Title

Managing Uncertainty: Scenario Generation and Control in Renewable Energy Systems

Author

Goujard, Guillaume

Publication Date

2023

Peer reviewed|Thesis/dissertation

Managing Uncertainty: Scenario Generation and Control in Renewable Energy Systems

By

Guillaume Goujard

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Scott J. Moura, Chair
Professor Maria Laura Delle Monache
Professor Alper Atamturk
Professor Daniel Arnold

Fall 2023

Managing Uncertainty: Scenario Generation and Control in Renewable Energy Systems

Abstract

Managing Uncertainty: Scenario Generation and Control in Renewable Energy Systems

By

Guillaume Goujard

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Scott J. Moura, Chair

This dissertation studies renewable energy systems, addressing the pressing challenges posed by their inherent uncertainty and the need for adapted control and forecasting strategies. With the increasing reliance on renewable sources to mitigate climate change impacts, understanding and optimizing these systems becomes crucial. This work presents new methodologies in four areas, each contributing to the efficient management and integration of renewable energy resources in the evolving energy market landscape.

The first chapter introduces a novel approach to the siting, sizing, and bid scheduling of a price-maker battery in a nodal wholesale market. This framework, developed through mixed-integer optimization, demonstrates how strategic positioning and sizing of battery storage can influence prices and alleviate congestion, turning a profit in markets traditionally challenged by high capital costs. By applying this model to New Zealand, the study offers practical insights into maximizing battery storage profitability in nodal markets.

In the second chapter, the focus shifts to the modeling and state estimation of Lithium-Sulfur (Li-S) batteries using a Piecewise Affine (PWA) system. Addressing the complexity of differential algebraic equations in standard battery models, this chapter presents a more efficient approach for real-time state estimation, crucial for applications demanding high energy density. The integration of a PWA model within a moving horizon estimation framework significantly improves state of charge estimates. This chemistry is particularly useful for sectors like electrified aviation and heavy-duty transport.

The third chapter explores the dynamic field of Airborne Wind Energy Systems (AWEs) control using Information-Directed Sampling (IDS). This control strategy adeptly balances the exploration-exploitation trade-off in a partially observable environment, enhancing the efficiency of AWEs in harnessing wind energy at varying altitudes. By implementing an IDS controller and validating it with real-world data, this research offers a new design of AWEs control, contributing to the more effective utilization of airborne wind energy.

Finally, the fourth chapter introduces 'MapeMaker', an open-source software package creating probabilistic scenarios for renewable power production. This tool is capable of generating scenarios that reflect both historical forecast accuracy and potential future advancements in forecasting technologies. By providing a means to create realistic alternative scenarios based on historical data, 'MapeMaker' is a useful tool in renewable energy planning and operations, particularly valuable for simulation-based analysis methods such as stochastic unit commitment and economic dispatch.

In summary, this dissertation presents a comprehensive collection of methodologies and tools that contribute to the optimization and integration of renewable energy resources in modern electricity markets. Through uncertainty-aware control strategies, forecasting techniques, and the development of practical tools, this work addresses the critical challenges of managing uncertainty in renewable energy systems.

Contents

List of Figures

List of Tables

Acknowledgments

I am thankful for the people have met over the last four years as part of my academic journey.

Professor Scott Moura, my advisor, has been a pivotal figure throughout my PhD journey, guiding me through various research projects and the challenges of a pandemic.

Professor David Woodruff, my advisor during two internships, provided me with my initial research experience at UC Davis in 2019.

Professor Alper Atamturk, Professor Maria Laura Delle Monache, and Professor Daniel Arnold, thank you for your support and feedback during my qualifying exam and for your thorough review of this dissertation.

Shout-out to Chitra, Dylan and Big Guy. Our weekly meetings have not only been productive research-wise but also a breeding ground for many startup ideas – perhaps an (ad)venture awaits us in the future.

My friends Preet, Laurel, the Oakland's bro Jos, Quentin, Maxime and Alexia enriched my research and social life.

Rami, your exceptional hands-on engineering work at the ranch, from Berkeley to Colorado, provided me with insights on feedback control.

Presenting my work at the 2022 Conference of the Cedar Scholars Association (CCSA22) was a nerve-wracking experience, but I couldn't have done it without the support of Frank, Adam, and Facu Fabian.

Of course, my deepest gratitude goes to my parents, Brigitte and Etienne, my sisters Clothilde, Aude, and Emeline, and my grandmothers, Madeleine and Eliette. Thank you for everything!

Finally, Adele, you have been an invaluable source of motivation since our paths crossed.

Chapter 1

Introduction

This dissertation explores the integration of renewable energy and battery systems into the power grid, with a specific focus on the uncertainties this integration carries. It unfolds over four research projects, each examining a unique aspect.

In this first chapter, I discuss the main challenges currently facing power systems. The second chapter explores what motivates my research by distinguishing between types of uncertainties. This leads to a review of common control tools used in this field and how this dissertation contributes overall. The layout of this first chapter is as follows:

1. **Power System Overview**: This section offers a definition of power systems, setting the stage for the dissertation.
2. **Uncertainty in Energy Systems**: It explores the various dimensions of uncertainty inherent in modern energy systems, particularly those involving power and battery systems.
3. **Control under Uncertainty**: It discusses the strategies and methodologies for managing and controlling battery systems admit these uncertainties.
4. **Contributions and Dissertation Structure**: This section highlights the unique contributions and outlines the structure and content of the subsequent chapters.

1.1 A changing power grid

An invention of cultural and historical significance

While electrical energy exists naturally e.g., in lightning or static energy, its most widespread manifestation is a distinctive feature of modern and industrial societies. We engage with other forms of energy frequently. A fire, for example, is a combustion reaction that transforms chemical energy into thermal energy. Riding down a hill allows one to experience mechanical energy. Electricity, however, is -hopefully- never experienced directly. Engineers build safety

systems to prevent such occurrences. Rather, in our societies, electricity acts as a transitional form of energy, as an energy carrier. Converted from a primary source, it is efficiently delivered for practical work such as heating, pulling, pushing, lighting, etc. It is a marvel of engineering and has now become so pervasive that its presence is sometimes taken for granted.

Figure 1.1: Electricity hall at the 1893 Columbus Chicago Fair. Both Edinson's General Electric and Westinghouse's companies were waging the 'current war' that would decide on the winning electrical standard: Alternating or Direct current.

The power grid started as a private undertaking in the United States. In 1882, the Pearl Street Station delivered its first kW of electrical power. It was funded by famous investors such as J.P. Morgan and Vanderbilt under the leadership of Thomas Edinson [1]. Electricity has then increasingly become intertwined with the concept of a public good. In the United States, a milestone in the historical development of the power grid was the passing of the Rural Electrification Act by the federal government in 1936. It provided federal loans for the installation of electrical distribution to serve electricity rural areas. This has led to the widespread adoption of the technology, namely through the creation of cooperative power companies. Most of them are still in place today. In the United States, utilities, as regulated entities are required to ensure that everyone in their area receives service, regardless of location or economic status [2]. In many countries, electricity is officially deemed a public good and its delivery is considered fundamental. For instance, India's Electricity Act of 2003 establish a framework to provide access for all [3]. In France, EDF, the primary supplier, operates under a mandate of public service, which includes ensuring access for all consumers,

again irrespective of their location or income [4]. Thus, electricity and its distribution have become elements of our collective unconscious.

In the U.S. alone, access to electricity is closely related to the American way of life. Aside from being a crucial vector of growth and industrialization, it is culturally associated to a range of novelties and amenities of the post-war roaring 60s. Electricity transformed the American households introducing conveniences such as light, refrigeration and television to the masses [5].

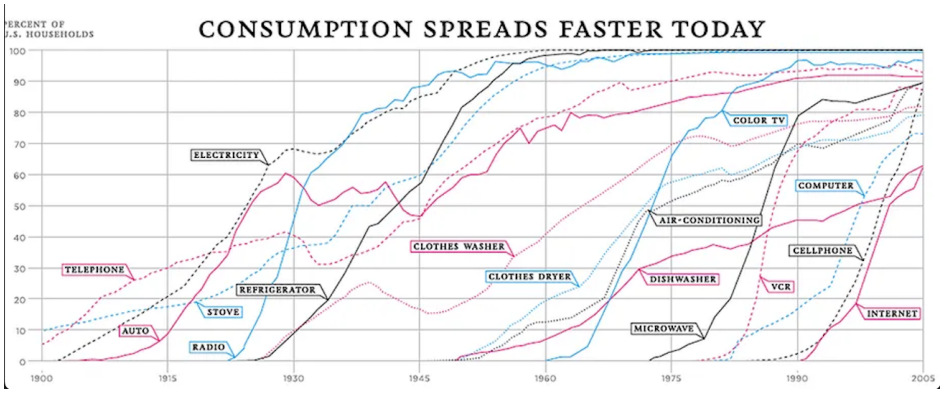


Figure 1.2: Percent of households reached by a variety of services. Note the pace of adoption of electricity post WWI which is higher than most of the products until TV. Electricity is also necessary for all the other products, except Auto. (Source: Ritchie and Roser, 2018)

Like any engineering achievement, pervasive use of electricity is the result of a combination of factors. Economies of scale are the main motivator for a natural monopoly. The power utilities sector makes no exception. Gradual improvement of efficiency on transmission lines, transformers and generators lowered costs. In parallel, an ever growing development of the distribution grid allowed a vast outreach to millions of customers, in turn increasing revenues. The pace of adoption of electricity is remarkable and was unmatched at the time. This is highlighted on figure 1.2. This success can be partly attributed to the dedication of engineers, practitioners, technicians, and academicians [6].

Structure of the power grid

The power grid operates as a vast, intricate machine. Its primary purpose is to efficiently transform, convey, and distribute electrical energy at lowest cost. Its complexity stems from the tight electro-magnetic coupling connecting each rotating machine and locking them in step in the so-called the swing equations [7]. The diversity of elements at different scales makes its analysis and modeling complicated. As detailed in [8], our understanding of these components often employs a top-down perspective. From power plants, power flows through an elaborate network encompassing transmission lines, substations, transformers, and wires.

This extensive system ultimately ensures the delivery of electricity to consumers, a process depicted in figure 1.3.

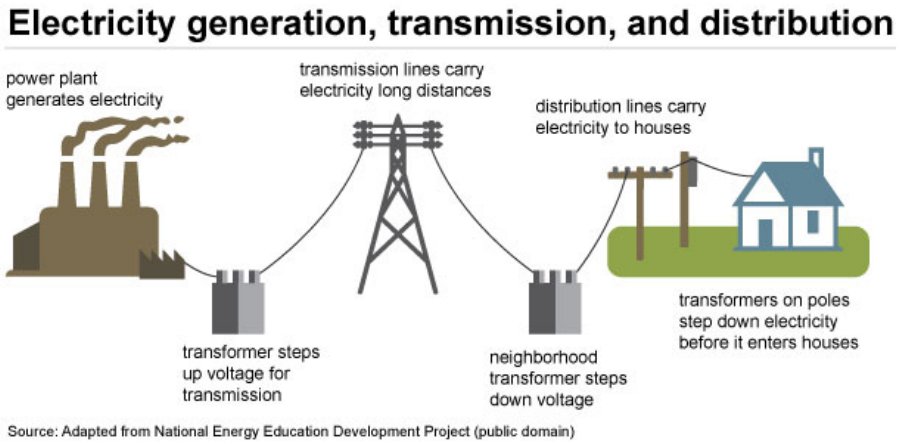


Figure 1.3: The power grid can be decomposed into Generation, Transmission, Distribution and Consumption (Source: EIA)

Interestingly, the majority of the power grid's components were invented and developed over a century ago, as highlighted in table 1.1. These inventions enabled the power to be delivered over long distances at high efficiency. For example, the hydroelectric dams of the U.S. North-west have facilitated the provision of low-cost energy to a wider customer base [9]. Inventions were critical to ensure reliability and safety of power delivery. Today, while these objectives continue to be relevant, efficient technologies have matured, and aggregated power demand has mostly plateaued since 2006 as shown on figure 1.4. In the next section, we'll discuss how this situation is ready for disruption.

Technology	Role	First Use	Average Age	Remarks
Transformers	Trans.	1885	25 for 70%	Vital for long-distance, minimal-loss power transmission.
AC Generator	Gen.	1890	29 years	Efficiently converts mechanical to electrical energy.
Electricity Meter	Cons.	1888	N/A	Facilitates accurate energy usage billing and tracking.
Circuit Breaker		1879	30+ for 60%	Ensures safety by protecting circuits from overload.
Power Lines	Dist.	1884	25+ for 70%	Enables widespread electricity distribution.
HVDC	Trans.	1954	N/A	Suitable for efficient long-distance and underwater transmission.

Table 1.1: Power grid main technologies as reported (ABB. 2014 [10])

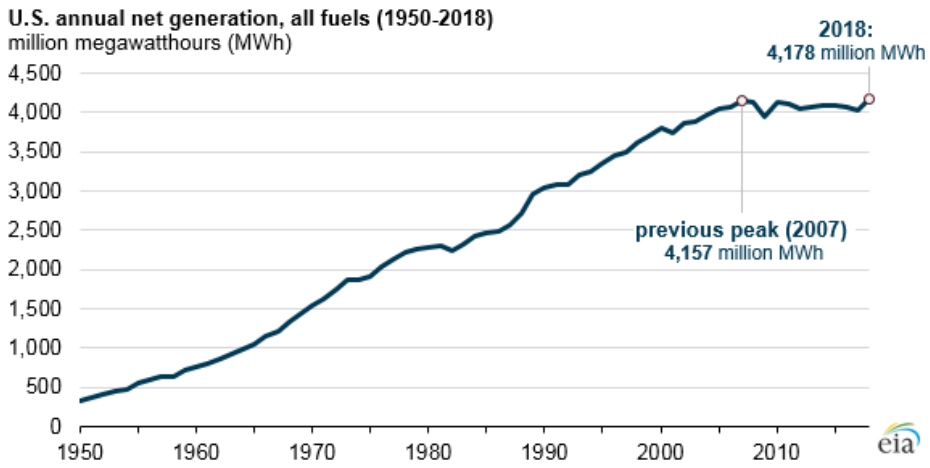


Figure 1.4: Power demand has plateaued since 2006 in the United States (Source: EIA)

Upcoming challenges

The power grid is presently witnessing a renewed surge of interest, bolstered by considerable investments from both public and private entities. In certain projections, investments in transmission-scale projects are expected to double in the coming years [11]. Additionally, an unparalleled expansion in capacity is anticipated, as illustrated in Figure 1.5.

These developments mark both exhilarating and challenging times for engineers tasked with navigating complex problems. In this section, I aim to encapsulate and provide an

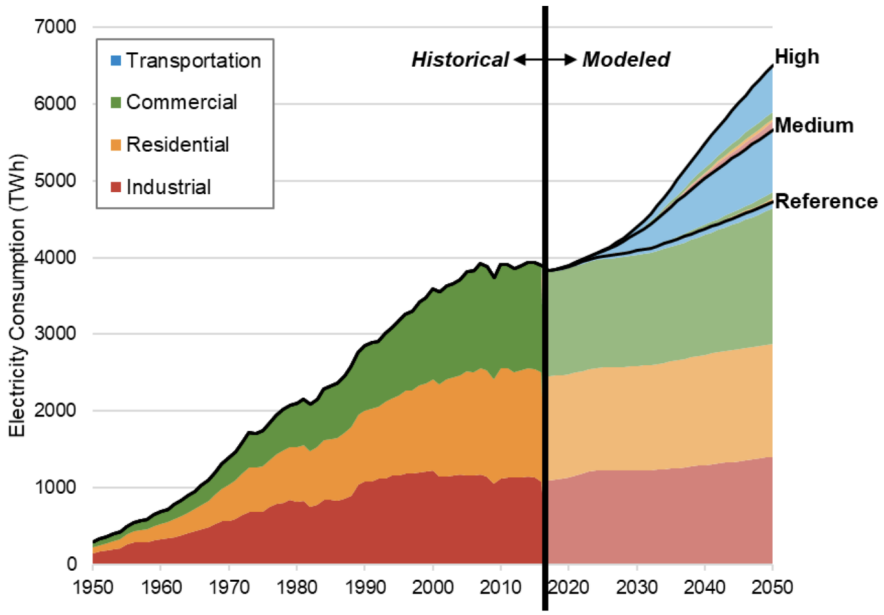


Figure 1.5: Historical and projected annual electricity consumption [12]

overview of the underlying reasons for this increased focus, thereby situating this work within its contemporary context. As of 2023, there are three primary factors driving this attention towards the power system.

1. **Fortifying** the existing grid against aging infrastructure and increasing climate risks
2. **Electrifying** most uses that currently rely on fossil fuels
3. **Cleaning** the grid so that electrification comes with reduced carbon emissions

Fortifying

As indicated in table 2.1, over 70% of the U.S. electricity grid is now more than 25 years old. Replacing every component of the U.S. power system could cost upwards of 5 trillion dollars, with the expenditure distributed across power plants (56%), transmission systems (9%), and distribution systems (35%) [13]. Recognizing this problem, the U.S. Department of Energy has pledged over $20 billion to upgrade the aging components in 2021, through the "Building a Better Grid" initiative [14]. The strength of an interconnected network is often limited to that of its weakest link. As the equipment deteriorates due to wear and tear, leading to an increased failure rate, the overall performance of the system is jeopardized. These failures cause not only service interruptions and costly emergency repairs but also erode confidence in the infrastructure's resilience. Maintaining a reliable and high-quality service

System	Temperature exposure
Generation	Reduced efficiency (internal losses solar, gas), derating (higher temperature of river for nuclear)
Distribution/Transmission	Derating and increased line losses, rapid equipment aging, increased congestion
Consumption	Increased consumption (HVAC)

Table 1.2: Impact of increasing average temperatures on power systems [16]

builds trust which is essential to encourage behaviors including electric vehicles adoption, or flexible consumption [15].

While the power grid components are becoming outdated and fragile, climate change is also altering weather patterns and increasing the frequency of extreme weather events. This escalates the probability of faults and makes black swan events more likely, as climatic variables contribute to a range of failure risks. Figure 1.6 shows a range of causes that triggered blackout events and highlights natural events as the main culprits. Table 1.2 illustrates the compounded risks behind an increase in temperature in power systems following Brockway and Dunn's review [16]. During my PhD, I observed several power challenges, like the Texas Power crisis of 2021 and California's wildfires and heatwaves. These events led to notable power outages and inspire research into strengthening the grid [17][18][19].

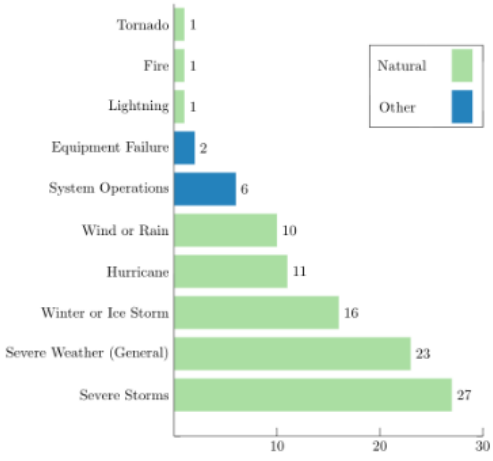


Figure 1.6: Causes of Large U.S. Electric Disturbance Events Affecting at Least 50,000 Customers (% 2000-2016) [20]

Nevertheless, viable solutions are currently investigated to incorporate climatic risks into future grid planning [21] and to manage operational risks in the present [22]. Both approaches necessitate a detailed quantification of uncertainty.

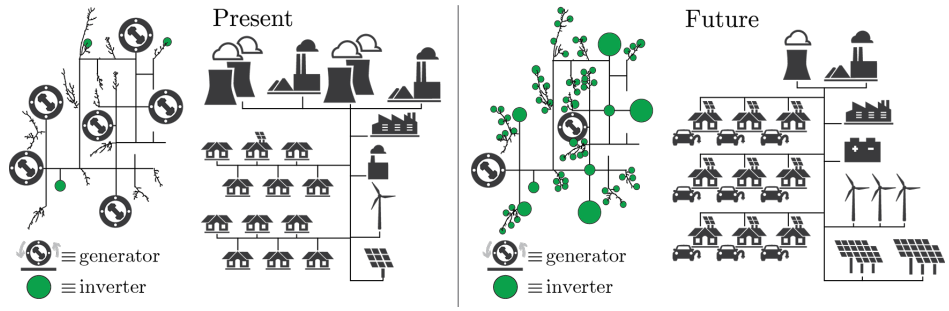


Figure 1.7: A vision of a future power grid: more decentralized supply is providing power on a distribution scale, minimizing the need for costly upgrades. (Source: NREL)

Electrifying

In the near future, we can anticipate significant change in demand patterns, primarily driven by a global trend towards electrification and sustainability. The transportation sector is on the cusp of a major shift with the increasing adoption of electric vehicles (EVs). Current projection of EV adoption accounts for as much as 76% of traveled vehicle miles by 2050 [12]. These vehicles, more efficient than their internal combustion counterparts in terms of energy/miles [23], play a pivotal role in reducing fossil fuel dependence while increasing the demand for electricity. This demand will be met by an expanding network of EV charging infrastructure, leveraging renewable energy sources more effectively. Recent research looks at building incentives around planning EV charging along with user mobility [24][25].

In parallel, the residential and commercial sectors are expected to transition to a more energy-efficient system. The integration of heat pumps for heating and cooling is under way. They are markedly more efficient than conventional gas-based systems. The IEA estimates that globally, heat pumps have the potential to reduce global carbon dioxide emissions by at least 500 million tonnes in 2030 (the annual CO_2 emissions of all cars in Europe today) [26]. This will increase significantly electricity usage as shown by figure 1.9. On the other hand, industrial processes, especially those that require high temperatures, also increasingly rely on electrification or hydrogen. They increasingly use methods such as electric arc furnaces or heat batteries [27][28]. Meeting the emerging needs of the energy transition presents several challenges, with cost being a primary concern. Additionally, constraints in raw materials and manufacturing processes can hinder the timely implementation of renewable energy solutions [29].

Cleaning

Addressing the increasing demand for electricity in the coming years will involve a multi-faceted approach. The expansion and enhancement of renewable energy sources such as solar panels and wind turbines are crucial. Recent trend shows a global decrease in the cost of electricity of those sources as shown on figure 1.8. Still, increasing bottlenecks and

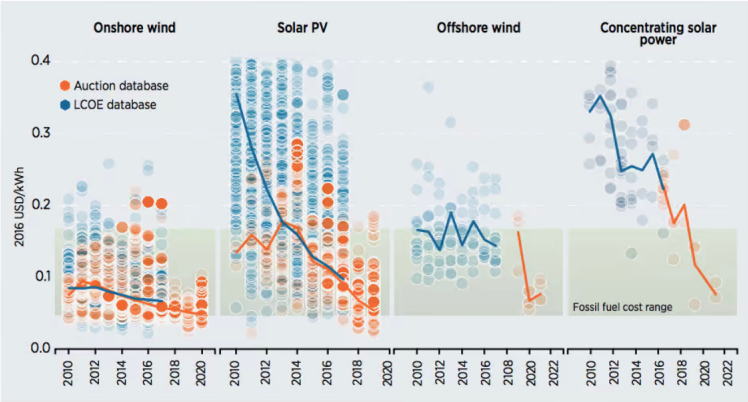


Figure 1.8: Levelized cost of electricity 2010-2022 (Source: IRENA)

competition for raw resources are likely to drive up costs. Time is critical, as the energy transition needs to occur promptly. These factors are driving research and development towards a variety of new technologies and battery chemistry [30]. Unlike the legacy grid, which has a relatively consistent composition of components, the future grid is anticipated to combine multiple systems at multiple scales through an increasingly smarter grid. Figure 1.7 depicts this transition from a top-down to a more decentralized approach. This evolution is expected to not only reduce costs but also enhance reliability [31].

