate ℓ is built.

Operation variables

$p^{\mathcal{G}}_{b,g,t,\omega}$: output of generator of type g at bus b during period t of scenario ω, in MW.

$p^{\mathcal{S}}_{b,s,t,\omega}$: energy stored (level) in storage facility of type s at bus b during period t of scenario ω, in MWh.

$p^{\mathcal{S}\text{-ch}}_{b,s,t,\omega}$: power input (charging) of storage facility of type s at bus b during period t of scenario ω, in MW.

$p^{\mathcal{S}\text{-dch}}_{b,s,t,\omega}$: power output (discharging) of storage facility of type s at bus b during period t of scenario ω, in MW.

$p^{\text{sh}}_{b,t,\omega}$: load shed at bus b during period t of scenario ω, in MW.

$f_{\ell,t,\omega}$: power flow through branch ℓ during period t of scenario ω, in MW.

$f^{+}_{\ell,t,\omega}$: power flow through branch ℓ during period t of scenario ω, in direction $o(\ell) \to d(\ell)$, in MW.

$f^{-}_{\ell,t,\omega}$: power flow through branch ℓ during period t of scenario ω, in direction $d(\ell) \to o(\ell)$, in MW.

p^{NC}_{ω}: RPS non-compliance in scenario ω, in MWh.

Auxiliary variables and expressions

C^{inv}: Annualized total investment cost in \$/y.

C^{op}_{ω}: Operation cost of scenario ω excluding RPS non-compliance, expressed in \$/y.

$C^{\text{op-RPS}}_{\omega}$: Operation cost of scenario ω including RPS non-compliance, expressed in \$/y.

Notation

Unless otherwise specified, the vector is noted by omitting the corresponding index, e.g. $f = [f_\ell]_{\ell \in \mathcal{L}}$

4.3.3 Objective

The objective is to minimize the total cost of expanding, maintaining and operating the system, which we break into an investment cost C^{inv} and an expected operation cost $\mathbb{E}_\zeta\left[C^{\text{op}}_{\omega}\right]$. The random variable ζ is used to represent in compact form all the scenario-dependent data.

$$C^{\text{inv}} = \sum_{\ell \in \mathcal{L}^\star} C_\ell^{\mathcal{L}\text{-cap}} x_\ell^{\mathcal{L}} + \sum_{b \in \mathcal{B}} \left(\sum_{g \in \mathcal{G}} \left(C_g^{\mathcal{G}\text{-cap}} + C_g^{\mathcal{G}\text{-FOM}} \right) P_g^{\mathcal{G}} x_{b,g}^{\mathcal{G}} + \sum_{s \in \mathcal{S}} \left[\left(C_s^{\mathcal{S}\text{-cap}} + C_s^{\mathcal{S}\text{-FOM}} \right) x_{b,s}^{\mathcal{S}} + \left(C_s^{\mathcal{S}\text{-PC-cap}} + C_s^{\mathcal{S}\text{-PC-FOM}} \right) x_{b,s}^{\mathcal{S}\text{-PC}} \right] \right) \quad (4.1a)$$

$$\mathbb{E}_\zeta \left[C_\omega^{\text{op}} \right] = \sum_{\omega \in \Omega} \pi_\omega C_\omega^{\text{op}} = \sum_{\omega \in \Omega} \left(\pi_\omega T^{\text{rep}} \sum_{t \in \mathcal{T}} \sum_{b \in \mathcal{B}} \tau \cdot \left[C^{\text{sh}} p_{b,t,\omega}^{\text{sh}} + \sum_{g \in \mathcal{G}} \left(C_g^{\mathcal{G}\text{-fuel}} + C_g^{\mathcal{G}\text{-VOM}} \right) p_{b,g,t,\omega}^{\mathcal{G}} + \sum_{s \in \mathcal{S}} \left(C_s^{\mathcal{S}\text{-dch}} + C_s^{\mathcal{S}\text{-VOM}} \right) p_{b,s,t,\omega}^{\mathcal{S}\text{-dch}} \right] \right) \quad (4.1b)$$

The operation cost during each representative horizon ω is multiplied by T^{rep} so that all costs are in \$/year.

Note that to obtain the total cost of maintaining and operating the system, we should also include the fixed costs of existing generation and storage resources. However, since we do not include retirements in our decision variables, these costs do not affect the optimization problem, so we omit them here.

4.3.4 Decision variables

Investment variables

Transmission expansion is considered through a set of candidate investments $\mathcal{L}^\star$, which can be either built or not built, so the corresponding decision variables are binary. Some storage and generation technologies have a modular characteristic with small module size, like inverter-based technologies (solar photovoltaic, wind, battery storage). We thus consider that any size can be built at a linear cost, and model expansion decisions with continuous variables. Traditional generation technologies, based on a turbine and synchronous generator set, do not have such a modular behavior with small module size. To avoid having an investment plan with unrealistically small generation buildouts, those technologies are modeled with integer variables, so that only integer multiples of commercial unit sizes can be built. We thus obtain the domains of (4.2).

$$x_{b,g}^{\mathcal{G}} \in \mathbb{Z}^+ \ \forall b \in \mathcal{B}, \forall g \in \mathcal{G}^{\mathbb{Z}} \quad (4.2a)$$

$$x_{b,g}^{\mathcal{G}} \in \mathbb{R}^+ \ \forall b \in \mathcal{B}, \forall g \in \mathcal{G}^{\mathbb{R}} \quad (4.2b)$$

$$x_{b,s}^{\mathcal{S}\text{-PC}}, x_{b,s}^{\mathcal{S}} \in \mathbb{R}^+ \ \forall b \in \mathcal{B}, \forall s \in \mathcal{S} \quad (4.2c)$$

$$x_\ell^{\mathcal{L}} \in \{0, 1\} \ \forall \ell \in \mathcal{L}^\star \quad (4.2d)$$

Operation variables

All operation variables are continuous (we do not include unit commitment or ramping in the second-stage problem). The outputs of all generators are non-negative. Storage charge and discharge are modeled separately in order to account for conversion losses, so the corresponding variables are non-negative too. Load shedding and Renewable Portfolio Standards (RPS) non-compliance are also non-negative continuous variables.

Branch flows Transmission branches can be modeled as lossless or lossy. Power flow through lossless transmission branches is represented by a single, sign-unconstrained variable f_ℓ, while lossy transmission branches are assigned one non-negative variable for flow in each direction, so that net flow in the $o(\ell) \to d(\ell)$ direction is $f_\ell^+ - f_\ell^-$. See section 4.3.6 for more details. The domains of operational variables are those of (4.3).

$$p^{\mathcal{G}}_{b,g,t,\omega} \in \mathbb{R}^+ \; \forall b \in \mathcal{B}, g \in \mathcal{G}, t \in \mathcal{T}, \omega \in \Omega \tag{4.3a}$$

$$p^{\mathcal{S}}_{b,s,t,\omega}, p^{\mathcal{S}\text{-ch}}_{b,s,t,\omega}, p^{\mathcal{S}\text{-dch}}_{b,s,t,\omega} \in \mathbb{R}^+ \; \forall b \in \mathcal{B}, s \in \mathcal{S}, t \in \mathcal{T}, \omega \in \Omega \tag{4.3b}$$

$$p^{\text{sh}}_{b,t,\omega} \in \mathbb{R}^+ \; \forall b \in \mathcal{B}, t \in \mathcal{T}, \omega \in \Omega \tag{4.3c}$$

$$p^{\text{NC}}_{\omega} \in \mathbb{R}^+ \; \omega \in \Omega \tag{4.3d}$$

$$f^+_{\ell,t,\omega}, f^-_{\ell,t,\omega} \in \mathbb{R}^+ \; \forall \ell \in \mathcal{L}^\circ, t \in \mathcal{T}, \omega \in \Omega \tag{4.3e}$$

$$f_{\ell,t,\omega} \in \mathbb{R} \; \forall \ell \in \mathcal{L}^\bullet, t \in \mathcal{T}, \omega \in \Omega \tag{4.3f}$$

4.3.5 Constraints

Construction constraints

We consider construction limits for new generation and storage resources. Limits are considered for each generation type at each bus (4.4a), each generation type across all buses (4.4b), and each bus across all generation types (4.4c). For storage, limits are considered for the energy capacity only (not for power conversion), for each storage type at each bus (4.4d) and each storage type across all buses (4.4e). Other combinations and limits for power conversion capacity were not considered in this formulation, but could easily be added to the model.

$$P_g^{\mathcal{G}} x_{b,g}^{\mathcal{G}} + X_{b,g}^{\mathcal{G}} \leq K_{b,g}^{\mathcal{BG}} \ \forall b \in \mathcal{B}, g \in \mathcal{G} \tag{4.4a}$$

$$\sum_{b \in \mathcal{B}} \left(P_g^{\mathcal{G}} x_{b,g}^{\mathcal{G}} + X_{b,g}^{\mathcal{G}} \right) \leq K_g^{\mathcal{G}} \ g \in \mathcal{G} \tag{4.4b}$$

$$\sum_{g \in \mathcal{G}} \left(P_g^{\mathcal{G}} x_{b,g}^{\mathcal{G}} + X_{b,g}^{\mathcal{G}} \right) \leq K_b^{\mathcal{B}} \ b \in \mathcal{B} \tag{4.4c}$$

$$x_{b,s}^{\mathcal{S}} + X_{b,s}^{\mathcal{S}} \leq K_{b,s}^{\mathcal{BS}} \ \forall b \in \mathcal{B}, s \in \mathcal{S} \tag{4.4d}$$

$$\sum_{b \in \mathcal{B}} \left(x_{b,s}^{\mathcal{S}} + X_{b,s}^{\mathcal{S}} \right) \leq K_s^{\mathcal{S}} \ s \in \mathcal{S} \tag{4.4e}$$

Operational constraints

Physical limits of generation and storage The output of each generator cannot exceed the available power, which is dictated by the installed capacity and the availability of intermittent resources (4.5a). Similarly, the charge and discharge of storage facilities cannot exceed the installed power conversion capacity (4.5b),(4.5c), and the availability of storage must be respected (4.5d). Note that this model does not consider a minimum operational storage level other than 0, but this could easily be added to the model if deemed appropriate.

$$p_{b,g,t,\omega}^{\mathcal{G}} \leq \alpha_{b,g,t,\omega} \left(X_{b,g}^{\mathcal{G}} + P_g^{\mathcal{G}} x_{b,g}^{\mathcal{G}} \right) \ \forall b \in \mathcal{B}, g \in \mathcal{G}, t \in \mathcal{T}, \omega \in \Omega \tag{4.5a}$$

$$p_{b,s,t,\omega}^{\mathcal{S}\text{-ch}} \leq x_{b,s}^{\mathcal{S}\text{-PC}} + X_{b,s}^{\mathcal{S}\text{-PC}} \ \forall b \in \mathcal{B}, s \in \mathcal{S}, t \in \mathcal{T}, \omega \in \Omega \tag{4.5b}$$

$$p_{b,s,t,\omega}^{\mathcal{S}\text{-dch}} \leq x_{b,s}^{\mathcal{S}\text{-PC}} + X_{b,s}^{\mathcal{S}\text{-PC}} \ \forall b \in \mathcal{B}, s \in \mathcal{S}, t \in \mathcal{T}, \omega \in \Omega \tag{4.5c}$$

$$p_{b,s,t,\omega}^{\mathcal{S}} \leq x_{b,s}^{\mathcal{S}} + X_{b,s}^{\mathcal{S}} \ \forall b \in \mathcal{B}, s \in \mathcal{S}, t \in \mathcal{T}, \omega \in \Omega \tag{4.5d}$$

Energy storage The change in energy storage level at each facility is driven by its charge and discharge (4.6a). To avoid end-of-horizon effects, we use the last period of the horizon as the initial storage state (4.6b). A constraint to impede simultaneous charging and discharging of storage units is omitted. We discuss more about this in Section 4.5.

$$p_{b,s,t,\omega}^{\mathcal{S}} = p_{b,s,t-1,\omega}^{\mathcal{S}} + \tau \left(\eta_s^{\mathcal{S}\text{-ch}} p_{b,s,t,\omega}^{\mathcal{S}\text{-ch}} - p_{b,s,t,\omega}^{\mathcal{S}\text{-dch}} \right) \ \forall b \in \mathcal{B}, s \in \mathcal{S}, t \in \mathcal{T} \setminus \{|T|\}, \omega \in \Omega \tag{4.6a}$$

$$p_{b,s,0,\omega}^{\mathcal{S}} = p_{b,s,|\mathcal{T}|-1,\omega}^{\mathcal{S}} + \tau \left(\eta_s^{\mathcal{S}\text{-ch}} p_{b,s,t,\omega}^{\mathcal{S}\text{-ch}} - p_{b,s,t,\omega}^{\mathcal{S}\text{-dch}} \right) \ \forall b \in \mathcal{B}, s \in \mathcal{S}, \omega \in \Omega \tag{4.6b}$$

Load shedding Shedded load cannot exceed demand (4.7).

$$p^{\text{sh}}_{b,t,\omega} \leq D_{b,t,\omega} \; \forall b \in \mathcal{B}, t \in \mathcal{T}, \omega \in \Omega \tag{4.7}$$

Power balance In this chapter, we consider a nodal representation of power flow, so the energy balance is ensured at each individual bus (4.8). Note that for each lossy branch ℓ, it is assumed that only a fraction $\eta^{\mathcal{L}}_{\ell}$ makes it to the other end of the branch, in each direction. See section 4.3.6 for more details.

$$\begin{aligned}\sum_{g \in \mathcal{G}} p^{\mathcal{G}}_{b,g,t,\omega} + \sum_{s \in \mathcal{S}} \left(\eta^{\mathcal{S}\text{-dch}}_{s} p^{\mathcal{S}\text{-dch}}_{b,s,t,\omega} - p^{\mathcal{S}\text{-ch}}_{b,s,t,\omega}\right) - \sum_{\ell \in \mathcal{L}^{\circ}: o(\ell)=b} f_{\ell,t,\omega} + \sum_{\ell \in \mathcal{L}^{\circ}: d(\ell)=b} f_{\ell,t,\omega} \\ - \sum_{\ell \in \mathcal{L}^{\bullet}: o(\ell)=b} \left(f^{+}_{\ell,t,\omega} - \left(1-\eta^{\mathcal{L}}_{\ell}\right) f^{-}_{\ell,t,\omega}\right) - \sum_{\ell \in \mathcal{L}^{\bullet}: d(\ell)=b} \left(f^{-}_{\ell,t,\omega} - \left(1-\eta^{\mathcal{L}}_{\ell}\right) f^{+}_{\ell,t,\omega}\right) \\ + p^{\text{sh}}_{b,t,\omega} = D_{b,t,\omega} \; \forall b \in \mathcal{B}, t \in \mathcal{T}, \omega \in \Omega\end{aligned} \tag{4.8}$$

Power flow & transmission limits In this chapter, we adopt a simplified version of power flow where it is assumed that power can be routed through the network as desired. This model is often called a transportation, or pipeline, formulation of power flow, and it ignores the physics that dictate how electricity flows in the grid. In chapter 5, improved linear representations of power flow are considered, and the impact of their inclusion on the investment plan obtained are assessed. The transportation formulation corresponds to a classic capacitated network flow problem, for which we only need to consider the nodal balance constraints (4.8), and the thermal capacity of the branches. Limits for existing branches are enforced by (4.9a) and (4.9b), while (4.9c) and (4.9d) ensure them for built branches, as well as that unbuilt branches have no flow. Note we consider separately lossless and lossy branches.

$$-F_{\ell} \leq f_{\ell,t,\omega} \leq F_{\ell} \; \forall \ell \in \mathcal{L}^{\circ} \cap \mathcal{L}^{\dagger}, t \in \mathcal{T}, \omega \in \Omega \tag{4.9a}$$

$$f^{+}_{\ell,t,\omega} + f^{-}_{\ell,t,\omega} \leq F_{\ell} \; \forall \ell \in \mathcal{L}^{\bullet} \cap \mathcal{L}^{\dagger}, t \in \mathcal{T}, \omega \in \Omega \tag{4.9b}$$

$$-F_{\ell} x^{\mathcal{L}}_{\ell} \leq f_{\ell,t,\omega} \leq F_{\ell} x^{\mathcal{L}}_{\ell} \; \forall \ell \in \mathcal{L}^{\circ} \cap \mathcal{L}^{\star}, t \in \mathcal{T}, \omega \in \Omega \tag{4.9c}$$

$$f^{+}_{\ell,t,\omega} + f^{-}_{\ell,t,\omega} \leq F_{\ell} x^{\mathcal{L}}_{\ell} \; \forall \ell \in \mathcal{L}^{\bullet} \cap \mathcal{L}^{\star}, t \in \mathcal{T}, \omega \in \Omega \tag{4.9d}$$

Planning constraints

Renewable Portfolio Standards (RPS) constraints represent policies in place in certain jurisdictions, whereby it is mandated that at least some fraction ρ^{RPS} of annual energy produced

must be generated with qualified renewable resources. This is the only constraint that couples operational variables from different scenarios, so it needs special treatment for a successful decomposition approach. We discuss more about this constraint in the next section.

$$\sum_{\omega\in\Omega}\pi_\omega\cdot\left(\sum_{t\in\mathcal{T}}\sum_{b\in\mathcal{B}}\sum_{g\in\mathcal{G}^{\mathcal{R}}}p^{\mathcal{G}}_{b,g,t,\omega}\right)\geq\rho^{\text{RPS}}\sum_{\omega\in\Omega}\pi_\omega\cdot\left(\sum_{t\in\mathcal{T}}\sum_{b\in\mathcal{B}}\sum_{g\in\mathcal{G}}p^{\mathcal{G}}_{b,g,t,\omega}\right) \tag{4.10}$$

4.3.6 Differences with respect to base model

In the following paragraphs, we describe in more detail the set of investment candidates, our model for transmission losses, and the RPS constraints, all of which deviate from the model in [35].

Investment candidates

In the model description of [35], the set of generation and storage investment candidates consists of all combinations of candidate technologies and power system buses. This set is later reduced to an explicit list of candidate units when the problem data is specified. We make this methodology precise by considering all combinations of technologies and buses subject to three *maximum potential capacity* constraints. The three constraints limit the total installed capacity for each storage/generation type at each bus, for each type across all buses, and for each bus across all types. In actual instances of the model, the maximum potential capacity is zero for many generator type, location combinations, allowing the number of variables included in actual instantiations of the model to remain tractable.

Transmission losses

As is common when performing planning studies, the model in [35] disregards transmission losses. However, ignoring these losses can significantly affect the solution obtained from the model, especially for systems with renewable generation, in which power may need to be transported over very large distances [32, 74]. Accurately capturing the quadratic nature of transmission losses without losing the model's linearity requires adding several linear segments for each transmission line, which greatly increases the number of variables in the model. To avoid this increase, we consider a single-segment linear approximation. Specifically, we consider each transmission branch ℓ to have a constant efficiency $\eta^{\mathcal{L}}_\ell$ given by

$$\eta^{\mathcal{L}}_\ell=1-F_\ell r_\ell\left(1+\frac{r^2_\ell}{\chi^2_\ell}\right), \tag{4.11}$$

where χ_ℓ is the reactance, r_ℓ is the resistance, and F_ℓ is the long-term capacity rating of branch ℓ, all in p.u. This linear approximation, based on Fitiwi et al. [32], is the result of

taking a first-order approximation of the quadratic model around the point $f_{\ell,t,\omega} = F_\ell$ and forcing the intercept at 0. This will overestimate losses in transmission branches that are loaded below their capacity, leading to somewhat conservative solutions. In contrast, assuming lossless transmission across long distances will significantly underestimate transmission capacity needs.

In order to correctly assign losses to buses, we also deviate from [35] by using two non-negative variables f_ℓ^+, f_ℓ^- to represent branch flows, as illustrated in Fig. 4.3.1. Note that formulation (b) only represents formulation (a) accurately if $f_\ell^+ \cdot f_\ell^- = 0$. Otherwise, model (b) will have higher losses than the corresponding model (a) with same net outgoing branch flow.

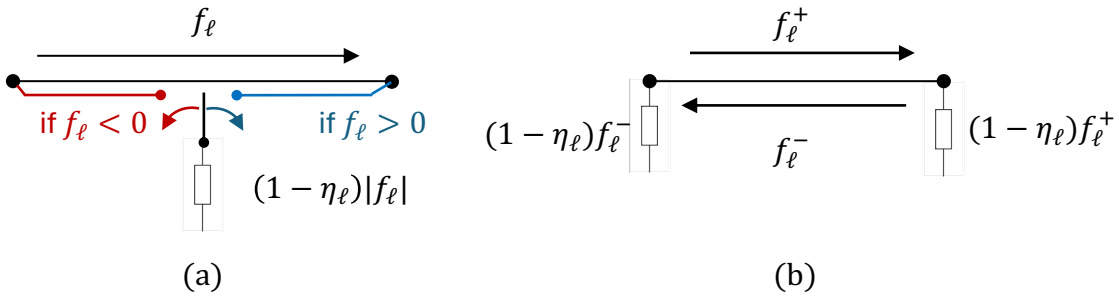

Figure 4.3.1: Proposed representation of flow in transmission branches to model transmission losses. In model (a), the flow is represented by sign-unconstrained variable f_ℓ. The losses can be computed as a fraction of $|f_\ell|$. The switch determines the end of the branch to which losses are assigned depending on the direction of flow. In model (b), the same is achieved by breaking the sign-unconstrained variable into positive and negative parts, representing flow in each direction, so that $f_\ell = f_\ell^+ - f_\ell^-$. Model (b) and model (a) are equivalent if $f_\ell^+ \cdot f_\ell^- = 0$.

Simplifying transmission losses Using one variable for flow in each direction as described above adds $|\mathcal{L}| \cdot |\mathcal{T}|$ variables to each scenario subproblem (over 260,000 variables in our test case), which can have a large impact on subproblem solution time. Furthermore, note that for branches with negligible losses, i.e. $\eta_\ell^{\mathcal{L}} \approx 1$, replacing f_ℓ with f_ℓ^+, f_ℓ^- increases degeneracy, as described next. Observe that for given $\hat{f}_\ell^+, \hat{f}_\ell^-$, any f_ℓ^+, f_ℓ^- such that $f_\ell^+ - f_\ell^- = \hat{f}_\ell^+ - \hat{f}_\ell^-$, and such that $f_\ell^+ + f_\ell^- \leq F_\ell$, will be a feasible solution with same objective value. Degeneracy can extend the crossover time, i.e. the time needed to obtain a basic feasible solution from an interior solution with same cost, which can be problematic in Mixed Integer Linear Program (MILP) solvers.