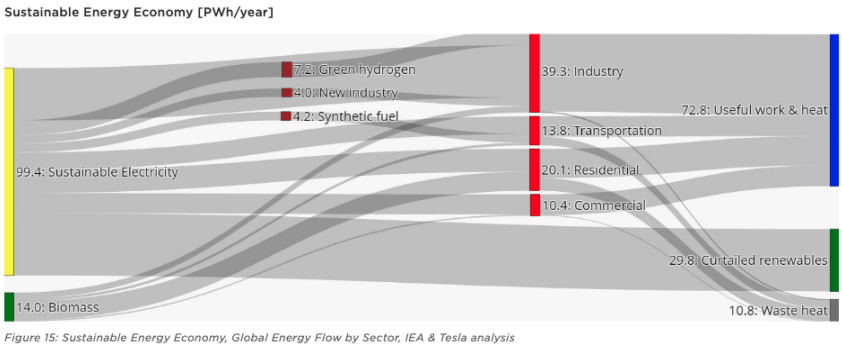


Figure 1.9: Sustainable economy scenario: the world economy demand is fulfilled by clean electricity (Source: Tesla Master Plan 3 with IEA data)

Finally, encouraging flexibility can also help keep costs at bay. This involves implementing new tariffs and billing strategies to modify user load patterns, enabling development of prosumers with behind-the-meter battery and solar [32]. Well managed, this reduces the need for extensive grid upgrades [33][34]. Transitioning to a new grid system will ultimately require thorough analysis and careful planning. Together, these solutions represent a concerted effort

to transition to a more sustainable, reliable, and flexible energy system as shown by figure 1.9.

1.2 Uncertainty, everywhere!

State-space systems: two examples

The power grid, like any machine, encounters different types of disturbances. They fit into three categories. Naturally present in stochastic processes, inherent noise is filtered and disregarded, e.g., voltage measurement noise. Perturbations in the uncontrollable inputs should be accommodated, e.g., an added load (kW). Finally, some disturbances must be rejected and cleared, like a surge of current due to a fault.

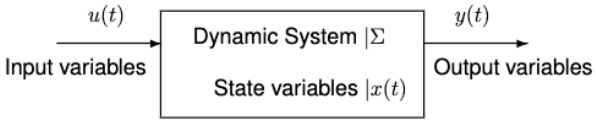


Figure 1.10: General dynamic system model (Source: Pr. Scott Moura's CE295 Lecture Notes)

Figure 1.10 presents a block-diagram schematic of a typical dynamical system model while the equations 1.1 introduce its respective state-space representation. This system, denoted by Σ, is defined by state variables $x(t)$ evolving according to a state transition function f. To put it concretely, a car can be modeled as a point-mass translating on a 1-dimensional road. Its minimal states are its position and velocity. They evolve in response to input variables $u(t)$ and are subject to input and state process noise ω. These inputs are the actions, either controlled or uncontrolled, of the system's environment. From Newton's second law of dynamics, a car acceleration equals to the sum of the applied forces. Let the forces be the motor torque -known up to measurement noise- but also the drag and wheel friction. Those last two items are not directly measured and hence are uncertain quantities. Meanwhile, the output variables $y(t)$ represent the observations of the system's response according to the observable g. Here, error in the measurements and model mismatch are encoded into the random variables ϵ. The location of the car can be tracked by GPS which itself introduces measurement noise and lagged information. The state-space approach enables us to distinguish between uncertainties in the plant model, states, inputs, and outputs. I will now introduce uncertainties along these four elements over the two main systems which are at the heart of my dissertation.

$$\begin{cases} x_{t+1} &= f(x_t, u_t, \omega_t) \\ y_t &= g(x_t, u_t, \epsilon_t) \end{cases} \tag{1.1}$$

	Battery Pack	**Power Grid**
Input	Current in amperes (A)	Real, reactive power injections (MW, MVar).
Output	Voltage (V)	Nodal voltages (kV), angles at measured substation
Model	Equivalent Circuit Model	AC/DC power flow models
States	State of charge percentage (%), capacitor voltages (V)	Bus voltages and phase angles, line flows (MW)

Table 1.3: Definitions of Input, Output, Model, and States for Battery Systems and Power Flow Models

To illustrate this representation, we give the examples of a battery and power systems that will get further developed in the next section on control under uncertainty. Table 1.3 summarizes this development. In the context of a battery system, we control the current (input) flowing into or out of the cells and we measure voltage (V) across the terminals. Voltage is a function of the input current and the internal state. Selecting a suitable model for battery cells depend on trading-off complexity, accuracy and practicality. In any case, a battery model tracks the state of charge (SoC) such as in an equivalent circuit model. In the context of a power system, the injections and voltage from generation sources are usually controllable while nodal demands (active and reactive) are generally taken as uncontrollable input or data [35]. Supervisory control and data acquisition (SCADA) record measurements at the substation level and report nodal voltage and phases to the operators. The model is a set of power flow equations, which could be based on alternating current (AC) or direct current (DC) formulations. These equations describe the physical laws governing the flow of electricity through the network. Finally, the states in the grid are typically the voltages at the buses (nodes), the phase angles and the power flow over the lines. Figure 1.11 presents a helping illustration: each one of the six buses has 4 states (voltage, angle, active and reactive power). These models are clear simplifications but are necessary to introduce relevant challenges around uncertainty.

Uncertainty in those systems and challenges

1. **Uncertainty in the output** y:

   - *battery system*: the output voltage can vary due to temperature changes, aging, or variations in the internal resistance not accounted for in the model. Incorrectly filtering the voltage output can lead to inaccuracies in estimating the state of charge. It can also affect the performance and health diagnostics [36][37]. The level of observability in different chemistry informs practical decisions. For instance, NMC chemistry exhibits higher observability than LFP and may be

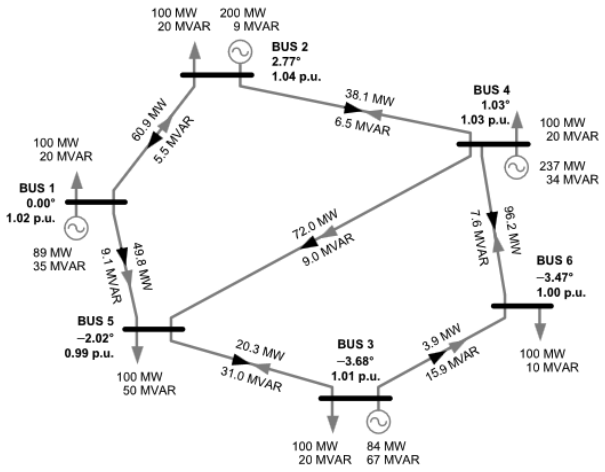


Figure 1.11: Six-bus power flow example from [35]

preferable in safety-critical applications such as electric aircraft [38]. Chapter 3 approaches this issue for a lithium-sulfur battery.

- *power system*: the output uncertainty is typically related to correctly sensing and compressing information relative to magnitude, frequency and angles from measurement units such as Phasor Measurement Units [39][40] or SCADA, see figure 1.12. Historically, the steady-state assumption made around the power flow equations meant that measuring at a few nodes on an hourly basis was enough to get a general sense of the system behavior. In recent years, with the increase in penetration of inverter based resources on the distribution grid, adequately measuring a highly variable voltage signal at a large scale of buses has become more challenging [41]. It is worth noting that the accurate measurement of frequency depends on the fundamental model that is fitted, i.e. usually a pure sinus wave. Hence, some of the most novel challenges around estimating frequency under predominant penetration of inverter-base resource may be more due to *uncertainty in the model* than in the output. During my PhD, I had the chance to collaborate with Professor Sascha Von Meier on identifying sub-synchronous oscillations from PMU data to alert on potential power electronics malfunctioning and resonating.

2. **Uncertainty in the input** u:

- *battery systems*: The input current to a battery pack is uncertain and usually present a systemic bias [42]. If not accounted for, the bias can have a cumulative effect and degrade SOC estimates after a few cycles. As I submitted this dissertation, Chitra Dangwal is studying the robustness of our approach in Chapter 3 to input bias.

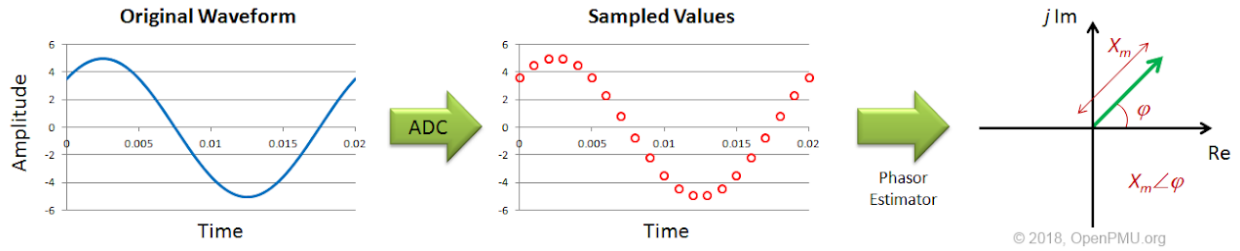


Figure 1.12: Phasors encode the voltage magnitude and angle with respect to a gps reference. It relies on the assumption of steady state that may not hold in practice. It is yet another application of state estimation.

- *power system*: input uncertainty comes predominantly from unpredictable load demand and the variable output from renewable sources [43][44]. The challenge lies in balancing generation with uncertain demand to avoid overloading the grid or wasting energy and this at lowest cost. This topic has seen a lot of interest as adding more renewable energy bring on more cost of uncertainty for the rest of the system [45][46]. This is the focus of chapter 5.

3. **Uncertainty from unknown or approximated Plant Model** Σ:

   - *battery systems*: most control-based plant models are derived from approximations in the electrochemical processes within the cell, which are complex and difficult to model accurately. This leads to challenges in predicting battery life and optimizing charging strategies. Characterization is typically approached through parameter estimation techniques and improving the fidelity of models. Our lab has detailed multiple ways to reduce those higher order models and validate them in the past [37][47].
   - *power system*: the plant model uncertainty arises from simplifications in the representation of the physical grid, such as linearizing the power flow equations [48]. This is problematic for system planning and real-time operations. Characterization involves validation against historical data, sensitivity analysis, and development of more accurate computational models that can capture the grid's nonlinear nature. Recent work has looked at identifying the topology of the distribution grid using advanced smart meters and PMUs [49][50].

4. **Uncertainty in the model states**:

   - *battery systems*: the state of charge or state of health estimation can be highly inaccurate due to the aforementioned output and model uncertainties, as well as from errors in initial state estimation. This affects the reliability of the battery management system [51].

- *power system*: system states such as bus voltages and phase angles, critical for operational decisions can also be inaccurate. Estimation methods often employ redundant measurements to enhance accuracy and guard against erroneous data [52]. These also entail estimating hidden information such as household load [53]. Additionally, uncertainties in electricity markets, characterized by game-theoretic behavior, are notable. Market participants place bids without knowing their competitors' strategy. This aspect of electricity markets, explored in Mathilde Badoual's dissertation [50], was a primary motivation for my PhD and is a central theme in chapter 2.

1.3 Control under uncertainty in power and battery systems

Decision making under uncertainty involves controlling a *modeled* system whose evolution is not entirely determined by its present *estimated* state and future actions. In this context, the framework laid out in the equations 1.1 is typically used by practitioners to study and analyse policy. Following Kumar and Varaya's foundational textbook on control under uncertainty [54], two main classes of problems stem from studying those systems. *State estimation* and *system identification* extract information from the past observations to reduce the uncertainty about the system behavior. On the other hand, *optimal control* deals with selecting the most attractive policy for the problem at hand.

State Estimation and System Identification

State estimation in dynamic systems can be achieved through various methods, each tailored to specific characteristics [55]. Kalman Filters (KF) offer optimal solutions for linear systems with Gaussian noise [56]. For nonlinear models, Extended Kalman Filters (EKF) and Unscented Kalman Filters (UKF) provide enhanced capabilities. On the one hand, EKF linearizes around the current state [57]. On the other, UKF employs a deterministic sampling technique for better accuracy in highly nonlinear environments [58]. Ensemble Kalman Filters (EnKF) are used in high-dimensional settings like weather forecasting, utilizing a Monte Carlo approach [59]. Particle Filters (PF), another nonparametric method, are suitable for nonlinear, non-Gaussian systems, capable of handling a diverse range of distributions [60]. Lastly, Moving Horizon Estimation is effective in constrained environments, relying on an optimization approach over a moving window of data [61]. Each method has its strengths, chosen based on dynamics, noise properties, and computational considerations. Chapter 3 identifies a piecewise affine system that governs the behavior of lithium-sulfur batteries. The chapter funnels this system within a MHE framework. Chapter 4 reduces a wind speed model and derives the optimal distribution of state estimates with a Kalman Filter for immediate control application. This chapter exemplifies state estimation, where the primary goal is to obtain accurate, real-time insights for control purposes.

Optimal control

Selecting a control policy will differ depending on assumptions on the system's nature, represented by f, g, on the uncertainties ϵ, ω and constraints on both the states and inputs x, u. Hardware limitations, such as computational speed, typically influence implementation choices. The control objective may take various forms. We characterize them following figure 1.13. Transient performance is commonly measured by speed (rise time, settling time), accuracy (overshoot, settling error) and implies that the system is stable [62]. It answers important questions. Does the system settle into a steady state? For a specific perturbation, how long does it take to reach stability? When adjusting to inputs or countering disturbances, by how much does it exceed the target? These are the objectives of low-levels regulators. These aspects are typically managed using feedback control loops like proportional-integrative-differential (PID) control and are approached by numerous textbooks on control theory [63][64]. In linear systems, controls can be provably fine-tuned for desired performance using techniques like pole placement [65].

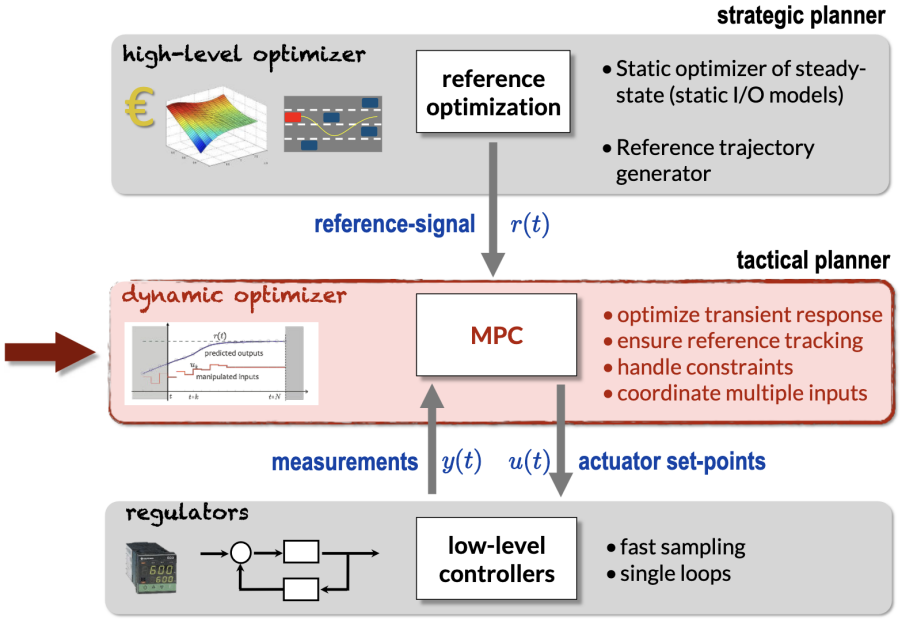


Figure 1.13: Ranking controllers from high to low level (Source: "Model Predictive Control", A. Bemporad, 2023)

Now that the system can safely stabilize and reach a setpoint, other classes of controllers, e.g., Linear Quadratic Regulators, are often employed to track updated references. They usually minimize a running objective that involves finding a tradeoff between reaching the setpoint rapidly, with large inputs or later, with smaller inputs[66]. More generally, optimization-based controllers design a sequence of setpoints that achieve lowest cost, which be potentially arbitrary here (e.g., revenues, energy, etc.). These problems can be tackled

using optimization methods and model predictive control (MPC) [67][68]. MPC involves predicting future system states and making decisions that minimize cost over a specified time horizon. Another approach is Dynamic Programming (DP) [69], which is particularly relevant when both the actions and state variables are discrete. DP breaks down the main program into smaller sub-problems and solves them recursively, making it efficient for certain classes of formulations. Reinforcement Learning (RL) [70][71], on the other hand, requires substantial amounts of data and learns optimal policies through a process of trial and error. RL is particularly useful in environments where the model is unknown or too complex, as it learns from interaction with the environment. These methods all aim to minimize a cost function in distinct ways. However, the specific approach, data requirements, and applicability can vary significantly. The exploration of these methods and their application to control challenges forms a core part of this dissertation and of collaborative project I had during my years of PhD.

As we have seen, different time scales, distinct objectives, modeling assumptions and hardware specifications will lead to a range of strategies. Often, a hierarchical control approach is used to manage these complexities in large systems, as shown in Figure 1.13. Each level deals with uncertainties at different scales, allowing for more effective and organized control across the entire system.

Chapter 1 of this dissertation focuses on battery sizing and siting, aligning with the role of a "strategic planner". Chapter 2, which develops a state observer for lithium-sulfur batteries, is categorized as "tactical planner". Chapter 3, dedicated to designing an altitude controller for an airborne wind energy system, also falls into this category. Lastly, Chapter 4, which involves designing scenarios for stochastic model predictive control at a system-wide level, is again associated with the "strategic planner" tier.

Control of battery systems over a power system

As a PhD student, I had the chance to collaborate with two research groups. The first group, eCAL, specializes in battery modeling and control. The second, the IEEE Power & Energy chapter, focuses on integrating inverter-based resources into the grid. In this section, I aim to guide the reader through the various control systems, essential for successfully integrating battery systems into power networks, particularly at the transmission level. The discussion will be structured based on a traditional time scale analysis, as depicted in Figure 1.14.

Microsecond Scale: Current Injection - Power Electronics Controls - At this scale, battery systems (DC side) interface with the grid (AC side), either by injecting active and reactive power (when controlled as a current source) or by setting active power and voltage magnitude (when controlled as a voltage source). A review from NREL explains the fundamental differences for those two modes [72] and figure 1.16 illustrates the grid-inverter-battery coupling. Today most of the inverters are grid following, while developing stable grid-forming inverter control is a hot research area. Since the battery discharges energy as direct current, it needs to be converted to alternate at the same frequency as the grid. This is typically achieved with a half-bridge converters and active filters. Pulse Width Modulation

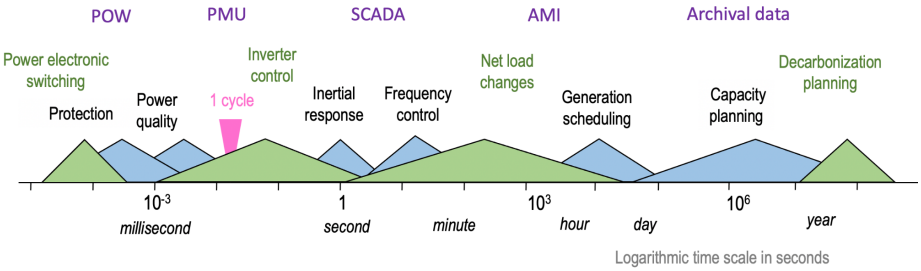


Figure 1.14: Time scales

(PWM) is a common technique that involves turning switches on or off in response to a regulating signal. This generates a stepped current input, which is then smoothed of its high-frequency components through an active filter.

Millisecond Scale: Current, Voltage Correction - Power Electronics Controls - At this scale, the battery system aims to follow a reference, usually an active/reactive power setpoint. For effective control, the battery system needs to inject a waveform that is correctly synchronized with the grid's voltage. This synchronization is achieved with a phase-locked loop (PLL), which locks onto the voltage phase and frequency, ensuring that the injection is accurately timed and phased.

- Current and Voltage Control Loops: they are often implemented using integrated PID controllers, providing the necessary tracking guarantee [73]. Some aspects of droop-control or power-sharing can also be part of the framework at this level to help maintain power balance and grid frequency. In power systems, this layer is called primary control [74][75].

- State Estimation: In a grid-following mode, the inverter must synchronize the injected current waveform with the voltage as detected at its terminals. A phase-locked loop is utilized to estimate these parameters despite uncertainties in measurement, ensuring the correct alignment of the signal [76].

Second Scale: Real-Time Control - At this time scale, a layer of secondary control can be deployed to restore the frequency and voltage to their respective nominal value. In the case of a microgrid, this can be achieved through various methods. Balance can be reached by using PID [77], LQR [78], MPC [79], etc. During my PhD I interned at Heila Technology, where I contributed to the development of decentralized control systems operating at this frequency. Alternatively, if the battery is located over a transmission grid, it could involve tracking a regulation signal from an Independent System Operator (ISO). For example, ERCOT (Electric Reliability Council of Texas) sends a signal to follow every second [80]. A controller is employed to maximize market revenue while managing the financial trade-off of not providing this service for energy/SOC managemenet[81][82].

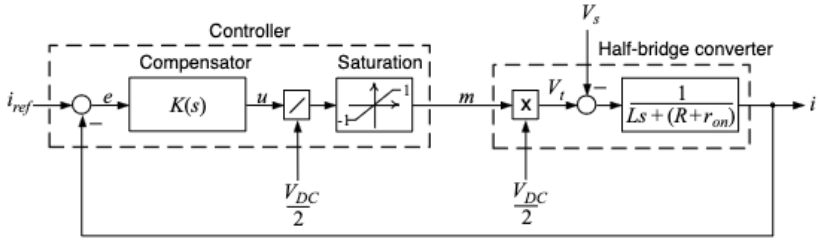


Figure 1.15: Control block diagram of a closed-loop half-bridge converter system from [73]. The compensator is proportional-integral (PI).

Second to Minute Scale: Operational Control - the Battery Management System (BMS) ensures that the battery remains within safe operational parameters by monitoring cell states. The BMS estimates states such as State of Charge (SOC) [83], State of Power (SOP) [84]. It regulates the pack temperature to avoid surpassing critical thresholds (e.g., 90°C), thus preventing decomposition in elements like the electrolyte and cathode [85][86]. It detects faults, clears or disengages the pack [87]. Finally, the BMS also balances voltage, charge, and capacity across cells, maintaining overall system health and efficiency. Those numerous functions are illustrated on figure 1.16.

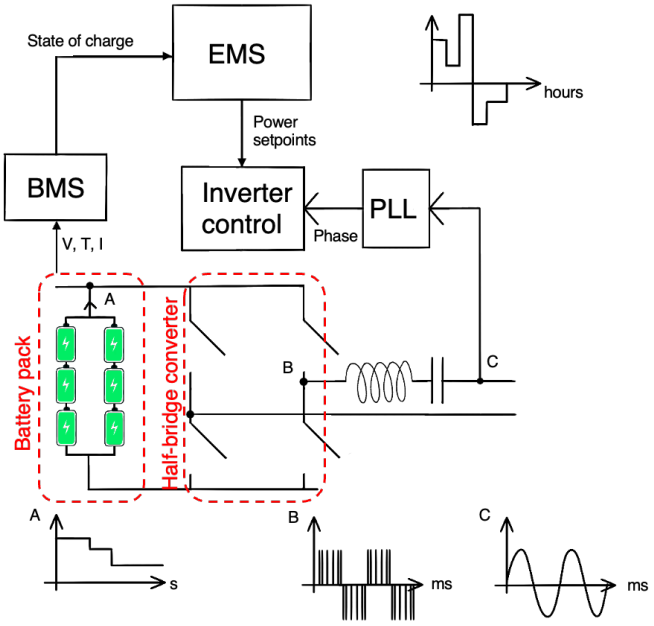


Figure 1.16: BMS-Converter control integration. Figure by Guillaume Goujard.

Hourly Scale: Optimization and Planning - On this scale, battery operations are steered by the Energy Management System (EMS). It employs higher-level control methods. For example, an MPC can minimize operational costs through effective prediction and management of demand and supply. Examples of specific applications range from market participation [88] with various optimization techniques [89], residential tariff optimization [90] to microgrid operations [91]. At this scale, the battery management system estimates state of health [92], and remaining useful life [93].