To mitigate these undesired effects, we propose partitioning the set of transmission branches into *lossy* branches $\mathcal{L}^\bullet$ and *lossless* branches $\mathcal{L}^\circ$. Lossy branches are modeled as described above, while flow through lossless branches is modeled with a single sign-unconstrained variable f_ℓ. A threshold r^{thresh} is defined to assign each branch ℓ to one

of the two sets:

$$\ell \in \begin{cases} \mathcal{L}^{\circ} & \text{if } 1 - \eta_{\ell}^{\mathcal{L}} > r^{\text{thresh}} \\ \mathcal{L}^{\bullet} & \text{otherwise} \end{cases} \tag{4.12}$$

Testing across different loading configurations in our test case, we have found that $r^{\text{thresh}} = 5 \cdot 10^{-3}$ p.u. reduces the number of additional variables by over 50% and the subproblem solution time by nearly 50%, while still accounting for more than 80% of the losses in the system. This is discussed in more detail in Chapter 5.

Spurious losses As seen in Fig. 4.3.1, making f_{ℓ}^{+} and f_{ℓ}^{-} simultaneously nonzero artificially increases the losses in a transmission element. We call these artificial losses spurious losses. This phenomenon can be prevented, at high computational cost, by introducing a binary variable for each branch and time period to enforce $f_{\ell}^{+} \cdot f_{\ell}^{-} = 0$. This is normally considered unnecessary, because spurious losses usually increase costs and would hence be naturally avoided by the optimization model. However, during certain periods, if the Locational Marginal Prices (LMPs) at the ends of a transmission element are negative, creating spurious losses will be optimal.

In the absence of policy incentives, negative LMPs are related with congestion during unusual loading conditions and are rare [32]. In [74], they have been found to occur in less than 0.2% of operating hours. The introduction of RPS constraints can create incentive for artificial load loss (through transmission and storage loops) if not implemented properly. A way of addressing this issue without increasing the computational burden is discussed next.

Renewable Portfolio Standards

The RPS constraint dictates that at least a fraction ρ^{RPS} of total energy generated during a certain horizon (usually, annualy) should come from qualified renewable sources. Given the nature of the data whence we obtain the scenarios in our stochastic program, scenarios correspond to representative days of the target year, not to different realizations thereof. Therefore, the RPS constraint should not be enforced for each scenario. Rather, it is formulated using the expectation across scenarios as presented in (4.10). This constraint presents a challenge for both computational tractability and model validity.

First, the expectation constraint couples the second-stage variables across scenarios. This is not an issue in [35] because the extensive form of the problem is solved. However, when implementing decomposition algorithms as we do in this work, second-stage decision variables must be independent across scenarios. A possible solution is to dualize the constraint and add the corresponding Lagrangian term with multiplier λ^{RPS} to the objective, to create a reformulation that is again decomposable by scenario. This approach is justified by the existence of a Renewable Energy Certificates (REC) market where utilities may trade surplus and shortfall of renewable energy production across days to satisfy the RPS mandate. Nonetheless, it poses the additional problem of finding the right value for λ^{RPS}, the REC price.

In [71], where this approach is taken, three methods are proposed: (a) using the value from real-world REC markets, (b) using the value of the dual variable of (4.10) in the solution to the linear relaxation of the stochastic program in extensive form (impractical for very large instances like ours), and (c) performing a sensitivity analysis until the achieved proportion of renewable energy in total production is satisfactory. We implement the third option.

The second challenge, to which we alluded in the previous section, is inducing spurious losses through negative LMPs. During days with low demand and high renewable availability, the REC market can depress LMPs and make them negative in certain parts of the grid. If this is widespread, the accuracy of the model is compromised, since there is an incentive to under-invest in renewable generation facilities, and meet RPS targets by way of creating artificial loads during such days instead. Tests in our model showed that if left uncontrolled, spurious losses could increase to over 50% of total demand during certain operating hours. Such an unrealistic behavior would put the validity of the whole model into question.

To address this issue while avoiding the computational burden of adding binaries to prevent spurious losses, we make the RPS constraint a soft constraint enforced at each scenario, so constraint (4.10) now takes the form of (4.13).

$$p_{\omega}^{\mathrm{NC}} + \sum_{\substack{b\in\mathcal{B},g\in\mathcal{G}^{\mathcal{R}}\\ t\in\mathcal{T}}} p_{b,g,t,\omega}^{\mathcal{G}} \geq \rho^{\mathrm{RPS}} \sum_{\substack{b\in\mathcal{B},g\in\mathcal{G}\\ t\in\mathcal{T}}} p_{b,g,t,\omega}^{\mathcal{G}} \;\forall\; \omega \in \Omega \tag{4.13}$$

$p_{\omega}^{\mathrm{NC}} \geq 0$ is the soft constraint's slack variable, i.e. i.e. the RPS target non-compliance of each scenario, when there is some. A term $\lambda^{\mathrm{RPS}} p_{\omega}^{\mathrm{NC}}$ is added to the operational costs of each scenario, as shown in (4.14).

$$C_{\omega}^{\text{op-RPS}} = C_{\omega}^{\mathrm{op}} + \lambda^{\mathrm{RPS}} p_{\omega}^{\mathrm{NC}} \;\forall \omega \in \Omega \tag{4.14}$$

This model corresponds to a case where there is no REC market; instead, penalties are assessed at the end of each representative horizon (day) and charged to utilities found to be out of compliance. Our model is likely to over-incentivize renewable investments and over-estimate expected operational costs, compared to an implementation of the RPS policy as an expectation constraint, but is an acceptable compromise to include transmission losses in the model.

4.3.7 CEP model

The two-stage stochastic, nodal capacity expansion planning problem that we consider in this thesis is thus given by putting all of the above together, shown in compact form in (4.15).

$$
\begin{aligned}
& \min C^{\text{inv}} + \mathbb{E}_{\zeta}\left[C_{\omega}^{\text{op-RPS}}\right] \\
\text{s.t.} \quad & A^{\diamond} x \geq b^{\diamond} && \text{(4.15a)} \\
& M_{\omega}^{\diamond} x + N_{\omega}^{\diamond} y_{\omega} \geq \xi_{\omega}^{\diamond} \ \forall \omega \in \Omega && \text{(4.15b)} \\
& x \in \mathcal{X}, y_{\omega} \in \mathcal{Y}_{\omega} \ \forall \omega \in \Omega && \text{(4.15c)}
\end{aligned}
$$

x represents a concatenation of all investment variables $x^{\mathcal{G}}, x^{\mathcal{S}}, x^{\mathcal{S}\text{-PC}}, x^{\mathcal{L}}$, with $\mathcal{X}$ an abstract representation of their domains (4.2). y_{ω} represents a concatenation of all operation variables $p_{\omega}^{\mathcal{G}}, p_{\omega}^{\mathcal{S}}, p_{\omega}^{\mathcal{S}\text{-ch}}, p_{\omega}^{\mathcal{S}\text{-dch}}, p_{\omega}^{\text{sh}}, f_{\omega}, f_{\omega}^{+}, f_{\omega}^{-}, p_{\omega}^{\text{NC}}$ for scenario ω, with $\mathcal{Y}_{\omega}$ an abstract representation of their domains (4.3). Constraint (4.15a) corresponds to construction limits constraints (4.4). Finally, constraint (4.15b) represents operation constraints (4.5)-(4.9), and the soft RPS constraint (4.13). We have used superscript $\diamond$ in all symbols to avoid confusion with other variables and parameters that use the same letters.

4.4 Solution approach with parallel computing

Problem (4.15) is a MILP, so in principle, one could try solving it by passing it to a commercial solver. In the stochastic programming field, this is called solving the problem in *extensive form*. Our problem, however, suffers from the curse of dimensionality: we are interested in solving an instance of the problem on a system with almost 9,000 buses, on at least hundreds of scenarios. At this scale, solving our MILP in extensive form without some decomposition can be very difficult.

Instead, we use an implementation of the PHA, and make adaptations as needed to solve the capacity expansion problem posed. Detailed and rigorous descriptions of the PHA can be found in [30, 33, 46, 87, 110], among many other references. Here, we provide only a high-level description of the strategy of the decomposition and the algorithm, applied to our two-stage stochastic program, followed by a discussion of the main features of our implementation that differ from conventional implementations of the PHA.

4.4.1 The Progressive Hedging Algorithm

We first offer an overview of the strategy behind the scenario decomposition in the PHA, and then present the algorithm itself. This is meant as illustration for the unfamiliar reader and not as a rigorous exposition. The interested reader is encouraged to consult the aforementioned references for more details.

Decomposition strategy

Three steps are taken to decompose the two-stage problem (4.15) into a set of subproblems, one for each scenario:

1. Replicate first-stage variables
2. Relax and dualize the non-anticipativity constraint
3. Add a penalty to induce convergence

Replicating first-stage variables The first step is replicating the first-stage (investment) variables x, creating one copy x_ω for each scenario ω. x is substituted with x_ω in constraints (4.15a) and (4.15b), as well as in the expression of investment cost C^{inv}, which we now call C^{inv}_ω. To ensure that the resulting optimization problem (4.16) is equivalent to (4.15), a constraint is added to make all copies equal to each other (4.16c). This constraint is called the non-anticipativity constraint, because if it is absent, a different investment x_ω could be chosen for each scenario ω, *anticipating* the output of the uncertain second stage.

$$\begin{aligned}
&\min \sum_{\omega\in\Omega} \pi_\omega \left(C^{\text{inv}}_\omega + C^{\text{op-RPS}}_\omega\right) \\
\text{s.t.}\quad & A^\diamond x_\omega \geq b^\diamond \ \forall\omega\in\Omega && \text{(4.16a)}\\
& M^\diamond_\omega x_\omega + N^\diamond_\omega y_\omega \geq \xi^\diamond_\omega \ \forall\omega\in\Omega && \text{(4.16b)}\\
& \pi_\omega x_\omega = \pi_\omega x \ \forall\omega\in\Omega && \text{(4.16c)}\\
& x, x_\omega \in \mathcal{X} \ \forall\omega\in\Omega, y_\omega \in \mathcal{Y}_\omega \ \forall\omega\in\Omega && \text{(4.16d)}
\end{aligned}$$

Dualizing non-anticipativity constraint and decomposing into subproblems The next step is dualizing the non-anticipativity constraint (4.16c), assigning it multipliers w_ω. Without (4.16c), there are no longer any constraints coupling different scenarios. The consensus first-stage decision variable x is replaced with a fixed parameter $\hat{x}$, which will be updated iteratively. Optimization problem (4.17) results, which can then be decomposed into the subproblems given in (4.18).

$$\begin{aligned}
&\min \sum_{\omega\in\Omega} \pi_\omega \left(C^{\text{inv}}_\omega + C^{\text{op-RPS}}_\omega + w_\omega\left(x_\omega - \hat{x}\right)\right) \\
\text{s.t.}\quad & A^\diamond x_\omega \geq b^\diamond \ \forall\omega\in\Omega && \text{(4.17a)}\\
& M^\diamond_\omega x_\omega + N^\diamond_\omega y_\omega \geq \xi^\diamond_\omega \ \forall\omega\in\Omega && \text{(4.17b)}\\
& x_\omega \in \mathcal{X} \ \forall\omega\in\Omega, y_\omega \in \mathcal{Y}_\omega \ \forall\omega\in\Omega && \text{(4.17c)}
\end{aligned}$$

$$
\begin{aligned}
& \min C_\omega^{\text{inv}} + C_\omega^{\text{op-RPS}} + w_\omega \left(x_\omega - \hat{x}\right) \\
\text{s.t.} \quad & A^\diamond x_\omega \geq b^\diamond && \text{(4.18a)} \\
& M_\omega^\diamond x_\omega + N_\omega^\diamond y_\omega \geq \xi_\omega^\diamond && \text{(4.18b)} \\
& x_\omega \in \mathcal{X}, y_\omega \in \mathcal{Y}_\omega && \text{(4.18c)}
\end{aligned}
$$

Adding a quadratic penalty to induce convergence Finally, the quadratic penalty term $\frac{1}{2}\rho \left\|x_\omega - \hat{x}\right\|^2$ is added to the objective function. In the convex case, this term is important to obtain strongly convex subproblems, so that optimal solutions are unique, and guarantee convergence of the algorithm. Note that this term vanishes when consensus is reached, i.e. when $x_\omega = \hat{x}\ \forall\omega$.

$$
\begin{aligned}
& \min C_\omega^{\text{inv}} + C_\omega^{\text{op-RPS}} + w_\omega \left(x_\omega - \hat{x}\right) + \frac{1}{2}\rho \left\|x_\omega - \hat{x}\right\|^2 \\
\text{s.t.} \quad & A^\diamond x_\omega \geq b^\diamond && \text{(4.19a)} \\
& M_\omega^\diamond x_\omega + N_\omega^\diamond y_\omega \geq \xi_\omega^\diamond && \text{(4.19b)} \\
& x_\omega \in \mathcal{X}, y_\omega \in \mathcal{Y}_\omega && \text{(4.19c)}
\end{aligned}
$$

The algorithm

The classic PHA consists in iteratively solving each subproblem (4.19) and updating the multipliers w_ω and consensus variable $\hat{x}$ until some desired level of consensus is reached for all first-stage variables across all subproblems (scenarios). A description is provided in Algorithm 6.

Caveats in the nonconvex case The convergence of the PHA, as well as the optimality of the result obtained at convergence, are only guaranteed in the convex case. However, since our capacity expansion problem has mixed integers, it is not convex, and has no such guarantees. The PHA for stochastic MILP is thus considered only a heuristic methodology. It can be, however, a good heuristic, as discussed next.

4.4.2 Acceleration heuristicsis

The PHA is a very popular algorithm with numerous extensions and variants. Again, more extensive reviews can be found in [33, 46, 110, 109]. Here, we do a quick overwiev of some features implemented in our software to overcome the two challenges mentioned above: accelerating convergence and obtaining optimality guarantees.

/* 0-th iteration */
$k \leftarrow 0$
for $\omega \in \Omega$ **do**
  $x_\omega^{(k)} \leftarrow \arg\min_{x_\omega, y_\omega} \left(C_\omega^{\text{inv}} + C_\omega^{\text{op-RPS}}\right)$:(4.19a)-(4.19c)
end
$\hat{x}^{(k)} \leftarrow \sum_{\omega \in \Omega} \pi_\omega x_\omega^{(k)}$
for $\omega \in \Omega$ **do**
  $w_\omega^{(k)} \leftarrow \rho\left(x_\omega^{(k)} - \hat{x}^{(k)}\right)$
end
/* main PHA algorithm */
repeat
  $k \leftarrow k + 1$
  for $\omega \in \Omega$ **do**
    $x_\omega^{(k)} \leftarrow \arg\min_{x_\omega, y_\omega} \left\{\left(C_\omega^{\text{inv}} + C_\omega^{\text{op-RPS}} + w_\omega^{(k)}\left(x_\omega - \hat{x}^{(k)}\right) + \frac{1}{2}\rho\left\|x_\omega - \hat{x}^{(k)}\right\|^2\right)\right.$
    $\left.:(4.19a)\text{-}(4.19c)\right\}$
  end
  $\hat{x}^{(k)} \leftarrow \sum_{\omega \in \Omega} \pi_\omega x_\omega^{(k)}$
  for $\omega \in \Omega$ **do**
    $w_\omega^{(k)} \leftarrow \rho\left(x_\omega^{(k)} - \hat{x}^{(k)}\right)$
  end
until $\sum_{\omega \in \Omega} \pi_\omega \left\|x_\omega^{(k)} - \hat{x}^{(k)}\right\|^2 < \epsilon$ *or* $k >$ *iteration limit*

Algorithm 6: Classic Progressive Hedging Algorithm

Accelerating convergence

In the PHA, the stochastic program is decomposed by scenario. As for any decomposition algorithm, the key to a successful implementation is that each subproblem can be solved fast. This section highlights measures undertaken to ensure that individual subproblems can be solved in a relatively short time and convergence of first-stage variables (consensus) occurs quickly.

Linearize quadratic term Despite recent advances in commercial solvers, quadratic mixed integer programs remain harder to solve than MILPs and the quadratic (also called *proximal*) terms (introduced in step 3 of the PHA scenario decomposition) can considerably delay the subproblem solution times. To reduce this effect, we linearize the proximal terms in the PHA subproblems. For binary variables, the quadratic term automatically reduces to a linear term, so no special linearization is necessary. For integer and continuous variables, an extra variable is added to represent the squared variables, and the quadratic penalty

is under-approximated by successive linear cuts added as needed during the execution of the algorithm [109]. It may seem inconsistent to take this approach here but discard doing something similar to approximate quadratic transmission losses. However, note that we only add linear sections for active first-stage variables, which is a much smaller number than the number of lossy transmission branches in each scenario subproblem. This linearization procedure is already included as a feature in the current mpi-sppy distribution.

Accelerating crossover The branch and cut algorithm implemented by commercial MILP solvers like Gurobi requires a basic feasible solution to the linear relaxation of the problem – the so-called root relaxation on which the algorithm branches. By design, the Simplex Algorithm returns such a solution. For problems of large size, like the one considered here, interior point algorithms are much faster, but require a crossover phase to convert their solution into an extreme point optimal solution. The crossover process can be accelerated by reducing degeneracy in the problem formulation. In our case, a few simple model improvements like adding a reference bus and simplifying the model for branches with negligible losses as described in section 4.3 were used to improve crossover time.

Setting ρ The value chosen for the parameter ρ, the coefficient associated with the quadratic penalty term and the weight update step, is known to have a great impact on the convergence of the PHA. Although the standard description of the PHA assumes that the same parameter ρ is used for all non-anticipative variables, it has been found that better results can be obtained if the parameter is made variable-dependent. In [71], generator investment variables are assigned different values of ρ from transmission investment variables, so as to maintain an approximately constant ratio of penalty/(investment \$). In [110], it is found that even better results can be obtained if the value of ρ is determined individually for each variable after assessing the level of agreement in the optimal solution across scenarios at the first PHA iteration. The authors call this method *sep-rho*. Here, we implement the *sep-rho* method as the ρ-setting policy.

Fixing variables A common heuristic in the PHA literature is to fix variables when there has been agreement across scenarios for a number of iterations. This can significantly reduce the time to convergence, especially in the presence of binary or integer variables. The caveat associated with this heuristic is that it may lead to suboptimal solutions, since variables might get fixed to suboptimal values early in the algorithm. Although there is no guarantee that this can be avoided, the availability of optimality gaps somewhat alleviates this concern. This heuristic is already included in mpi-sppy, so its implementation, as well as experimentation with different values for the number of iterations that triggers the fixing, is greatly simplified. In our problem, the largest impact was observed when fixing investment variables, in particular transmission lines, after agreement around the lower bound, i.e. eliminating candidates that are not built in any of the scenarios considered.

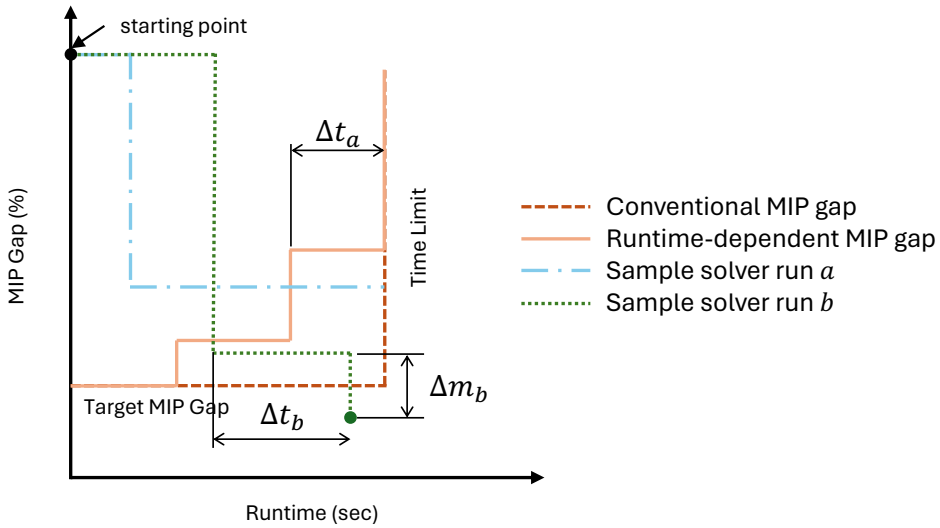

Figure 4.4.1: Runtime-dependent MIP gap. Solver execution starts from the point in the top-left corner and finishes the first time that the MIP gap curve is crossed. Conventional solver behavior allowing a fixed target MIP gap or a runtime limit is represented by the red dashed curve. By implementing the solid orange curve instead, a time Δt_a on sample run a can be saved; a time Δt_b can be saved in sample run b, sacrificing Δm_b of the final optimality gap obtained.

Runtime-dependent MIP gap To simultaneously avoid spending too much time on one individual subproblem and avoid sacrificing MIP quality at later PHA iterations, a time-dependent MIP gap scheme was implemented, where the MIP gap was increased from the initial target value after some time of solver execution. This is depicted in Fig. 4.4.1.

4.4.3 Obtaining optimality gaps

The mpi-sppy package includes various types of inner and outer bound spokes. Several combinations were tested, with satisfactory results being obtained when using the *Lagrangian* outer bound spoke and *looper* inner bound spoke. Our tests have not been exhaustive, so other spoke combinations may behave better as more scenarios or other power systems are considered in the continuation of this work.

4.5 Computational study

At the time of development of this research, the interested consisted in verifying that we can solve a large instance of our model by the method described in the previous section in a reasonable time. To do that, we implement the stochastic optimization model on a large, realistic power system and vary the number of scenarios included.

4.5.1 Test case and scenarios

CATS system We test our model on an extension of the CATS system developed in [101], adapted for capacity expansion planning [72]. The CATS is a geographically realistic representation of the California generation and transmission system, which includes over 8,000 buses and 10,000 transmission branches, with locations based on publicly available data and simulated, but realistic values for electrical parameters that are not publicly available.

RPS target Current California legislation states an RPS target of 60% by 2035, with a goal of 100% renewable energy by 2045 [15]. With current technology, such a target, if not interpreted as a *net* target, can only be obtained at prohibitively expensive costs. For the tests conducted here, an intermediate but still ambitious RPS target of 70% by 2045 was chosen.

Climate-informed scenarios Load and generation profiles used are generated based on projected weather and climate data from a regional refinement for California of the E3SM global circulation model for the year 2045, using a 3km-resolution grid [116]. A probability space can be constructed from the projected data, where each element corresponds to one day of hourly data for the target year. The space is reduced to a sample space via clustering using the *k-means* algorithm on a normalized feature vector of load and generation profiles using the default implementation of *k-means* available in *scikit* [80]. This process is similar to that of [71], but using projected weather data rather than historical data. The size of the sample space is varied in different tests cases. Each sample in the sample space corresponds to a scenario of the stochastic optimization problem.