These scales and control loops have been represented on figure 1.16: the EMS steers the inverter control by sending power references that minimize a cost objective over the long run. It is informed on the battery states by the BMS. The inverter control tracks the setpoints by locking onto the phase angle and frequency of the voltage at its terminal. It then modulates its switching frequencies so that the battery DC current (A) can be converted into an AC signal (B). Finally, the high frequency components are filtered so that the current injection is well behaved with a phase angle that -potentially- maximize its active power output (C).

1.4 Research Overview, Contributions and Outlines

The two main research directions explored in this dissertation are renewable energy systems and uncertainty. Each project, directly or indirectly, delves into aspects of control under uncertainty for energy systems. The rest of the thesis is organized as follows:

Chapter 2 addresses the emerging market for transmission-scale storage, focusing on how such a system can be profitable in a nodal wholesale market by optimally selecting its location and size. It introduces a mixed-integer optimization framework tailored for a price-maker storage system, which maximizes its revenue by considering its influence on market prices and congestion. A distinctive feature of this approach is that it focuses on optimizing both the siting of the battery and its bidding schedule on a grid with transmission constraints. The chapter utilizes simulations based on data from New Zealand to validate the need for such a comprehensive approach. Contributions include:

- Development of a mixed-integer linear programming framework that captures the price-maker behavior of a battery operator in a nodal electricity market, factoring in equilibrium constraints.
- Comparative analysis of price-maker and price-taker approaches, emphasizing the significance of the price-maker assumption in maximizing profitability, especially in congested areas.
- Application of the developed framework to real-world data, offering insights into optimal strategies in terms of siting, sizing, and bid scheduling.

- A novel perspective on battery as a strategic asset in nodal markets, highlighting its dual role in alleviating congestion and enhancing earnings by leveraging locational marginal pricing differences.

The chapter contributes to the understanding of battery storage systems in wholesale electricity markets, offering practical strategies for operators to optimize their investments in the face of evolving dynamics.

Chapter 3 delves into the identification and state estimation challenges of Lithium-Sulfur (Li-S) batteries, particularly important for high energy density applications. It introduces a novel approach to constrained estimation for Li-S batteries by integrating a piecewise affine (PWA) system within a moving horizon estimation (MHE) framework. This approach overcomes the limitations of differential algebraic (DAE) equations used in standard Li-S battery litterature, which are difficult to apply for real-time state estimation due to their constrained and stiff nature. The chapter outlines the development of a PWA system using a linear tree algorithm, formulated based on data from a calibrated DAE model. The PWA is then utilized to create a mixed-integer quadratic program for efficient online estimation. The method's accuracy and computational efficiency are validated through simulations, demonstrating effectiveness even under high observational noise. The primary contributions of this chapter include:

- Development of a learning algorithm for a piece-wise affine system that accurately captures the dynamics of a 0D electrochemical DAE model for Li-S batteries.
- Integration of the PWA into a MHE framework, facilitating constrained state estimation of Li-S batteries and overcoming the limitations of traditional DAE-based models.
- Demonstration of the method's accuracy and computational efficiency, particularly in simulating and conducting state estimation under real-world conditions like high observational noise.

This chapter contributes to the field of Li-S battery modeling and state estimation, offering an innovative approach that balances the need for accuracy and computational efficiency. It provides valuable insights for further research and practical applications in the area of battery technology and management.

Chapter 4 focuses on the control of Airborne Wind Energy Systems (AWEs), an emerging technology in renewable generation. The informed altitude regulation maximizes power output in an environment with time-varying and partially observable wind speeds. The chapter leverages Information-Directed Sampling (IDS) with a Kalman Filter to optimize the balance between exploration (gaining information about wind conditions at different altitudes) and exploitation (maximizing immediate power output). The wind profile distribution is estimated recursively with an online procedure to online-estimate the noise covariance. This enables the computation of the expected wind-output distribution and expected

entropy reduction. The final controller minimizes regret per bit of information gained about the optimum. The chapter validates this approach using real-world data and compares the performance of the IDS with baseline controllers. The primary contributions of this chapter include:

- The development of a persistent forecasting model for AWEs that minimizes assumptions on wind profile evolution, improving upon previous forecasting models.
- The integration of a Gaussian process structure into the wind profile evolution model, coupled with an online learning algorithm for process covariance function estimation.
- Application of Kalman Filter recursive equations to balance the trade-off between maximizing immediate wind power output and gaining information for future operations.
- Introduction of an IDS strategy that efficiently manages the exploration-exploitation trade-off in AWE altitude control.

This chapter advances the control methodology for AWEs by providing a comprehensive approach that combines statistical inference, model order reduction, and innovative control strategies. It contributes significantly to the field of renewable energy in optimizing the performance of AWE systems in a complex and variable environment.

Chapter 5 presents algorithms for creating probabilistic scenarios for renewable power production, focusing on tailoring forecast uncertainty. Scenarios are constructed to either reflect historical forecast accuracy or project potential improvements in forecasting technology. This is particularly relevant for future studies where forecasting methodologies are expected to evolve. The chapter's methodology also enables the generation of alternative realizations of renewable energy production consistent with historical accuracy, proving critical in simulation-based analysis methods. These scenarios are illustrated using real data for day-ahead wind forecasts. The chapter's contributions include:

- Development of a method to create probabilistic scenarios for renewable power production, which can adapt to varying levels of forecast accuracy. This is significant for modeling future improvements in forecasting technologies.
- Introduction of a novel approach to generate synthetic realizations that align with historical forecast accuracy, providing a means to create realistic alternative scenarios.
- Implementation of these methodologies in an open-source software package, 'mape_maker', enhancing accessibility and applicability in the renewable energy sector.
- Illustration and validation of the proposed methods using actual day-ahead wind forecast data, demonstrating the practical utility of the approach.

Overall, this chapter contributes significantly to the field of renewable energy forecasting, offering tools and methodologies that help model and anticipate the variability and uncertainty inherent in its generation. The methods and software introduced here are used to solve stochastic unit commitment at a large scale [94].

Chapter 2

Optimal Siting, Sizing and Bid Scheduling of a Price-Maker Battery on a Nodal Wholesale Market

The market for battery storage is set to boom in the coming years. This trend may be explained by a combination of factors ranging from a falling cost of technology to a growing need to address uncertain and unflexible renewable energy generation. It is well known that capital cost remains too high for a large-scale lithium-ion battery storage to be profitable solely by arbitrating a wholesale market. However, in this report, we show that by carefully selecting its location and its size with respect to its influence on the prices and on the congestions, a battery storage can still be profitable on a nodal wholesale market. To that end, we develop a price-maker mixed-integer optimization framework that maximizes a depreciated battery storage revenue to yield the optimal location and size of a battery storage. Furthermore, it can be used to optimize the bidding schedule of a battery storage in a nodal transmission-constrained wholesale market. We conducted multiple simulations to illustrate and confirm the need for such an approach. Namely, we compared the price-maker results with price-taker results on actual data from the New Zealand nodal wholesale market.

2.1 Introduction

Background & Motivation

Deployment of Grid-Scale batteries is surging in energy markets. In the United States alone, prospective studies suggest that installed capacity will increase by 3.8 GW by 2023 [95] [96]. Different factors explain this trend: a falling cost of this technology, more adapted market rules, new regulations requesting storage, and an increase in penetration of renewable energies in the electricity mix. Batteries are especially well tailored to improve the grid reliability by participating in ancillary and regulation markets [97] [98].

There are three different ways to site a battery on the grid: on the distribution network near the load center, co-located with various renewable energy generators (solar, wind) or on the transmission network [99]. We can broadly classify the literature pertaining to the optimization of operation and sizing of a battery in three groups. The first group takes the Independent System Operator's (ISO) perspective having the duty of optimizing system reliability and reducing cost of transmission [100]. The second one focuses on finding the best joint operation of a battery with other resources such as solar [101] [102], wind [103] or both [104], and increasingly with electric vehicles. Finally, the last group deals with the optimization of a standalone storage system from a merchant operator's perspective and at a transmission level. Our contribution adds to the last group.

While many of the existing batteries earn the majority of their revenues from ancillary and regulation markets [105][106], the competition is fierce, and the market limited. As an example, the Hornsdale Power Reserve (100 MW/129 MWh) has taken up to 55% in shares of the South Australian Frequency Control Ancillary Service (FCAS) and lowered price by 90% since its entry[107]. Frequency Regulation market signals are also power-intensive and therefore require the battery to operate short-time cycles. These cycles rapidly lead to battery degradation. In fact, a recent finding [108] showed that a battery with an expected lifetime of 90 months could only last for 56 months by cycling aggressively. While it is well known that a battery operating solely on a wholesale market is hardly competitive, a company seeking to install a battery on a nodal network should nonetheless seek to amortize its investment on the wholesale market rather than solely on the ancillary and regulation markets. In fact, despite getting most of its revenue from FCAS, the Hornsdale Power Reserve still offers 119 MWh of its capacity directly to the wholesale market for energy arbitrage [109] [110]. In light of new regulations, utilities may have no choice but to plan to deploy large-scale battery units, putting the depreciation of the investment second to the operational aspects. For instance, the California Public Utilities Commission adopted a 1.325 GW procurement mandate for electricity storage for 2020. These entities need to plan and participate in the existing wholesale market. To that end, the battery operator's challenge will not only consist in finding the best capacity and location, but also in following an optimal bidding schedule.

A widespread approach to generate an optimal bidding schedule for battery storage is to assume that the storage will be a *price-taker*, and devise an optimization program with exogenous market prices. Some articles have looked into finding an optimal sizing and operation but under exogenous real-time prices [111]. Some have focused on finding the optimal dispatch over multi-markets (day-ahead market, reserve and regulation market)[112]. Additionally, the impact of degradation on the optimal bidding schedule was investigated in [113]. On nodal markets, feasibility studies concerning implementation of battery storage have been conducted but always under the *price-taker* assumption [114] [115]. However, Mohsenian-Rad laid a framework for a price-maker large-scale battery storage program over a nodal-wholesale market [116]. His article only considers large-scale battery storage and does not provide a comparative price-taker approach to illustrate the importance of endogenous prices. It also does not consider the siting and sizing planning problem.

Price-taker vs Price-maker approach

When bidding on the wholesale market, there are two different approaches: considering the new battery as a *price-taker* or as a *price-maker*. where the prices are pay-as-cleared

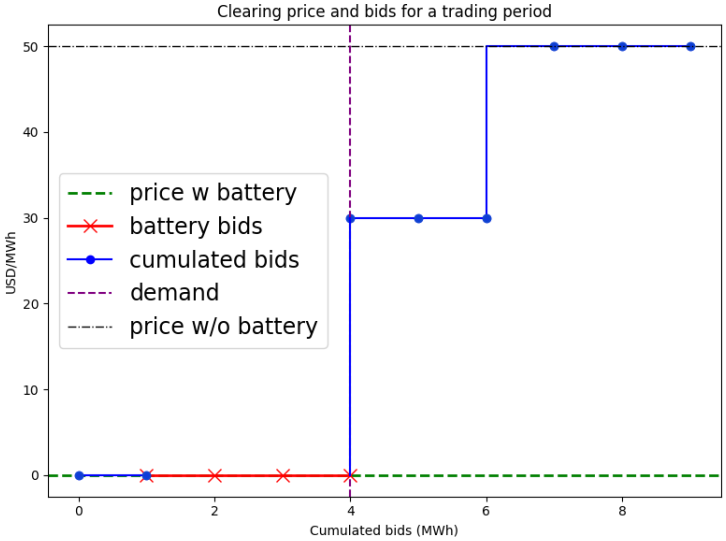


Figure 2.1: *Price-Taker* strategy : self-scheduling

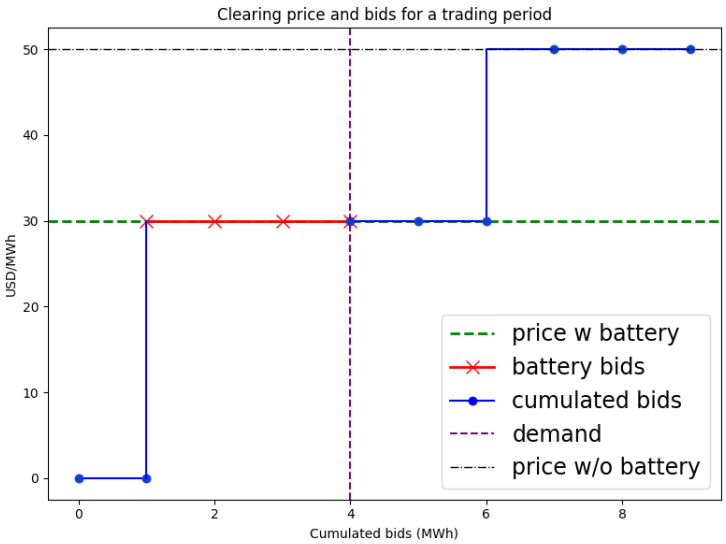


Figure 2.2: *Price-Maker* strategy

Over a wholesale-market, an agent is said to be *price-taker* if its bidding behavior does not influence the market's clearing price. This assumption is usually taken when the agent bids small volumes with respect to the demand of the market. When an agent bids large volumes, its bid volume and price can both impact market prices to its advantage, it has market-power. When congestion occurs over a node, the nodal demand is disconnected from cheap generation. As a result, the nodal price increases. The battery bids in this case, even if small, are not negligible to the single-node-market demand. As a result, the battery is a *price-maker* on the congested area, i.e, its bids will move the nodal cleared price.

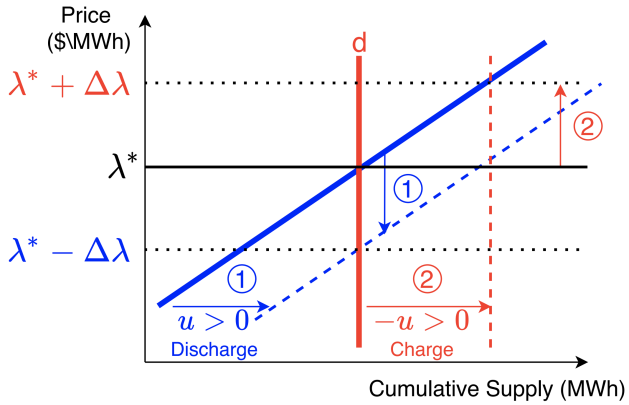


Figure 2.3: Discharge (1) and charge (2) effects on prices on single node market

Assuming the battery node is congested, we represented on Fig 2.3 two main cases of how the battery injection u can impact the prices in both directions. The blue thick supply curve idealizes the accumulated accessible bids and the red thick vertical curve is the demand d. If the battery injects energy into the grid (1), the supply curve shifts the cleared price λ^* to the right, defined as the intersection of supply and demand. Then the cleared price falls by $\Delta\lambda$. In an extreme case, injecting too much energy can remove congestion and the price would fall even further to meet the market marginal price. If it withdraws energy (2), demand d increases, the cleared price will increase. In the following, a *price-taker* (*price-maker* resp.) program is defined as a battery profit maximization optimization program with the local marginal prices (LMPs) taken as exogenous (resp. endogenous).

Focus of this article and contribution

In this article, we examine the situation where a battery located at the transmission level and bidding over a wholesale nodal market is a *price-maker* when congestion occurs. Over a congested area, a well-operated battery can both alleviate congestion and maximize its earning by selling energy at a higher price than the system marginal price. To the best of our knowledge, our article is the first to study the optimal sizing and siting for a battery on a nodal wholesale market under endogenous market prices. This allows for taking into

account the depth of congestion patterns over the grid. We further illustrated this approach by comparing it to a simpler price-taker strategy on the New Zealand grid. To achieve these results, we develop a mixed-integer linear program to yield the optimal sizing under market equilibrium constraints. We finally exhaustively run our program on the major nodes of the network using the bidding data of the New Zealand wholesale market. We compare our results with a simple *price-taker* program to illustrate the importance of the *price-maker* assumption when congestion occurs for an operator wishing to invest into energy storage.

2.2 Problem notations and formulation

Battery storage problem

We consider a standalone battery located at the transmission level and bidding solely on a nodal wholesale market. The grid's buses are described by a set of nodes $\mathcal{N}$ numerated from 1 to n. The battery is located at node $i_b \in \mathcal{N}$, and has capacity z^{cap}. At each trading period t, the operator submits a bid price c_t and volume q_t on the market and is charged with the cleared price. The objective is to set up an optimization program with three main purposes:

1. Find the optimal siting of the battery on the network,
2. Find the optimal capacity z^{cap},
3. Find the best bidding schedule (c, q),

We will devise an optimization program with variables $z^{cap}, (c, q)$ over a time horizon of T. The optimal siting is a parameter that will be found by iterating the optimization program over the nodes of the network. In the following sections, we describe each part of the optimization problem.

State of Charge (SOC) Constraints

Denote $u \in \mathbb{R}^T$, the storage power injection vector. $u_t > 0$ (resp. $u_t < 0$) if the battery is injecting (resp. withdrawing) energy to the grid at time period t. Denote $z \in \mathbb{R}^T$, the SOC profile of the battery. The SOC profile is constrained by the capacity and some minimum level set to zero in this article (2.1). The power injection vector is bounded by an upward and downward charging rate constraint to limit short and deep cycles (2.2). An ideal linear model for the SOC profile is used with an assumed efficiency of 1 (2.3). Finally, to prevent a bias in the revenues of the battery, the first SOC is imposed to be equal to the last (2.4). Note that more realistic values or linear constraints could be added to the framework without loss of generality.

$$\begin{aligned}
\forall t, \quad 0 \leq z \leq z^{cap} & \qquad \text{SOC constraint} & (2.1)\\
u^{min} \leq u \leq u^{max} & \qquad \text{Charging rate constraint} & (2.2)\\
z_{t+1} = z_t - u_t & \qquad \text{SOC update equation} & (2.3)\\
z_{T+1} = z_1 & \qquad \text{SOC boundary conditions} & (2.4)
\end{aligned}$$

Denoting 1_n as the unit vector of size n, z can be expressed as a function of u and the initial state: $z = z_1 \cdot 1_T + Su$. Where S is a sub-triangular matrix filled with -1. The 4 constraints can be rewritten under compact form and as a function of u:

$$\tilde{S}u \leq \bar{u} + z^{cap}e \qquad \text{Battery Power Injection Feasibility} \qquad (2.5)$$

Where we have :

$$\begin{aligned}
\tilde{S} &= \begin{bmatrix} S & -S & I & -I \end{bmatrix}^\top \\
\bar{u} &= \begin{bmatrix} -z_1 \cdot 1_{T-1} & 0 & -z_1 \cdot 1_{T-1} & 0 & u^{max} \cdot 1_T & u^{min} \cdot 1_T \end{bmatrix}^\top \\
e &= \begin{bmatrix} 1_{T-1} & 0 & 0_T & 0_T & 0_T \end{bmatrix}^\top
\end{aligned}$$

The demand is assumed to be inelastic. To enforce that the battery cannot sell and buy at the same time, we introduce a binary variable $r^c \in \{0,1\}^T$ and the three following constraints:

$$\begin{aligned}
c_t &\geq 0 \\
(r^c - 1)L &\leq q^u \\
c_t &\leq r^c L
\end{aligned}$$

With this formulation if $q^u < 0$ then $c_t = 0$ which removes the battery contribution to the economic dispatch bids. If $q^u \geq 0$, c_t is positively unconstrained.

Optimization problem

Let $\lambda_t \in \mathbb{R}^n$ be the local marginal prices (LMP) for time period t. The LMP are nodal clearing prices. At each period, we earn ($u_t > 0$) or pay ($u_t < 0$) the following periodic profit: $\lambda_t^{i_b} \cdot u_t$, where $\lambda_t^{i_b}$ is the LMP of the battery node. Since the battery charging cycles imply dynamic arbitrage, our program will optimize the bids from $t = 1$ to $t = T$.

In this paper, operating cost is restricted to capital cost. Charging cost is directly accounted for in the revenues of the battery. Given a time horizon T, the battery's profits must exceed the adjusted battery cost to ensure a positive net present value over T. The cost of battery capacity is denoted as b \$/MWh and the battery lifetime as y years. If we note $b_T = \frac{b}{365 \cdot 24 \cdot y}$ as the one-period adjusted unit capacity cost, then the adjusted battery

cost over T period is simply : $b_T \cdot T \cdot z_{cap}$. Further we note $B = b_T \cdot T$. This quantity is going to appear as a linear penalty on z^{cap} in the battery program.

Two programs are now developed: one under *price-taker* assumption and the other under *price-maker* assumption. Both frameworks will be compared in an offline setup and with perfect foresight in the results section (4.7). The blue color is used hereinafter to highlight optimization variables.

Price-taker framework

Under a price-taker framework, the LMP are assumed not to depend on the battery bids. They are exogenous. Our bidding strategy will be a succession of self-schedule bid volumes sent to the market operator.

$$\begin{aligned} \max_{z_{cap},u} \quad & \Big(\sum_{t=1}^{T} \lambda_t^{i_b} u_t\Big) - b_T \cdot z_{cap} \\ \text{subject to} \quad & \tilde{L}u \leq \bar{u} + z^{cap} e \end{aligned} \tag{2.6}$$

Price-maker framework

Under the price-maker framework, the LMP are no longer exogenous. Both the LMP and the storage power injection are the results of the market clearing operation denoted as E.D and introduced in (2.8). Hence $(\lambda_t, u_t) = \text{E.D}_t(c_t^u, q_t^u)$, and the bidding strategy (c^u, q^u) is solution of:

$$\begin{aligned} \max_{z_{cap},c^u,q^u,u,\lambda} \quad & \Big(\sum_{t=1}^{T} \lambda_t^{i_b} \cdot u_t\Big) - B \cdot z_{cap} \\ \text{subject to} \quad & \tilde{L}u \leq \bar{u} + z^{cap} e \\ & (\lambda_t, u_t) = \text{E.D}_t(c_t^u, q_t^u), \forall t \leq T \end{aligned} \tag{2.7}$$

Nodal Wholesale Market

Topology of the network and DC Power flow equations

Denote $p_t \in \mathbb{R}^n$ as the nodal power injection vector. The l transmission lines have capacities $h_i, i \in l$. To obtain conditions on the feasibility of p_t, we use a linearization of the AC power flow equations known as DC Power Flow equations for computational simplicity. On the one hand, power is balanced over a lossless grid: $1^T p_t = 0$. On the other, flows over lines are purely real and linear in the voltage angles. Denoting $H \in \mathbb{R}^{g,2l}$ the shift-factor matrix, line flows can be linked to the nodal power injections p_t to finally obtain the constraint $Hp_t \leq h$.[117]

Time Frame of the market

Bids and constraints of market participants are considered in a general multi-periodic setting : $\mathcal{T} = \{1, ..., T\}$. Multiple-time blocks of energy are not considered.

Demand

At each time step, $d_t \in \mathbb{R}^n_+$ is assumed given and inelastic.

Producer

Let $g_t \in \mathbb{R}^g_+$ be the generation vector dispatched by the market. More precisely, $g_{t,i}$ is the generation dispatched by the market operator to the unit i and at time step t. A simple Wholesale market with one bidding band per generator and per time-step is considered to simplify notations. At each time period t, each producer sends three information to the operator :

- $P^{min}_{t,i}$, its must-run capacity which is self-scheduled,
- $P^{max}_{t,i}$, the maximum capacity it is willing to offer in the market,
- $a_{t,i}$, the bidding price associated to the volume $P^{max}_{t,i} - P^{min}_{t,i}$

It follows that the cost for the dispatch of g_t is $a_t^T(g_t - P^{min}_t)$. With g_t feasible if and only if : $I^g_t g_t \leq q^g_t$.

$$I_g = \begin{bmatrix} I \\ -I \end{bmatrix}; \qquad q^g_t = \begin{bmatrix} P^{max}_t \\ P^{min}_t \end{bmatrix}$$

Since there may be multiple generators at the same node, we denote $M_g \in \mathbb{R}^{n,g}$, the producer-node adjacency matrix. $M_{g_{i,j}} = 1$ if and only if producer j belongs to node i, with $i \in \{1, ..., g\}$ and $j \in \mathcal{N}$. Otherwise $M_{g_{i,j}} = 0$.

Battery

The battery buys on the wholesale market at the clearing price. In that case, the bid volume q^u_t is negative, therefore u should be negative and act as a supplement of demand. Both buying and selling behaviors are ensured with the following constraint:

$$u_t \leq q^u_t$$

By construction (see (2.2)), if $q^u_t < 0$, $c_t = 0$ and, $u^*_t = q^u_t$ to minimize the objective function. Otherwise, $\epsilon = q^u_t - u^*_t$ is an extra MWh of demand that has to be met by a generator, resulting in increasing cost.
Similarly to the generators, $M_u \in \mathbb{R}^n$ is the battery-node adjacency matrix, all its elements are zero except for :

$$M_{ui_b} = 1$$

Economic Dispatch

At a time period t, the market operator dispatches the generation g_t and the storage u_t in order to minimize the cost for the network while satisfying demand, respecting power balance as well as line and bids constraints. The nodal power injection vector is defined as the nodal difference between what is produced $M_g g_t$ and what is consumed d_t. $M_u u_t$ can either act as a supplement of production or consumption. This leads to $p_t = M_g g_t + M_u u_t - d_t$.

$$\begin{aligned} \min_{p_t, g_t, u_t} \quad & a_t^T g_t + c_t \cdot u_t \\ \text{subject to} \quad & \gamma : 1^T p_t = 0 \\ & \lambda : p_t = M_g g_t + M_u u_t - d_t \\ & \beta : H p_t \leq h \\ & \sigma_g : I_g g_t \leq q_t^g \\ & \sigma_u : u_t \leq q_t^u \end{aligned} \tag{2.8}$$

Mathematical Program under Equilibrium Constraints

As shown in program (2.7), λ and u are solutions of the Economic Dispatch. In other words, they must satisfy the KKT conditions of the LP program (2.8). First, the Lagrangian of the Economic Dispatch and its dual function are devised, then a linearization of the KKT conditions and of the objective function are developed.

Lagrangian and dual function

The Lagrangian of the problem is defined as $\mathcal{L}(p_t, g_t, u_t, \omega)$ where $\omega = (\gamma, \lambda, \beta, \sigma^g, \sigma^u)$ belongs to the set of dual variables $\Omega = \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}_+^{2l} \times \mathbb{R}_+^g \times \mathbb{R}_+$. For clarity, since the time period t is fixed, the dependency of the optimization variables in t is dropped. However, the dependency in t of the exogenous input is kept.

$$p = \min_{p,g,u} \max_{\omega \in \Omega} \quad \mathcal{L}(p, g, u, \omega)$$

$$\begin{aligned} \mathcal{L}(p, g, u, \omega) &= a_t^T g + c_t \cdot u + \gamma 1^T p + \beta^T (Hp - h) \\ &\quad + \lambda^T (d_t + p - M_g g - M_u u) \\ &\quad + \sigma_g^T (I_g g - q_t^g) + \sigma_u (u - q_t^u) \\ &= (a_t^T - \lambda^T M_g + \sigma_g^T I_g) g \\ &\quad + (c_t - \lambda^T M_u + \sigma_u) u \\ &\quad + (\gamma 1^T + \beta^T H + \lambda^T) p \\ &\quad - \beta^T h + \lambda^T d_t - \sigma_g^T q_t^g - \sigma_u q_t^u \end{aligned}$$

We define $\mathcal{A}$ s.t :

$$\begin{aligned}\mathcal{A} = \{\omega \in \Omega, & a_t^T - \lambda^T M_n + \sigma_g^T I_g \geq 0; \\ & c_t - \lambda^T M_u + \sigma_u \geq 0; \\ & \gamma 1^T + \beta^T H + \lambda^T \geq 0\}\end{aligned}$$

Then, the dual of the problem can be devised under a closed-form expression:

$$g(\omega) = \min_{p,g,u} \mathcal{L}(p, g, u, \omega)$$

$$g(\omega) = \begin{cases} \lambda^T d_t - \beta^T h - \sigma_g^T q_t^g - \sigma_u q_t^u & \text{if } \omega \in \mathcal{A} \\ -\infty \quad \text{otherwise} \end{cases}$$

Since the problem is a linear optimization problem, it is convex. Furthermore, since the economic dispatch has a feasible solution without the battery bids, there exists a feasible point into the set. Therefore, Slater's conditions are satisfied, and strong duality holds:

$$a_t^T g^* + c_t u^* = d_t^T \lambda^* - h^T \beta^* - q_t^{gT} \sigma_g^* - q_t^u \sigma_u^* \tag{2.9}$$

KKT Conditions for the Economic Dispatch

In this section, a set of constraints $\mathcal{C}$ is devised such that if there exists $(p, g, u, \omega) \in \mathcal{C}$ then $\lambda, u = E.D_t(c_t, q_t^u)$. Let (p, g, u, ω) satisfies the KKT conditions, they are optimal and verify:

- **Stationary conditions** for the Lagrangian

$$\nabla_p L = \gamma 1 + H^T \beta + \lambda = 0 \tag{2.10}$$
$$\nabla_g L = a_t - M_g \lambda + I_g^T \sigma_g = 0 \tag{2.11}$$
$$\nabla_u L = c_t - M_u \lambda + \sigma_u = 0 \tag{2.12}$$

- **Complementary Slackness conditions**

$$\beta^T (H p_t - h) = 0 \tag{2.13}$$
$$\sigma_g^T (I_g g - q_t^g) = 0 \tag{2.14}$$
$$\sigma_u (u - q_t^u) = 0 \tag{2.15}$$

- **Primal Constraints**
  All the other constraints present in program (2.8) must be satisfied at the optimum.

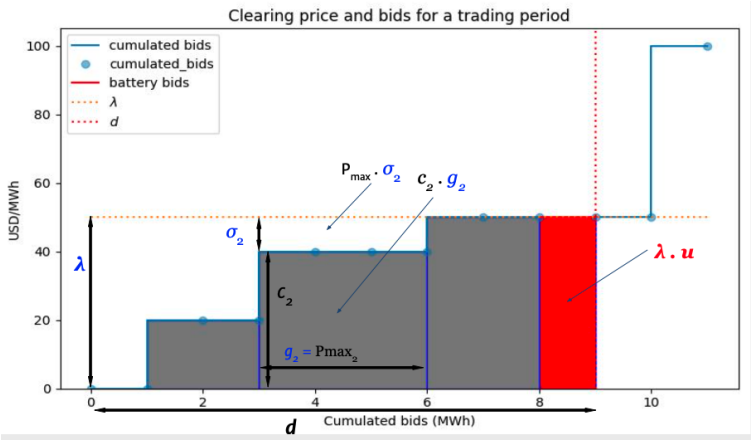


Figure 2.4: Linearization of objective function for single-node wholesale market

Linearization of the program

The KKT conditions cannot be re-injected as such into our battery program (2.7).

Slack constraints The 3 equations of complementary slackness are not affine but can be linearized using integers following Fortuny-Amat and McCarl linearization [118]. A new set of integer variables is introduced: $r^{\beta} \in \{0,1\}^{2l}, r^{\sigma_g} \in \{0,1\}^{2g}, r^{\sigma_u} \in \{0,1\}$. L is set to be sufficiently large. If there exists a set of variables satisfying the following linear constraints, then complementary slackness conditions are satisfied.

$$\beta \leq (1 - r^{\beta})L \tag{2.16}$$
$$Hp_t - h \leq r^{\beta}L \tag{2.17}$$
$$\sigma_g \leq (1 - r^{\sigma_g})L \tag{2.18}$$
$$(I_g g - q_t^g) \leq r^{\sigma_g}L \tag{2.19}$$
$$\sigma_u \leq (1 - r^{\sigma_u})L \tag{2.20}$$
$$(u - q_t^u) \leq r^{\sigma_u}L \tag{2.21}$$

Objective function The objective function of the battery program is not linear and can be transformed following [[118], section 6.4.3.1]

First, by multiplying (2.12) by u :

$$c_t u - M_u \lambda u + \sigma_u u = 0$$

Then from the complementary slackness condition (2.15):

$$\sigma_u u = \sigma_u q_t^u$$

Finally from strong duality (2.9) :

$$-c_t u = a_t^T g_t + \beta^T h + \sigma_g^T q_t^g + \sigma_u q_t^u - \lambda^T d_t$$

The resulting objective function is :

$$\begin{aligned}
\lambda^T u &= -c_t u - \sigma_u u && \text{(from (2.12))} \\
&= -c_t u - \sigma_u q_t^u && \text{(from (2.15))} \\
&= a_t^T g + \beta^T h + \sigma_g^T q_t^g + \sigma_u q_t^u - \lambda^T d_t && \text{(from (2.9))} \\
&= a_t^T g + h^T \beta + q_t^{gT} \sigma_g - d_t^T \lambda
\end{aligned}$$

The objective function is now a linear function in the variables: $g, \beta, \lambda, \sigma_g$

Illustration of the objective function linearization for a single-node wholesale market The above-linearization can be understood as a geometrical transformation as highlighted by 2.4, where we wish to express the red rectangle as a combination of rectangles linear in the optimization variables. To illustrate it, a single node wholesale market is considered with no self-schedule bids ($P_{min} = 0$). The one-time profit made by the battery is:

$$\lambda \cdot u = \underbrace{\lambda \cdot d}_{\text{paid by the loads}} - \underbrace{\sum_{i \leq g} \lambda \cdot g_i}_{\text{paid to the other agents}}$$

The stationary equation (2.11) multiplied by the generator output g_i yields:

$$\lambda g_i = a_i g_i + \sigma_i^{max} g_i$$

This corresponds to the splitting of the generator profit rectangle into two sub-rectangles : its surplus (above its bidding price) and its costs (below its bidding price).

Finally, the slack constraints impose that $\sigma_i^{max} g_i = \sigma_i^{max} P_i^{max}$. As a result, the objective function is linear :

$$\lambda \cdot u = \underbrace{\lambda \cdot d}_{\text{paid by loads}} - \sum_{i=1}^{g} \underbrace{a_i g_i}_{\text{cost of generator } i} + \underbrace{\sigma_i^{max} P_i^{max}}_{\text{surplus of generator } i}$$

MILP program

The subscripts relative to the time period are reintroduced to introduce the multi-period price-maker mixed-integer linear program for the battery revenues maximization. The optimization variables are finally:

$$z^{cap}, q^u, r^c \quad g, p, u \quad \gamma, \beta, \lambda, \sigma^g, \sigma^u \quad r^{\sigma_g}, r^\beta, r^{\sigma_u}$$

And the program is:

$$
\begin{aligned}
\min \quad & \Big(\sum_{t=1}^{T} d_t^T \lambda_t - (a_t^T g_t + h^T \beta_t + q_t^{gT} \sigma_t^g)\Big) \quad -B z^{cap} \\
\text{s.t.} \quad & \tilde{L} u \leq \bar{u} + z^{cap} e \\
& (r_t^c - 1) L \leq q_t^u && \forall t \leq T \\
& c_t \leq r_t^c L && \forall t \leq T \\
& 1^T p_t = 0 && \forall t \leq T \\
& H p_t \leq h && \forall t \leq T \\
& I_g g_t \leq q_t^g && \forall t \leq T \\
& u_t \leq q_t^u && \forall t \leq T \\
& p_t = M_n g_t + M_u u_t - d_t && \forall t \leq T \\
& \gamma_t 1 + H^T \beta_t - \lambda_t = 0 && \forall t \leq T \\
& a_t + M_n \lambda_t + I_g^T \sigma_t^g = 0 && \forall t \leq T \\
& c_t + M_u \lambda_t + \sigma_t^u = 0 && \forall t \leq T \\
& \beta_t \leq (1 - r_t^\beta) L && \forall t \leq T \\
& H p_t - h \leq r_t^\beta L && \forall t \leq T \\
& \sigma_t^g \leq (1 - r_t^{\sigma_g}) L && \forall t \leq T \\
& (I_g g_t - q_g) \leq r_t^{\sigma_g} L && \forall t \leq T \\
& \sigma_t^u \leq (1 - r_t^{\sigma_u}) L && \forall t \leq T \\
& (u_t - q_t^u) \leq r_t^{\sigma_u} L && \forall t \leq T
\end{aligned}
$$

Case Study: New Zealand

New Zealand's Energy Market (NZEM) is relevant to compare the price-maker (PM) and price-taker (PT) approaches as the clearing ED algorithm is a nodal wholesale market using DC powerflow equations [119]. Note that all data, code and graphs presented in this section can be found at [120]. New Zealand's daily wholesale market operates over 48 trading periods of 30 minutes. The Electrical Authority gives open access to a number of relevant data resources regarding the NZEM, including loads, bids and network characteristics.

Scenario

While the program should be ran over at least a year of data to yield results consistent with seasonal patterns of the power flow, we here arbitrarily decided to work on one day (09/02/2019) to illustrate how the constraints affect the results. We made some assumptions

to simplify the bidding data and topology, to favor congestion over the network while ensuring the plausibility of the LMP and preventing infeasible situations.

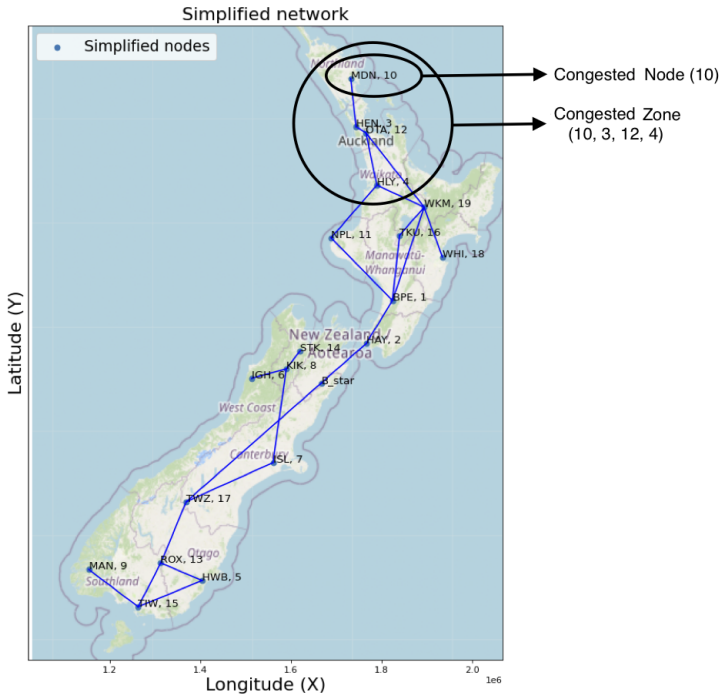


Figure 2.5: Simplified New Zealand power grid

- **Topology:** Most of the nodes being relatively minor, we aggregate the original 258 nodes in 19 main nodes and we connect them with transmission lines forming the "backbone" of the transmission system as per [121]. A swing node (B*, 0) is set between the two islands.

- **Loads:** We multiplied the load at node 10 by 12, so that for some hours, the demand exceeds the capacity of the single line connecting nodes 10 to 3 (see Fig 2.5, North end). We applied a 1.3 factor to the rest of the loads to favor congestion and to prevent infeasible formulation due to high self-schedule volumes.

- **Bids:** For each generator, P_{min} was taken as the volume bid at \$0 and P_{max} as the maximum of the cumulative volume offered on the market. Finally a was set to the weighted average of the bidding prices for bids over \$0. However, with these bids the simulated LMPs are unrealistic. We adjust the bidding prices a so that the simulated marginal cost fits the realized marginal cost.

- **Marginal prices and congestion charge:** Figure 2.6 displays γ, the marginal cost of the system. The congestion charge is reported at a node as the difference between its LMP λ, and γ. λ_{BL} (resp. λ_{PT}) are the baseline price without the battery (resp. with the PT operated battery).

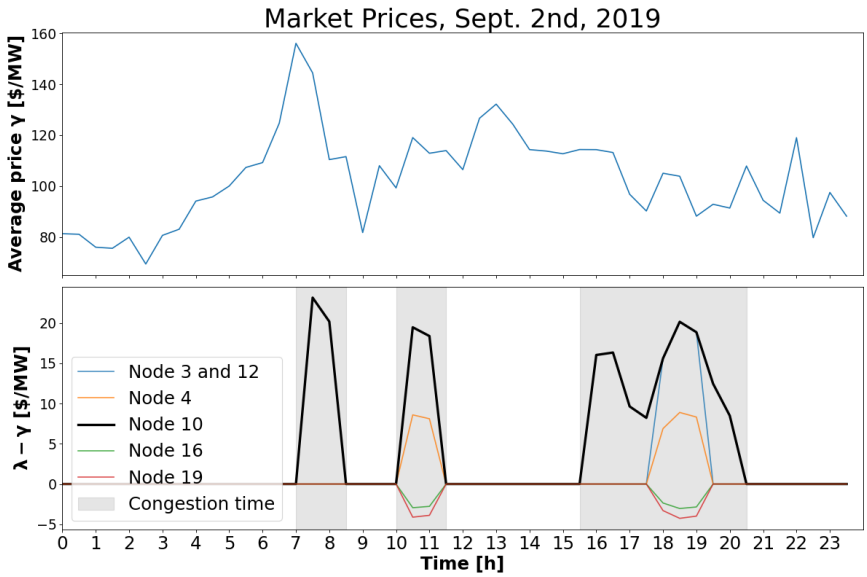


Figure 2.6: Gamma prices and congestion charge for the scenario.

Results over the congested zone

The power capacity is set so that the battery can fully discharge in an hour. Varying this parameter can significantly impact the results as increasing power capacity relaxes the feasible set and thus delivers a better arbitrage for the same installed energy capacity. Note that we could introduce a linear penalty in the power capacity to obtain the optimal parameter. The normalized unit capacity penalty is set to 180$/MWh, i.e. each MWh has to yield a profit of more than 180$ over the scenario. Considering a current capacity cost of around 600$/kWh for a 1-hour duration battery [122], this penalty corresponds to a roughly 9 year-investment period. Selecting a lower penalty leads to abnormally large capacity because of the existence of a significant arbitrage with perfect foresight. Moreover, it is inconsistent with the current 7-10 year lifetime of Li-ion battery storage [123]. A larger penalty would prevent any investment in the battery.

We implemented the price-maker program (PM) following (2.2) and reported on optimal capacity and profits for different sitings. A summary of these results is presented on Table 2.1. The nodes outside the congested area yield to a null capacity by design of the penalty and are not reported in the table. That is, the revenue for 1MWh of energy capacity on

those nodes is worth \$173 which is smaller than its associated cost of \$180. The unit profit (profits for 1MWh of capacity) is reported at the top row of the table and would be the measure chosen to install a battery under a price-taker assumption. We note that node 10 would be chosen by PM to install a 35MWh battery. It would also be chosen by PT without considering information about its size. PT would next select node 4 instead of 3 (or 12) because of higher unit profit. Interestingly, PM informs us that node 4 would yield to a smaller profit since node 4 must be less congested.

Battery Node	**3 or 12**	**4**	**10**
Unit Profit (z_{cap} = 1MWh, in \$)	17	22	39
Energy capacity $\mathbf{z^*_{cap}}$ (in MWh)	45.5	16.4	35.0
Price-Maker Profits (Revenue-Cost, in \$)	575.7	63.6	**769.6**

Table 2.1: Optimal storage size, and profits on congested nodes and for unit capacity cost of $B = 180\$/\text{MWh}$.

Results over Node 10

The PT Losses (third subplot) are defined as the expected one-time revenue -if baseline prices did not change- minus the actual revenue -since prices change (second subplot)-. By providing too much power (e.g. box B at 04:00 p.m., 4th subplot), the price-taker battery loses market power as the congestion is solved. The LMP falls with respect to baseline, to meet the system marginal price (2nd subplot). It sells at a low price, hence incurring large losses (3rd subplot). By withdrawing too much power (e.g. box A at 11:30 a.m.), the battery creates a congestion, the LMP increases (2nd subplot). It buys at a high price, hence incurring large losses (3rd subplot). On the contrary, the same battery operated under PM program, yields higher final revenues (first subplot) as it does not fully solve congestion events. While this happens for a relatively large capacity (accounting for *20%* of the congested node load but only *2%* of the total load) we show that in our scenario, prices react rapidly to even small additions of storage capacity. In Fig 2.8, we plot the average of the absolute difference of the PT and PM LMP at node 10 λ_{10}^{PT}, λ_{10}^{PM} and the baseline LMP λ_{10}^{BL}, as a function of installed capacity. We notice that the price-taker assumption is invalidated even for a small capacity. We deduce from this figure that as soon as a congestion event occurs on node 10, a price-taker assumption cannot hold.

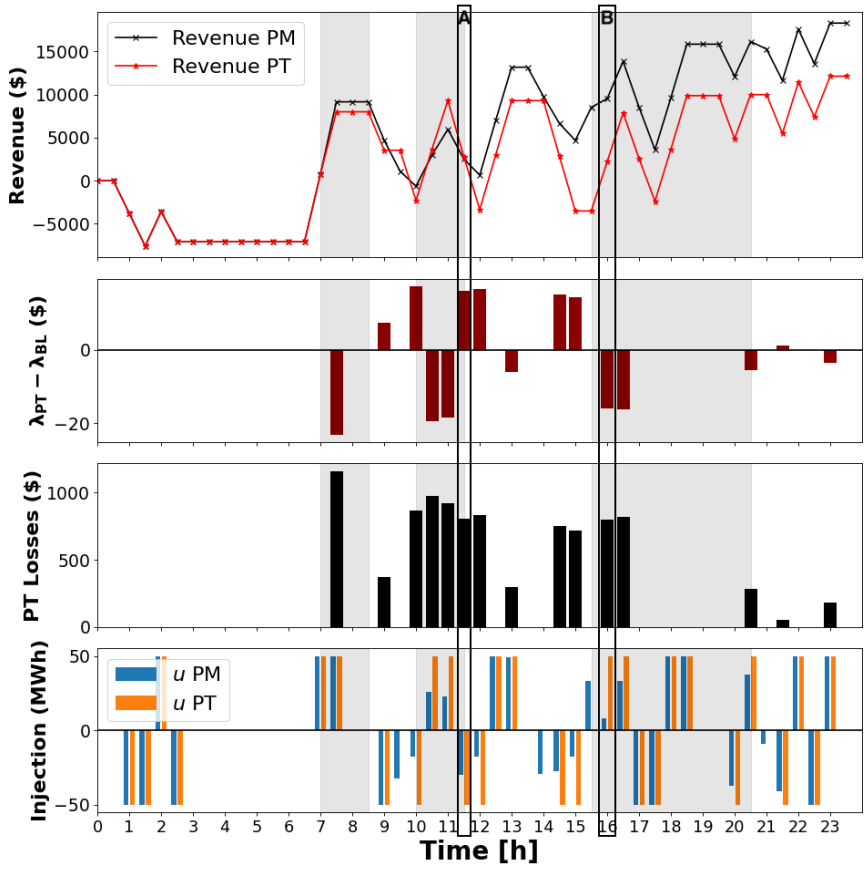


Figure 2.7: PM vs PT comparison for accumulated profits, deviation from baseline prices for PT battery, PT losses and power injection for a 100MW/100MWh battery. The PT assumption can unexpectedly collapse prices or create congestion, which produces less actual revenues than expected.

2.3 Conclusion

We developed a mixed-integer linear program that sites and sizes storage to maximize profits over a nodal wholesale market with the cleared price being endogenous to the program. This is unique from previous work that assume cleared prices are exogeneous (the "price-taker" assumption) and/or do not consider optimal siting. To express the market constraints, we formulated and then exploited the KKT conditions of an Economic Dispatch and linearized the objective function using strong duality.

Finally, we tested our program on the New Zealand Energy Market. We believe this market-aware approach to be new and paramount to correctly size a battery as the assumption of price exogeneity collapse when congestion occurs. While these results hold for a specific scenario of congestion, further work should look into characterizing the influence of

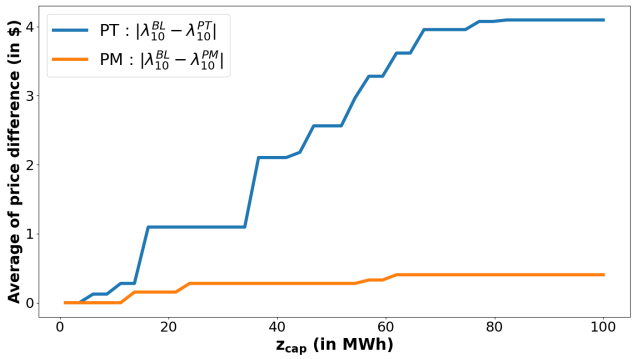


Figure 2.8: Average absolute difference between the PM and PT LMPs with the BL LMP, as a function of z_{cap}. A non-zero value indicates the presence of storage bids have altered the clearing price.

a battery bidding strategy on the prices and on surpluses of the market. If even a small battery can use its price-maker position over a congested area to benefit from high-price, this anti-competitive behavior when operated by larger-scale batteries could impact social welfare.