Algorithm and computational parameters All tests were done on the *quartz* HPC cluster at LLNL. Each node of the cluster has 36 2.1GHz Intel Xeon cores and 128GB RAM. Within each cylinder, 3 CPU cores are dedicated to each optimization subproblem, with each subproblem corresponding to a scenario of the stochastic optimization problem. Three cylinders were used for each test: one hub, one lower-bounder and one upper-bounder, so that each test instance uses 9 CPU cores per scenario. Our model was implemented using mpi-sppy 0.11.1.2 and *Pyomo* 6.2. Gurobi 10.0.2 was used to solve the MILPs.

The different techniques described here that have been used to solve this problem introduce numerous parameters, and hence degrees of freedom, into our algorithm. For the purpose of this study, parameters have only been varied from their default values to the extent necessary to obtain a solution within the time limit imposed. Although there is room for further exploration of the interaction between solution time and solution quality across test instances, fixed parameter settings across test instances were found to be sufficient to achieve the desired MIP gap within the time limit imposed.

4.5.2 Results and discussion

Table 4.5.1 summarizes computational results for all instances tested. As expected for problems of this size, all instances hit the iteration limit before PHA convergence. In all cases, a solution with an optimality gap below 5% is obtained in under 4h. This gap is in the vicin-

ity of 2% for the cases with over 100 scenarios, which may be of more interest to decision makers. These results suggest that when maintaining a constant ratio of cores per scenario subproblem, the growth in total computation time as more scenarios are included remains tractable.

We expect that accounting for the uncertainty in climate projections will require including many more scenarios in our pool. Nevertheless, we find these results encouraging for two reasons. First, they suggest that reasonably low computation times can be obtained by keeping the ratio of cores per scenario subproblem constant. It is worth pointing out that our largest scenario instance only used about 3% of the computational capacity of the HPC cluster, so a more than tenfold increase in the number of scenarios seems well within reach. Second, scenario reduction techniques are considered outside of the scope of this thesis, but it is clear that as many more scenarios are included, some scenario reduction methodology will be necessary. In [71], with computational limitations in mind, the initial space of samples is reduced to the order of 50 scenarios. We expect that being able to handle a much larger number of scenarios than that, together with improved scenario reduction techniques, will allow working with a larger climate uncertainty set.

As mentioned before, one important feature of mpi-sppy is providing optimality bounds during the excution of the PHA. The evolution of the bounds for each case is shown in Fig. 4.5.1. All tests were run for 60 iterations, but the progression of optimality bounds shows that, if desired, termination could have been triggered earlier if a looser optimality target had been set. Moreover, these results suggest that, as is common for this type of problems, good-quality feasible solutions can be obtained relatively quickly, within 10 or 20 iterations. After that, most of the gap reduction is provided by tightening lower bounds. In our tests, computational resources were allocated symmetrically to all cylinders. Some improvement could likely be obtained by dedicating more computational power to the more demanding, but more impactful, outer-bounding cylinder, especially after the initial iterations are finished.

Table 4.5.1: Optimality gaps achieved, total computation time and number of nodes dedicated, for various numbers of scenarios.

Number of scenarios	CPU Cores	Wall-clock time (min)	PHA Iterations	Optimality gap achieved (%)
10	90	85	60	4.8
20	180	102	60	4.1
50	450	132	60	3.8
100	900	170	60	2.1
200	1,800	200	60	2.8
360	3,240	206	60	2.4

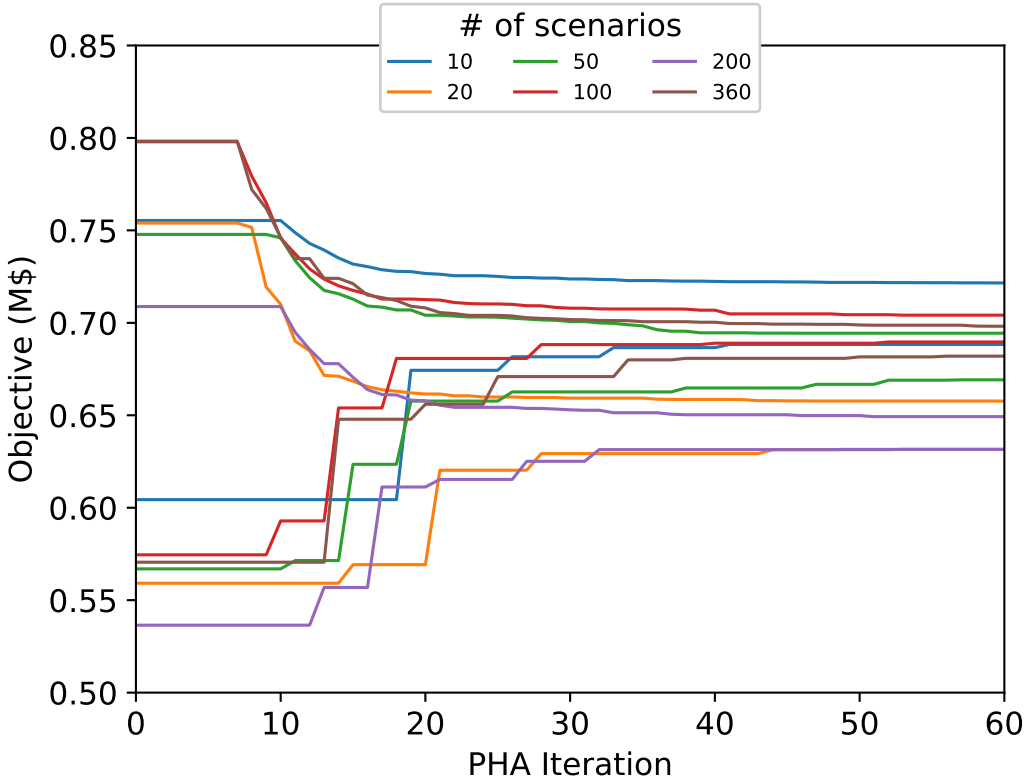

Figure 4.5.1: Optimality bounds provided by mpi-sppy during the execution of the algorithm for different number of scenarios.

In Fig. 4.5.2, we compare the optimal capacity expansion plan across instances tested. It is interesting to note that as more scenarios are considered, both the optimal cost and the optimal investment plan continue to vary. This behavior could perhaps be mitigated by more sophisticated scenario selection methods, but nonetheless highlights the importance of using a sufficiently large scenario set in the stochastic program.

It is also noteworthy that no storage facilities were built in any of our test cases, although it is imaginable that this could change if the RPS target were increased. While our focus in this chapter is on the computational aspects of our method and we leave a thorough analysis of the obtained results for future work, it seems clear that at least part of the value of storage may fail to be captured in the current model. This could be improved by including planned retirements of the existing generation fleet, as well as model modifications that highlight the value of flexibility brought by storage units: for example unit commitment and ramping, representation of security constraints, and improved power flow representation.

RPS penalty and achieved RPS A sensitivity analysis on the value of the soft constraint penalty λ^{RPS} in (4.13) was performed using the case with 20 scenarios (Fig. 4.5.3).

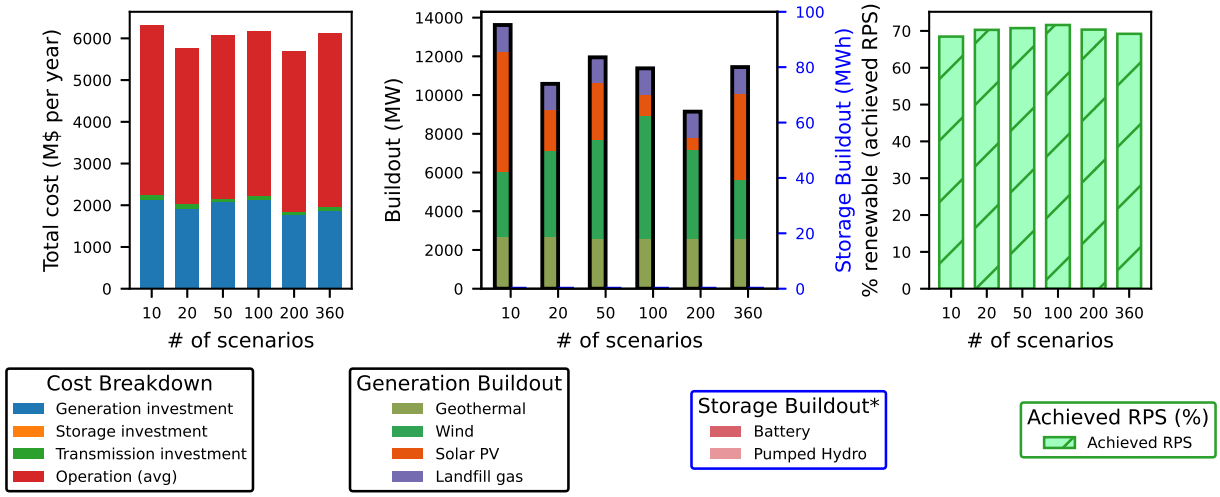

Figure 4.5.2: Comparison of solution results for various numbers of scenarios. *Left*: Breakdown of optimal expected cost. *Center*: Generation and storage buildout. *Note that no storage was built in any of our test cases. *Right*: Expected proportion of energy generated with renewables.

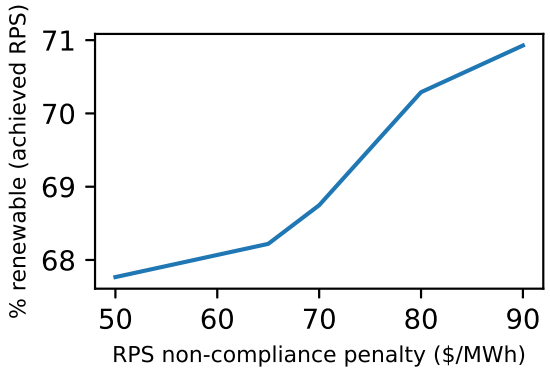

Figure 4.5.3: Sensitivity of proportion of total energy served with renewables to RPS penalty λ^{RPS}. Tests were performed using the instance with 20 scenarios. Based on these results, a penalty of \$80/MWh was used for the value of λ^{RPS} for all test instances.

A value around \$80/MWh resulted in produced renewable energy near the target of 70%, so this was the value used for all the other test cases. This result appears to be relatively stable as the number of scenarios vary, as can be seen in Fig. 4.5.2, although the investments required to maintain it do change.

Spurious losses It is worth discussing why spurious transmission losses are a problem but spurious storage losses from simultaneous charging and discharging do not seem to be.

This difference can be understood by comparing the *total* potential spurious losses and the *available* potential spurious losses of each type. By total potential spurious losses we refer to the total losses incurred if all resources were used at full capacity[9]: in our test system, this is in the order of 200MW for storage facilities, and 16GW for transmission assets. So the potential for impact is orders of magnitude higher with transmission than storage. To illustrate the second factor, note that the negative LMPs that drive spurious losses also encourage using storage facilities to store energy for future use (in a genuine, non-spurious way). Thus, using storage facilities to create spurious losses only makes sense if the storage facility is already full, which reduces the potential for spurious storage losses. This does not happen for transmission elements. The combination of these two factors explains why considering transmission losses has such a high impact on the relevance of spurious losses on our model.

As stated in section 4.3.6, the change in the RPS constraint eliminates the induced negative LMPs, other than those potentially caused by congestion. As expected, spurious losses associated with these were found to be negligible in all tests.

4.6 Conclusions

We have extended the stochastic capacity expansion model of [35], with the goal of representing a power system of sufficiently high-resolution to adequately capture climate variability and geographically realistic generation, storage, and transmission expansion decisions. To handle the high-resolution geographical and temporal data required for this improved accuracy, we implemented a variant of the model that could be decomposed by scenario and solved in an HPC environment using an implementation of PHA specifically tuned for our problem. mpi-sppy was used to facilitate the parallelization of the algorithm, as well as for obtaining optimality bounds during algorithm execution. Tests were conducted on a full-resolution, synthetic but realistic representation of the California generation and transmission system, with scenarios based on a high-resolution climate projection for 2045. Instances with up to 360 representative days were satisfactorily solved to under 5% optimality gaps within a few hours.

We envision the continuation of this research along two main directions: validation and improvement of the model, and improvement of the computational solution. On the former, a sensitivity analysis of the impact of key model parameters and modeling assumptions (e.g. the chosen representation of power flow and the presence of resource flexibility constraints) on the obtained investment plan needs to be better understood. Part of this work is addressed in Chapter 5, where we look into the impact of different power flow formulations on the solution quality and time.

On the latter, our tests suggest that a significant performance improvement can be achieved by adopting an asymmetric management of computational resources across cylin-

[9]This is not a feasible operation status, but serves as an upper bound for the total losses in the system, and hence for spurious ones.

ders. In addition, addressing the difference in computational time among scenarios by way of an asynchronous decomposition of the optimization problem through scenarios, could have a significant impact on both solution time and quality. This is a functionality that has been recently added to mpi-sppy [30] and would constitute a natural extension of this work in the short term.

Chapter 5

Improved linear power flow representation for Capacity Expansion Planning

Abstract

In this chapter, we consider four additional versions of the Capacity Expansion Planning (CEP) model of Chapter 4, obtained adding representations of power flow constraints using the $b\theta$ and Power Transfer Distribution Factors (PTDF) formulations of DC power flow, with and without transmission losses. We present special modeling performed to tractably include transmission losses in the PTDF formulation of CEP. The performance in terms of computational time and solution quality of all formulations is tested on 365 different loading and generation conditions of a 500-bus realistic testcase, to obtain comparisons and issue a recommendation for usage on the Progressive Hedging Algorithm (PHA) implementation of the stochastic CEP in future work. We find that the PTDF should be preferred over the $b\theta$ formulation, although neither significantly outperforms the network flow representation of Chapter 4 in the tests performed.

Acknowledegment of individual contributions

The work presented in chapters 4 and 5 is the result of a collaboration with Lawrence Livermore National Laboratory (LLNL), during two summer internships at LLNL and subsequent LLNL-funded graduate student research appointments at UC Berkeley. As is natural in such collaborations, the efforts of different people contributed to the resulting work. I do not intend to claim that everything presented in this chapter is solely the result of my work. This section clarifies individual contributions of the LLNL team members to the work presented here (see also the analogous section in Chapter 4).

Minda Monteagudo procured the climate projection data and the electricity load forecast data used as baseline to obtain the weather-adjusted load timeseries. The model to estimate weather impacts on electricity demand is also her work [67]. Matthew Signorotti developed the software tool that produces timeseries of potential generation availability for wind and solar power from the output of downscaled projected climate data [95], and provided the generation availability timeseries used in all tests.

The individual contributions of the author of this thesis include:

- the extension of the PTDF model of [85] to the joint CEP model, as well as the conceptualization of the representation chosen for transmission losses in the PTDF formulation;
- the elaboration of the testcase used in this chapter, based on the power system data of [8] and the cost data gathered by Amelia Musselman for the California Test System (CATS) testcase [72] ;
- and all software implementation and experiment design.

Jean-Paul Watson was the project lead; he and Amelia Musselman contributed numerous useful suggestions during weekly conversations.

5.1 Introduction

It is common in generation CEP models to use a zonal representation of the power grid, in which only the aggregate transmission capacity between zones is modeled, not the actual transmission network. In this work, we include transmission expansion in the expansion planning process, so a more accurate representation of the network is needed. In chapter 4, a nodal model where individual transmission lines are represented was proposed. However, this model used a regular network flow formulation, often called in the literature a transportation formulation. In the network flow model, nodal power balance and thermal capacity of transmission lines are satisfied, but the physics that dictate how electricity flows in a grid are disregarded. In other words, it is assumed that at each node of the network, the operator may route power as desired through each connected element. The transportation formulation is thus a relaxation of a more accurate model that considers power flow constraints, and can potentially be overly optimistic, i.e. it may avoid necessary transmission investments by routing power through the grid in unrealistic ways, leading to suboptimal investment plans. Neglecting transmission losses can have similar consequences. Ignoring them may result in overestimating the capacity of transmission lines, thereby producing a suboptimal plan that ultimately leads to unforeseen load shedding and thus overcosts.

To maintain the linearity, and hence computational tractability, of the optimization model, the so-called DC representation of power flow is often used in CEP models (see [22] and references therein). This includes the models used as base for our model [35, 71]. However, as is often the case in Mixed Integer Linear Programs (MILPs), computational

tractability is not guaranteed at scale. Even this conventional representation can significantly increase the solution time of Transmission Expansion Planning (TEP) models, driving researchers to look for alternatives, like the PTDF formulation.

Incorporating the switching of transmission lines into the PTDF model has been a standard practice in power systems for contingency analysis, through so-called line outage distribution factors and compensated PTDFs (see e.g. [112]). The strategy there is to use a PTDF formulation that can be used for the normal conditions of the grid, as well as scenarios where single (or several) transmission elements are taken out of service. However, in those cases, the switching of transmission elements occurs because of contingencies, i.e. they are parameters of the analysis, not a decision variable. In [89], a model for optimal transmission switching is proposed where virtual power injections are added to cancel power flows in outaged branches, the same principle of compensated PTDF matrices in line outage analysis.

The only example we have found where this methodology is extended to transmission expansion planning is [85], where only transmission expansion, not generation and storage, are considered. The PTDF formulation of DC power flow is used in security-constrained generation capacity expansion in [43], but only generation, not transmission and storage, is considered. In this chapter, we incorporate the formulation of [85] into our joint generation, storage and transmission CEP model.

The effect of including transmission losses in TEP models has been studied in the literature before [32, 74]. In [74], a joint generation, transmission and storage CEP problem is solved for the European interconnection, finding differences of as much as 66% between energy transmitted long-distance vs. generated locally when transmission losses are included in the model, and associated overcosts in the vicinity of 6%. They use however a zonal representation of power flow, with 250 nodes for the entire continent.

The model in [85] is a lossless model. We are not aware of any examples in the literature of lossy PTDF formulations for capacity expansion. In [31], a methodology is proposed to include losses in a PTDF formulation of power flow for an operational problem: finding local marginal prices. In that methodology, a linearization of power flow is obtained starting from a solution to the AC power flow problem. This approach would require iterative updating when the loading conditions change, which would make a computational implementation of this method harder for a glscep problem.

In this chapter, we propose improvements to the model presented in the previous chapter by considering all combinations of the two dimensions mentioned in the paragraphs above: we test the $b\theta$ and PTDF formulations of power flow, in their variants neglecting and considering transmission losses, and analyze their impact on the solution quality and solution time of the investment plan obtained.

The main contributions of this chapter are two joint generation, transmission and storage CEP models using the PTDF formulation of DC power flow, of which one includes transmission losses. We provide justification for the heuristics used in the modeling of transmission losses based on a systematic study of numerous load and generation conditions. Finally, our tests comparing the performance of the different power flow representations along two

dimensions: solution quality and time to reach Mixed Integer Program (MIP) gaps are also a contribution of this work.

Structure of this chapter

This chapter constitutes a self-contained sequential unit with Chapter 4; familiarity with the contents of Chapter 4 is assumed. The other chapters of this dissertation are not a prerequisite for understanding this content.

In Section 5.2, we cover some preliminaries of the two formulations of DC power flow used in our models: $b\theta$ and PTDF. We also cover the models for switching transmission elements in both formulations, with focus on the less common flow-cancelling model for switching in the PTDF formulation. In Section 5.3, we present the models proposed in this work, as well as some empirical justification for the approximations used in the lossy PTDF formulation. In Section 5.4 we briefly present the testcase developed for all tests. In Section 5.5, we present and discuss the results of the performance tests. Finally, we conclude in 5.6

5.2 Preliminaries

In this section, we present an overview of some necessary preliminaries for introducing the model improvements of linear power flow in CEP models that we propose in this chapter. We first overview the $b\theta$ and PTDF formulations of power flow, and then present how each can be incorporated into the CEP model introduced in Chapter 4.

5.2.1 Nomenclature

We list here only the nomenclature introduced for the formulations introduced in this chapter, i.e. those not already included in section 4.3.2, as well as any notation that deviates from the previous chapter.

Set indices

$\mathcal{B}$: To avoid confusion with a branch's susceptance, in this chapter we avoid using b to index the set of buses $\mathcal{B}$. i, m, n are used instead as noted.

Parameters

Electrical parameters

χ_ℓ: reactance of branch ℓ, in p.u.

r_ℓ: resistance of branch ℓ, in p.u.

b_ℓ: susceptance of branch ℓ, in p.u.

Sensitivity matrices

Ψ: PTDF matrix of the network,

L: sensitivity matrix of transmission branch losses to nodal power injections,

Decision variables

Operation variables

$\theta_{i,t,\omega}$: voltage angle at bus i during period t of scenario ω, in radians.
$\tilde{f}_{\ell,t,\omega}$: virtual injection to cancel flow through candidate branch ℓ during period t of scenario ω, in MW
Auxiliary variables
$p^{\text{inj}}_{i,t,\omega}$: Net power injection at bus i during period t of scenario ω, in MW.
$f^{\mathcal{L}^\bullet}_{\ell,t,\omega}$: Power lost to thermal dissipation in branch ℓ during period t of scenario ω, in MW.
Auxiliary parameters
σ_ℓ: direction of flow on restricted lossy branch $\ell \in \mathcal{L}^\bullet$,
$\varrho_{\ell,\omega}$: metric for the bidirectionality of branch ℓ across the time periods of scenario ω,
δ^d: threshold for metric $\varrho_{\ell,\omega}$, above which branch ℓ is restricted to one direction,
r^{thresh}: threshold for branch efficiency. For $\eta^{\mathcal{L}}_\ell$ above $1 - r^{\text{thresh}}$, branches are considered lossless.
Notation
Unless otherwise specified, the vector is noted by omitting the corresponding index, e.g. $f = [f_\ell]_{\ell \in \mathcal{L}}$. A colon (:) is sometimes used as a subscript to signify all the elements in a dimension of a matrix, to emphasize that an entire row or column is sliced. For example, $\Psi_{:,b}$ is the b-th column of matrix Ψ, and $\Psi_{\ell,:}$ is the ℓ-th row of matrix Ψ.

5.2.2 Lossless linear power flow

We start with a quick overview of DC power flow in the lossless case. We follow here the exposition of [31] and briefly overview the DC power flow model. A complete derivation of it from Ohm's and Kirchhoff's laws is omitted; we refer the interested reader to the aforementioned reference or classic textbooks like [37, 100, 113] for a more detailed analysis.