Chapter 3

Modeling and State Estimation for Lithium Sulfur Batteries as a Piecewise Affine System

Lithium-sulfur (Li-S) is a promising battery chemistry for applications demanding high energy densities, such as electrified aircraft and heavy-duty trucks, among others. A critical challenge in modeling the Li-S chemistry lies in the use of differential algebraic (DAE) equations for representing the electrochemical dynamics. Due to their constrained and stiff nature, these equations are not conducive to real-time state estimation. In this study, we propose a novel approach to constrained state estimation for Li-S batteries by integrating a piecewise affine (PWA) model into a moving horizon estimation (MHE) framework. We begin by deriving the PWA model using a linear tree algorithm based on data obtained from simulations of a calibrated DAE model. We further leverage the unique structural advantages of the proposed PWA model to formulate a real-time state estimation algorithm grounded in a mixed-integer quadratic program. Overall, our initial findings, based on a single constant current trajectory, demonstrate that our approach offers an accurate and computationally efficient method for modeling and state estimation of Li-S batteries. The coupled PWA-MHE framework effectively captures the dynamics of the DAE system, even in the presence of high observational noise (20mV).

3.1 Introduction

With growing interest in Li-S batteries, significant effort has been directed toward material and chemistry research. Such thrusts include better understanding of fundamental reactions, improvement of cell chemistry, and design of the cell electrolyte and electrodes [124, 125, 126]. While these efforts are critical for advancing Li-S technology, there is a need for more focus on modeling and control research to accelerate Li-S commercialization.

Battery models in literature range from high fidelity - computation intensive models to

low fidelity - computation efficient models [127, 128]. Model selection largely depends on specific use case. First principle models provide insights about the system thermodynamics, kinetics, and transport behavior, but they suffer from drawbacks such as requiring prior knowledge of electro-thermochemistry, mathematical complexity, high computational costs, and model calibration difficulties. On the other hand, spatially averaged lumped models like equivalent circuit models (ECMs) or single particle model (SPM) are easier to simulate, calibrate and can be easily applied for battery management system (BMS) applications. In practice the tradeoffs between computational complexity, interpretability and accuracy are key factors when choosing the best model for a given problem.

Fotouhi et al. give an overview of the technological readiness of Li-S batteries from the standpoint of modeling and estimation [129] . Estimation techniques using ECMs for Li-S batteries are well studied in the literature. The key limitation of using ECMs is that they lack of information of electrochemical processes occurring inside the Li-S cell. Electrochemical models, on the other hand, give better insight into the governing physics inside the battery. Zero-Dimensional [127] models give insight into the reaction kinetics of the Li-S cell along with phenomena like precipitation and shuttle effect. 1D models [127] capture both the reaction kinetics and transports dynamics. Both 0D and 1D models capture the distinctive shape of a low discharge current voltage profile accurately in both "plateau" regimes. The modeling and state estimation for these electrochemical models are challenged by complicated reaction pathways, complex DAE system dynamics, weak local observability, and the lack of an estimator design tool for DAE. Some attempt has been made to circumvent some of these challenges while retaining the interpretability provided by electrochemical models. In [130], Xu reduced the 0D electrochemical DAE model to an ODE which was further used for state estimation using an unscented Kalman filter (UKF). A key challenge is that the resulting ODE model has a very stiff structure that can cause issues in its numerical implementation. In [131], state estimation for a 0D DAE model using an extended Kalman filter (EKF) is shown. Despite good performance of the estimator in this work, the measurement noise studied was very low compared to a real life setting, suggesting high measurement noise may pose an issue for EKF estimation.

To address the above challenges associated with standard DAE electrochemical models, we propose a PWA Li-S model that well-approximates the dynamics of a 0D DAE model. To this effect, a learning algorithm is developed in the same spirit as decision and classification tree methods [132]. This PWA framework simplifies the nonlinear DAE model equations into a set of disjoint linear systems. The key advantage is to retain the advantageous structure of linear systems while simultaneously capturing the evolution of electrochemical states. We show that this model can not only simulate states as accurately as the DAE models, but its structure makes it amenable to more advanced state estimation techniques like MHE [133][61]. The most similar work might be [134], which uses a PWA approximation of a Newman-style lithium-ion battery model for the output equation for model predictive control.

The main contributions of this paper are summarized as follows:

- Development of a piece-wise affine learning algorithm that accurately approximates the 0D electrochemical DAE a Li-S battery cell.
- Constrained state estimation of Li-S battery using moving horizon estimation on our piece-wise affine model.

One of the limitations in the field of Li-S battery is the availability of experimental data. Most of the existing literature analyses the battery behavior in low current discharge conditions [127, 128, 130]. Due to this, we focus on modeling cell behavior during discharge.

The paper is organized as follows. In section II, we first introduce the DAE model taken from the literature to simulate the dynamics of a Li-S battery cell. After analyzing various related shortcomings of using this model for state estimation, we introduce PWA systems. In section III, a learning algorithm is then detailed to train a "PWATree" over a dynamical system dataset. The tree is then encoded into a moving horizon estimation framework. In section IV, we finally present results related to our use case: both the offline fit and online estimation with noise undergoing a full discharge cycle under constant current are presented.

3.2 DAE Model

Model

This section summarizes the Li-S model chosen as the baseline formulation for this study. We chose a zero-dimensional electrochemical model developed in our previous work [131] and originally adopted from [128]. The 0D model captures the reaction kinematics at the cathode while the transport dynamics are ignored. We chose the 0D electrochemical model as a baseline, since it captures the voltage dynamics and information about the chemical state of the cathode.

During the discharge process, Li-ions liberated from the anode move toward cathode, and reduce the sulfur species to different Li-polysulfide species via a series of complex electrochemical reactions. The zero-dimensional model in this study considers a 3-step electrochemical reaction given by :

$$\frac{3}{8}S_8^0 + e^- \longleftrightarrow \frac{1}{2}S_6^{2-} \tag{3.1}$$

$$S_6^{2-} + e^- \longleftrightarrow \frac{3}{2}S_4^{2-} \tag{3.2}$$

$$\frac{1}{6}S_4^{2-} + e^- \longleftrightarrow \frac{2}{3}S^{2-} \downarrow \tag{3.3}$$

In the fully charged condition, the cathode is composed of elemental sulfur S_8^0. During discharge S_8^0 reduces to high order polysufides, which further reduce to low order sulfide Li_2S,

which further precipitates. The high order polysufides in a Li-S battery are soluble in the electrolyte. This causes a parasitic loss called as the 'shuttle effect', which is the movement of these soluble species back and forth between the electrodes. It is a key phenomenon to monitor in Li-S batteries.

The 0D electrochemical model takes the form of a differential algebraic equation (DAE) system (3.4).

$$\begin{aligned} \dot{x} &= f(x, z, u) \\ 0 &= g(x, z, u) \\ y &= h(x, z, u) \end{aligned} \tag{3.4}$$

where $x \in \mathbb{R}^5_+$, is the mass of sulfur species $[S_8^0,\ S_6^{2-},\ S_4^{2-},\ S^{2-}, S_p]$ involved in the reaction, $z \in \mathbb{R}^3$ are the algebraic states that represent the currents involved in each electrochemical reaction step $[i_1,\ i_2,\ i_3]$ in (3.1)-(3.3). Finally, $u \in \mathbb{R}$ is the input current and $y \in \mathbb{R}$ is the output voltage. The differential function $f(x, z, u)$ is shown in (3.5).

$$f(x,z,u) = \begin{bmatrix} -\frac{3}{8}\frac{n_{S8}M_S}{n_eF}i_{H1} - k_s x_1 \\ \frac{1}{2}\frac{n_{S6}M_S}{n_eF}i_{H1} - \frac{n_{S6}M_S}{n_eF}i_{H2} - k_s x_2 \\ \frac{3}{2}\frac{n_{S4}M_S}{n_eF}i_{H2} - \frac{1}{6}\frac{n_{S4}M_S}{n_eF}i_L \\ \frac{2}{3}\frac{n_S M_S}{n_eF}i_L - k_p x_5(x_4 - S_*^{2-}) \\ k_p x_5(x_4 - S_*^{2-}) \end{bmatrix} \tag{3.5}$$

The algebraic constraints of the system, $0 = g(x, z, u)$, are summarized in the equations (3.6), (3.7), (3.9), (3.10) below. The currents $[i_1,\ i_2,\ i_3]$ associated with each electrochemical reaction (3.1)-(3.3) are modeled using Butler-Volmer kinetics:

$$\begin{aligned} i_j = -i_j^0 a_r \{ &\prod_i \left(\frac{x_i}{x_i^0}\right)^{s_{i,j}} e^{\frac{F}{2RT}\eta_j} \\ &- \prod_i \left(\frac{x_i}{x_i^0}\right)^{-s_{i,j}} e^{-\frac{F}{2RT}\eta_j} \end{aligned} \tag{3.6}$$

The sum of currents in each reaction i_j equals the total input current flowing through the battery:

$$I = \sum_j i_j \tag{3.7}$$

Each overpotential term η_j in (3.6) is related to the output voltage and the standard reduction potential for each reaction in (3.1)-(3.3).

$$h(x, z, u) = V(t) = E_j + \eta_j \tag{3.8}$$

The standard reduction potential is modeled using Nernst equation (3.9).

$$E_j = E_j^0 - \frac{RT}{n_j F} \sum_i s_{i,j} ln \left(\frac{x_i}{n_{S_i} M_{S8} v} \right) \tag{3.9}$$

The active reaction area a_r in the cathode decreases as sulfur precipitate S_p increases, and is modeled as follows:

$$a_r = a_r^0 (1 - \omega \cdot x_5)^\gamma \tag{3.10}$$

Further details about the DAE model and its parameters associated with (3.1)-(3.3) can be found in [135].

Challenges

Modeling

The DAE system and its reduced ODE form (3.4) from [130] are numerically challenging to solve. Instabilities arise when states approach zero or take complex values in equation (3.6). Therefore, initial conditions for low-order polysulfides $S_1 - S_4$ must be non-zero, even when starting fully charged with S_8 or S_6. The model is notably unstable when states $S_1 - S_4$ are near zero. Fig. 3.1 shows sulfur species evolution during constant current discharge. After 4000 seconds, S_8 is depleted, and lower-order sulfides emerge. Independent Gaussian noise ($\sigma = 10^{-6}$ gm) was added to the states. Despite its small spread, it considerably affects the voltage output plotted in red in the lower subplot. This shows the stiffness of the model output function [130],as small noise in state gets considerably amplified in the output. Model sensitivity to noise is here evident. However, the piecewise affine output function proposed in section (3.3) and plotted in blue mitigates this voltage noise by reducing the voltage output to its best affine approximation.

Estimation

Various algorithms like Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) are tempting options for Li-S battery state estimation. EKF is suitable for systems allowing local linear approximations, while UKF is better for nonlinear systems.

EKF requires solving algebraic equations for new state estimates, a computationally intensive step. Also as shown in Fig. 3.1, the output function is highly sensitive to state estimates, which makes EKF not ideal for this application. UKF, while more apt for nonlinear systems, struggles with ensuring non-negative sigma points, which is critical given the system's sensitivity to negative states for sulfur species. This limitation is also acknowledged in [130]. The paper addresses this fundamental challenge.

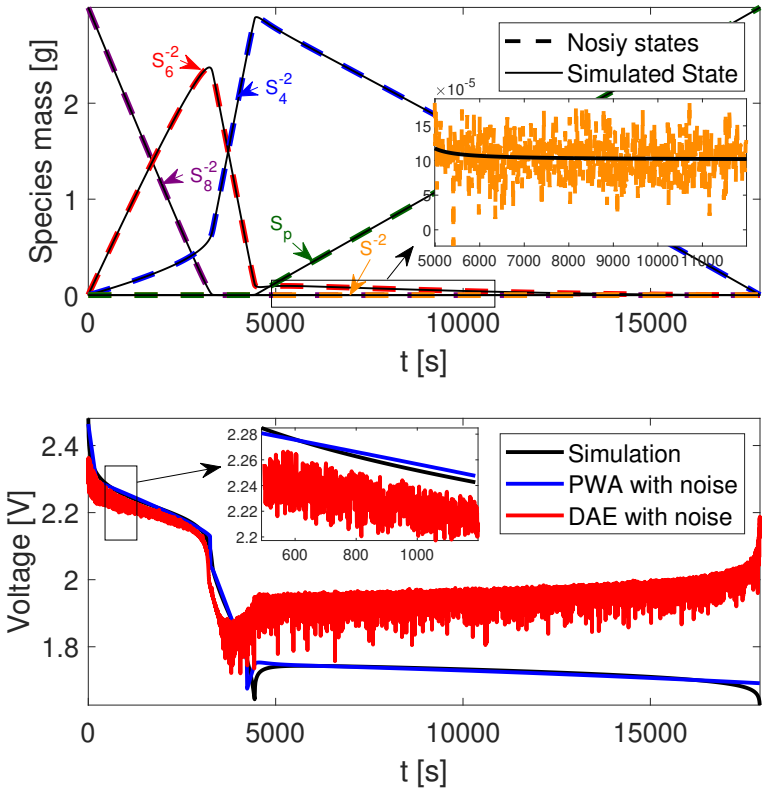


Figure 3.1: Effect of small-magnitude (standard deviation of 10^{-6}) process noise on output voltage $h(x, u)$ for DAE vs PWA system. While noise is magnified in the DAE case, it is of the same order in the PWA case.

3.3 A piece-wise affine approximation

Although the state dynamics presented in (3.4) are a general nonlinear system, they have a particular structure. For example, observe how the mass of S_8 in Fig. 3.1 decreases linearly from start to time 0.4e4s. More generally, (3.5) behaves linearly along certain regions of the state-space if the (i) shuttle effect and (ii) algebraic constraints are neglected. In this section, we introduce a data-driven tree-based PWA approximation to the nonlinear DAE system that leverages this particular structure. Then, the resulting system and its mathematical structure are exploited for state estimation. A PWA system defines both state update and observation output equations as PWA over a partition $\mathcal{M} = \{\mathbb{E}_i\}_{i \leq n_M}$ of the space.

$$\begin{aligned} x_{t+1} &= A_i x_t + B_i u_t + f_i \\ y_t &= C_i x_t + D_i u_t + g_i \\ \forall (x_t, u_t) &\in \mathbb{E}_i \end{aligned} \tag{3.11}$$

Note 1: f_i, g_i are constant offsets (different from functions f, g in (3.4)), thus why we call this system piecewise *affine*. *Note 2:* We wrote the following section for a controllable PWA system to keep the method general to accommodate varying current. However, in the results (section 4.7), we trained a PWATree over a constant current discharge profile. This boils down to learning a PWA autonomous system where B_i, D_i equal zero.

Notation

This section relates equivalences between linear algebra (dynamical systems) and machine learning (binary decision tree) concepts. It serves as reference to understand the relationship between a PWA system and PWATree. The polyhedral (resp. discrete index) notation will be useful in the state estimation (resp. learning) section.

- *Dataset*: our dataset is sampled trajectories of associated states, controls, and outputs of the system (3.4). Denote it as $\mathcal{X} = \big(X, U, Y\big)$. Each element is a timeseries e.g. $X = (x_t)_{t \leq n}$. A state x_t (resp. control u_t) has dimension p (resp. m). For the 0D Li-S model, $p = 5$.

- *Split*: A split (or orthogonal hyperplane) is parameterized by a state (resp. control), an index j and a threshold τ_j. It separates the state and control space into two orthogonal half-spaces. Upper and lower half-spaces are distinguished by their direction $\alpha \in \{-1, 1\}$.
$$\Lambda_j = \left\{ \begin{bmatrix} x & u \end{bmatrix} \in \mathbb{R}^{p+m} \;\middle|\; \alpha \begin{bmatrix} x & u \end{bmatrix}_j \leq \tau_j \right\} \tag{3.12}$$

- *Rectangle*: A rectangle is a polytopic set defined as a non-empty intersection of orthogonal hyperplanes.
$$\mathbb{E} = \bigcap_{j=1}^{k} \Lambda_j \tag{3.13}$$

- $\mathcal{M} = \{\mathbb{E}_i\}_{i \leq n_M}$ partitions the space if for all (x, u), there exists a unique rectangle $\mathbb{E}_i$ where (x, u) belongs. Introducing the indicator binary variable $\delta^i = \mathbb{1}((x, u) \in \mathbb{E}_i)$, it is equivalent to:
$$\sum_{i=1}^{n_M} \delta_t^i = 1 \tag{3.14}$$

- The projection of dataset $\mathcal{X}$ to hyper-rectangles yield discrete index subsets: $R_i = \{t \in [n], (x_t, u_t) \in \mathbb{E}_i\}$.

- The parameter $\theta_i = (A_i, B_i, f_i, C_i, D_i, g_i)$ stores the linear system parameter of equation (3.11).

- *PWA System*: Finally, a PWA system is abstracted by the general parameter $\Theta = \{\mathbb{E}_i, \theta_i\}_{i \leq n_M}$ that contains the partitions and parameters.

  When $(x(t), u(t))$ does belong to $\mathbb{E}_i$, the state and observation agrees to the equations (3.11) subjected to parameters $\theta_i = (A_i, B_i, f_i, C_i, D_i, g_i)$. An example of a PWA function F is presented on Fig. (3.2).

Our objective is to find the best PWA system that minimizes a certain objective function, detailed in the next subsection.

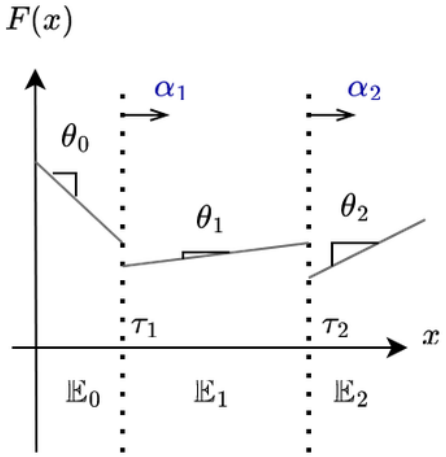


Figure 3.2: Piecewise function F function. Two splits (hyperplanes) are decomposing the state space in three hyper rectangles $\mathbb{E}_i$. Each is associated to a parameter θ_i and a direction (α) with respect to those split $(0, \tau)$.

Learning objective

Given a PWA system Θ, an initial condition x_0 and a timeseries of exogenous controls U, a trajectory of states and observation, a.k.a. dataset, $\hat{\mathcal{X}} = (\hat{X}, \hat{Y}, U)$ is simulated by simply *pushing forward* through (3.11). Hence we write: $\hat{\mathcal{X}} = \hat{\mathcal{X}}(\Theta, x_0, U)$. To facilitate SOC estimation, it is critical for our PWA system to be accurate at estimating the true nonlinear state trajectory. This is what the scoring function (3.15) measures. On the one hand, it minimizes the simulated state tracking error $x_t - \hat{x}_t$ while minimizing the voltage output error defined as $y_t - \hat{y}_t$.

$$\mathcal{S}(\mathcal{X}, \hat{\mathcal{X}}) = \sum_{t=0}^{T} ||x_t - \hat{x}_t||^2 + ||y_t - \hat{y}_t||^2 \tag{3.15}$$

Note: if our objective was to learn a decision tree, the scoring function (or splitting criterion) would simply be the training variance of the resulting tree (see [136]).

Learning algorithm

Theorem 1. A PWA system can be represented under the form of a PWATree.

A sequence of k non-identical splits: $\mathcal{S} = \{(j_i, \tau_i)\}_{i \leq k}$ generates $2k$ half-spaces Λ following equation 3.12. We can define a sequence of rectangles parameterized by their location with respect to these splits (α): $E(\alpha) = \bigcap_{j=1}^{k} \Lambda_j(\alpha_j)$. Finally, we can show that their intersection $\mathcal{M} = \Pi_{\alpha \in \{-1,1\}^k} E(\alpha)$ is indeed a partition.

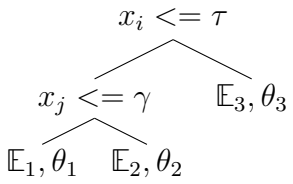


Figure 3.3: Tree representation of a PWA system

Any PWA system defined over hyper-rectangles (3.13) can be encoded into a tree format, like illustrated in Fig. 3.3. This representation facilitates a tree-based learning algorithm where splits are selected to minimize a certain objective function (3.15). The parameters in each rectangle $\mathbb{E}_i$ can be estimated via a convex optimization sub-routine $\mathcal{O}$. Sub-routine is given by the mathematical program (3.16), which regresses the parameters θ of a linear system over a dataset $\mathcal{X}$. An L_1 penalty encourages sparsity in the estimated parameters. This sub-routine is a quadratic program and will be called to quickly evaluate the quality of a split which generates a new partition of the state-control space. The proposed method in Algorithm 1 follows the training algorithm of classification and regression trees. We recursively loop over *candidate* binary splits that partition the state-control space. To reduce computational burden, the evaluated splits are sampled from the quantiles of the distributions of each state. The single split which reduces the scoring function the most is selected. This generates a PWA system that yields the most faithful representation of the nonlinear system. The split is then enacted by recursively calling on the procedure for the left and right children $\mathbb{E}_l, \mathbb{E}_r$. Finally, Δ^{max} fixes the maximum depth of the tree while n_{min} makes sure that enough data is collected in the leaves to run the subroutine.

$$
\begin{aligned}
\mathcal{O}(x, y) = \min_{\theta} \quad & \sum_{t=1}^{N} ||w_t||^2 + \sum_{t=1}^{N} ||v_t||^2 + \lambda ||\theta||^1 \\
\text{subject to} \quad & x_{t+1} = Ax_t + Bu_t + f + w_t \\
& y_t = Cx_t + Du_t + g + v_t \\
& \theta = (A, B, f, C, D, g)
\end{aligned} \tag{3.16}
$$

where w_t, v_t represent the error between the (resp.) state and output trajectories in dataset $\mathcal{X}$ and trajectories predicted by the linear model parameterized by θ.

Algorithm 1 PWATree

Input: Data $\mathcal{X} = \{(x_i, x_{i+1}, y_i)\}_{i \in I}$ current PWA Tree Θ_0, current depth Δ
Exogenous Data: initial dataset $\mathcal{X} = \{(x_t, x_{t+1}, y_t)\}$
Hyperparameters: $\{n_{min}, \Delta^{max}\}$, function $\mathcal{S}$, subroutine $\mathcal{O}$
PWA system $\Theta = \{\mathbb{E}_i, \theta_i\}_{i \leq n_M}$ Compute simulation dataset $\hat{\mathcal{X}}$ from Θ_0
Compute initial score $v = \mathcal{S}(\mathcal{X}, \hat{\mathcal{X}})$
Set $v_{min} \leftarrow v$, SPLIT $\leftarrow$ False

if $\Delta < \Delta^{max}$ and $|I| \geq n_{min}$ **then**
  for Feature j and threshold τ **do**
    Left child $R_l = \{i \in [n] : x_{i,j} < \tau\}$

    Compute affine parameters $\theta_l = \mathcal{O}(\mathcal{X}_{R_l})$

    Right child $R_r = \{i \in [n] : x_{i,j} > \tau\}$

    Compute affine parameters $\theta_r = \mathcal{O}(\mathcal{X}_{R_r})$

    Generate tentative PWATree $\hat{\Theta} = \Theta_0 \cup \{(\mathbb{E}_l, \theta_l), (\mathbb{E}_r, \theta_r)\}$

    Compute simulation dataset $\hat{\mathcal{X}}$ from $\hat{\Theta}$

    Compute simulation score $v = \mathcal{S}(\mathcal{X}, \hat{\mathcal{X}})$

    if $v < v_{min}$ **then** Update $j^* \leftarrow j$, $\tau^* \leftarrow \tau$
Update $v_{min} \leftarrow v$, SPLIT $\leftarrow$ True Update $\Theta^* \leftarrow \hat{\Theta}$

  if SPLIT is True **then**
    $\Theta_l = \text{PWATree}(\mathcal{X}_{R_l}, \Theta^*, \Delta + 1)$

    $\Theta_r = \text{PWATree}(\mathcal{X}_{R_r}, \Theta^*, \Delta + 1)$

    $\Theta_0 \leftarrow \Theta_0 \cup \{(\mathbb{E}_l, \theta_l), (\mathbb{E}_r, \theta_r)\}$ **return** Θ_0

Moving horizon estimation

Moving horizon estimation (MHE) is an optimization technique that is employed to estimate the unknown system states given a state update, observation and potentially algebraic func-

tions. It has been widely studied in diverse contexts[133][61]. While this method has the disadvantage of being computationally expensive relative to state observers, it has particular advantages in our application. First, it imposes no assumptions on the noise process. Second, one can include linear constraints thus enabling constrained state estimation, a particularly helpful feature for the 0D Li-S model. The formulation of (3.17) introduces a baseline MHE formulation. Different flavors of MHE exist in the literature, but usually MHE presents 3 objectives to minimize. First, minimize the error between measured (y_τ) and modeled observations. Then, fit the state trajectory to the dynamical system model. Finally, fit the last known estimate (or initial condition) $\bar{x}_{t-h}$. The third term is an optional relaxation of the algebraic constraints $g(x) = 0$. Horizon parameter $h \in \mathbb{N}$ (not to be confused with function h in (3.4)) is the window size and can be tuned. The larger h is, the more accurate the estimation will be but at the expense of computation time. Note that the state estimates are made to belong to a polytopic set $\mathcal{C}$. Finally, some weights can be associated to each of those objectives according to prior knowledge of the system.

$$\begin{aligned}\text{argmin}_{\mathbf{x}\in\mathcal{C}} \sum_{\tau=t-h}^{t} &||y_\tau - h(\mathbf{x}_\tau)||_R^2 + ||\mathbf{x}_{\tau+1} - f(\mathbf{x}_\tau, u_\tau)||_Q^2 \\ &+ ||g(\mathbf{x}_{t-h})||_\xi^2 + ||\mathbf{x}_{t-h} - \bar{x}_{t-h}||_\Pi^2 \end{aligned} \tag{3.17}$$

The prospect of funneling a PWA system inside a MHE framework is appealing. MHE transforms to a Quadratic Program in the linear time invariant case. Other works [133][137] have shown that a PWA-MHE can be formulated as a Mixed Integer Quadratic Program (MIQP). This formulation can be solved near-optimally by using the full extent of modern branch and bound solvers such as GUROBI.

PWA-MHE for Li-S battery

Replacing the nonlinear h and f functions from program (3.17) by our PWA system yields program (3.26). The algebraic constraints we consider are two-fold:

- As detailed in the notations, the sum of the indicators must equal one.
- Mass conservation is verified by constraining the sum of the normalized states to one. We later relax this constraint to allow for the Li-S shuttle effect.

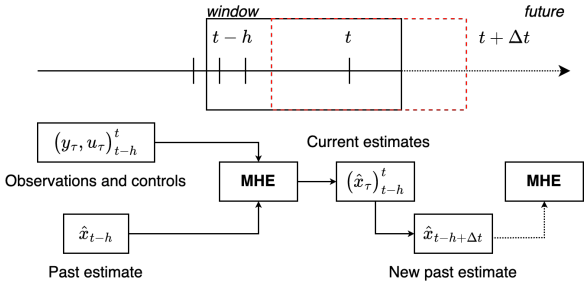


Figure 3.4: Operational framework for moving horizon estimation: Two hyper-parameters are the window length h and the update period Δt. The MHE algorithm estimates states based on the previous h observations, controls, and state estimate at time $t-h$. Finally, we reuse the first index of the next window in the next MHE calculation.

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & \sum_{\tau=t-h}^{t} ||\mathbf{v}_\tau||_R^2 + ||\mathbf{w}_\tau||_Q^2 + ||1 - 1^\top \mathbf{x}_\tau||_\xi^2 \\
& + ||\mathbf{x}_{t-h} - \bar{x}_{t-h}||_\Pi^2 \\
\text{subject to} \quad & \mathbf{x}_{\tau+1} = \sum_{i=1}^{n_M} \boldsymbol{\delta}_{\boldsymbol{\tau}}^{\boldsymbol{i}} \cdot \left(A_i \mathbf{x}_\tau + B_i u_\tau + f_i\right) + \mathbf{w}_\tau \\
& y_\tau = \sum_{i=1}^{n_M} \boldsymbol{\delta}_{\boldsymbol{\tau}}^{\boldsymbol{i}} \cdot \left(C_i \mathbf{x}_\tau + D_i u_\tau + g_i\right) + \mathbf{v}_\tau \\
& \sum_{i=1}^{n_M} \boldsymbol{\delta}_{\boldsymbol{\tau}}^{\boldsymbol{i}} = 1, \qquad \mathbf{x}_\tau \in [0,1]
\end{aligned}
\tag{3.18}
$$

This program can be reformulated into linear form by introducing new binary optimization variables and additional constraints. We refer the reader to these articles for more information[133][137].

We finally illustrate the pipeline tying the estimation algorithm to the datasets of observations and controls in Fig 3.4.

Linearization

Following an extensive literature on linearizing such systems, we can introduce one continuous optimization variables z_i per region and force them to zero when δ is 0 in the following way.

$$
\begin{aligned}
-M_i \delta_i \leq \mathbf{z}_i \leq M \cdot \delta_i \\
A_i \mathbf{x} + B_i u - M \cdot (1 - \delta_i) \leq \mathbf{z}_i \leq A_i \mathbf{x} + B_i u + M \cdot (1 - \delta_i)
\end{aligned}
\tag{3.19}
$$

In that case:

$$\begin{aligned} \mathbf{z}_{i,t} &= \delta_i \cdot \big(A_i \mathbf{x}_\tau + B_i u_\tau\big) \\ \mathbf{x}_{t+1} &= \sum_{i=1}^{n_M} \mathbf{z}_{i,t} \end{aligned} \tag{3.20}$$

The rest of this section deals with enforcing constraints s.t. $\delta = \mathbb{1}\big((x,u) \in \mathbb{E}\big)$. The first challenge is transform an indicator function of a closed interval to a sequence of linear constraints and the second is to write an algorithm that translates a PWATree into an union of constraints and optimization variables.

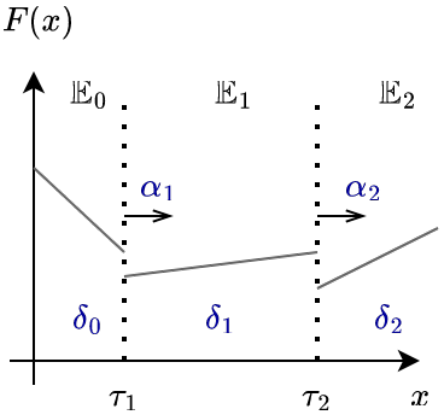


Figure 3.5: Piecewise function F function with $x \in \mathbb{R}$. The optimization variables α are equal to 1 if x is bigger than their respective thresholds. Hence $\delta_0 = 1 \iff \alpha_1 = 0, \alpha_2 = 0$ and $\delta_1 = 1 \iff \alpha_1 = 1, \alpha_2 = 0$

We introduce another set of binary variables α that indicates if (x,u) is on the left (resp. right) side of the orthogonal hyperplane (see notations (3.12)). Fig.(3.5) highlights these two sets.

This equivalence can easily be encoded as constraints in an optimization problem. Let M be a large number and consider the following inequalities:

$$M \cdot (\alpha - 1) <= (x,u)_i - \tau \leq \alpha \cdot M \tag{3.21}$$

Let E_k be a hyper-rectangular region of the space, its boundaries are such that there exist a set I and J such that, if $(x,u) \in E_k$

$$\begin{aligned} \forall j \in J, (x,u)_{\kappa_j} \leq \tau_j &\iff \alpha_j = 0 \\ \forall i \in I, (x,u)_{\kappa_i} \geq \tau_i &\iff \alpha_i = 1 \end{aligned} \tag{3.22}$$

Hence, we can conclude that:

$$\delta_k = 1 \iff \bigcap_{i \in I_k} \{\alpha_i = 1\} \bigcap_{j \in J_k} \{\alpha_j = 0\} \tag{3.23}$$

In the optimization, we enforce:

$$\bigcup_{i\in I_k}\{\alpha_i = 0\} \bigcup_{j\in J_k}\{\alpha_j = 1\} \implies \delta_k = 0 \tag{3.24}$$

And the reverse implications will be logically enforced ... more information here. Finally we can add the following constraints the problem:

$$\begin{aligned} \delta_k &\leq 1 - \alpha_j,\ \forall j \in J_k \\ \delta_k &\leq \alpha_i,\ \forall i \in I_k \end{aligned} \tag{3.25}$$

We finally obtain:

$$\begin{aligned} \min_{\mathbf{v},\mathbf{w},\mathbf{x},\boldsymbol{\delta_\tau}} \quad & \sum_{\tau=t-h}^{t} ||\mathbf{v}_\tau||_R^2 + ||\mathbf{w}_\tau||_Q^2 + ||\mathbf{x}_{t-h} - \bar{x}_{t-h}||_\Pi^2 \\ & \sum_{i=1}^{n_M} \boldsymbol{\delta_\tau^i} = 1 \\ & \mathbf{x}_\tau \in \mathcal{C} \\ & \text{Equations (3.25) (3.21) (3.19) (3.20)} \end{aligned} \tag{3.26}$$

3.4 Results

The PWAtree and MHE algorithms are both tested on the same full discharge cycle at low discharge current. Our future research interest looks to expand this appproach to varying current. The dataset is simulated by the high-fidelity DAE model (3.4). Key statistics are summed up in Table 3.1. The dynamics were previously graphed on Fig. 3.1. Additional Gaussian noise was added to the voltage output equation to mimic real-life conditions. Its magnitude (standard deviation of 20mV) is an order of magnitude greater than the usual distribution of errors of modern voltage sensors (order of 1mV). This magnitude was selected to showcase the robustness of the estimator design. Estimating states for this chemistry and profile is already a significant contribution, as previous papers have predominantly looked at estimating the states (i) in the high plateau region only where observability is high and (ii) for low observation noise to avoid infeasibility in the observation function (see Section 3.2). In this section, we will first assess the offline fit of the PWATree to the data, then examine the estimation results.

Piecewise Affine Fitting

We implemented Algorithm 1 on the aforementioned dataset. This ultimately returns a PWATree of depth 2. Each iteration of the algorithm finds the best splits that minimizes the scoring function (3.15). Table 3.2 shows how the scoring function (analogous to training

Table 3.1: Simulation Parameters

Parameter	Value
Constant current u (A)	1
Time to discharge (s)	$89,634$
Capacity (Ah)	24.9
Sub-sample Period (s)	500
Timesteps in simulation dataset	179
Standard deviation of observation noise (mV)	20
Initial state	$\begin{bmatrix}1 & 0 & 0 & 0 & 0\end{bmatrix}$

loss) decreases as we grow the tree depth, in terms of mean square state estimation error and scoring function (3.15). As expected, the tree yields better score as the tree grows in depth.

Table 3.2: PWA Scores with depth/iteration of Algorithm 1

Depth/Iteration	MSE State	MSE Output	Score
0	3.94	0.75	4.69
1	0.90	0.28	1.18
2	0.15	0.13	**0.28**

Fig. 3.6 depicts the final PWATree. The fitting algorithm remarkably learns *interpretable* divisions in the dynamics. For example, the splits on S_8 and S_6 mark the depletion of those species which are key transitional steps in the dynamics of the chemical reactions. Fig. 3.9 highlights with vertical dotted lines the moments in time when the PWA system switches.

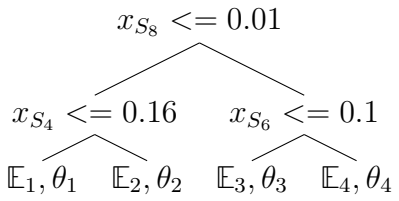


Figure 3.6: PWAtree for Li-S battery constant current discharge

The state and output trajectories for the learned tree is presented Fig. 3.9. After three iterations of Algorithm 1, the PWA-trajectory is indistinguishable from the original nonlinear DAE model. Remarkably, even the characteristic lowest dip at time 20,000 sec is well approximated by a linear function in the states. This is due to a switch from $\mathbb{E}_2$ to $\mathbb{E}_1$ that

allows for the output equation to change its form and in particular its intercept (precisely from $g_2 = 1.27$ to $g_1 = 1.16 \in \theta_1$ from Fig. 3.6). Thus, we see PWATree accurately captures the dynamics of the states and the output function, and offers a compelling model approximation for state estimation.

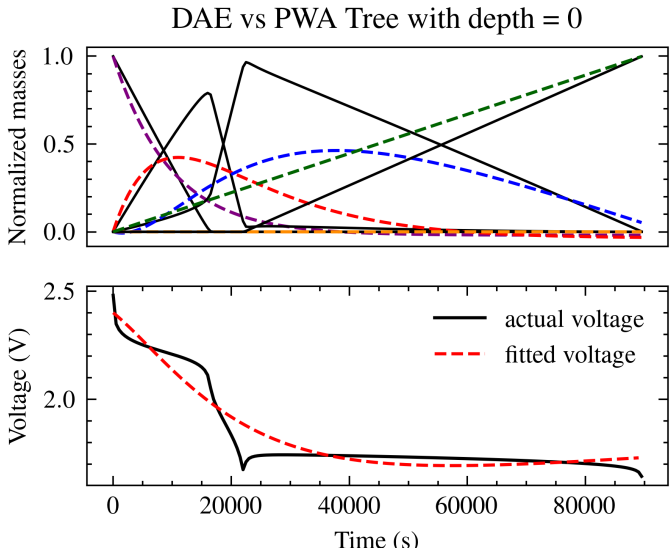


Figure 3.7: Model to Model comparison for a PWATree of depth 0. This represents the best linear time invariant fit to the dataset.

Moving Horizon Estimation results

In this, we run a sequence of MHEs from a perturbed initial state and under noisy observations. Table 3.3 details the hyper-parameters used, Section 3.3 and Fig. 3.4 provide more detail on the timeline of this sequence. The value of the weights were hand-tuned to approximately equal the contribution of each objective function (10^{-4}) detailed in Program 3.17.

The results are presented in Fig. 3.10, including the noised voltage timeseries (lower subplot). Despite an offset initial condition, the MHE framework quickly converges within a small neighborhood of the actual values. This is explained by the high voltage output measured at start. The only way for the MHE to match this high voltage input is to assign state S_8 to 1 which explains this fast convergence to the actual states. Another feature of the PWATree approach is the ability to locate very precisely the switch time (dip time) of the transition from high to low plateau.

To showcase this feature, we run an instance of MHE over window A. of Fig. 3.10. We again start from an offset initial condition and and feed in 30 sampled observations from Time 15,000s to 30,000s. The internal states $\hat{x}_{t-h,t}$ are plotted along with the filtered

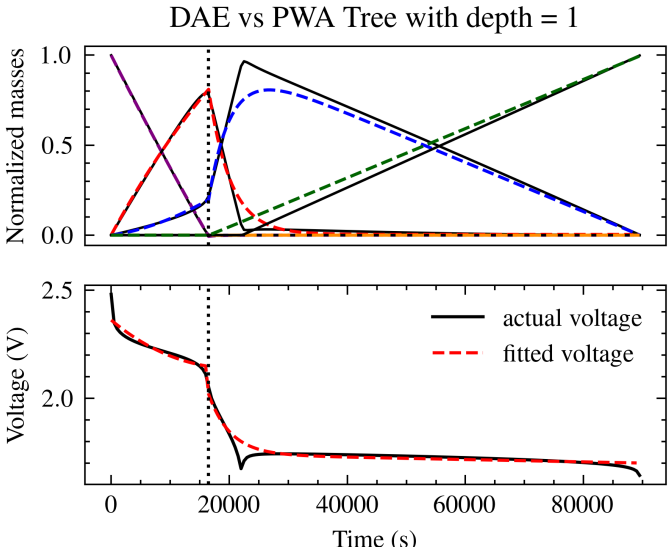


Figure 3.8: Model to Model comparison for a PWATree of depth 1. The first split selected $(x_{S_8}, 0.01)$ reduces considerably the fitted state error and the output error by separating the high and low plateau.

Table 3.3: MHE Parameters

Parameter	Value
state constraint $\mathcal{C}$	$x_t \in (0,1), \sum_i x_{i,t} = 1$
weights for output error R	1
state update error Q	10
past-estimate error Π	1
Length of window h	15
Update period Δt	5
Average of computation time (s)	3.4 (std 2.9)
Initial state	$\begin{bmatrix}0.7 & 0.2 & 0.1 & 0 & 0\end{bmatrix}$

observation $\hat{y}_t^{MHE}$ on Fig. 3.11. The states converges to their actual values thanks to the good identification of the dip at time 22,000s (dotted lines) vs actual dip time of 21,500s. For information, only state b. would be reported to the user after such a run.

After the dip, we rely on the accuracy of the linear approximation in the low plateau region to deliver accurate state estimates. Finally, it is important to highlight that the PWA-MHE can produce sulfur masses that are precisely zero (e.g., see the initial condition in Table 3.3) yet constrained to be non-negative. This is not mathematically possible with the original DAE model and a Kalman-based filter.

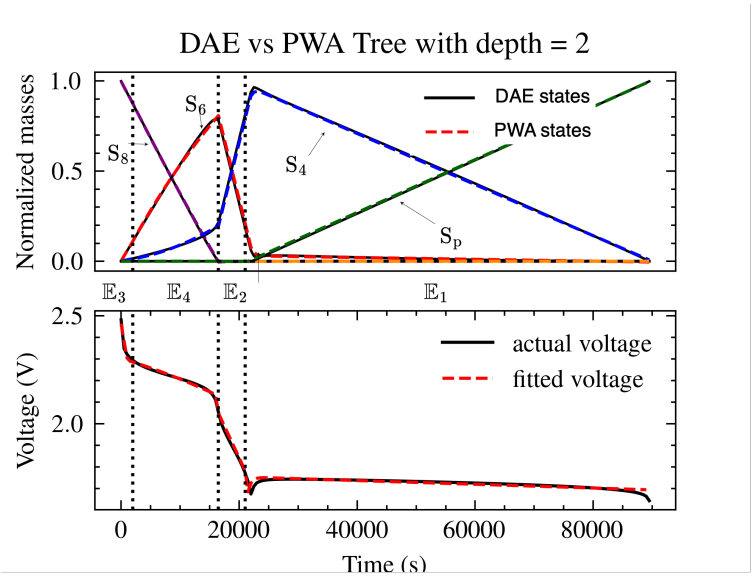


Figure 3.9: Model to Model comparison for simulated dynamics for ground-truth DAE model (black solid) and optimal PWATree (dashed color). The dynamics from the fitted tree closely follow those resulting from the nonlinear DAE model. The different splits where the PWA systems switches parameters are represented in dotted lines and the hyper-rectangles where the states lie are written between the subplots. The states leading to the transition from high to low plateau (time 20,000s) belong to the same PWA hyper-rectangle. This region is crucial for state estimation, as observability is lost in the low-plateau region, as analyzed in [130, 131].

3.5 Conclusion and future work

In this article, we demonstrated how learning a PWA system through the use of a binary tree algorithm can yield an interpretable and tractable approximation to nonlinear dynamics. In the case of a Lithium Sulfur battery, a PWA system was learned with high accuracy. We further demonstrate the application of the PWATree into a moving horizon estimation (MHE) algorithm. We use MHE to estimate states given noisy observations. This approach could be generalized to more challenging current profiles and other types of complex dynamical systems. Our future work will first involve generalizing the learning phase over a sequence of diverse charging and discharging profiles. This step is crucial to allow our estimation method to generalize to more challenging current profiles.

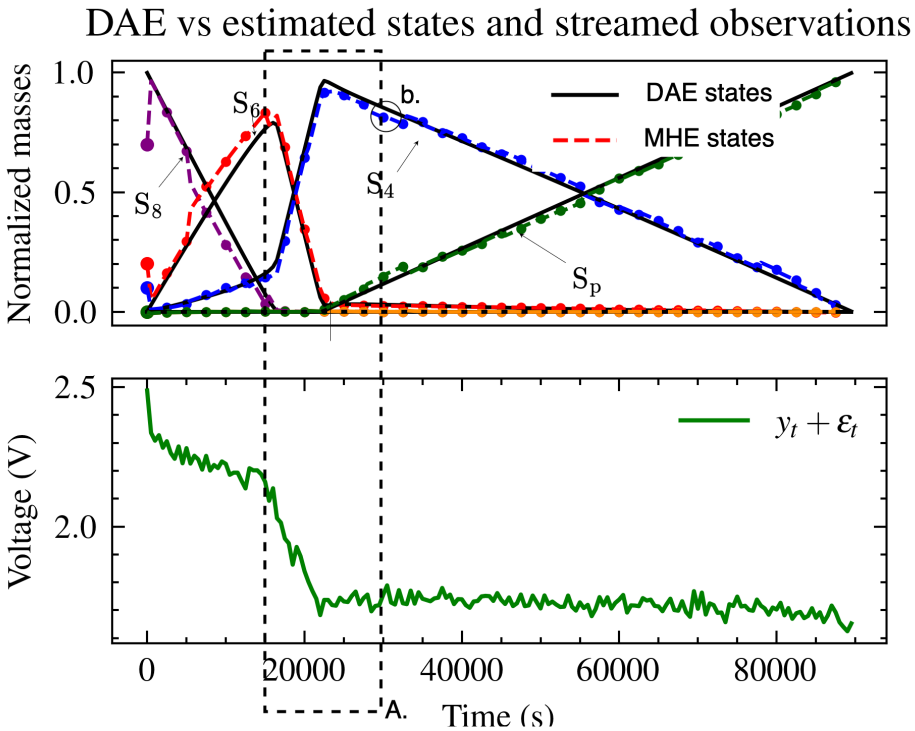


Figure 3.10: State estimation results using PWA-MHE. The true states from the DAE model (in solid lines) are estimated by the algorithm despite the large measurement noise (standard deviation 20mV) and offset initial condition (first time step points). This figure is the result of running a sequence of MHE until the end while reporting the last estimate. In particular to estimate the circled state b., an MHE streamed a window A. of the past h-measurements.

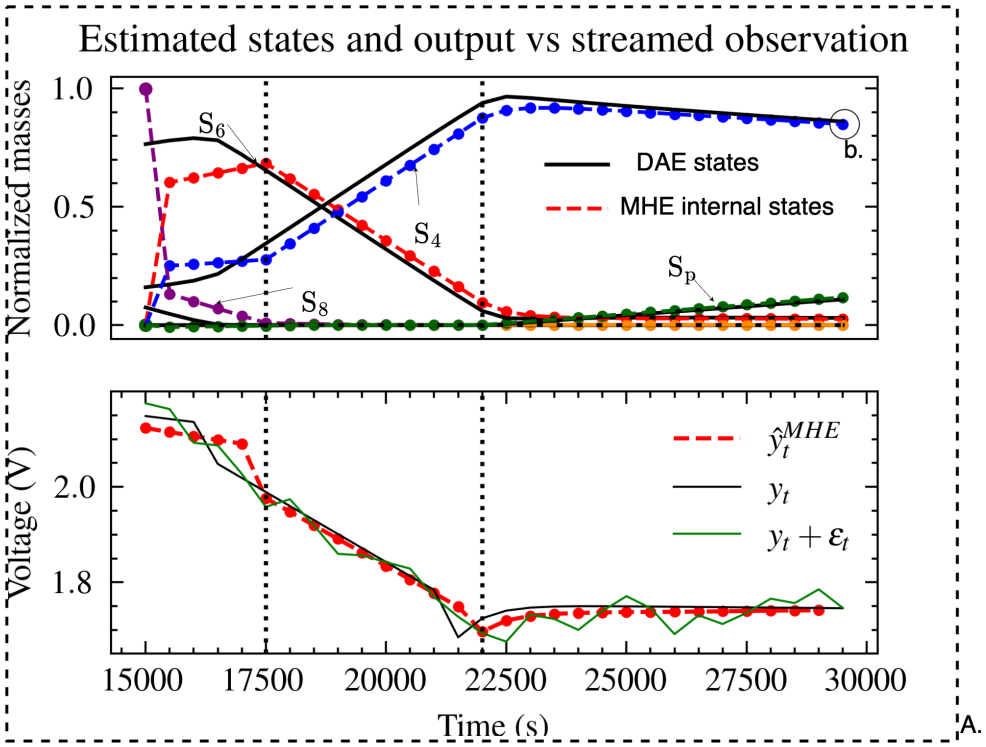


Figure 3.11: One time-step MHE output. Despite a poor initialization (State S_8 concentrates all the mass), MHE is fitting the voltage observation (lower subplot in green) with the red curve over the window A. of past data. The state dynamics (upper subplot) are also estimated, such that their evolution minimizes the plant model error and deviation from initial condition. Finally, the last time step state b. is reported. Estimated states converge to true states.