The power flow problem is posed to identify, given a set of balanced power injections and withdrawals in an electricity network, how electricity flows through the transmission network. Power flow in electricity grids depends on voltage magnitudes and angles at each node of the network, according to a nonlinear relation. Given the computational challenge of solving a large, nonlinear mixed integer optimization problem, it is common in TEP to sacrifice some model accuracy by adopting three key assumptions:

- Voltage magnitudes are very close to their rated values (1 p.u.),
- Voltage angle differences between branch ends are small ($\sin \Delta\theta \approx \Delta\theta$),
- Resistance of transmission elements is negligible compared to their reactances: $r_\ell \ll \chi_\ell$, so that losses in transmission elements can be neglected.

These assumptions allow some approximations that result in a linear representation of power flow, commonly called DC power flow. We next present the conventional formulation of DC power flow, often called the $b\theta$ formulation, and the equivalent (in the lossless case) PTDF formulation, which presents some advantages in certain circumstances.

bθ formulation

To lighten the notation, we omit subscripts ω and t for scenario and period in this section. Consider a transmission element ℓ, with end nodes which are arbitrarily labeled *origin* and *destination*, and denoted by $o(\ell)$ and $d(\ell)$ respectively. Let θ_m be the voltage angle at node m, and b_ℓ the susceptance of the branch. Then, the flow f_ℓ through ℓ, in direction $o(\ell) \to d(\ell)$, is given by (5.1a). The origin and destination ends were chosen arbitrarily; a negative value of f_ℓ indicates that power actually flows in the reverse direction. As is common in linear programming, the sign-unrestricted variable f_ℓ can be represented in split form with two non-negative variables as $f_\ell^+ - f_\ell^-$, as in (5.1b). The $b\theta$ DC power flow problem is obtained by simply adding (5.1) to the network flow formulation introduced in chapter 4. Note the right hand side of the equations, which gives this formulation its name. We will see how this equation modifies the CEP problem, but let us first present the PTDF formulation.

$$f_\ell = b_\ell \left(\theta_{o(\ell)} - \theta_{d(\ell)}\right) \;\; \forall \ell \in \mathcal{L}^\dagger \cap \mathcal{L}^\circ \tag{5.1a}$$

$$f_\ell^+ - f_\ell^- = b_\ell \left(\theta_{o(\ell)} - \theta_{d(\ell)}\right) \;\; \forall \ell \in \mathcal{L}^\dagger \cap \mathcal{L}^\bullet \tag{5.1b}$$

Reference bus Note that only angle differences matter in (5.1), so there are infinitely many values of voltage angles that describe the same state of the power system, as long as angle values relative to each other are maintained. To obtain a unique solution, the angle at an arbitrary bus, called the *reference* bus is fixed to a given value (without loss of generality, 0).

PTDF formulation

The purpose of the PTDF methodology is to obtain a matrix Ψ such that the flow in each branch of the network induced by an injection of 1MW at a certain node m compensated by a withdrawal of 1MW at bus n is given by[1]

$$f = \Psi \left(e_m - e_n\right),$$

where $f = [f_\ell]_{\ell \in \mathcal{L}}$ is the vector of all branch flows and e_i is the canonical unitary vector with a 1 at position n and zeros everywhere else. By considering the superposition of multiple such unitary injections/withdrawals, the effect of any set of balanced injections and withdrawals p^{inj} as defined in (5.2) is given by (5.3). We can thus restate the goal of the

[1] This statement is not true in general. In fact, the PTDF methodology is a linearization of power flow around a given point. So what takes the form given here is the *change* in branch flow induced by an *additional* 1MW power injection/withdrawal with respect to an initial point of operation. Since here we consider a purely linear power flow, the initial point of operation does not affect the linearization and our statement is true.

PTDF methodology as finding a matrix Ψ such that the flows in all branches of the system for any given set of balanced nodal power injections and withdrawals is given by (5.3). The matrix Ψ is called the matrix of Power Transfer Distribution Factors (PTDF), which gives the method its name.

$$p_i^{\text{inj}} = \sum_{g \in \mathcal{G}} p_{i,g}^{\mathcal{G}} + \sum_{s \in \mathcal{S}} \left(\eta_s^{\mathcal{S}\text{-dch}} p_{i,s}^{\mathcal{S}\text{-dch}} - p_{i,s}^{\mathcal{S}\text{-ch}} \right) + p_i^{\text{sh}} - D_i \ \forall i \in \mathcal{B} \tag{5.2}$$

$$f = \Psi p^{\text{inj}} \tag{5.3}$$

Since energy balance must always be satisfied, all power injections and withdrawals in p^{inj} must be compensated by injections and withdrawals of equal magnitude and opposite sign somewhere else in the network. In the PTDF methodology, this is assumed to occur at a reference bus called the *slack* bus. Note that compensations are of equal magnitude because transmission losses are neglected for now.

To derive the expression of the PTDF matrix, we start from the $b\theta$ equation (5.1), which can be written for all branches in matrix form as

$$f = -B_d A \theta, \tag{5.4}$$

where A is a network incidence matrix[2], B_d is a diagonal matrix with $[b_\ell]_{\ell \in \mathcal{L}}$ in its main diagonal, and $\theta = [\theta_n]_{n \in \mathcal{B}}$ is the vector of all bus angles. The nodal balance constraint (4.8) can be written in matrix form as

$$p^{\text{inj}} = A^T f. \tag{5.5}$$

Combining (5.4) and (5.5), we obtain (5.6).

$$p^{\text{inj}} = -A^T B_d A \theta, \tag{5.6}$$

from which we can obtain an expression of θ that can be substituted into (5.4), which yields:

$$f = -B_d A \left[A^T B_d A \right]^{-1} p^{\text{inj}}, \tag{5.7}$$

so that the desired PTDF matrix is given by the expression (5.8).

$$\Psi = -B_d A \left[A^T B_d A \right]^{-1} \tag{5.8}$$

The previous presentation is a very quick overview for the reader already familiar with the PTDF methodology and is not indended as thorough or rigorous. Unfamiliar readers are encouraged to consult the aforementioned references for a rigorous derivation. To conclude this brief overview, we enumerate a few important remarks.

[2] i.e. an $\mathcal{L} \times \mathcal{B}$ matrix where for each row ℓ, $A_{\ell,o(\ell)} = -1$, $A_{\ell,d(\ell)} = +1$, and $A_{\ell,i} = 0$ for all other $i \in \mathcal{B}$.

Disconnected components In this analysis, we assume the network consists of a single connected component, since it is otherwise not true that injections anywhere in the system can be compensated at a single, arbitrary, slack bus. If there is more than one connected component in the grid, a PTDF matrix must be obtained (and a slack bus must be defined) for each one of them.

Reference bus and validity of the inversion Matrix $A^T B_d A$ is in fact a singular matrix. This is a consequence of system (5.4) being an underdetermined system because, as mentioned before, only angle differences matter in (5.1), so there are infinitely many values of voltage angles that describe the same state of the power system. A unique solution, is obtained by fixing the angle at the reference bus at 0. The reference bus and the slack bus are usually chosen to be the same, but this does not need to be the case in general. If chosen to be the same, $A^T B_d A$ can be made invertible by deleting the row and column corresponding to the reference (and slack) bus.

Choice of the slack bus A natural question to ask is whether the choice of the slack bus affects the results obtained with the PTDF method. Neglecting transmission losses, it does not. This can be easily verified observing that a superposition property is satisfied for PTDF expressions. Suppose an injection or withdrawal at a bus n_1 is in fact compensated at a bus n_2, instead of at the slack bus n_s that was assumed during derivation of Ψ. The new compensation can be expressed as the superposition of an injection at n_1 compensated at n_s and a withdrawal of same magnitude at n_2, compensated at n_s. In consequence, as long as the network consists of a single connected component, the choice of the slack bus does not affect the results. This is not true in the lossy case, as will be seen in Section 5.3.2. Another related question is how to consider an injection at the slack bus. Since under the PTDF methodology, it will be compensated at the same slack bus, there will be no effect on transmission branches. If in reality there is an injection at the slack bus, compensated at some other bus n_1 of the system, the effect of this injection can be seen as the effect of a withdrawal of equal magnitude at n_1.

Topology dependence and nodal representation Given a PTDF matrix Ψ, expression (5.3) directly gives flows in all branches, so thermal limit constraints (4.9a) can be included in the model without the need for a nodal representation of the network flow. In other words, nodal power balance constraints (4.8) can be omitted from the power flow formulation. The topology of the system is thus captured within the PTDF matrix Ψ. This allows a more compact formulation, since voltage angle variables and nodal balance equations are no longer necessary, but also pose an additional problem: in transmission expansion, the topology of the system changes depending on which transmission candidates are built. How can the PTDF methodology be adapted to that case? We address that question in the following section.

5.2.3 CEP model with $b\theta$ power flow constraints

We now present the result of including the $b\theta$ power flow constraints of the previous section into the CEP model of Chapter 4.

In a network under expansion, there are two types of branches: existing and candidates for construction. Constraints (5.1) should be enforced for all existing branches, as well as for candidate branches that are selected for construction, but should be ignored for candidate branches that are not built. This logic can be easily implemented using a traditional Big-M formulation:

$$-\left(1 - x^{\mathcal{L}}_{\ell}\right) M \leq f_\ell - b_\ell \left(\theta_{o(\ell)} - \theta_{d(\ell)}\right) \leq \left(1 - x^{\mathcal{L}}_{\ell}\right) M \;\forall \ell \in \mathcal{L}^\star \cap \mathcal{L}^\circ \tag{5.9a}$$

$$-\left(1 - x^{\mathcal{L}}_{\ell}\right) M \leq f^+_\ell - f^-_\ell - b_\ell \left(\theta_{o(\ell)} - \theta_{d(\ell)}\right) \leq \left(1 - x^{\mathcal{L}}_{\ell}\right) M \;\forall \ell \in \mathcal{L}^\star \cap \mathcal{L}^\bullet, \tag{5.9b}$$

where $x^{\mathcal{L}}_{\ell}$, as before, is the binary variable that represents the decision to build candidate transmission element ℓ. Recall that (4.9c) and (4.9d) ensure that flow is zero for unbuilt candidate branches. Assuming the system is connected, we can use the graph of the power system $G = (\mathcal{B}, \mathcal{L})$, defining F_ℓ/b_ℓ as the length of edge ℓ, to obtain valid values for the big-M. For each candidate ℓ, a valid M_ℓ is given by the length of the shortest existing path between the candidate's ends. For more details, see [6].

Domain of $\boldsymbol{\theta}$ As mentioned before, the angle at the reference bus is fixed to 0. If we have an upper bound M^Θ on the maximum angle difference between any pair of buses, then a valid domain for all bus angles would be $\left[-M^\Theta, M^\Theta\right]$. Assuming the system consists of just one connected component, we can obtain a valid M^Θ by finding the longest path on the graph of the power system $G = (\mathcal{B}, \mathcal{L})$, using again F_ℓ/b_ℓ as the length of edge ℓ. M^Θ would be the length of the longest path. For our tests, however, we have found a value of 2π to be sufficient.

$$-M^\Theta \leq \theta_{i,t,\omega} \leq M^\Theta \;\forall i \in \mathcal{B}, t \in \mathcal{T}, \omega \in \Omega \tag{5.10a}$$

$$\theta_{i,t,\omega} \in \mathbb{R} \;\forall i \in \mathcal{B}, t \in \mathcal{T}, \omega \in \Omega \tag{5.10b}$$

To obtain the nodal stochastic capacity expansion model with the $b\theta$ formulation of DC power flow, it suffices to add the $b\theta$ constraints to the model obtained before. In compact form, it has the same general two-stage stochastic MILP form of problem (4.15), show here in (5.11). Note, however, that incorporating the $b\theta$ formulation into the model requires the addition of voltage angle variables $\theta_{t,\omega}$, and one more power flow constraint of type (5.1) or (5.9) for each branch, so $|\mathcal{B}|$ more variables and $|\mathcal{L}|$ more constraints for each scenario and period are added to the model. So considering power flow constraints indeed restricts and complicates the model.

$$\begin{aligned} & \min C^{\text{inv}} + \mathbb{E}_{\zeta}\left[C_{\omega}^{\text{op-RPS}}\right] & \\ \text{s.t.} \quad & A^{\diamond}x \geq b^{\diamond} & (5.11a)\\ & M_{\omega}^{\triangleleft}x + N_{\omega}^{\triangleleft}y_{\omega} \geq \xi_{\omega}^{\triangleleft}\ \forall\omega\in\Omega & (5.11b)\\ & x \in \mathcal{X}, y_{\omega}^{\triangleleft} \in \mathcal{Y}_{\omega}^{\triangleleft}\ \forall\omega\in\Omega & (5.11c) \end{aligned}$$

The objective is defined in (4.1), as before. As in (4.15), x represents the concatenation of all investment variables $x^{\mathcal{G}}, x^{\mathcal{S}}, x^{\mathcal{S}\text{-PC}}, x^{\mathcal{L}}$, with $\mathcal{X}$ an abstract representation of their domains (4.2). The concatenation of all operation variables now also includes θ_{ω}, and is denoted $y_{\omega}^{\triangleleft}$, with $\mathcal{Y}_{\omega}^{\triangleleft}$ a compact representation of their domains, which now besides (4.3) also includes (5.10). Constraint (5.11a) is an abstract representation of construction limits constraints (4.4), and is the same as in model (4.15). Finally, constraint (5.11b), the abstract representation of operation constraints, now includes (4.5)-(4.9), (4.13), (5.1), and (5.9). We have used superscript $\triangleleft$ in all abstract symbols that differ from model (4.15) to avoid confusion.

5.2.4 CEP model with PTDF power flow constraints

In the PTDF formulation presented before, voltage angle variables have been eliminated from the model, so power flow constraints (5.9) for candidate branches cannot be used. Instead, the changing topology is addressed through so-called virtual injections. The modeling strategy consists in first considering an extended power system where all candidate transmission elements are added to the grid. Then, the branches that are not chosen for construction are switched off by adding appropriately calculated virtual injections and withdrawals of power at the branch's ends.

Virtual power injections

We follow the exposition of [85] ; [112] can also be consulted for a detailed formulation with examples. Consider a branch ℓ with ending buses $m = o(\ell)$ and $n = d(\ell)$, which we want to switch off. We will achieve that by considering virtual injections of magnitude $\tilde{f}_{\ell}$ at m and n. This is depicted in Fig. 5.2.1.

The flow through ℓ before the addition of the virtual injections is, from (5.8):

$$f_{\ell} = \sum_{i\in\mathcal{B}} \Psi_{\ell,i} p_i^{\text{inj}},$$

so that after adding the virtual injections, we get

$$f_{\ell} = \sum_{i\in\mathcal{B}} \Psi_{\ell,i} p_i^{\text{inj}} + \Psi_{\ell,m}\tilde{f}_{\ell} - \Psi_{\ell,n}\tilde{f}_{\ell}.$$

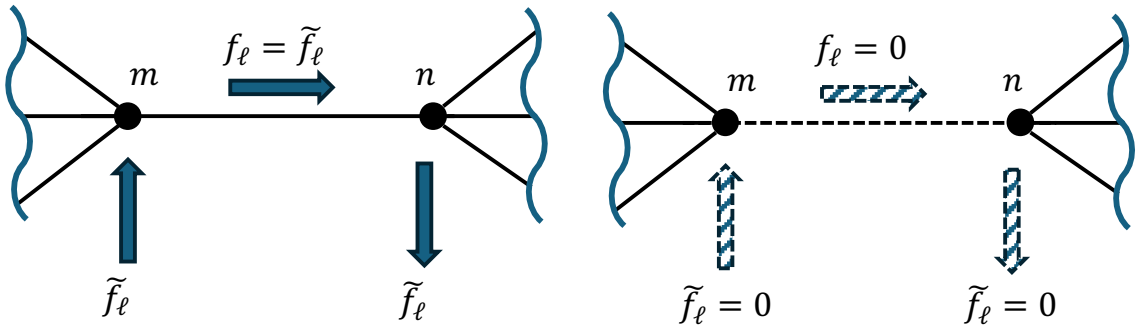

Figure 5.2.1: Virtual injections to switch a transmission branch off. On the left, situation with virtual injections $\tilde{f}_\ell$ such that $\tilde{f}_\ell = f_\ell$. Observe that the resulting flows in the rest of the network are the same as when switching the branch off, as shown in the situation on the right.

Note, however, that all other branches are also affected by the virtual injections caused by switching off ℓ, so (5.3) becomes:

$$f = \Psi p^{\text{inj}} + (\Psi_{:,m} - \Psi_{:,n})\, \tilde{f}_\ell$$

If the virtual injection $\tilde{f}_\ell$ is chosen to match the value of f_ℓ, the power flowing through branch ℓ entirely comes from and goes to the virtual injection and withdrawal. So, from the perspective of the rest of the grid, the effect is the same as what would be obtained if branch ℓ was not connected to the grid (see Fig. 5.2.1). To switch branch ℓ off, one therefore needs to enforce

$$\tilde{f}_\ell = \sum_{i \in \mathcal{B}} \Psi_{\ell,i} p_i^{\text{inj}} + (\Psi_{\ell,m} - \Psi_{\ell,n})\, \tilde{f}_\ell.$$

When considering the simultaneous disconnection of all the elements in a set $\mathcal{L}'$, one can apply linear superposition to the results above. A pair of virtual injections would thus be added for each disconnected element, obtaining (5.12) and (5.13).

$$f = \Psi p^{\text{inj}} + \sum_{\ell \in \mathcal{L}'} \left(\Psi_{:,o(\ell)} - \Psi_{:,d(\ell)}\right) \tilde{f}_\ell \tag{5.12}$$

$$\tilde{f}_\ell = \Psi_{\ell,:} p^{\text{inj}} + \sum_{\ell' \in \mathcal{L}'} \left(\Psi_{:,o(\ell')} - \Psi_{:,d(\ell')}\right) \tilde{f}_{\ell'} \;\; \forall \ell \in \mathcal{L}' \tag{5.13}$$

Transmission expansion planning with PTDF

As mentioned before, the plan is to only enforce (5.13) for lines that are not selected for construction, so that they are switched off. Lines selected for construction should not be switched off, so they should not have any virtual injection, but they must satisfy the thermal limit constraint (4.9c). This can be enforced with help of a big-M formulation as in (5.14) and (5.15).

$$-Mx_\ell^{\mathcal{L}} \leq \tilde{f}_\ell - \Psi_{\ell,:}p^{\text{inj}} - \sum_{\ell' \in \mathcal{L}^\star} \left(\Psi_{:,o(\ell')} - \Psi_{:,d(\ell')}\right) \tilde{f}_{\ell'} \leq Mx_\ell^{\mathcal{L}} \; \forall \ell \in \mathcal{L}^\star \tag{5.14}$$

$$-M'\left(1 - x_\ell^{\mathcal{L}}\right) \leq \tilde{f}_\ell \leq M'\left(1 - x_\ell^{\mathcal{L}}\right) \; \forall \ell \in \mathcal{L}^\star \tag{5.15}$$

It is shown in [85] that the combined result of enforcing (4.9c), (5.14) and (5.15) can also be obtained from (5.15) and (5.16). Valid values for M' can be obtained from a shortest-path problem as described for the $b\theta$ formulation. Given M_ℓ for branch ℓ, a valid M'_ℓ would be given by $M'_\ell = b_\ell M_\ell$. As in [85], however, we find that using 10 p.u. is sufficient for all our tests.

$$-F_\ell x_\ell^{\mathcal{L}} \leq \tilde{f}_\ell - \Psi_{\ell,:}p^{\text{inj}} - \sum_{\ell' \in \mathcal{L}^\star} \left(\Psi_{:,o(\ell')} - \Psi_{:,d(\ell')}\right) \tilde{f}_{\ell'} \leq F_\ell x_\ell^{\mathcal{L}} \; \forall \ell \in \mathcal{L}^\star \tag{5.16}$$

As mentioned before, thermal limits in existing branches can also be enforced without explicitly defining f_ℓ, as in (5.17).

$$-F \leq \Psi p^{\text{inj}} + \sum_{\ell \in \mathcal{L}^\star} \left(\Psi_{:,o(\ell)} - \Psi_{:,d(\ell)}\right) \tilde{f}_\ell \leq F \tag{5.17}$$

, where $F = [F_\ell]_{\ell \in \mathcal{L}^\dagger}$ is the vector of rated capacities of all existing branches.

The domain of the virtual injections $\tilde{f}$ is thus

$$-M'_\ell \leq \tilde{f}_\ell \leq M'_\ell \; \forall \ell \in \mathcal{L}^\star \tag{5.18a}$$

$$\tilde{f}_\ell \in \mathbb{R} \; \forall \ell \in \mathcal{L}^\star \tag{5.18b}$$

5.3 Power flow model improvements for CEP

We are now ready to introduce the power flow model improvements for CEP proposed in this thesis. The first is a straightforward extension of the PTDF approach for TEP of [85] presented before to the joint CEP that we consider in this work. We then propose some modeling choices based on statistical analysis to include transmission losses in the PTDF representation of power flow.

5.3.1 Joint generation, storage and transmission CEP with PTDF

In the model proposed for transmission expansion planning in [85] and presented above, the entries in the bus injection vector p^{inj} contain only the output from existing generation

facilities. However, the validity of the approach is not affected by considering generation and storage candidates in those entries. We can thus incorporate this formulation of power flow constraints in lieu of the $b\theta$ formulation, while keeping the rest of our capacity expansion model.