Chapter 4

Exploration vs. Exploitation in Airborne Wind Energy Systems via Information-Directed Sampling Control

Airborne Wind Energy systems (AWEs) are an emerging wind generation technology. They differ from conventional turbines in that they are attached to the ground by a tether and can evolve from low to high altitudes (approx. 1km). Informed altitude control of AWEs is key to track favorable wind speed and maximize power output in a time-varying and partially-observable environment. Leveraging recent advances in Multi-Armed Bandit problems, we recursively estimate the wind profile distribution and use the residuals to fit the noise covariance in an online fashion. This filtering approach paves the way for the computation of (i) the distribution of the wind-output given past observations and (ii) the expected reduction in entropy in the optimum distribution with respect to the potential future altitude set-point. We implement an *Information Directed Sampling* controller that minimizes the ratio of squared-regret per bit of information gained about the optimum. We finally compare our controller with different baseline controllers using real-world data.

4.1 Introduction

Background

Decarbonizing the energy supply relies on the emergence and development of carbon-free generation sources that are competitive in terms of both cost and performance. While wind energy is already one of the most prevalent renewable resources in the United States[138], conventional turbines suffer from low capacity factor due to variability in wind speed. Airborne wind energy systems (AWEs) differ from conventional wind turbines by employing

both a lifting body (kite, rigid wing, or aerostat) and adjustable length tethers, offering intriguing advantages. AWEs can harvest wind at higher altitudes [139] and across a range of operating altitudes. At a first approximation, wind speed increases monotonically with height [140] [141]. It follows that, AWEs can achieve a higher capacity factor compared with conventional stationary turbines by operating at higher altitudes, or by making adjustments to track favorable wind speeds across a wide range of altitudes. Higher altitude wind patterns - besides being stronger - are generally more temporally consistent and less turbulent [142]. Other advantages include low visual & acoustic disturbance, ease of deployment and cheaper cost, as they do not need expensive aluminum towers [143].

Literature Review

Past research on AWE altitude control focused on methods that find and stabilize at an optimal operating altitude [144][145]. The notion of optimality is, however, difficult to characterize. No meteorological model exists to accurately forecast the whole wind shear profile at high altitudes for the short-term time horizons most relevant to AWE controls.

Characterizing the distribution of the maximum wind energy output is challenging. The wind profile is relatively high dimensional, and at each time step the observation is partial. The reason for this is that instrumentation for monitoring wind speeds is co-situated with the turbine itself; thus the wind speed is only measured at the current operating altitude. The challenge of estimating a distribution of hidden variables can be found in many different scientific fields. A traditional way to address this issue is to use model reduction, filter theory or Gaussian processes. Application of those methods are especially common in meteorology [146][147][148]. We particularly recommend the readings of both [149]–where an extended Kalman Filter is used to estimate the wind speed distribution and fit the model parameters – and [150] – where a Kalman Filter estimates states corresponding to a projection of wind speed on carefully designed basis functions. These fundamental ideas of low-dimensional state estimation of a spatio-temporal system under sparse observations will be further illustrated in this work.

AWE altitude control differs from these works in that there is a causal relationship between action and observation. Moving the AWEs to a specific altitude harvests wind power at the new location, contributing directly to the objective function (exploitation) that we seek to maximize. Yet, control actions also elicit information about the state estimate (exploration). This is the dual role of control as described by P. R. Kumar and P. Varaiya [54]. This problem intervenes frequently in reinforcement learning. Assuming a structure in the wind speed evolution, our problem can be framed as a stochastic *restless* multi-armed bandit where the states (wind speeds) of all arms (altitudes) can change at each step according to a known stochastic transition function. Restless bandits are notoriously intractable. To overcome these challenges, we leverage the problem's specific structure and use Information-Directed Sampling, which is known to perform well in these instances[151].

Recent efforts have recognized the stochastic nature of the spatio-temporal evolution of the wind profile. Bin-Karim et al. developed a model predictive controller (MPC) [152]

and Bafandeh et al. [153] used a Lyapunov-based extremum seeking, using surrogate power deficit as the objective. Both approaches rely on the statistical accuracy of the underlying forecasting model trained offline under full observability (assuming knowledge of the whole profile), but operating under sparse observation. As reckoned by Baheri et al. in [154] and [155], the statistical properties of the wind shear should be learned online to avoid a long and costly period of data collection. To address this shortcoming, they designed Gaussian Process forecasting models and used Bayesian Optimization (expected improvement) to deal with the trade-off between exploitation and exploration. The first potential limit to this method, as Dunn et al. noted in [156], is that the power production function is nonlinear in the wind-speed. Hence, there is little reason that the power output would be normally distributed. Another unaddressed problem is that the power output is not simply a function of the wind speed, but also depends on the altitude adjustment. Third, the expected improvement is a greedy controller which focuses its sampling effort near the estimated optimum. It has been proven to perform poorly in a best-arm identification problem [157].

In previous work [158], we assumed that the wind profile follows a vector auto-regressive process. This approach highlights that the forecast model is difficult to fit in practice, as one needs to estimate the past lagged profiles from sparse observations to forecast future profiles. Finally, Dunn et al. examined the impact of sensor configurations on power output for different controllers. This work revealed that partial observability can significantly degrade performance, due to poor forecasting accuracy [156] [159].

Contribution

Our work seeks to complement previous efforts. We define a persistent forecasting model (as in [156]) which assumes little on the wind profile evolution law. This defines a vectorial stochastic process governing the wind profile evolution as in [158]. To address the problem of estimating the whole profile given sparse observation, we reduce the system dimensions by (i) projecting onto a 4-dimensional subspace, and (ii) leveraging a Gaussian process structure on the wind profile evolution similar to [154] to develop an online learning algorithm for the process covariance function. Finally, (iii) the Kalman Filter (KF) recursive equations -by forecasting the impact of an altitude adjustment on the state estimates- allow us to finely balance regret minimization and information gain on maximizing wind power output. We finally illustrate the methodology via simulations.

This paper is organized as follows. Section 4.2 introduces the wind evolution model, and problem formulation. Section 4.3 describes the model order reduction framework. Section 4.4 outlines the Kalman Filter representation for a near-persistent model and its state distribution under partial observation. Section 5.2 details an online Expectation-Maximization algorithm used to continuously learn the covariance matrix of the KF. Section 4.6 formulates the Information Directed Sampling controller. Finally, Section 4.7 discusses results and provides a comparative analysis.

4.2 Problem Formulation

Wind dynamics

Denote as $w(t,h)$ the wind *speed* at time t and altitude $h \in \bar{H} = (0,H)$; and $w_t \colon h \to w(t,h)$ is the wind *profile* belonging to $H = L^2(\bar{H},\mathbb{R})$ of square integrable functions. We assume the existence of a Partial Differential Equation governing the evolution of w_t,

$$\frac{\mathrm{d}w_t}{\mathrm{d}t} = \mathbf{f}(w_t, \boldsymbol{\lambda}) \tag{4.1}$$

The mis-specification of the law $\mathbf{f}, \boldsymbol{\lambda}$ governing the evolution is the first source of uncertainty: the *model error*.

Using Euler's forward method, finite elements and assuming a linear PDE, one can show the previous PDE reduces to a simple persistent dynamical system with state $x_t \in \mathbb{R}^p$ and random noise ϵ_t, where x_t are the coefficients of w_t projected onto a p-dimensional basis of $L^2(\bar{H},\mathbb{R})$ [150].

$$x_{t+1} = x_t + \epsilon_t$$

There exists an observation function $\phi : \mathbb{R} \to \mathbb{R}$ from which we can approximately recover the wind profile form, which further introduces a second source of uncertainty: the *observation error*.

$$w_t(h) = \phi(h)^\top x_t + \gamma_t(h)$$

Finally, due to the problem structure we only observe one wind speed at one altitude per time step. This defines the following *partially-observable* Markov model, which induces a third source of uncertainty, see Fig 4.1.

Since we do not observe the hidden vector state x_t, we must estimate its distribution over a probability space $(\Omega, \Sigma, \mathbb{P})$, where the uncertainty arises from the *state error*, *observation error* and *partial observability*.

By designing a proper controller, we can influence the *partial observability* to reduce the entropy of the state estimate's distribution.

AWE Controller

At time t, the decision maker sequentially chooses actions u_t from a finite action set $\bar{U}$ (of the accessible altitudes) and observes the corresponding outcomes (wind speed) $(\mathbf{Y}_{t,u_t})_{t\in\mathbb{N}}$. With our previously defined notations:

$$\mathbf{Y}_{t,u_t} = \mathbf{w}_t(u_t)$$

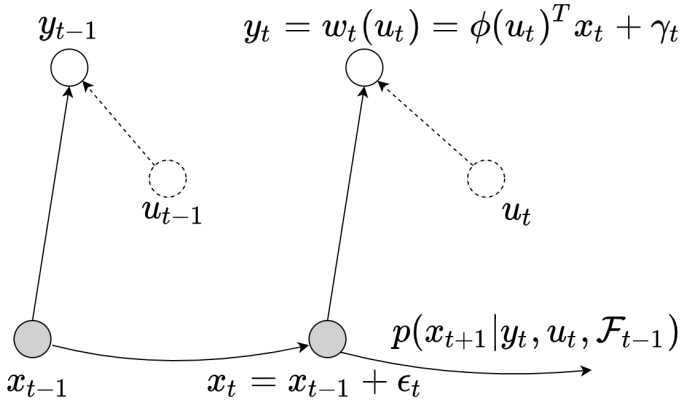


Figure 4.1: Hidden Markov Chain. The coefficient vector x_t is hidden, the observation y_t depends on the altitude set-point u_t. The transition density depends on the history $\mathcal{F}_t$.

The agent associates a reward $(y, u) \to R(y, u - u_{t-1})$ which is the wind power output for an outcome y and action u. The function is fixed and known [144] as,

$$R(y, \Delta u) = \underbrace{c_1 \cdot \min\left(y, V_{rated}\right)^3}_{\text{Generated energy}} - \underbrace{(c_2 y^2 + c_3 y^2 |\Delta u|)}_{\text{Energy lost}} \tag{4.2}$$

where $\Delta u = u - u_t$ is the adjustment with respect to the past altitude. The reward function accounts for both the energy generated or lost by adjusting altitude. In section 4.6, this will be seen as a penalty for exploring. Parameters c_1, c_2, c_3, V_{rated} are constants imposed by the manufacturer. Figure 4.9 shows the power output with regards to wind speed for different altitude changes Δu. It is nonlinear, non-convex and non-monotonic. Note, that we should not specifically track the greatest wind speeds, but rather wind speed regions close to the rated speed.

Uncertainty about the vector state $\mathbf{x_t}$ induces uncertainty about the true optimal action $\mathbf{u_t^*}$, which we denote by: $\mathbf{u_t^*} \in \text{argmax}_u(\mathbb{E}[\mathbf{R}_{t,u}|\mathbf{x_t}])$. Thus, we define the T–period regret of the sequence of actions $u_1, \cdots, u_T$ as the random variable:

$$\mathbf{Regret}(T) = \sum_{t=1}^{T} \mathbf{R}_{t,\mathbf{u_t^*}} - \mathbf{R}_{t,\mathbf{u_t}}$$

In this work we are interested in minimizing the Bayesian notion of regret $\mathbf{Regret}(T)$ conditioned to available information. It is important to stress that the expectation is taken over the randomness of action $\mathbf{u_t}$, over the system state $\mathbf{x_t}$, and outcomes $\mathbf{Y_t}$. Finally, we define the filtration (or history) of observations and actions as $\mathcal{F}_t = \sigma((\mathbf{Y_k}, \mathbf{U_k}) \mid k \leq t)$ which is a sigma algebra of $\mathcal{F}$. To simplify notations we denote: $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_t]$.

Our final objective is to select a sequence of policies $(\pi_t)_{t\in\mathbb{N}}$ over action space: $\pi(\cdot) := \mathbb{P}_t(u = \cdot)$ that minimizes $\mathbb{E}[\mathbf{Regret}(\pi, T)]$. This is addressed in Section 4.6.

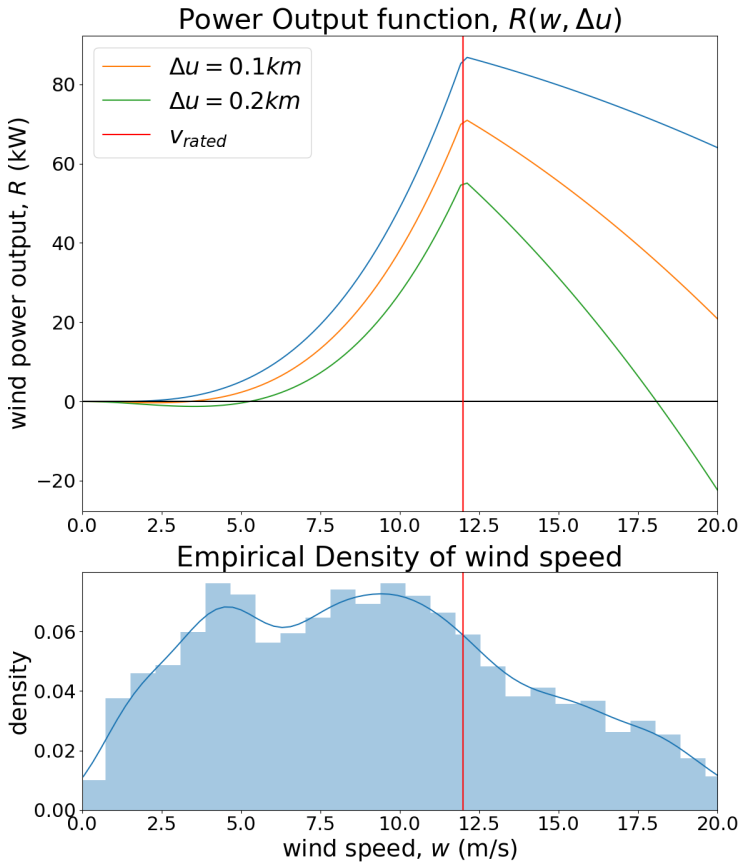


Figure 4.2: Power output function of wind-speed and altitude adjustment $\Delta u \in \{0, 0.1, 0.2\}$ & Empirical density of wind speed. High wind speed is likely and induce lower power output. Altitude adjustments Δu come with an increasing cost with wind speed (in $w^2 \cdot |\Delta u|$).

Offline setup

We consider the data set from [160], which was collected from an experimental campaign at Cape Henlopen State Park in Lewes, Delaware. In this work, the data used consists of wind speed measurements recorded by a 915-Mhz wind profiler between July 1, 2014 and August 31, 2014. The profiler records wind speeds in $\Delta h = 50$ meter altitude increments from 150 m to 1000 m, in 30 minute intervals.

The wind profile is sampled at constant altitudes $\mathcal{H} = \{i\Delta h\}_{i \leq n}$ such that $y_t \in \mathbb{R}^n$ with $n = 18$, and $(y_t)_i = w_t(i\Delta h)$. We use this dataset to both select the functional basis and set up the simulation environment to evaluate our online controller.

4.3 Model order Reduction

We aim to select p basis functions with $p < n$, $\Phi = (\phi_i)_{i \leq p}$ with $\phi_i \in \mathcal{L}^2(\bar{H}, \mathbb{R})$, so that the projected wind profile is accurate and its dynamics are simple to track. Practically, the choice of Φ will decompose w_t into a linear combination of elements of Φ multiplied by coefficient vector (the system state) $x \in \mathbb{R}^p$, i.e.

$$w_t(h) = \sum_{i=1}^{p} x_{t,i}\phi_i(h) + \gamma_t(h) \tag{4.3}$$

Fitting the coefficients is accomplished by least squares:

$$\begin{aligned} x_t &= \text{argmin}_x \sum_{j=1}^{n} \left(w_t(j\Delta h) - \sum_{i=1}^{p} x_i \phi_i(j\Delta h) \right)^2 \\ &= \text{argmin}_x (y_t - B(\phi)^\top x)^2 \end{aligned}$$

Hence, we can write that:

$$x_t = \left(B(\phi)^\top B(\phi)\right)^{-1} B(\phi) y_t, \text{ where } [B(\phi)]_{i,j} = \phi_i(j\Delta h)$$

In section 4.4, we will formalize a persistent dynamical system as a Linear Gauss-Markov model such that:

$$\begin{cases} x_{t+1} &= x_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, Q) \\ y_t &= B(\phi)^\top x_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, R) \end{cases} \tag{4.4}$$

Ideally, we seek a functional basis which (i) provides inertia, (ii) has good model accuracy and (iii) has good observability. Our methodology is the following: we restrict the family of functions to polynomials of degree smaller than n:

$$\Phi = (\Phi_i)_{i \leq n} = \left\{ \phi_i(h; \alpha) := \sum_{k=0}^{i} \alpha_i (H - h)^i; \ i < n \right\}$$

A high inertia system (i) evolves slowly with time i.e. $\mathbb{E}[(x_{t+1} - x_t)(x_{t+1} - x_t)^\top] = ||Q||_F$ is small. A projection is accurate (ii) if its reconstruction error is small: $\mathbb{E}[(y_t - y_t^r)(y_t - y_t^r)^\top] = ||R||_F$. Finally, the system has good observability (iii) if observing y gives high information content about the state. That is, the mutual information between X and Y is high: $I(X;Y) = H(Y) - H(Y|X)$. Both quantities can be estimated with the data and written as a function of R and $Cov(Y_i, Y_j)$.

Upon selecting a polynomial order $p = 4$, we obtain a covariance error matrix $\hat{R}$ such that $\gamma_t \sim \mathcal{N}(0, \hat{R})$. Note that we could build this approximation with some general notion of the wind profile and hence this offline step does not require collecting data in an actual test bed. We denote $B(\phi)$ as Φ in the following.

4.4 Kalman filter formulation of forecast model

Past research approaches have focused on developing forecast algorithm that can be learned over the history of sparse observations. However, even with full observation the task is difficult. The challenge in building a forecast model lies in the complex dynamics of the wind shear's physics. Consider Fig. 4.4. Note the average wind speed (over each profile) has jumpy behavior occurring at seemingly random times. This suggets weak seasonality and heteroskedastic variance. Finally, the partial auto-correlation analysis would lead us to select a model with two past-time steps as a auto-regressive input. However, we cannot access, nor accurately estimate the past wind profiles from partial observations. Even if a complete and accurate forecasting model was devised, then it would not be immediately useful in an online setting because one needs to infer the full profile's distribution before forecasting the next time steps. Let alone fitting a covariance matrix over the process noise which in itself is a difficult task. However, one would need the past profile which can only be estimated using learning and estimating a fast-moving profile under sparse observations (1 altitude per time step) is difficulty tractable. Hence the use of Gaussian Processes to reduce the dimensions of the problem in past approach. (1.5*3*18*18) always more than $1.5 * 3n^2$ parameters.

The challenge in building a forecast model lies in the complex dynamics of the wind shear's physics. This can be highlighted by plotting the standard deviation of the profile which is heteroskedastic (see Fig. 4.4).

An alternative to a complex-to-fit and estimate-dependent forecasting model is to formulate a persistent model following (4.4). Persistent models are known to perform relatively well [156]. In this case, the computational effort involves estimating the current profile and the required prior knowledge of the model is minimal. Note that the persistent state equation (4.4) is linear in the state dynamics, as is the observation process. A significant difference with a plain vanilla Kalman Filter is that the observation function is directly controllable since it relates to the altitude.

$$\begin{cases} x_{t+1} & = x_t + \epsilon_t \\ y_t & = \phi(u_t)^\top x_t + \nu_t \end{cases} \tag{4.5}$$

Recall the central assumptions of the Kalman Filter (KF):

1. The sequence of random variables $x_0, \omega_0, \nu_0, \cdots, \omega_T, \nu_T$ is jointly Gaussian and mutually independent,

2. Noise ϵ_t is independent and normally distributed with $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{E}[\epsilon_t \epsilon_t^\top] = Q_t$. Note that we do not assume Q is time-invariant since we later develop an online adaptive learning algorithm for Q,

3. Noise ν_t is independent, identically and normally distributed with $\mathbb{E}[\nu_t] = 0$ and $\mathbb{E}[\nu_t \nu_t^\top] = \phi(u_t)^\top R \phi(u_t)$. Matrix R can be estimated offline using the procedure explained in Section 4.3, and corresponds to the distribution of approximation error.

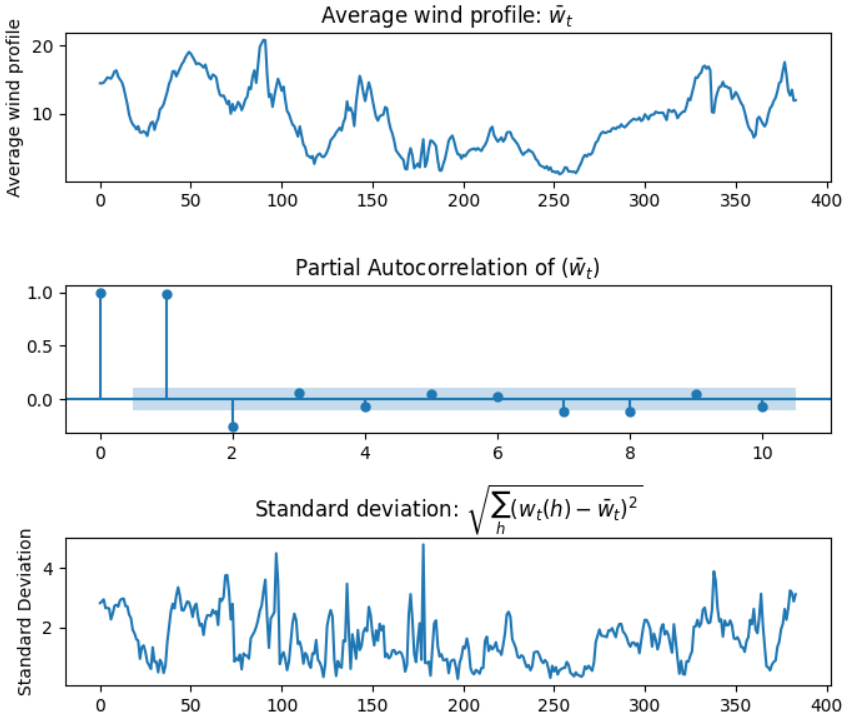


Figure 4.3: Subplot 1: Wind Profile average presents some jumpy behaviour and no clear seasonality. Subplot 2: Partial Auto-correlation function: in full-observability a Vectorial Auto-Regressive model with 2 lags would be adapted. Subplot 3: Heteroskedastic variance
Even with full-observability, learning an accurate forecasting model is challenging due to non-stationnarity of the mean and variance of the profile

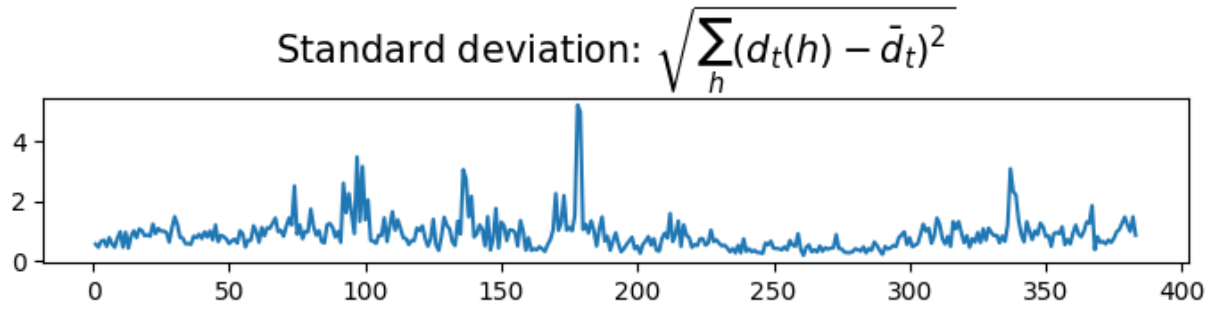


Figure 4.4: Wind Profile Evolution $d_t = w_t - w_{t-1}$ presents a Heteroskedastic variance

The objective of a filter is to obtain the state density x_T given past observations up to a certain time $\mathcal{F}_t$. Since all random quantities are Gaussian and the state and observation equations are linear, the filter optimally blends new information by simply applying a lin-

ear feedback to the prior state. We encourage interested readers to consult [54] for more information. The recursive KF equations are summarized as follows:

1. The state x_t is normally distributed and given by density $p_t(x_t)$, defined by its first two moments: $x_{t|t} = \mathbb{E}_t[x_t]$ and $\Sigma_{t|t} = \mathbb{E}_t[(x_t - x_{t|t})(x_t - x_{t|t})^T]$.