Power balance Nodal balance constraints (4.8) are not part of the PTDF model, so we need to add an aggregate power balance constraint (5.19).

$$\sum_{i\in\mathcal{B}}\left(\sum_{g\in\mathcal{G}} p^{\mathcal{G}}_{i,g,t,\omega} + \sum_{s\in\mathcal{S}}\left(\eta^{\mathcal{S}\text{-dch}}_{s} p^{\mathcal{S}\text{-dch}}_{i,s,t,\omega} - p^{\mathcal{S}\text{-ch}}_{i,s,t,\omega}\right) + p^{\text{sh}}_{b,t,\omega}\right) = \sum_{i\in\mathcal{B}} D_{i,t,\omega} \;\forall t \in \mathcal{T}, \omega \in \Omega \quad (5.19)$$

In compact form, our nodal stochastic capacity expansion model with the PTDF formulation of DC power flow takes the form of (5.20). Again, it has the same general two-stage stochastic MILP form of our two previous formulations (4.15) and (5.11). Here, voltage variables θ_ω and flow variables f_ω are not necessary, but virtual injections $\tilde{f}_\omega$ are added, so this formulation has $|\mathcal{B}| + |\mathcal{L}| - |\mathcal{L}^\star|$ less variables than the $b\theta$ formulation for each scenario and period. Moreover, nodal balance and power flow constraints are substituted by a single global balance constraint and the virtual injection constraints (5.15),(5.16). Therefore, for each scenario and period, this formulation has $|\mathcal{B}| + |\mathcal{L}| - 2\,|\mathcal{L}^\star| - 1$ less constraints than the $b\theta$ formulation.

$$\begin{aligned}
& \min C^{\text{inv}} + \mathbb{E}_\zeta\left[C^{\text{op-RPS}}_\omega\right] \\
\text{s.t.}\quad & A^\diamond x \geq b^\diamond && (5.20a)\\
& M^\triangleright_\omega x + N^\triangleright_\omega y_\omega \geq \xi^\triangleright_\omega \;\forall \omega \in \Omega && (5.20b)\\
& x \in \mathcal{X}, y^\triangleright_\omega \in \mathcal{Y}^\triangleright_\omega \;\forall \omega \in \Omega && (5.20c)
\end{aligned}$$

Once again, the objective is defined in (4.1). As in the two previous formulations, x represents the concatenation of all investment variables, with $\mathcal{X}$ an abstract representation of their domains (4.2). The concatenation of all operation variables now only includes $p^{\mathcal{G}}_\omega, p^{\mathcal{S}}_\omega, p^{\mathcal{S}\text{-ch}}_\omega, p^{\mathcal{S}\text{-dch}}_\omega, p^{\text{sh}}_\omega, p^{\text{NC}}_\omega, \tilde{f}_\omega$, and is denoted $y^\triangleright_\omega$, with $\mathcal{Y}^\triangleright_\omega$ a compact representation of their domains (4.3a)-(4.3d), (5.18). Constraint (5.20a) is an abstract representation of construction limits constraints (4.4), and is the same as in the other two models. Finally, as mentioned before, constraint (5.20b), which represents operation constraints, excludes nodal balance constraints (4.8) and power flow equations (5.1),(5.9), so it only includes (4.5)-(4.7), (4.13), (5.15)-(5.19). We have used superscript $\triangleright$ in all abstract symbols that differ from the previous models to avoid confusion.

Advantages and shortcomings of the PTDF approach

Comparing models (5.20) and (5.11), it is not obvious that either is a better formulation of the capacity expansion problem than the other. In fact, there are tradeoffs between the two formulations. We discuss here some of them to motivate why it can be a sensible choice for our capacity expansion model.

Compactness vs sparsity There is a tradeoff between two important attributes for linear computational solvers: model size and model sparsity. The $b\theta$ formulation requires more variables and constraints, which leads to a significantly larger model: $|\mathcal{B}|$ voltage angle variables, $|\mathcal{B}|$ nodal power balance constraints and $|\mathcal{L}|$ power flow constraints are part of the $b\theta$ formulation, but not of the PTDF one. However, enforcing power flow equations to satisfy the thermal constraint for each branch ℓ in the $b\theta$ formulation only involves the branch flow variable f_ℓ and the voltage angles of the corresponding end buses, so while larger, it is also a very sparse formulation, which solvers can handle better. On the other hand, it can be seen in (5.12) that the power flow constraint for each branch includes power injection variables (i.e. generation and storage outputs) from all buses in the system. So the resulting constraint matrix is smaller, but much more dense. There is a non-trivial tradeoff between the two, but in general, it is likely that considering the full model PTDF description (5.20), $b\theta$ formulations would outperform the PTDF one. This is however not a meaningful comparison because in practice, the PTDF formulation is used in conjuction with two impactful heuristics.

PTDF heuristics: thresholding and branch monitoring The first common heuristic is to consider all entries of Ψ with absolute value below some threshold (or absolute value, relative to the max within its row), to be zero. This can reduce the density of the resulting matrix, at small accuracy cost. The justification for this technique is that when considering a particular branch, injections very far away in the system will have a negligible impact on flows in said branch, and can thus be disregarded. The second, more impactful heuristic, is that in most applications, the list of transmission elements, for which thermal limit constraints will be tight at any given operational period is known from historical data. In other words, planners know which elements are likely to be the ones dictating reinforcement or expansion needs, and which ones can be excluded from the model without affecting the optimal plan. Explicitly defining a list of monitored branches and only including those elements in (5.17) in the model can further reduce the size of the system, and compensate for the increased density of the remaining lines. Note that even if the same approach is attempted in the $b\theta$ formulation, all nodal balance and power flow equations are necessary to include just a subset of thermal limits, so defining a list of monitored elements does not reduce the model size significantly.

Reduction of the MIP gap Capacity expansion planning models are in most cases, including ours, mixed integer problems of very large size, which cannot be solved to optimality

(i.e. until the MIP gap is reduced to zero) within a reasonable time. Since subproblem execution is often ended by reaching a predefined time limit (cf. Section 4.4.2), the rate at which MIP optimality gap is reduced is very relevant for determining a preferred method among different alternatives. An advantage found by Rahmani, Kargarian, and Hug is that solving the transmission expansion planning problem with a PTDF formulation reduces MIP gap more quickly than its $b\theta$ counterpart, so that time-limited suboptimal solutions are of better quality in the PTDF case. In our solution approach, we often interrupt the solver before reaching the predefined target optimality gap. The results of [85] suggest that PTDF is a better formulation for that approach.

Pre-solve computational cost A substantial part of the computational cost in the PTDF formulation consists of obtaining the PTDF matrix, which includes inversion operations, very costly for large systems like the ones that we concern ourselves with in this thesis. It has been found that this overhead is too long for smaller systems, but could become worth it when the system considered is larger [85]. On its own, that could be enough justification for the choice of a PTDF power flow formulation in our capacity expansion problem. In addition, if solving the problem with the PHA, the matrix is the same for all iterations, so the expensive computational step of calculating Ψ and posing the model is only incurred once. In subsequent iterations, the reduced and compact formulation could outperform the $b\theta$ formulation.

5.3.2 Transmission system losses

In Section 4.3.6, we incorporated transmission losses into the capacity expansion model with a transportation formulation and discussed some implications of this modeling choice. In the previous sections of this chapter, we have presented more accurate representations of power flow, so that the value of transmission assets is not overestimated by the capacity expansion model. However, these models have considered until now the lossless case, an inaccuracy that may maintain some overestimation of the system's flexibility, possibly at the expense of other flexibility-providing resources, like storage facilities, and ultimately leading to operational overcosts. In this section, we treat the incorporation of transmission losses to the capacity expansion model with power flow constraints without sacrificing the linearity of the model. We discuss some challenges and the methodology proposed to overcome them.

The strategy chosen is to utilize again the linear loss model presented in Section 4.3.6, where transmission losses are modeled as a constant proportion of branch flow, thus implying that each transmission element has a constant efficiency $\eta_\ell^{\mathcal{L}}$. Recall that the value of efficiency chosen for each transmission line, given by (5.22), is the branch's efficiency at its rated capacity, according to the common quadratic loss model [32, 31]. This choice implies that transmission losses are overestimated in most branches, during most time periods. In spite of its simplicity, implementing this linear model has some challenges, which we describe next.

Transmission system losses in the $b\theta$ formulation

Adding (5.1) and (5.9) to the network flow model (4.15) already results in a $b\theta$ model with transmission losses represented as in section 4.3.6. Just like in the transportation formulation, spurious losses may appear as a result of having $f_\ell^+ \cdot f_\ell^- > 0$ for some branch $\ell \in \mathcal{L}^\bullet$. We do not make any considerations other than those made for the transportation formulation, to address spurious transmission losses. See section 4.3.6 for more details about that.

Transmission system losses in the PTDF formulation

As derived above, and shown in several references (e.g. [85, 113]), the $b\theta$ and PTDF formulations are equivalent when transmission losses are disregarded[3].

The methodology proposed in [31] to include transmission losses in a PTDF formulation for nodal pricing requires initialization from a solution to the AC power flow problem. In our application, the state of the system can change significantly for different representative days and periods within each day, and throughout the optimization process depending on investment candidates built. Adopting a similar methodology in our case would require iteratively updating the calculation of the PTDF matrix during the optimization process, with challenges of convergence and computational tractability. As mentioned before, part of the justification for the PTDF version of the model is that the expensive PTDF calculation needs to occur only once. Thus, sacrificing some accuracy, we instead propose a linear model that can be maintained during the optimization process.

Finding a suitable loss distribution factors matrix The classic loss ditribution factors approach, and that followed in [31], is to obtain a sensitivity matrix L such that the losses $f^{\mathcal{L}^\bullet}$ in all transmission elements can be obtained as $f^{\mathcal{L}^\bullet} = Lp^{\text{inj}}$. In that general approach, the matrix L is obtained via a Jacobian-based sensitivity analysis, and thus depends on the state of the system on which the Jacobian is evaluated, i.e. on the starting operational point. The underlying assumption is that matrix L will be used for system states in the vicinity of that operational point. Using an initial state of $f_\ell = 0$ for all $\ell \in \mathcal{L}$ leads to a trivial $L = 0$, leading to a lossless model. If, on the other hand, the Jacobian is evaluated at a non-zero operational point, but the resulting L is used for all system states, unrealistic results like negative losses could be obtained, as illustrated next.
Consider the linearization of (4.11), which we rewrite below as (5.21) and (5.22), and the

[3]This statement holds for the PTDF formulation derived in section 5.2, but may not for the formulation that uses heuristic simplifications. The inclusion of only monitored lines in the thermal limits constraints and the approximatiotn of Ψ entries with low absolute value to 0 may break that equivalence.

original linearization (5.23) from [32].

$$f_\ell^{\mathcal{L}^\bullet} = \left(1 - \eta_\ell^{\mathcal{L}}\right) f_\ell \tag{5.21}$$

$$1 - \eta_\ell^{\mathcal{L}} = r_\ell F_\ell \left(1 + \left(\frac{r_\ell}{\chi_\ell}\right)^2\right) \tag{5.22}$$

$$1 - \hat{\eta}_\ell^{\mathcal{L}} = r_\ell \hat{f}_\ell \left(1 + \left(\frac{r_\ell}{\chi_\ell}\right)^2\right) \tag{5.23}$$

where $\hat{f}_\ell$ is the value of f_ℓ at the operational point chosen and $\hat{\eta}_\ell^{\mathcal{L}}$ is the branch's efficiency at that operational point. Note that the definition in (5.22) ensures $\eta_\ell^{\mathcal{L}} < 1$, so having $f_\ell < 0$ in (5.21) will lead to negative losses. This is not surprising since what we are attempting is representing a quadratic function with a linear function.

Negative losses are not only possible but likely to happen since they would compensate the cost of expensive generation, and would be favored by the optimization model. This will not happen if the point where the linearization is performed is updated, because updating the value of $\hat{f}_\ell < 0$ in (5.23) would lead to $\hat{\eta}_\ell^{\mathcal{L}} > 1$ and thus nonnegative losses in (5.21). An approach similar to that pursued in the $b\theta$ formulation could be attempted, i.e. adding separate, non-negative, branch flow variables f_ℓ^+, f_ℓ^- for flow in each direction. However, that would imply adding $2\left|\mathcal{L}\right|$ variables and non-negativity constraints, which undermines the compactness of the PTDF. We thus prefer a different solution that we describe next: modeling lossy transmission elements as unidirectional.

Limiting branch flow direction Suppose that for a specific scenario (representative day ω) and period (t), the set of all branches in the model is partitioned into two categories: bidirectional branches, through which power may flow in either direction, and unidirectional branches, through which power may only flow in one predetermined direction. Furthermore, lossy branches are a subset of unidirectional branches. In that case, negative losses are no longer an issue, since we can write them via (5.24), (5.25).

$$f_{\ell,t,\omega}^{\mathcal{L}^\bullet} = \sigma_{\ell,t,\omega} \cdot \left(1 - \eta_\ell^{\mathcal{L}}\right) f_{\ell,t,\omega} \tag{5.24}$$

$$\sigma_{\ell,t,\omega} = \begin{cases} 0 & , \text{ if } \ell \in \mathcal{L}^\circ \\ +1 & , \text{ if } \ell \in \mathcal{L}^\bullet \text{ and flow is } o(\ell) \to d(\ell) \\ -1 & , \text{ if } \ell \in \mathcal{L}^\bullet \text{ and flow is } o(\ell) \leftarrow d(\ell) \end{cases} \tag{5.25}$$

Note that this approach does come at a computational cost. Flow variables f are not explicitly part of the formulation, so direction restriction is added to the model through constraints (5.26). These constraints must be added for all lossy lines, including non-monitored ones. As mentioned before, these constraints add very dense rows to the problem, so they come at a high computational cost. This approach is therefore only appealing if the size of $\mathcal{L}^\bullet$ is not too large.

$$\sigma_{\ell,t,\omega} \cdot \left(\Psi_{\ell,:} p^{\text{inj}} + \sum_{\ell' \in \mathcal{L}^\star} \left(\Psi_{\ell,o(\ell')} - \Psi_{\ell,d(\ell')} \right) \tilde{f_{\ell',t,\omega}} \right) \geq 0 \; \forall \ell \in \mathcal{L}^\bullet, t \in \mathcal{T}, \omega \in \Omega \tag{5.26}$$

Assigning losses to buses In the derivation of the PTDF matrix Ψ, the nodal balance constraint (5.5) assumed that no energy was lost in the transmission network. If part of the energy does not make it to the end of the branch, (5.5) must be updated to (5.27)

$$p^{\text{inj}} = A^T f - \tilde{A}^T \text{diag}(1 - \eta^{\mathcal{L}}) f, \tag{5.27}$$

where $\tilde{A}$ is an $|\mathcal{L}| \times |\mathcal{B}|$ non-negative matrix, where each row adds to 1. Each entry $\tilde{A}_{\ell,m}$ is the proportion of losses in branch ℓ that are assumed to occur at bus m. A common approximation, adopted in [74] and the Pyomo-based power system tool *egret* [51], is to assign half of the losses to each end of the branch, i.e. $\tilde{A} = (1/2) \cdot |A|$. This is a reasonable approximation if the flow direction is not known a priori. Since in our model, flow direction is known for lossy branches, we can directly assign them as in (5.28) and (5.29).

$$\tilde{A}_{\ell,o(\ell)} = \begin{cases} 1 & \text{if } \sigma_\ell < 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.28}$$

$$\tilde{A}_{\ell,d(\ell)} = \begin{cases} 1 & \text{if } \sigma_\ell > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.29}$$

Substituting (5.27) for (5.5) in (5.8), we get

$$\Psi = -B_d A \left[\left(A^T - \tilde{A}^T \text{diag}\left(1 - \eta^{\mathcal{L}}\right) \right) B_d A \right]^{-1}. \tag{5.30}$$

Choice of the slack bus In the previous section, we argued that the choice of the slack bus is not consequential in the lossless case, because superposition guarantees that the effect of choosing different slack buses cancel each other. This is no longer true in the lossy case. A lengthier and rigorous exposition can be found in [31] and [55] in the context of locational marginal pricing, but for illustration, it may suffice to imagine an injection of 1MW at a bus A with two possible locations for the compensating withdrawal: B and C. Because of transmission losses, the withdrawal at either B or C will be < 1MW, call them u_B and u_C respectively. Suppose that B is chosen as the slack bus when calculating the PTDF matrix, but C is actually where the injection is compensated. Then, the result of a compensation at C can be expressed as the superposition of an injection of 1MW at A and a negative injection u_C/u_B at C, both compensated at B. In other words, in order to make the superposition work, we would need to adjust all injections based on the ratio of losses, which depends on which bus is actually the slack bus at a given state of the system, and which

bus was assumed to be the slack bus in the PTDF calculation. In our model, the actual slack is likely to change for different periods and throughout the optimization process, as different candidates are considered, which would significantly complicate the computational optimization process. A common approximation is to use a distributed slack bus, where a constant fraction of losses is assumed to be compensated at each of a handful of buses. See [55] for more details about this approach. We present our approach for determining the distributed slack buses in the next section.

Spurious losses Spurious losses in the sense of Section 4.3.6 are not an issue here, since bidirectional flow is never permitted for lossy branch elements.

Power balance with losses The power balance constraint (5.19) must be updated, since losses now affect the balance. Recall that variable $f_{\ell,t,\omega}$ is no longer part of the formulation; it is used in (5.31) as shorthand for the right-hand side of (5.12).

$$\sum_{i\in\mathcal{B}}\left(\sum_{g\in\mathcal{G}} p^{\mathcal{G}}_{i,g,t,\omega} + \sum_{s\in\mathcal{S}}\left(\eta^{\mathcal{S}\text{-dch}}_{s} p^{\mathcal{S}\text{-dch}}_{i,s,t,\omega} - p^{\mathcal{S}\text{-ch}}_{i,s,t,\omega}\right) + p^{\text{sh}}_{b,t,\omega}\right) - \sum_{\ell\in\mathcal{L}^{\bullet}} \sigma_{\ell,t,\omega}\left(1-\eta^{\mathcal{L}}_{\ell}\right) f_{\ell,t,\omega} = \sum_{i\in\mathcal{B}} D_{i,t,\omega}\ \forall t\in\mathcal{T}, \omega\in\Omega \quad (5.31)$$

Implementation and empirical justification of PTDF heuristics

In the methodology presented in the two previous sections, we have introduced two heuristic methodologies to exploit the compactness of the PTDF formulation and obtain a better capacity expansion model. However, we have left two key details about the heuristics unaddressed: how are monitored lines selected, and how are directed/undirected lines partitioned and assigned a direction. We cover those two questions in the following paragraphs.

Selecting monitored branches Monitored branches are selected following the same methodology as [85]: the linear relaxation of the capacity expansion problem (5.11) is solved first. For each scenario ω, if the flow through a transmission element ℓ during period t is above a predefined threshold (expressed as a proportion of the branch's rated capacity), the pair ℓ, t is added to the list of monitored branches for that representative day. The value of the threshold was set after some iterative testing, and was set at 0.8 for all our tests.

Partitioning lossy and directional branches The approach of partitioning branches into lossy and lossless, and restricting lossy branches to be unidirectional is predicated on two assumptions:

- Transmission losses are not evenly distributed across branches. In other words, it suffices to consider a small fraction of all branches to capture most of the losses in the system.

- In most lossy branches, power flows in just one direction in most states of the system for which a PTDF matrix will be shared.

The first assumption can be checked by analyzing the distribution of losses across branches on a sample of power flow solutions. To check the second assumption, a metric of directionality is necessary. We propose the directionality metric of (5.32), which we will use to restrict the direction of flow in lossy branches for all periods $t \in \mathcal{T}$ of each scenario $\omega \in \Omega$. Different grouping of periods are conceivable, and the metric would need to be modified accordingly.

$$\varrho_{\ell,\omega} = \frac{\left|\sum_{t\in\mathcal{T}} f_{\ell,t,\omega}\right|}{\sum_{t\in\mathcal{T}} |f_{\ell,t,\omega}|} \tag{5.32}$$

Note that $0 \leq \varrho_{\ell,\omega} \leq 1$. It is equal to 1 if the flow through branch ℓ has the same direction for all periods t in the scenario ω considered. It is 0 if the cumulative energy through ℓ in one direction throughout the day is the same as the cumulative energy in the opposite direction. In this sense, the closer that $\varrho_{\ell,\omega}$ is to 1, the more unidirectional the branch is in the periods considered.

The implementation of the heuristic that we propose is inspired in the one for monitored branches. The linear relaxation of the CEP (5.11) is first solved. The solution is analyzed to obtain directionality metrics for all branches in the system. Only branches with inefficiencies above a certain threshold and enough unidirectionality are considered lossy and directed. This is made precise for each $\omega \in \Omega$ in (5.33), with δ^d and r^{thresh} as thresholds.

$$\ell \in \begin{cases} \mathcal{L}^{\bullet} & \text{if } \varrho_{\ell,\omega} > \delta^d \text{ and } 1 - \eta^{\mathcal{L}}_{\ell} > r^{\text{thresh}} \\ \mathcal{L}^{\circ} & \text{otherwise} \end{cases} \tag{5.33}$$

Remark. In (5.32), $\varrho_{\ell,\omega}$ is a metric of how unidirectional branch ℓ is across the time periods of representative day ω. Accordingly, its value is used to determine whether branch ℓ will be restricted to one direction for all time periods of ω. It is conceivable that periods are grouped in some other way, e.g. peak hours of similar representative days (in the same season) are grouped together, and the same for mid-day off-peak hours and midnight off-peak hours. This approach can be easily extrapolated for that use case as well.

Empirical validation To asses the empirical validity of the aforementioned assumptions, we introduce two more statistical metrics. Consider the inefficiency threshold r^{thresh} and unidirectionality threshold δ^d as used in (5.33). We define $\eta^{\mathcal{L}\star}(r^{\text{thresh}})$ as the proportion of total losses in the system incurred in branches with efficiencies below $1 - r^{\text{thresh}}$. We define $\varrho^{\star}(\delta^d)$ as the proportion of total losses in the system incurred in branches with directionality metric $\varrho > \delta^d$. These two quantities are illustrated in Fig. 5.3.1.

The empirical validity of these two assumptions was tested on a realistic power system[4]. The test conducted consisted in solving 100 instances of the multiperiod Optimal Power

[4]See Section 5.4 for details about the system.

Flow (OPF) problem (i.e. solving only the operational part of the CEP problem without expansion candidates, for 100 values of ω) and obtaining the values of the two metrics defined above.

The results are at the bottom of Fig. 5.3.1. Consistently across the 100 instances tested, over 95% of system losses occur in branches with ϱ metric ≥ 0.99. So the first assumption holds. Similarly, around 85% of system losses occur in branches with efficiencies under $1 - 5 \cdot 10^{-3}$, which correspond to about 45% of the total branches.

Caveats of heuristic In these paragraphs, we highlight some limitations of the previous empirical test in regard to the optimality of the resulting model, and argue why this heuristic model can be of value nonetheless. First, consider the directionality restriction. By restricting the direction of flow in some branches, we are strictly reducing the size of the feasible set of optimization model (5.20). The previous empirical analysis suggests that the optimal solution would have had flow through restricted branches in the chosen direction anyway. In other words, it suggests that the optimal solution was not in the region deleted by restricting flow direction, and thus, the restriction comes at no cost. There is, however, no guarantee that this is in fact the case. The effect of this inaccuracy can be mitigated by tuning the value of the threshold δ^d.

Secondly, consider the disregard of losses in branches with bi-directional flow. A key fact is that we only need to consider in the model the transmission losses whose inclusion will lead to obtaining an optimal investment plan with lower cost than the one obtained with the lossless model. Thus, if disregarding the losses in bidirectional branches does not change the optimal investment plan obtained, this simplification of the model is not problematic. One of the possible consequences of neglecting transmission losses in the expansion model is overbuilding cheaper generation in remote places and underbuilding it in buses closer to the load, as a result of overestimating the actual energy that can make it to the demand from the remote source. In this type of situation, the losses causing the change in the optimal investment plan correspond to flows that are always in the same direction, and would therefore be captured in our model. There is no guarantee, however, that these are the only consequential transmission losses. If some of the transmission losses driving a change in the optimal investment plan occur in bidirectional branches, then our methodology fails to capture those losses, and can also lead to suboptimal investment plans comparable to those of the lossless model.

Finally, we revisit our analysis of the results in Fig. 5.3.1. The conclusion that we can disregard losses in the most efficient transmission branches because $\eta^{\mathcal{L}\star}(r^{\text{thresh}}) > 0.85$, i.e. 85% of total losses would still be captured is predicated on the assumption that all losses are equally consequential. In other words, it is assumed that a transmission element with lower absolute losses is unlikely to cause a change in optimal investment plan with respect to the lossless case. Once again, this is reasonable, but not guaranteed. A numerical test is necessary to gauge better how effective our heuristic is on realistic power system data. We consider this outside the scope of this work.

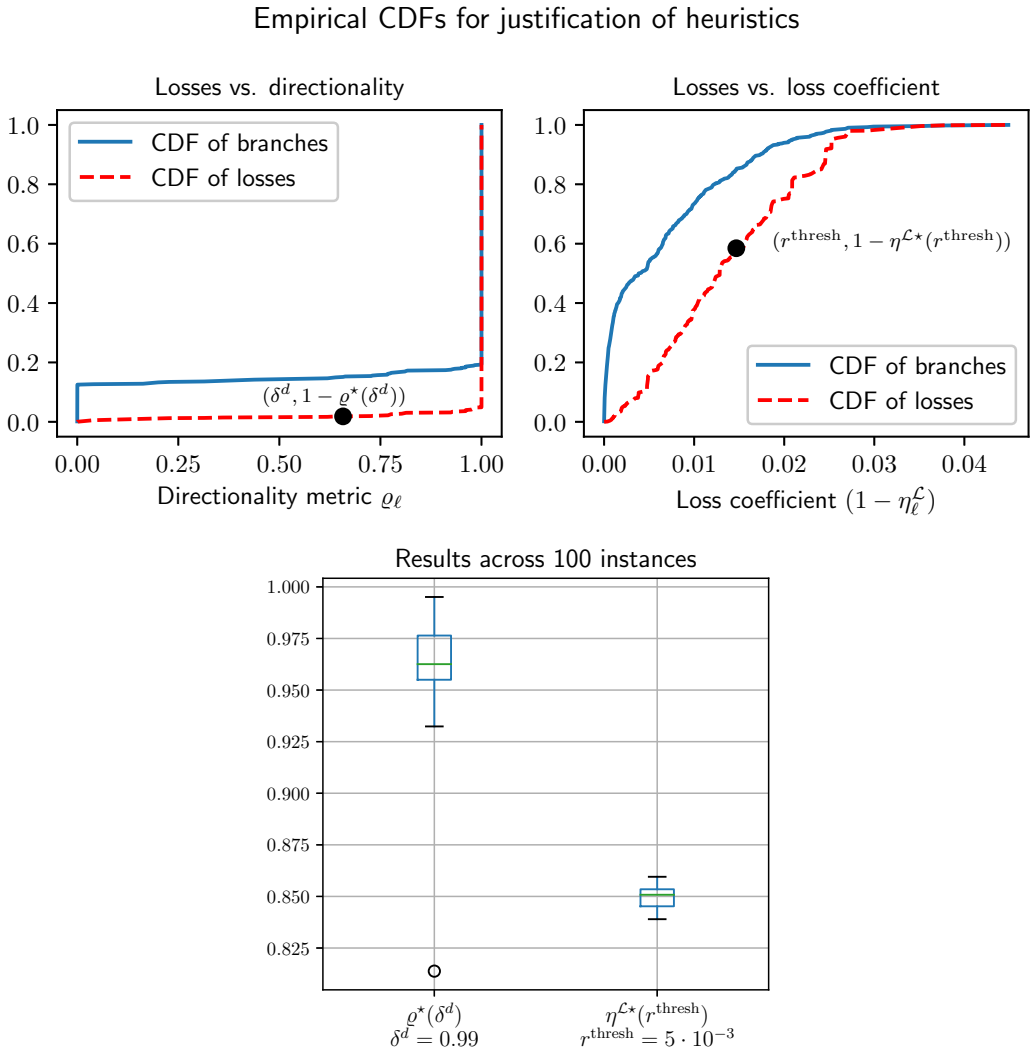

Figure 5.3.1: Empirical test of assumptions for heuristics in lossy PTDF formulation. *Top*: Illustration of validation metrics $\eta^{\mathcal{L}\star}$ and $\varrho^\star$ on a particular scenario ω. *Left*: Empirical Cumulative Distribution Function (CDF) of branch losses vs. ϱ is obtained by sorting all branches in ascending order by $\varrho_{\ell,\omega}$, and summing total branch losses across periods in the same order. The marker shows an example of $\varrho^\star(\delta^d)$ for $\delta^d = 0.65$. *Right*: Empirical CDF of branch losses vs. $(1-\eta^{\mathcal{L}})$ is obtained by sorting all branches in ascending order by inefficiency, and and summing total branch losses across periods in the same order. The marker shows an example of $\eta^{\mathcal{L}\star}(r^{\text{thresh}})$ for $r^{\text{thresh}} = 0.015$. *Bottom*: Distribution of the values of $\varrho^\star(0.99)$ and $\eta^{\mathcal{L}\star}(0.005)$ across the 100 instances tested.

5.4 Test system

To test the methodology proposed, a 500-bus test system was constructed based on the South Carolina ACTIVs test case [7, 8]. That system was complemented with capacity expansion data as described next.

Power system and geographical data Power system data was downloaded from [7]. The dataset includes coordinates for all substations in the system. All coal power plants were retired from the test system. This was done to induce a need for expansion in a realistic way: carbon-intensive technologies are increasingly being retired in North American power systems.

Availability of solar and wind generation At each of the substation coordinates found in the dataset, timeseries of relevant hourly parameters (temperature, humidity, solar irradiance, etc.) were collected from the ERA5 climate reanalysis model [42], for each day of the years 2019 and 2020. These timeseries were fed to a data-processing pipeline based on the System Advisory Model tool [75], which outputs, for each day in the time range, a timeseries with potential availability of solar and wind power, expressed as a fraction of installed capacity ($\alpha_{b,g,t}$ in model (4.15)) [5].

Wind timeseries post-processing The ERA5 model provides data on a grid with a resolution of 31km, which can be too coarse for accurate estimation of wind data. As mentioned in the previous chapter, downscaling that data, i.e. obtaining reliable data with a finer resolution, is a non-trivial task. A regionally refined model similar to the one used for California [116] was not readily available for South Carolina, and it was found that the data obtained had very low wind availability, with average availability below 20%. The purpose of the present chapter is to test our computational approach, for which we need a range of plausible generation candidates. So in order to keep wind generation competitive in the optimization model, all timeseries of wind availability were scaled up uniformly until the site with highest availability reached an average availability of 35%. In addition, a wind "super-site" is created by scaling the wind timeseries of substation *Saluda* until its average availability reaches 45%.
The postprocessing done to the wind availability timeseries makes the data less realistic, and may affect the validity of the correlation with solar availability and load. However, the scope of this chapter is to test the computational performance of the PTDF approach, and not to draw conclusions on the implementability of the investment plans obtained. Overall, the data being used is sufficiently realistic for some analysis of the computational performance to be meaningful. Admittedly, results could vary when tests on more realistic data are conducted,

[5] ERA5 timeseries were provided by Minda Monteagudo of Lawrence Livermore National Lab. These timeseries were processed and translated to timeseries of wind and solar generation availability by Matthew Signorotti of Lawrence Livermore National Lab. I am grateful to both for their invaluable help.

especially regarding the validity of the heuristics implemented, but we believe that the results conducted here are still of interest to researchers in the field. An implementation on a more recent version of the California Test System is foreseen for the near future, but could not be conducted before the conclusion of this doctoral thesis.

Electricity demand An hourly timeseries of electricity demand for the entire system was obtained by taking as baseline the publicly available data of the Energy Information Agency for the two largest utility companies in the area: Dominion and Duke Energy. That baseline, in conjuction with the ERA5 timeseries mentioned before was then processed following the methodology in [67] to obtain an estimated timeseries of demand for each day in years 2019 and 2020. The load timeseries was then distributed across buses using the normalized values in the ACTIVS test system as load distribution factors. [6]. The baseline used is for the aggregate system of the two utilities, which span a larger system than our 500-bus testcase. Load timeseries were therefore scaled down uniformly so that the peak demand across all days in the dataset was 1.6 times the installed generation capacity, accounting for 35% of peak demand growth by 2045[7] and 25% of capacity reserve margin [23]. Note again that the mismatch between the target years for climate data and the target demand growth used as baseline may bring the correlation into question, but in this chapter we are only interested in the computational aspects of the method, so we are satisfied with this approach.

Generation and storage investment candidates All substations are considered as candidate sites for solar and battery storage construction. To reduce the number of variables in the model, within each substation, only the bus with highest incident transmission capacity is considered as candidate. The 20 sites with highest average availability in the timeseries are considered as candidates for construction of wind power generation. New natural gas power plants are only allowed at the substations with existing natural gas power plants. No construction of nuclear or coal power plants is allowed in this test case.

Transmission expansion candidates All existing transmission lines can be reinforced. The reinforcement considered consists in installing a copy of the existing line (i.e. same capacity and electric properties). In addition, a new corridor between the wind "super-site" and one of the main centers of load, the city of Greenville is considered.

Costs All costs are taken from the testcase in Chapter 4, i.e. which will be available for consult [72].

[6]The work of obtaining the load baselines, ERA5 timeseries, processing the data and providing the timeseries of load at each bus in the test system was performed by Minda Monteagudo of LLNL.

[7]Based on 1%-1.6% annual growth [17, 56]

5.5 Numerical results and discussion

Tests

To test the performance of the CEP models with the different power flow formulations, we solve the single-scenario CEP problem for the 365 days of 2019 available in the data, following the methodology described in Section 5.3. In actual CEPs, models must be solved over several representative scenarios. In fact, in Chapter 4 we have argued that a large number of scenarios should be included in climate-resilient CEP. However, in the PHA framework of chapter 4, the problem is decomposed by scenario, so the actual optimization problems handled by the solver are single-scenario subproblems. We thus consider these tests to be the most meaningful for assessing the best formulation to be incorporated into our framework in the future.

Time performance As in chapter 4, all models are implemented in Pyomo and solved with Gurobi v10.0.2 on the *quartz* High Performance Computing (HPC) cluster of LLNL. Each node of the cluster has 36 2.1GHz Intel Xeon cores and 128GB RAM. Problems were run in parallel, but note that since we only solve single-scenario instances, there is no communication between scenarios. Each solver instance was assigned 4 CPU cores. We report here as *model creation time* the sum of the time required to create the Pyomo model and the time required to create the plugin to communicate with the Gurobi solver, which is also a function of model size and complexity. In an iterative algorithm like PHA, this time would only be incurred once. *Gurobi solver time* is reported separately, as it gives an idea of the time that would be incurred each iteration.

Benchmarking solution quality A full evaluation of our loss model should consider as benchmark an AC power flow formulation, so that the goodness of the approximation of transmission losses used can be assessed. This is outside of the scope of this thesis, but is part of the foreseen continuation of this work. Here, we use the lossy $b\theta$ model as benchmark. The CEP models considered differ in the description of the operational problem. Therefore, full two-stage solutions of one model cannot be directly evaluated in another model to evaluate solution quality. However, all models share the same investment variables, and satisfy the same construction limits. This is only natural since all models are proposed to define an optimal investment plan. To compare the total cost of different solutions, the lossy $b\theta$ formulation for each scenario is solved, fixing the first-stage variables to the values obtained by each model and solving the operational problem only. Observe that since construction limits are common for all models and load can be shedded, benchmarking problems are guaranteed to be feasible.
To make comparisons easier, the resulting total investment and operation costs are normalized with respect to the cost of the solution returned by the lossy $b\theta$ model. Normally, this would imply that all normalized costs should be ≥ 1. However, because solver executions

are time-limited, returned solutions may not be optimal, and thus normalized costs < 1 are possible, and indeed occur.

Results and discussion

Setup time vs solver execution time Table 5.5.1 confirms what was mentioned in Section 5.2: the use of the PTDF formulation leads to a more compact problem, with about half as many rows and columns, but much more dense, with over three times more non-zero entries in the constraint matrix. This tradeoff translates into a runtime tradeoff between model creation and solver solution, which is reported in Table 5.5.2. Again, we observe the expected tradeoff: PTDF formulations necessitate longer setup times, but lead to quicker solver execution times, about 50% faster. In an iterative method like PHA, this behavior can lead to significantly better performance.

PTDF vs. bθ The PTDF and $b\theta$ models are equivalent in the lossless case (before line-monitoring and thresholding heuristics), so we expect them to perform nearly identically. This is indeed what we observe, not just on average (Table 5.5.2) and in distribution (Fig. 5.5.1), but in almost all instances. This can be seen in the two top plots of Fig. 5.5.4. Our results suggest that the lossless PTDF formulation clearly dominates the $b\theta$ formulation: achieved costs are not significantly different for all instances, but solution times are much smaller, especially for instances with long solution times.
Adding transmission losses to the model complicates the optimization problem and comes at a computational cost, but should avoid increased operational costs that lossless formulations are blind to. Accordingly, we expect lossy models to reach better-quality solutions, but perform worse in terms of time than their lossless counterparts. This is indeed what we observe in the results. The tradeoff between time and solution quality is clearly visible in the bottom plots of Fig. 5.5.4 for the PTDF formulation, and in the boxplots for the $b\theta$ model. In our tests, the additional time required (often 3x increase) seems too high for the moderate solution quality gained, but availability of computational power and additional tests on other systems may change this conclusion.

MIP gap reduction In Fig. 4.4.1 we plot the distribution of times taken to reach different target MIP gaps for the formulations tested. Our results are in line with those of [85]: PTDF formulations seem to do a better job of reducing the gap during early iterations, hitting the targets of 3%, 1%, 0.5% and 0.1% consistently sooner than the corresponding $b\theta$ versions. However, if a very tight gap is required, the PTDF formulations time out in over half of the instances, while most of the $b\theta$ versions manage to reach the target within the time limit. In chapter 4, we showed how having a runtime-dependent target MIP gap can accelerate the convergence of the PHA algorithm. These results suggest that a PTDF formulation would behave better in that case, and is thus a better candidate for being used in a PHA implementation of a stochastic formulation of the CEP problem.

Transportation formulation A somewhat unexpected result of our tests is the comparatively good performance of the simple network flow, also called transportation, formulation of chapter 4. As expected, this model is much quicker than the others considered, both in setup time and solver execution time. While we expected the investment plans returned by the model to lead to significant operational overcosts, this does not seem to be the case. Solution quality performance is actually better on average for the network flow model than for all other models tested. While median cost performance is marginally better for the lossy PTDF formulation, a comparison on an instance-by-instance basis (Fig. 5.5.5) shows that neither formulation consistently outperforms the other one.
The lack of an advantage in considering explicit power flow constraints when compared to just a network flow model suggests that our test case may not be sufficiently meshed for power flow constraints to make a difference. Future tests on more meshed networks, like the CATS test system of chapter 4 may help in determining the value of these formulations.

Table 5.5.1: Comparison of model sizes and densities for different formulations. The model used for comparison is a single-scenario CEP model for the same day. *Note*: It may seem counterintuitive that the lossy $b\theta$ formulation necessitates less rows than its lossless counterpart. This is due to the branch thermal limit constraints. Recall that for lossless branches, flow is represented by a single, sign-unconstrained variable and thermal limits require a two-sided constraint (4.9a). For lossy branches, flow is represented by a pair of nonnegative variables and thermal limits requira single constraint (4.9b).

Model formulation	# rows	# columns	# non-zeros
Lossy transportation	78,412	71,395	227,817
$b\theta$	109,540	69,379	254,913
Lossy $b\theta$	95,524	83,395	296,961
PTDF	54,752	43,051	621,691
Lossy PTDF	54,990	43,051	1,122,273

5.6 Conclusions and future work

In the work presented in this chapter, we considered four new versions of the CEP model of Chapter 4, adding different representations of power flow constraints. Starting from the conventional $b\theta$ DC power flow neglecting transmission losses, three models were proposed: a $b\theta$ formulation with linear transmission losses, and PTDF formulations with and without transmission losses. While TEP models with PTDF had been proposed in the literature, the joint formulation of CEP proposed here, as well as the modeling done to include transmission losses, are contributions of this work.

Table 5.5.2: Comparison of time performance of different model formulations. Times shown are the average across 365 instances. The time to create the model instance, which includes the time to obtain the PTDF matrix, is larger for the PTDF formulation, but leads to significant time savings during solver execution.

Model formulation	Model creation time (sec)	Gurobi solver time (sec)	Normalized cost
Lossy transportation	16.60	13.22	1.012
$b\theta$	15.38	60.03	1.028
Lossy $b\theta$	15.80	118.75	1.0
PTDF	21.37	27.51	1.028
Lossy PTDF	33.71	68.52	1.024

Computational tests were performed solving single-scenario instances to assess the performance of the different formulations in terms of solution quality and computing time required to reach certain MIP gap targets. Our results suggest that in the lossless case, the PTDF formulation dominates the $b\theta$ formulation, obtaining nearly identical solution quality in a fraction of the time, and should thus be favored for scenario decomposition techniques. In the lossy cases, the PTDF formulation also performs better than $b\theta$ in terms of time, but cannot match its solution quality.

The tradeoff found in our tests between time and solution quality comparing lossy and lossless formulations does not seem to justify the inclusion of transmission losses in the model. We do qualify this statement with two caveats: First, although different loading and generation conditions were used, all tests were conducted on the same network. Tests on larger systems could modify this conclusion. Second, the MIP gap behavior of the lossy PTDF formulation is significantly better than its $b\theta$ counterpart, so the conclusion might again change when considering the iterative solution to the stochastic version of the problem.

In fact, and somewhat unexpectedly, the tradeoff between time and quality found in our tests does not seem to justify the inclusion of any power flow formulation on top of a simple network flow formulation. Again, this statement comes with the caveat that all tests were conducted on the same network; tests on the realistic CATS system, which could not be included in this thesis, might change this conclusion. This is interesting future work that we intend to pursue.

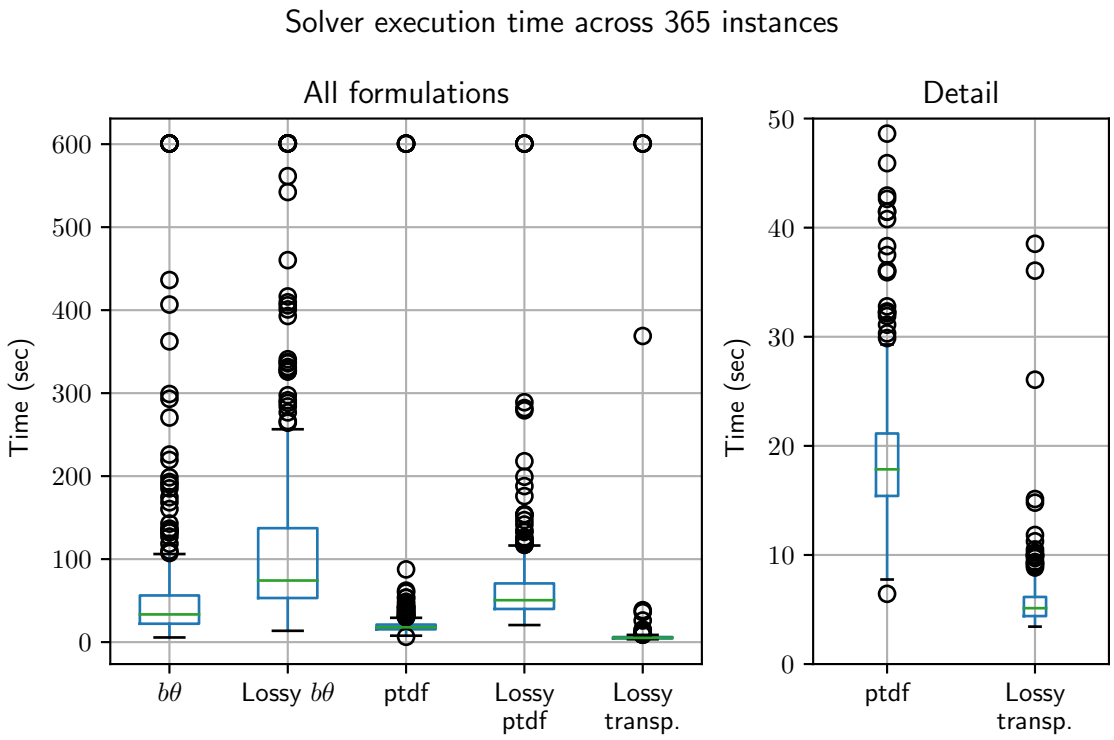

Figure 5.5.1: Box and whisker plots[a] showing distribution of solver execution times for different power flow formulations, over the 365 instances solved. The time limit was set at 600 sec.

[a] The box plot uses matplotlib's default options: "The box extends from the first quartile (Q1) to the third quartile (Q3) of the data, with a line at the median. The whiskers extend from the box to the farthest data point lying within 1.5x the inter-quartile range from the box. Flier points are those past the end of the whiskers."

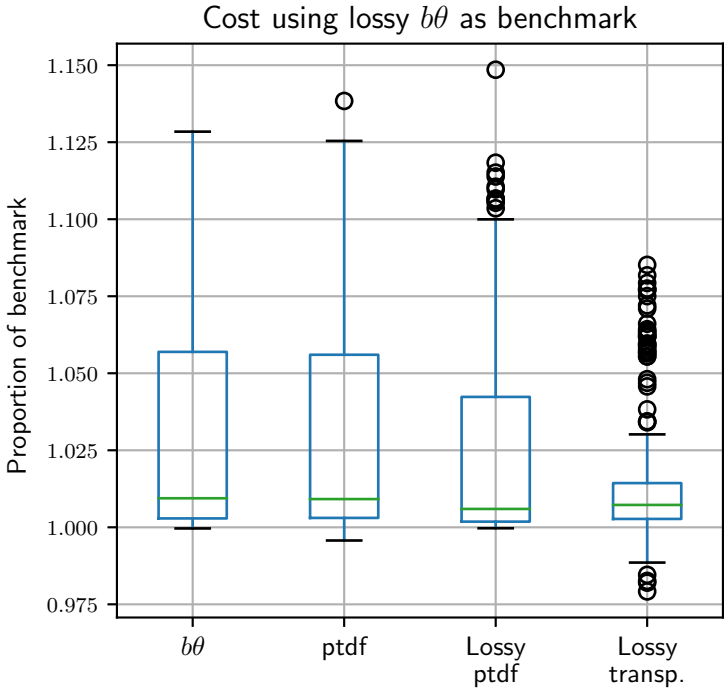

Figure 5.5.2: Box and whisker plots[b] showing distribution of achieved cost for different power flow formulations, over the 365 instances solved, benchmarking against the lossy $b\theta$ formulation. The lossy $b\theta$ formulation is also used for normalizing results, so its achieved cost is 1 for all instances and is omitted from the plot. Note that some instances of the lossy $b\theta$ formulation time out at 600 sec before reaching optimality, so in those cases, it is possible to reach a better objective than the benchmark, i.e. a normalized cost < 1

[b] See footnote [a] in Fig. 5.5.1, page 147.

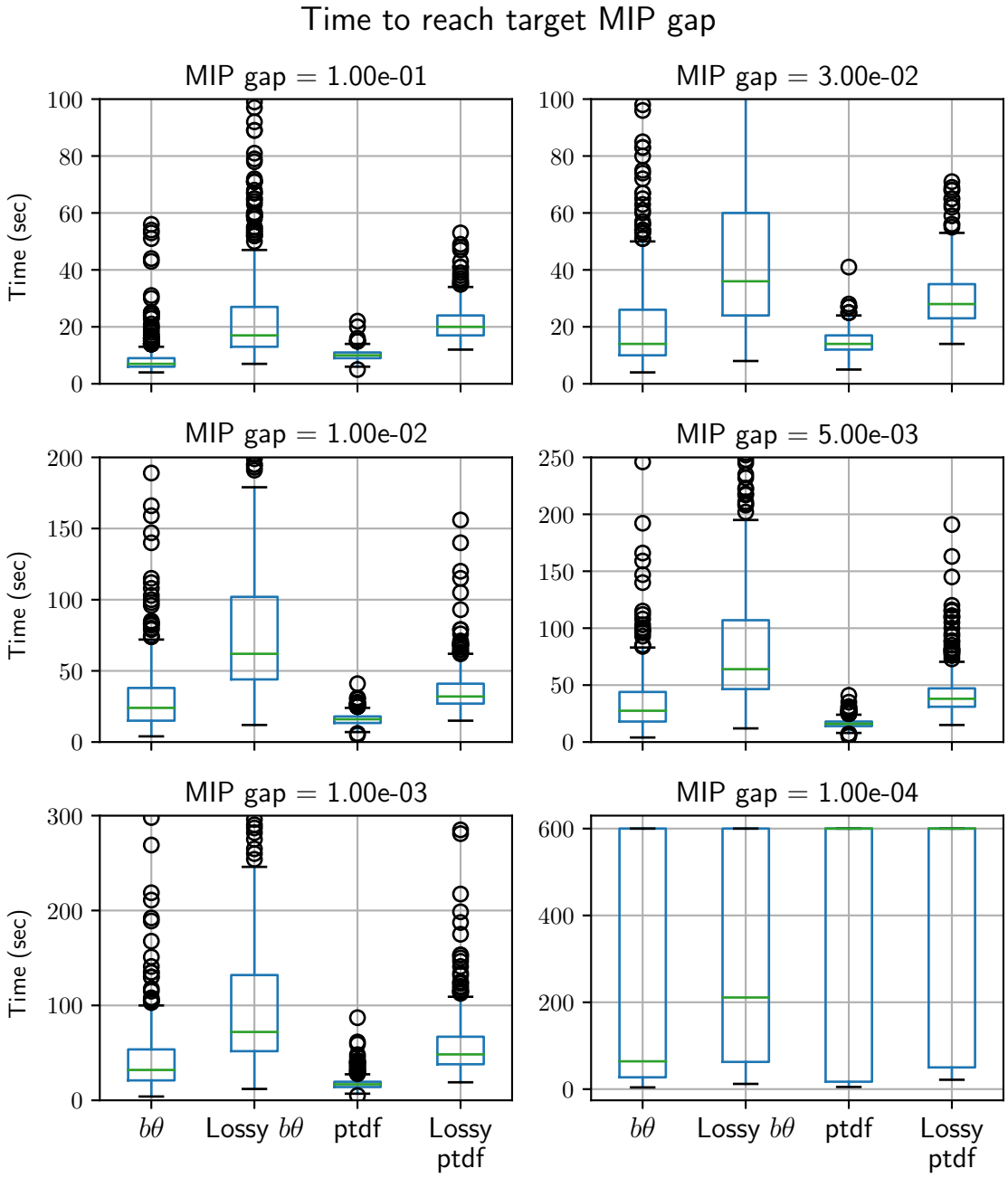

Figure 5.5.3: Box and whisker plots[c] showing distribution of time to reach certain target MIP gaps for different model formulations. In general, PTDF formulations seem to reduce the MIP gap quicker than the $b\theta$ versions, but take longer for the final stretch of gap reduction. 600 sec was the runtime limit, so target gaps may not have been reached at that time.

[c] See footnote [a] in Fig. 5.5.1, page 147.

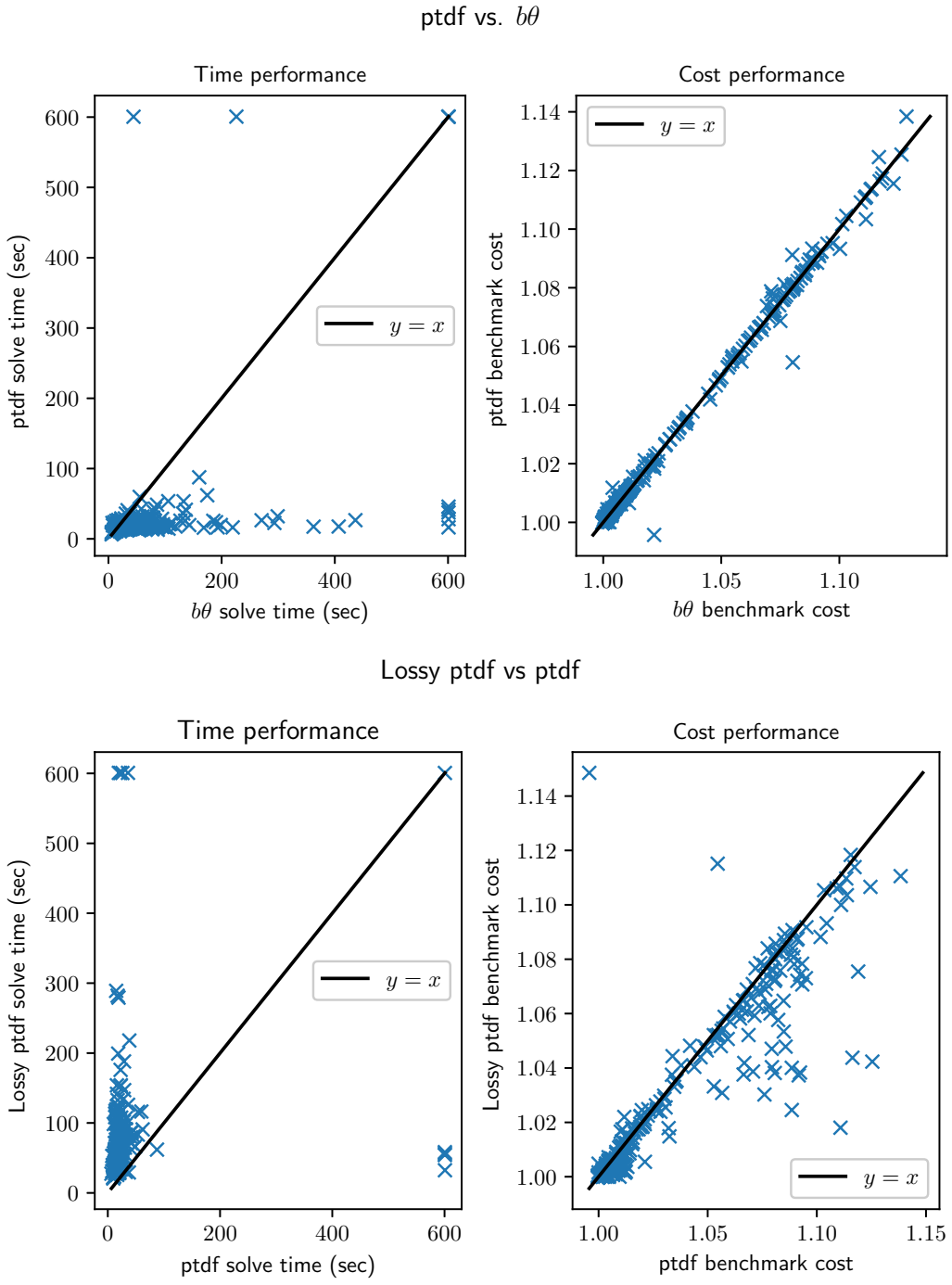

Figure 5.5.4: Scatter plots comparing solution time and solution quality performance of different formulations for each of the 365 instances tested. *Top*: Lossless PTDF vs. $b\theta$. *Bottom*: Lossy PTDF vs lossless PTDF. With nearly identical costs and significantly shorter times, the PTDF clearly dominates the $b\theta$ formulation in the lossless case. A tradeoff appears in the lossy case.

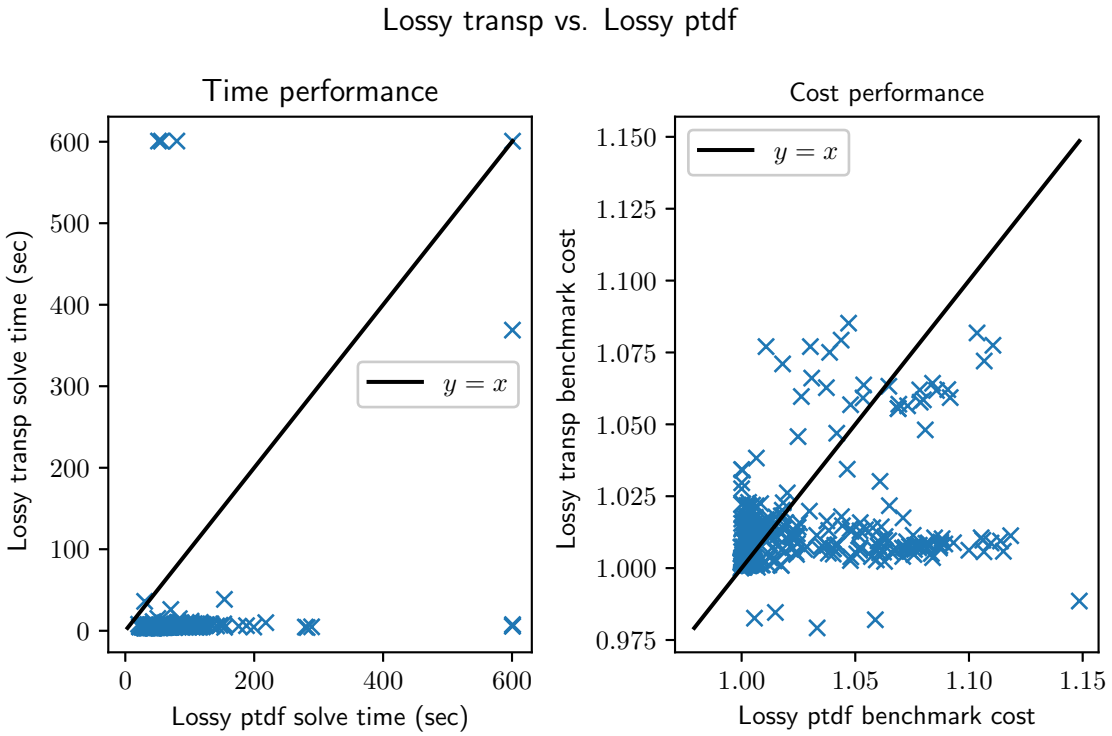

Figure 5.5.5: Scatter plot comparing the time performance and solution quality of lossy transportation and lossy PTDF formulations. These results do not seem to invalidate the simpler transportation formulation as a usable model. Solution time, as expected, are much shorter, but this does not systematically come at the cost of solution quality: neither lossy ptdf nor lossy transportation consistently outperforms the other in terms of solution quality.

Chapter 6

Future avenues of research

Abstract

In each of the three topics studied in this dissertation, interesting avenues of research were identified but could not be pursued because they were out of the timeframe and scope of the project, or interests changed. We devote this chapter to a brief discussion of a few of them.

6.1 Markovian model for battery sizing

Data-driven methods to address variable prices

The most direct continuation of the work on our Markovian model is the implementation of a solution method for the case with variable prices. While we provide initial steps in a path to develop a method for the variable price case in Section 3.5, the challenges identified may lead to a roadblock. An alternative to overcome that obstacle could be to incorporate data-driven methods into the model. One way to do that could be to propose a machine learning model that determines the optimal charging policy for the battery as a function of the state of the continuous-time Markov chain and the long-term commitment, leaving the rest of the model, i.e. finding a limiting distribution and numerically optimizing the resulting function, unchanged. A second option could be to pursue a pure reinforcement learning model instead of attempting to determine the limiting distribution of the state of charge of the battery like we do in the model presented here. Both would significantly extend the usability of our model and hence its interest to researchers and users.

6.2 Community Choice Aggregation

Critical size of the aggregation

The results of Owen [78] predict that asymptotically, all core allocations are dual allocations, which in our setting we have called shadow price allocations. Owen also finds that in the case without degeneracy, the convergence happens with a large enough but finite number of participants. In the tests that were conducted in our work, the number of participants was only increased to 50, but the shrinking of the uniform price core was already apparent.

If the uniform price core is a singleton set, the question of finding the most fair allocation therein becomes moot, so identifying if the core has collapsed is an important point. An interesting open question is thus a characterization of the critical size of the aggregation before the collapse of the uniform price core into the dual core. It is conceivable that the structure of the set of participants, i.e. how many agents have local generation and storage and of which sizes, as well as the degree of correlation between their generation and consumption, will affect the outcome. If this characterization can be obtained, interesting models and questions ensue, for instance: should a cap be put on the size of aggregations? Given a cap, large Community Choice Aggregations (CCAs) need to be broken down; how should tentative participants be clustered into CCAs? Can there be a market for CCA memberships? Would such a market promote efficiency? unfairness?

Trading stability for fairness

Throughout the work of Chapter 2, we treat the stability of the imputation returned by the mechanism as a hard constraint. We are only willing to relax it to a concept of approximate, highly-probable stability to recover the otherwise lost computational tractability. An interesting alternative would be to consider a model in which stability is explicitly traded for fairness. This can be achieved for example through a least-core formulation similar to that of [54, 106]. Again, numerous interesting questions follow: what would be the tradeoff between fairness and stability? Can we characterize the coalitions that would need to be "subsidized" to gain insights about implementability challenges?

6.3 Parallel computing for Capacity Expansion Planning (CEP)

Power flow representations and scenario selection

This work was conducted in collaboration with Lawrence Livermore National Laboratory (LLNL), as part of an ongoing project there, so some immediate extensions of the work performed as part of this thesis are already foreseen within that project. This includes implementing the power flow formulations of Chapter 5 on the California Test System (CATS) test case, as well as testing the computational limits reached when using more accurate,

convex formulations of AC power flow. Similarly, work on scenario selection and reduction is an ongoing branch of that project. Numerous interesting questions arise regarding that topic, since assigning probabilities to outlier scenarios with particularly high costs (extreme weather events) is not a trivial task. Is Robust Optimization the only option, or can the stochastic programming version of the problem still make sense in that case? Or perhaps a combination thereof, as in [68]?

Asynchronous Progressive Hedging Algorithm (PHA)

During our tests of solving the problem with the PHA, it was observed that in most iterations, only a handful of complicating scenarios were holding back the execution of the algorithm on all cores. We addressed this issue by proposing a runtime-dependent target Mixed Integer Program (MIP) gap, with satisfactory results. An alternative which we did not explore in this work is implementing the Asynchronous PHA instead [30], or variants thereof, which may lead to better time and quality performance, and an overall better usage of the available resources.

Extensions to mpi-sppy

During the implementation of our solution method in software using mpi-sppy, a couple of useful extensions to the tool have been identified, which if implemented could positively impact the CEP software developed. The first one is the possibility of assigning processes (*ranks* in Message Passing Interface (MPI) jargon) to cylinders asymmetrically. In the current implementation, each cylinder gets the same number of ranks, regardless of its nature. However, in problems like ours, where second-stage subproblems are pure Linear Programs (LPs), the subproblems solved by inner-bounder spokes represent a significantly lighter computational load than those of outer-bounder spokes. A second extension that could be of interest is a spoke that is given a different model, known to the modeler to be a relaxation (or otherwise valid outer-bounder) of the actual model. It is conceivable that it is easier to provide tight outer bounds on certain model relaxations than on the original model, and thus such a spoke could lead to better bounds in early iterations.

Bibliography

[1] Nail Akar and Khosrow Sohraby. "Infinite- and finite-buffer Markov fluid queues: a unified analysis". In: *Journal of Applied Probability* 41.2 (June 2004), pp. 557–569.

[2] Pablo Apablaza, Sebastián Püschel Løvengreen, Rodrigo Moreno, Sleiman Mhanna, and Pierluigi Mancarella. "Assessing the Impact of DER on the Expansion of Low-Carbon Power Systems under Deep Uncertainty". In: *2024 Power Systems Computation Conference*. June 2024.

[3] Eric Balkanski, Umar Syed, and Sergei Vassilvitskii. "Statistical Cost Sharing". In: *arXiv:1703.03111 [cs]* (Mar. 2017).

[4] Stephen M. Barrager and Edward G. Cazalet. *Transactive energy: a sustainable business and regulatory model for electricity*. eng. Reston: Public Utilities Reports, Inc, 2016. ISBN: 978-0-910325-47-9.

[5] Dimitri P. Bertsekas. *Dynamic programming and optimal control*. Fourth edition. Belmont, Mass: Athena Scientific, 2012. ISBN: 978-1-886529-08-3.

[6] S. Binato, M.V.F. Pereira, and S. Granville. "A new Benders decomposition approach to solve power transmission network design problems". In: *IEEE Transactions on Power Systems* 16.2 (May 2001), pp. 235–240.

[7] Adam B. Birchfield. *SouthCarolina 500-Bus System: ACTIVSg500*. URL: `https://electricgrids.engr.tamu.edu/electric-grid-test-cases/activsg500/` (visited on 05/25/2024).

[8] Adam B. Birchfield, Ti Xu, Kathleen M. Gegner, Komal S. Shetye, and Thomas J. Overbye. "Grid Structural Characteristics as Validation Criteria for Synthetic Networks". In: *IEEE Transactions on Power Systems* 32.4 (July 2017), pp. 3258–3265.

[9] John R. Birge and François Louveaux. *Introduction to Stochastic Programming*. Springer Science & Business Media, June 2011. ISBN: 978-1-4614-0237-4.

[10] Eilyan Y. Bitar, Ram Rajagopal, Pramod P. Khargonekar, Kameshwar Poolla, and Pravin Varaiya. "Bringing Wind Energy to Market". In: *IEEE Transactions on Power Systems* 27.3 (Aug. 2012), pp. 1225–1235.

[11] Mark Bolinger, Joachim Seel, Cody Warner, and Dana Robson. *Utility-Scale Solar, 2021 Edition*. Technical Report. Lawrence Berkeley National Laboratory, Oct. 2021. URL: https://emp.lbl.gov/sites/default/files/utility_scale_solar_2021_edition_slides.pdf (visited on 05/03/2022).

[12] Severin Borenstein, Meredith Fowlie, and James Sallee. *Designing Electricity Rates for An Equitable Energy Transition*. Tech. rep. Next 10 & Energy Institute at Haas, Feb. 2021. URL: https://www.next10.org/publications/electricity-rates.

[13] Kevin Brokish and James Kirtley. "Pitfalls of modeling wind power using Markov chains". In: *2009 IEEE/PES Power Systems Conference and Exposition*. Mar. 2009, pp. 1–6.

[14] Michael L. Bynum, Gabriel A. Hackebeil, William E. Hart, Carl D. Laird, Bethany L. Nicholson, et al. *Pyomo–optimization modeling in python*. Third. Vol. 67. Springer Science & Business Media, 2021.

[15] California Public Utilities Commission. *Renewables Portfolio Standard (RPS) Program*. URL: https://www.cpuc.ca.gov/rps/ (visited on 09/20/2023).

[16] Timothy Capper, Anna Gorbatcheva, Mustafa A. Mustafa, Mohamed Bahloul, Jan Marc Schwidtal, et al. "Peer-to-peer, community self-consumption, and transactive energy: A systematic literature review of local energy market models". In: *Renewable and Sustainable Energy Reviews* 162 (July 2022), p. 112403.

[17] Dominion Energy South Carolina. *2023 Integrated Resource Plan*. Integrated Resource Plan Docket No. 2023-9-E. Dominion Energy South Carolina, Jan. 2023, p. 45. URL: https://dms.psc.sc.gov/Attachments/Matter/ee0417c1-e32f-47f4-a9ee-fd3dc0725186 (visited on 05/25/2024).

[18] Xiangyu Chen, Ka-Cheong Leung, and Albert Y. S. Lam. "Power Output Smoothing for Renewable Energy System: Planning, Algorithms, and Analysis". In: *IEEE Systems Journal* 14.1 (Mar. 2020), pp. 1034–1045.

[19] Xin Chen and Jiawei Zhang. "A Stochastic Programming Duality Approach to Inventory Centralization Games". In: *Operations Research* 57.4 (Aug. 2009), pp. 840–851.

[20] California Public Utilities Commission. *CPUC Modernizes Multi-Tenant and Multi-Property Solar and Storage Tariffs and Strengthens Solar Consumer Protections*. Press Release. Docket number: R.20-08-020. Nov. 2023. URL: https://www.cpuc.ca.gov/news-and-updates/all-news/cpuc-modernizes-multi-tenant-and-multi-property-solar-and-storage-tariffs-2023 (visited on 05/09/2024).

[21] Antonio J. Conejo, ed. *Decomposition techniques in mathematical programming: engineering and science applications*. Berlin ; New York: Springer, 2006. ISBN: 978-3-540-27685-2.

[22] Antonio J. Conejo, Luis Baringo Morales, S. Jalal Kazempour, and Afzal S. Siddiqui. *Investment in Electricity Generation and Transmission*. Springer International Publishing, 2016. ISBN: 978-3-319-29499-5. DOI: 10.1007/978-3-319-29501-5.

[23] North American Reliability Corporation. *2022 Long-Term Reliability Assessment*. Tech. rep. Dec. 2022, p. 64. URL: https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_LTRA_2022.pdf (visited on 05/25/2024).

[24] Michael T. Craig, Jan Wohland, Laurens P. Stoop, Alexander Kies, Bryn Pickering, et al. "Overcoming the disconnect between energy system and climate modeling". In: *Joule* 6.7 (July 2022), pp. 1405–1417.

[25] Lisandro Dalcin and Yao-Lung L. Fang. "mpi4py: Status Update After 12 Years of Development". In: *Comput. Sci. Eng.* 23.4 (July 2021), pp. 47–54.

[26] Kevin De León. *Clean Energy and Pollution Reduction Act of 2015*. Oct. 2015. URL: https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160SB350 (visited on 05/26/2024).

[27] Chris J. Dent, Janusz W. Bialek, and Benjamin F. Hobbs. "Opportunity Cost Bidding by Wind Generators in Forward Markets: Analytical Results". In: *IEEE Transactions on Power Systems* 26.3 (Aug. 2011), pp. 1600–1608.

[28] Vivek Deulkar, Jayakrishnan Nair, and Ankur A. Kulkarni. "Sizing Storage for Reliable Renewable Integration". In: *2019 IEEE Milan PowerTech*. June 2019, pp. 1–6.

[29] Bhaskar Dutta and Debraj Ray. "A Concept of Egalitarianism Under Participation Constraints". In: *Econometrica* 57.3 (1989), pp. 615–635.

[30] Jonathan Eckstein, Jean-Paul Watson, and David L. Woodruff. "Projective Hedging Algorithms for Multistage Stochastic Programming, Supporting Distributed and Asynchronous Implementation". In: *Operations Research* (July 2023).

[31] Brent Eldridge, Richard O'Neill, and Anya Castillo. "An Improved Method for the DCOPF With Losses". In: *IEEE Transactions on Power Systems* 33.4 (July 2018), pp. 3779–3788.

[32] Desta Z. Fitiwi, L. Olmos, M. Rivier, F. de Cuadra, and I. J. Pérez-Arriaga. "Finding a representative network losses model for large-scale transmission expansion planning with renewable energy sources". In: *Energy* 101 (Apr. 2016), pp. 343–358.

[33] Dinakar Gade, Gabriel Hackebeil, Sarah M. Ryan, Jean-Paul Watson, Roger J.-B. Wets, et al. "Obtaining lower bounds from the progressive hedging algorithm for stochastic mixed-integer programs". In: *Math. Program.* 157.1 (May 2016), pp. 47–67.

[34] D.P. Gaver and R.G. Miller. "Limiting distributions for some storage problems". In: *Studies in applied probability and management science*. Ed. by Kenneth Arrow, Samuel Karlin, and Herbert Scarf. Stanford University Press, 1962, pp. 110–126.

[35] Roderick S. Go, Francisco D. Muñoz, and Jean-Paul Watson. "Assessing the economic value of co-optimized grid-scale energy storage investments in supporting high renewable portfolio standards". In: *Applied Energy* 183 (Dec. 2016), pp. 902–913.

[36] Isaac-Camilo Gonzalez-Romero, Sonja Wogrin, and Tomas Gómez. "Review on generation and transmission expansion co-planning models under a market environment". In: *IET Generation, Transmission & Distribution* 14.6 (2020), pp. 931–944.

[37] John J. Grainger, William D. Stevenson, and William D. Stevenson. *Power system analysis*. New York: McGraw-Hill, 1994. ISBN: 978-0-07-061293-8.

[38] Liyang Han, Thomas Morstyn, and Malcolm McCulloch. "Incentivizing Prosumer Coalitions With Energy Management Using Cooperative Game Theory". In: *IEEE Transactions on Power Systems* 34.1 (Jan. 2019), pp. 303–313.

[39] Pavithra Harsha and Munther Dahleh. "Optimal Management and Sizing of Energy Storage Under Dynamic Pricing for the Efficient Integration of Renewable Energy". In: *IEEE Transactions on Power Systems* 30.3 (May 2015), pp. 1164–1181.

[40] William E Hart, Jean-Paul Watson, and David L Woodruff. "Pyomo: modeling and solving mathematical programs in Python". In: *Mathematical Programming Computation* 3.3 (2011), pp. 219–260.

[41] Reza Hemmati, Rahmat-Allah Hooshmand, and Amin Khodabakhshian. "Comprehensive review of generation and transmission expansion planning". In: *IET Generation, Transmission & Distribution* 7.9 (Sept. 2013), pp. 955–964.

[42] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, et al. "The ERA5 global reanalysis". In: *Quart J Royal Meteoro Soc* 146.730 (July 2020), pp. 1999–2049.

[43] V. H. Hinojosa and J. Velásquez. "Improving the mathematical formulation of security-constrained generation capacity expansion planning using power transmission distribution factors and line outage distribution factors". In: *Electric Power Systems Research* 140 (Nov. 2016), pp. 391–400.

[44] Frédéric Hourdin, Thorsten Mauritsen, Andrew Gettelman, Jean-Christophe Golaz, Venkatramani Balaji, et al. "The Art and Science of Climate Model Tuning". EN. In: *Bulletin of the American Meteorological Society* 98.3 (Mar. 2017), pp. 589–602.

[45] Jesse D. Jenkins, Erin N. Mayfield, Eric D. Larson, Stephen W. Pacala, and Chris Greig. "Mission net-zero America: The nation-building path to a prosperous, net-zero emissions economy". English. In: *Joule* 5.11 (Nov. 2021), pp. 2755–2761.

[46] Valentin Kaisermayer, Daniel Muschick, Martin Horn, and Markus Gölles. "Progressive hedging for stochastic energy management systems". In: *Energy Syst* 12.1 (Feb. 2021), pp. 1–29.

[47] H. Emre Kankaya and Nail Akar. "Solving Multi-Regime Feedback Fluid Queues". In: *Stochastic Models* 24.3 (Aug. 2008), pp. 425–450.

[48] Jae Ho Kim and Warren B. Powell. "Optimal Energy Commitments with Storage and Intermittent Supply". In: *Operations Research* 59.6 (Dec. 2011), pp. 1347–1360.

[49] Daniel Sadi Kirschen and Goran Strbac. *Fundamentals of power system economics*. Second Edition. Hoboken: Wiley, 2019. ISBN: 978-1-119-30988-8.

[50] Bernard Knueven, David Mildebrath, Christopher Muir, John D. Siirola, Jean-Paul Watson, et al. "A parallel hub-and-spoke system for large-scale scenario-based optimization under uncertainty". In: *Math. Prog. Comp.* (Aug. 2023).

[51] Bernard Knueven, James Ostrowski, and Jean-Paul Watson. "On Mixed-Integer Programming Formulations for the Unit Commitment Problem". In: *INFORMS Journal on Computing* 32.4 (Oct. 2020), pp. 857–876.

[52] V.G. Kulkarni. "Fluid Models for Single Buffer Systems". In: *Frontiers in Queueing: Models and Applications in Science and Engineering*. Ed. by Jewgeni H. Dshalalow. CRC Press, Jan. 1997, pp. 321–338. ISBN: 978-0-8493-8076-1.

[53] L. Ruby Leung, David C. Bader, Mark A. Taylor, and Renata B. McCoy. "An Introduction to the E3SM Special Collection: Goals, Science Drivers, Development, and Analysis". In: *Journal of Advances in Modeling Earth Systems* 12.11 (2020), e2019MS001821.

[54] Jing Li, Yujian Ye, Dimitrios Papadaskalopoulos, and Goran Strbac. "Computationally Efficient Pricing and Benefit Distribution Mechanisms for Incentivizing Stable Peer-to-Peer Energy Trading". In: *IEEE Internet of Things Journal* 8.2 (Jan. 2021), pp. 734–749.

[55] E. Litvinov, Tongxin Zheng, G. Rosenwald, and P. Shamsollahi. "Marginal loss modeling in LMP calculation". In: *IEEE Transactions on Power Systems* 19.2 (May 2004), pp. 880–888.

[56] Duke Energy Progress LLC. *2023 Integrated Resource Plan. Annex D: Electric Load Forecast.* Integrated Resource Plan Docket No. 2023-8-E. Aug. 2023. URL: `https://dms.psc.sc.gov/Attachments/Matter/6c2bed37-eba2-4d58-8fec-16e694d9fea4` (visited on 05/25/2024).

[57] Vitor V. Lopes, Teresa Scholz, Ana Estanqueiro, and Augusto Q. Novais. "On the use of Markov chain models for the analysis of wind power time-series". In: *2012 11th International Conference on Environment and Electrical Engineering*. May 2012, pp. 770–775.

[58] Angeliki Loukatou, Paul Johnson, Sydney Howell, and Peter Duck. "Optimal valuation of wind energy projects co-located with battery storage". In: *Applied Energy* 283 (Feb. 2021), p. 116247.

[59] Sweta Malik, Maeve Duffy, Subhasis Thakur, Barry Hayes, and John Breslin. "A priority-based approach for peer-to-peer energy trading using cooperative game theory in local energy community". In: *International Journal of Electrical Power & Energy Systems* 137 (May 2022), p. 107865.

[60] Andreu Mas-Colell, Michael Dennis Whinston, and Jerry R. Green. "Social Welfare Functions and Social Optima". eng. In: *Microeconomic theory*. Section: 22.C. New York: Oxford University Press, 1995, pp. 825–831. ISBN: 0-19-507340-1.

[61] Malte Meinshausen, S. J. Smith, K. Calvin, J. S. Daniel, M. L. T. Kainuma, et al. "The RCP greenhouse gas concentrations and their extensions from 1765 to 2300". In: *Climatic Change* 109.1 (Aug. 2011), p. 213.

[62] Gianluigi Migliavacca, Marco Rossi, Dario Siface, Matteo Marzoli, Hakan Ergun, et al. "The Innovative FlexPlan Grid-Planning Methodology: How Storage and Flexible Resources Could Help in De-Bottlenecking the European System". In: *Energies* 14.4 (Jan. 2021), p. 1194.

[63] Paul Milgrom and Mitchell Watt. "Linear Pricing Mechanisms for Markets without Convexity". In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. EC '22. New York, NY, USA: Association for Computing Machinery, July 2022, p. 300. ISBN: 978-1-4503-9150-4.

[64] Kelsey Misbrener. *CPUC undermines the economic value of California schools, apartments going solar*. Nov. 2023. URL: `https://www.solarpowerworldonline.com/2023/11/cpuc-undermines-economic-value-california-schools-apartment-solar/` (visited on 05/09/2024).

[65] Debasis Mitra. "Stochastic Theory of a Fluid Model of Producers and Consumers Coupled by a Buffer". In: *Advances in Applied Probability* 20.3 (1988), pp. 646–676.

[66] Kendall Mongird, Vilayanur Viswanathan, Jan Alam, Charlie Vartanian, Vincent Sprenkle, et al. *2020 Grid Energy Storage Technology Cost and Performance Assessment*. Tech. rep. DOE/PA-0204. Pacific Northwest National Laboratory, Dec. 2020, p. 117.

[67] Minda M Monteagudo, Stephen Po-Chedley, and Jean-Paul Watson. "Population and Temperature Impacts on Electricity Demand in California". In: *AGU23* (2023).

[68] Alexandre Moreira, David Pozo, Alexandre Street, Enzo Sauma, and Goran Strbac. "Climate-aware generation and transmission expansion planning: A three-stage robust optimization approach". In: *European Journal of Operational Research* 295.3 (Dec. 2021), pp. 1099–1118.

[69] Hervé. Moulin. *Fair division and collective welfare / Hervé Moulin*. eng. Cambridge, Mass: MIT Press, 2003. ISBN: 0-262-13423-3.

[70] Francisco D. Muñoz, Benjamin F. Hobbs, Jonathan L. Ho, and Saamrat Kasina. "An Engineering-Economic Approach to Transmission Planning Under Market and Regulatory Uncertainties: WECC Case Study". In: *IEEE Transactions on Power Systems* 29.1 (Jan. 2014), pp. 307–317.

[71] Francisco D. Muñoz and Jean-Paul Watson. "A scalable solution framework for stochastic transmission and generation planning problems". In: *Computational Management Science* 12.4 (Oct. 2015), pp. 491–518.

[72] Amelia Musselman. *Climate-Resilient Stochastic Power System Expansion Planning for a Realistic California Test Case*. Working paper. Lawrence Livermore National Lab. (LLNL), June 2024.

[73] George Nemhauser and Laurence Wolsey. *Integer and Combinatorial Optimization*. 1st ed. Wiley, June 1988. ISBN: 978-0-471-82819-8. DOI: 10.1002/9781118627372.

[74] Fabian Neumann, Veit Hagenmeyer, and Tom Brown. "Assessments of linear power flow and transmission loss approximations in coordinated capacity expansion problems". In: *Applied Energy* 314 (May 2022), p. 118859.

[75] NREL. *PySAM - Python API for System Advisory Model*. Golden, CO, July 2023. URL: github.com/nrel/pysam (visited on 07/26/2023).

[76] NREL. *Wind generation profile in 2012*. Tech. rep. National Renewable Energy Laboratory, Mar. 2016. URL: https://%20www.dropbox.com/s/a67gw17wz1clk50/34843-2012.csv?dl=0.

[77] Shmuel S. Oren and Shao Hong Wan. "Optimal Strategic Petroleum Reserve Policies: A Steady State Analysis". In: *Management Science* 32.1 (Jan. 1986), pp. 14–29.

[78] Guillermo Owen. "On the core of linear production games". In: *Mathematical Programming* 9.1 (Dec. 1975), pp. 358–370.

[79] George Papaefthymiou and Bernd Klockl. "MCMC for Wind Power Simulation". In: *IEEE Transactions on Energy Conversion* 23.1 (Mar. 2008), pp. 234–240.

[80] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830.

[81] Ivan Penn. "California Panel Backs Solar Mandate for New Buildings". In: *The New York Times* (Aug. 2021).

[82] Federico Perea, Justo Puerto, and Francisco R. Fernández. "Avoiding unfairness of *Owen* allocations in linear production processes". In: *European Journal of Operational Research* 220.1 (July 2012), pp. 125–131.

[83] M. V. F. Pereira and L. M. V. G. Pinto. "Multi-stage stochastic optimization applied to energy planning". In: *Mathematical Programming* 52.1 (May 1991), pp. 359–375.

[84] David Pozo, Enzo E. Sauma, and Javier Contreras. "A Three-Level Static MILP Model for Generation and Transmission Expansion Planning". In: *IEEE Trans. Power Syst.* 28.1 (Feb. 2013), pp. 202–210.

[85] Mohsen Rahmani, Amin Kargarian, and Gabriela Hug. "Comprehensive power transfer distribution factor model for large-scale transmission expansion planning". In: *IET Generation, Transmission & Distribution* 10.12 (2016), pp. 2981–2989.

[86] Elena Raycheva, Jared B. Garrison, Christian Schaffner, and Gabriela Hug. *The Value of Flexibility in a Carbon Neutral Power System*. arXiv:2211.13625 [math]. Nov. 2022. DOI: 10.48550/arXiv.2211.13625. (Visited on 08/25/2023).

[87] R. T. Rockafellar and Roger J.-B. Wets. "Scenarios and Policy Aggregation in Optimization Under Uncertainty". In: *Mathematics of OR* 16.1 (Feb. 1991), pp. 119–147.

[88] Yu Ru, Jan Kleissl, and Sonia Martinez. "Storage Size Determination for Grid-Connected Photovoltaic Systems". In: *IEEE Transactions on Sustainable Energy* 4.1 (Jan. 2013), pp. 68–81.

[89] Pablo A. Ruiz, Aleksandr Rudkevich, Michael C. Caramanis, Evgenyi Goldis, Elli Ntakou, et al. "Reduced MIP formulation for transmission topology control". In: *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Oct. 2012, pp. 1073–1079.

[90] Enzo E. Sauma and Shmuel S. Oren. "Proactive planning and valuation of transmission investments in restructured electricity markets". In: *J Regul Econ* 30.3 (Nov. 2006), pp. 358–387.

[91] Teresa Scholz, Vitor V. Lopes, and Ana Estanqueiro. "A cyclic time-dependent Markov process to model daily patterns in wind turbine power production". In: *Energy* 67 (Apr. 2014), pp. 557–568.

[92] Bruno Sericola. "A Finite Buffer Fluid Queue Driven by a Markovian Queue". In: *Queueing Systems* 38.2 (June 2001), pp. 213–220.

[93] Bruno Sericola and Bruno Tuffin. "A fluid queue driven by a Markovian queue". In: *Queueing Systems* 31.3 (July 1999), pp. 253–264.

[94] Alexander Shapiro, Darinka Dentcheva, and Andrzej P. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. MPS-SIAM series on optimization 9. Philadelphia: Society for Industrial and Applied Mathematics : Mathematical Programming Society, 2009. ISBN: 978-0-89871-687-0.

[95] Matthew Signorotti. *Computational Pipeline Predicts Solar and Wind Energy Availability under Various Climate Change Projections*. Poster. Washington DC, Apr. 2024.

[96] Fereidoon P. Sioshansi, ed. *Evolution of global electricity markets: new paradigms, new challenges, new approaches*. 1st ed. Amsterdam ; Boston: Elsevier, 2013. ISBN: 978-0-12-397891-2.

[97] Mark Specht. *The Basics of Integrated Resource Planning in California*. Section: Energy. May 2019. URL: `https://blog.ucsusa.org/mark-specht/integrated-resource-planning-california/` (visited on 05/27/2024).

[98] Steven Stoft. *Power system economics: designing markets for electricity*. Piscataway, NJ : New York: IEEE Press ; Wiley-Interscience, 2002. ISBN: 978-0-471-15040-4.

[99] Jie Tang, Alexandre Brouste, and Kwok Leung Tsui. "Some improvements of wind speed Markov chain modeling". In: *Renewable Energy* 81 (Sept. 2015), pp. 52–56.

[100] Joshua Adam Taylor. *Convex Optimization of Power Systems:* 1st ed. Cambridge University Press, Feb. 2015. ISBN: 978-1-107-07687-7. DOI: 10.1017/CBO9781139924672.

[101] Sofia Taylor, Aditya Rangarajan, Noah Rhodes, Jonathan Snodgrass, Bernie Lesieutre, et al. "California Test System (CATS): A Geographically Accurate Test System based on the California Grid". In: *IEEE Transactions on Energy Markets, Policy and Regulation* (2023), pp. 1–12.

[102] Wayes Tushar, Tapan Kumar Saha, Chau Yuen, Paul Liddell, Richard Bean, et al. "Peer-to-Peer Energy Trading With Sustainable User Participation: A Game Theoretic Approach". In: *IEEE Access* 6 (2018), pp. 62932–62943.

[103] Nelson A. Uhan. "Stochastic linear programming games with concave preferences". In: *European Journal of Operational Research* 243.2 (June 2015), pp. 637–646.

[104] Tomas Valencia Zuluaga, Amelia Musselman, Jean-Paul Watson, and Shmuel Oren. "Parallel Computing for Power System Climate Resiliency: Solving a Large-Scale Stochastic Capacity Expansion Problem with mpi-sppy". In: *Electric Power Systems Research, Special issue: PSCC 2024* (2024).

[105] Tomas Valencia Zuluaga, Amelia Musselman, Jean-Paul Watson, and Shmuel S. Oren. "Parallel Computing for Power System Climate Resiliency: Solving a Large-Scale Stochastic Capacity Expansion Problem with mpi-sppy". In: *2024 Power Systems Computation Conference*. Paris, June 2024.

[106] Tomas Valencia Zuluaga and Shmuel S. Oren. "Cost Sharing Mechanism with Statistical Learning for Peer-to-Peer Energy Trading". In: *2023 IEEE Power & Energy Society General Meeting (PESGM)*. July 2023, pp. 1–5.

[107] Tomas Valencia Zuluaga and Shmuel S. Oren. "Data-Driven Sizing of Co-Located Storage for Uncertain Renewable Energy". In: *Policy and Regulation IEEE Transactions on Energy Markets* 1.4 (Dec. 2023), pp. 348–359.

[108] Tomas Valencia Zuluaga and Shmuel S. Oren. "Sizing Co-located Storage for Uncertain Renewable Energy Sold Through Forward Contracts". In: *Proceedings of the 11th Bulk Power Systems Dynamics and Control Symposium (IREP 2022)*. Banff, Canada, 2022.

[109] Fernando Badilla Veliz, Jean-Paul Watson, Andres Weintraub, Roger J.-B. Wets, and David L. Woodruff. "Stochastic optimization models in forest planning: a progressive hedging solution approach". In: *Ann Oper Res* 232.1 (Sept. 2015), pp. 259–274.

[110] Jean-Paul Watson and David L. Woodruff. "Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems". In: *Comput Manag Sci* 8.4 (Nov. 2011), pp. 355–370.

[111] Eric Wilson, Andrew Parker, Anthony Fontanini, Elaina Present, Janet Reyna, et al. *End-Use Load Profiles for the U.S. Building Stock: Methodology and Results of Model Calibration, Validation, and Uncertainty Quantification*. Tech. rep. NREL/TP-5500-80889, 1854582, MainId:78667. NREL, Mar. 2022, NREL/TP–5500–80889, 1854582, MainId:78667. DOI: 10.2172/1854582. URL: https://www.osti.gov/servlets/purl/1854582/ (visited on 11/02/2022).

[112] Allen J. Wood, Bruce F. Wollenberg, and Gerald B. Sheblé. "Calculation of Network Sensitivity Factors". In: *Power Generation, Operation, and Control*. 3rd. John Wiley & Sons, Dec. 2013, pp. 339–343. ISBN: 978-1-118-73391-2.

[113] Allen J. Wood, Bruce F. Wollenberg, and Gerald B. Sheblé. *Power Generation, Operation, and Control*. 3rd. John Wiley & Sons, Dec. 2013. ISBN: 978-1-118-73391-2.

[114] Kaigui Xie, Qinglong Liao, Heng-Ming Tai, and Bo Hu. "Non-Homogeneous Markov Wind Speed Time Series Model Considering Daily and Seasonal Variation Characteristics". In: *IEEE Transactions on Sustainable Energy* 8.3 (July 2017), pp. 1281–1290.

[115] Haoyu Yuan, Reetam Sen Biswas, Jin Tan, and Yingchen Zhang. "Developing a Reduced 240-Bus WECC Dynamic Model for Frequency Response Study of High Renewable Integration". In: *2020 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*. 2020, pp. 1–5.

[116] Jishi Zhang, Peter Bogenschutz, Qi Tang, Philip Cameron-smith, and Chengzhu Zhang. "Leveraging Regional Mesh Refinement to Simulate Future Climate Projections for California Using the Simplified Convection Permitting E3SM Atmosphere Model Version 0". English. In: *EGUsphere* (Oct. 2023), pp. 1–49.