2. **Prediction Step** Given $x_{t|t}$, $\Sigma_{t|t}$, the forward (or prediction) equations apply model dynamics in (4.5):

$$\begin{aligned} x_{t+1|t} &= \mathbb{E}_t[x_{t+1|t}] = x_{t|t} \\ \Sigma_{t+1|t} &= \Sigma_{t|t} + Q_t \end{aligned}$$

3. **Analysis Step** To ease notation denote ϕ_{t+1} as $\phi(u_{t+1})$. Given $x_{t+1|t}$, $\Sigma_{t+1|t}$, note that $\mathcal{F}_{t+1} = \{y_{t+1}\} \cup \mathcal{F}_t$ and hence using Bayes formula $x_{t+1}|\mathcal{F}_{t+1} = (x_{t+1}|\mathcal{F}_t)|(y_{t+1}|\mathcal{F}_t)$. We know:

$$\begin{bmatrix} x_{t+1|t} \\ y_{t+1|t} \end{bmatrix} = \mathcal{N}\Bigg(\begin{bmatrix} x_{t|t} \\ \phi_{t+1}^\top x_{t|t} \end{bmatrix}, \\ \begin{bmatrix} \Sigma_{t+1|t} & \Sigma_{t+1|t}\phi_{t+1} \\ \phi_{t+1}^\top \Sigma_{t+1|t} & \phi_{t+1}^\top(\Sigma_{t+1|t} + R)\phi_{t+1} \end{bmatrix} \Bigg)$$

   Setting the Kalman Gain as:

$$L_t = \Sigma_{t+1|t}\phi_{t+1}\left[\phi_{t+1}^\top(\Sigma_{t+1|t} + R)\phi_{t+1}\right]^{-1}$$

   After collecting y_{t+1}, the measurement update reads:

$$\begin{aligned} x_{t+1|t+1} &= x_{t|t} + L_t\left(y_{t+1} - \phi(u_{t+1})^\top x_{t|t}\right) \\ \Sigma_{t+1|t+1} &= \Sigma_{t+1|t} - L_t\phi(u_{t+1})^\top \Sigma_{t+1|t} \end{aligned}$$

The measurement update indicates how the density of the state estimate will change given the next altitude u_{t+1}. Importantly, note that the covariance update depends on u_{t+1}. We can leverage this structure to select the altitude that decreases the state entropy at the optimum altitude. This particular fact will be used to develop our controller.

Given A and Q, the Kalman recursion equations yield $x_{t|t} = \mathbb{E}_t[x_t]$ and $\Sigma_{t|t} = \mathbb{E}_t[(x_t - x_{t|t})(x_t - x_{t|t})^T]$.

For completion, we state the Kalman Filter equations underneath:

$$\boxed{\begin{aligned} \epsilon_k &= y_k - \hat{y}_{k|k-1} \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k\epsilon_k \\ P_{k|k} &= P_{k|k-1} - K_k R_{k|k-1} K_k^T \\ K_k &= P_{k|k-1}C^T R_{k|k-1}^{-1} \\ R_{k|k-1} &= C P_{k|k-1} C^T \end{aligned}}$$

4.5 Open-Loop online estimation of parameters

Parametric Process Covariance

As illustrated in Fig. 4.4, the process covariance Q_t should be considered as time-varying and hence needs to be learned. To address, we propose to (i) continuously estimate the covariance over a sample of s past observations, and (ii) reduce the number of free parameters to decrease the estimate's variance. For (ii), we consider a geo-physical structure for the process w_t. This follows previous work on this matter [154]. Denote $d_t = w_t - w_{t-1}$ as the process noise of the functional wind speed process. We can then assume that d_t is a Gaussian process with respect to space, i.e.,

$$d_t(z) \sim \mathcal{GP}(0, k(z, z'))$$

where $k(z, z') = \sigma_0^2 \exp(-\frac{1}{2}\frac{(z-z'))^2}{\sigma_1^2}) + \sigma_2^2$ is its covariance function. We chose the Squared Exponential (SE) kernel plus a constant kernel for its overall quality of fit (see Fig. 4.6). The only notable difference between the sample and kernel-based covariance is the higher uncertainty that reigns at the upper altitude layers. Effectively, it boils down to parameterize the covariance matrix by three parameters $\theta = (\sigma_0^2, \sigma_1^2, \sigma_2^2)$. In particular $d_t(h_i)$ has covariance K_θ where $[K_\theta]_{i,j} = k(h_i, h_j)$. Hence, after projection Q_t is itself parametrized by θ:

$$Q_{t,\theta} = \Phi^\top K_\theta \Phi$$

Estimating the covariance matrix offline by relying on the assumption of the process covariance time-invariance can end up in formulating an ill-conditioned Kalman Filter. This is know to drastically reduce the quality of the estimate of the KF and even lead it to diverge incorrect values of the noise[161][162].

On Figure 4.5, we estimate the sample covariance matrix using over a sliding window of 150 time periods so that the estimation can be deemed decent as the concentration ratio is small enough $\frac{parameters}{sample} = \frac{6}{4*150} < 1/100$ (section 3 of [163]) and plot its 2-norm which corresponds to its spectral radius on Figure 4.5. We see that the maximum eigenvalue of the covariance matrix is drastically decreasing (from 180 until the 100 to 70 around time period 200). Highlighting the necessity for an adaptive update of the Q matrix.

Fixed Horizon Expectation Maximization Algorithm

Our state space model is a Hidden Markov Chain. The log likelihood of a sequence of state and observation from $t = 0$ to $t = T$ conveniently has a separable form. We also note that

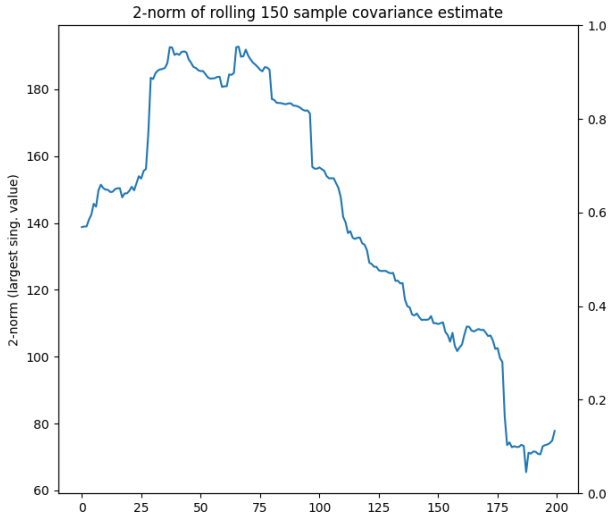


Figure 4.5: 2-norm of the sample process covariance matrix $\hat{Q}_t$ over a window of 150 samples. $\hat{Q}_t$ is not time-invariant.

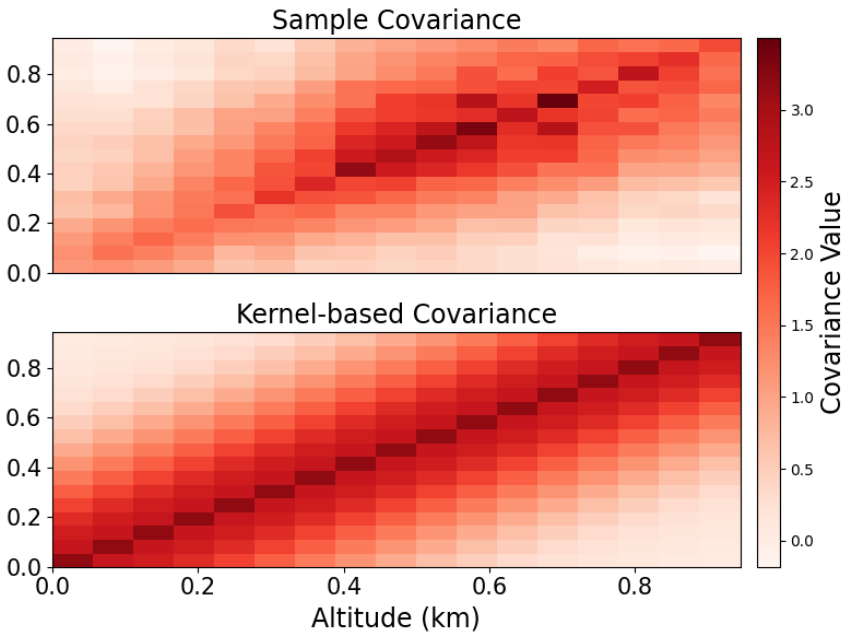


Figure 4.6: Sample $\hat{K}$ and Kernel-based covariance $K_{\hat{\theta}}$. The 3-parameters kernel is a reasonable estimation.

only the transitions from x_{t-1} to x_t depend on θ.

$$p(x_{0:T}, y_{0:T}; \theta) = p(x_0) \prod_{t=1}^{T} p(x_t|x_{t-1}; \theta) p(y_t|x_t)$$

$$\log p(x_{0:T}, y_{0:T}; \theta) = \sum_{t=1}^{T} \log \big(p(x_k|x_{k-1}; \theta)\big) + C$$

One traditionally estimates the covariance matrix of a Kalman Filter offline through an expected maximization (EM) type of algorithm. Given a set of observations, the algorithm iteratively generates a sequence of parameter estimates (θ^c) such that the log-likelihood of the observations is non-decreasing over the iterations. Let θ' be a parameter that we assume governs the evolution of x. Since $L(\theta)$ is not random, and thus $\mathbb{E}_{\theta'}[L(\theta)] = L(\theta)$, we have:

$$\begin{aligned} L(\theta) &= \log p_\theta(y_{1:T}) = \log p_\theta(y_{1:T}, x_{1:T}) - \log p_\theta(y_{1:T}|x_{1:T}) \\ \mathbb{E}_{\theta'}[L(\theta)] &= \mathbb{E}_{\theta'}[\log p_\theta(y_{1:T}, x_{1:T})] - \mathbb{E}_{\theta'}[\log p_\theta(y_{1:T}|x_{1:T})] \\ L(\theta) &= Q(\theta;\theta') - H(\theta,\theta') \end{aligned}$$

Iteratively increasing $\theta^{(c+1)} = \text{argmax}_\theta\ Q(\theta, \theta^{(c)})$ leads to increasing $(L(\theta^{(c)}))$ (see [164]). Hence the EM algorithm converges to a local maxima of the log-likelihood of the model. We now specify the expectation step and drop the conditions $\theta', y_{1:T}$ on the expectation, for ease of reading.

$$\begin{aligned} Q_T(\theta,\theta') &= \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_T[\log p(x_t|x_{t-1};\theta)] \\ &= -\frac{1}{2T}\sum_{t=1}^{T} \underbrace{\mathbb{E}_T[\epsilon_t^\top Q_\theta^{-1}\epsilon_t]}_{c_t(\theta)} - \frac{1}{2}\log\det(Q_\theta) \end{aligned}$$

Using the Rauch–Tung–Striebel smoother, we can determine: $x_{t|T}$, $\epsilon_{t|T}$, $\Sigma_{t|T}$, $\Sigma^t_{t-1|T} = \mathbb{E}_T[\epsilon_t\epsilon_t^\top]$, and

$$c_t(\theta) = \epsilon_{t|T}^\top Q_\theta^{-1}\epsilon_{t|T} + \text{Tr}\big(Q_\theta^{-1}(\Sigma_{t|T} + \Sigma_{t-1|T} - 2\cdot\Sigma^t_{t-1|T})\big)$$

Interested readers can refer to [165]. Practically, we use $s = 50$ past samples and the Scikit-Learn Gaussian Process kernels to minimize the log-likelihood using the box-contrained Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS-B) to return our sequence of estimates, $(\theta^*_{t-s:t})_t$.

$$\theta^*_{T-s:T} = \text{argmin}_\theta\ -\frac{1}{2s}\sum_{t=T-s}^{T} c_t(\theta) - \frac{1}{2}\log\det(Q_\theta)$$

4.6 Closed-loop Information Directed Sampling

From $x_{t|t}, \Sigma_{t|t}$, we can obtain a non-stationary distribution of the wind-profile at altitude u by applying:

$$w_t(u)|\mathcal{F}_t \sim \mathcal{N}(\phi(u)^\top x_{t|t}\ ,\ \phi(u)^\top\Sigma_{t|t}\phi(u))$$

Our objective is to select a sequence of policies $(\pi_t)_{t\in\mathbb{N}}$ that minimizes $\mathbb{E}[\mathbf{Regret}(\pi, T)]$. We consider policies from a set of binomial distributions, with n denoting the number of altitude bins and parameters $p_i = (\frac{i}{n})_{i\leq n}$ so that, for each altitude bin i, the distribution is centered on $i = p_i \cdot n$,

$$D = \Big(Binom(p_i, n) \Big)_{i\leq n} \quad \text{(Set of Randomized Policies)}$$

Let $\Delta_t(u) = \mathbb{E}_t[R_{t,U^*} - R_{t,u}]$ be the regret of taking action u and $g_t(u) := I_t(U^*; Y_{t,u})$ be the information gain about the optimum. $\mathbf{U}^*$ is the optimal altitude (and random) for gross power output, i.e. $R(y, 0)$ in (4.2). Note $\mathbf{U}^*$ is independent of altitude adjustment Δu. Its distribution is:

$$\alpha_t(u) = \mathbb{P}_t(U^* = u) = \mathbb{P}_t\big(\cap_{a\neq u} \{R(\mathbf{Y}_{t,u}, 0) \geq R(\mathbf{Y}_{t,a}, 0)\}\big)$$

$g_t(u)$ is also known as the mutual information between the wind speed at u and the maximum $\mathbf{U}^*$. The information directed sampling algorithm selects the probability which minimizes the information ratio:

$$\pi_t^{IDS} = \text{argmin}_{\pi\in D}\left\{ \Psi_t(\pi) = \frac{\Delta_t(\pi)^2}{g_t(\pi)} \right\} \quad \text{(IDS Policy)} \tag{4.6}$$

The controller minimizes the squared regret incurred per-bit of information acquired about the optimum [151]. This policy has sub-linear regret growth, and further work could devise a precise bound for our specific persistent model.

Estimation of Regret

As mentioned in the introduction, the distribution of the power output $R(\mathbf{Y})$ even under normally distributed wind speed has no closed-form distribution. We must rely on Monte Carlo simulations to estimate the regret. Hence, we take M $(x^m)_{m\leq M}$ samples from $\mathcal{N}(x_{t|t}, \Sigma_{t|t})$ and record their rewards (power output) $(r^m)_m$ given the last altitude $u_{t-1} \in \mathcal{F}_t$ and net optimizer u_m^*.

$$\begin{aligned}\Delta_t(u) &= \mathbb{E}_t[R_{t,U^*}] - \mathbb{E}_t[R(Y_{t,u}, u - u_{t-1})] \\ &\approx \frac{1}{M}\sum_{m=1}^{M} R(y^m_{t,u^*_m}, u^*_m - u_{t-1}) - \mathbb{E}_t[R(Y_{t,u}, u - u_{t-1})]\end{aligned}$$

The second term is a tractable integral with respect to a one-dimensional normal density.

Estimation of the information gain function

We choose U^*, the *gross optimum*, as the optimal altitude for gross power $R(y, \Delta u = 0)$ instead of net power $R(y, \Delta u = u - u_{t-1})$. This resolves the following issue. IDS optimizes

an instantaneous objective, which would penalize exploration via the third term in (4.2) without considering the potential long-term reward. Using gross power $R(y,0)$ side-steps this issue. The information gain is defined as the expected reduction in entropy $H(\alpha)$ of the posterior distribution of U^* for observing $Y_{t,u}$:

$$g_t(u) = \mathbb{E}_t\big[H(\alpha_t) - H(\alpha_{t+1})|u_t = u\big]$$

To estimate the first and second term, we again use Monte Carlo simulations to approximate α_t. Then, as explained in Section 4.4, taking action u will update the covariance function: $\Sigma_{t+1|t+1}(u) = \Sigma_{t+1|t} - L_t\phi(u)^\top\Sigma_{t+1|t}$. Hence, we can sample M scenarios from $\mathcal{N}(x_{t|t}, \Sigma_{t+1|t+1}(u))$ and approximate $\alpha_{t+1}(u)$.

$$\alpha_{t/t+1}(u) \approx \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}(u = \text{argmax}\, r_m^{t/t+1})$$

Finally, we select a policy π from D such that the information ratio is minimized: $\min_{\pi\in D}(\pi^\top\Delta_t)^2/(\pi^\top g_t)$.

4.7 Results

To illustrate the KF-IDS controller, we simulate different baseline controllers over the same wind profile and compare various performance measures. As presented in Section 4.2, the data comes from Cape Henlopen for days ranging from July 7, 2014 (time index 0) to July 11 (time index 96). The values for the power output function can be found in [144]:

Symbol	Value	Symbol	Value
h_{min}	0.15 km	Δt	30 min
h_{max}	1 km	c_1	0.0579 kW s^3/m^3
r_{max}	0.01 km/min	c_2	0.09 kW s^2/m^2
v_r	12 m s^{-1}	c_3	1.08 kW s^2/m^2 ·km

Table 4.1: AWE Model Parameters

Baseline controllers

The most fundamental performance metric to evaluate our controller is power generation. As in previous works, we also consider the following baseline controllers:

1. *Omniscient Dynamic Programming (Omniscient)*: We recursively solve the DP equations yielding the trajectory that maximizes overall power generation. This provides the upper bound of energy generation.

2. *Omniscient Fixed Altitude (Optimal Fixed)*: The optimal fixed altitude.

3. *Lowest Fixed Altitude (Fixed 100m)*: This baseline represents conventional wind turbines and helps us measure the incremental gains of AWE systems.

We additionally consider the following controllers. For notations see Section 4.6.

4. *Regret Minimization (Greedy)*: We sample from the distribution π that minimizes expected squared regret $(\pi^\top \Delta)^2$.

5. *Information Gain Maximization (Info.)*: We sample from the distribution π that maximizes expected information gain $\pi^\top g$.

6. *Information-Directed Sampling (IDS)*: We sample from the distribution π that maximizes the information ratio $(\pi^\top \Delta)^2/\pi^\top g$.

Discussion

In the following simulations, we assume wind speed is measured at ground level (this is inexpensive), in addition to measuring wind speed at the AWE's controlled altitude.

Qualitative Analysis - Profile of the trajectories

Sampled trajectories for 5 policies are displayed in Fig. 4.7. The IDS and Greedy trajectories are similar. High wind speed periods (e.g. between 24 and 60 hours) have relatively fixed trajectories. This results from the objective function structure: the penalty for adjusting scales with squared wind speed. Hence, the squared regret for exploring (adjusting) will be dissuasive. On the other hand, during lower wind speed periods the incentive to explore becomes stronger (e.g. 60 - 85 hours). A penalty-unaware controller, such as the Information controller, constantly adjusts altitude to improve its knowledge of the wind power optimum (see Fig. 4.10).

Qualitative Analysis - KF Estimation Performance

In Fig. 4.8, the actual wind profile (upper subplot) is visually compared to the recovered profile from the KF estimates, formally defined as the sequence $(\Phi^\top x_{t|t})_t$ (lower subplot). When an IDS controller is used, the filter catches most of the wind profile patterns except when it is static (e.g. between 24 to 60 hours). During these periods, the IDS controller sacrifices recovery of the actual wind speed profile (exploration) for performance (exploitation).

Figure 4.9 shows a snapshot of the actual wind profile and the KF-estimated profile $(\Phi^\top x_{t|t}) \pm$ one standard deviation under IDS control. The estimate is accurate and tight near the AWE's position. The probability distribution for the gross power α_t is shown to the left. The probability is high near the gross optima's position. This illustrates how KF-IDS control captures the wind profile trend to guide altitude adjustments.

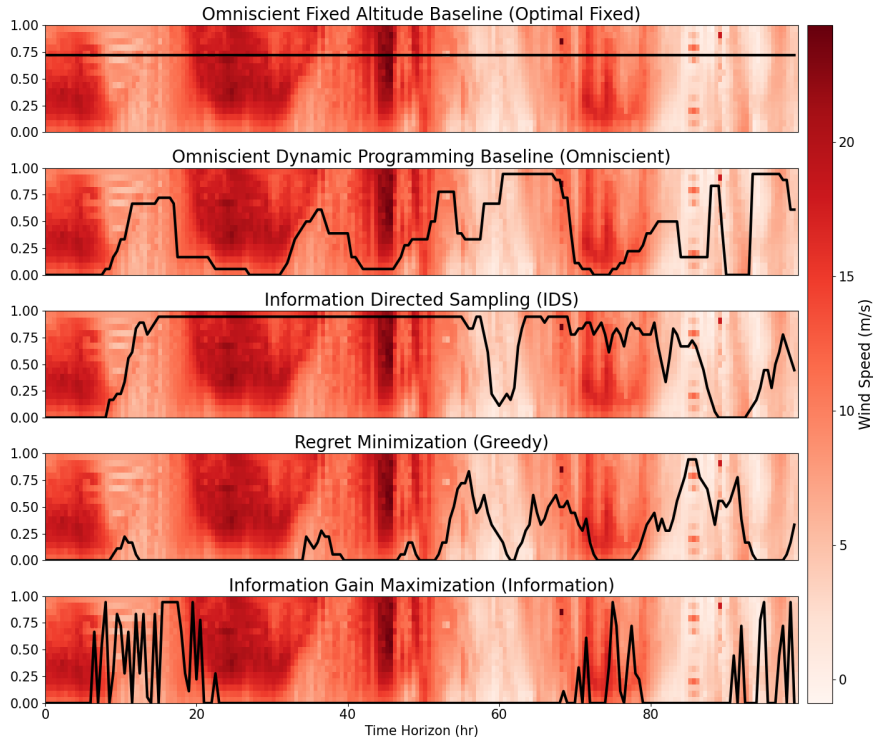


Figure 4.7: Samples of controller trajectories for 5 different policies over the same wind-profile. IDS and Greedy presents some qualitative similarities due to the large penalty associated with adjusting altitudes with high-speed wind.

Qualitative Analysis - Cumulative Output Curves

Figure 4.10 presents cumulative power output curves. We conclude that the IDS controller results in superior net energy production for this wind profile scenario. Note the fixed 100m benchmark still provides about 75% of the omniscient controller's total energy output. Intelligent altitude control closes this gap, but to varying degrees.

Quantitative Analysis - filter and policies performance

The previous analyses are limited to sampled roll-outs and specific wind profiles. By reporting other performance measures in Table 4.2, we illustrate that IDS can be superior to a fixed or regret-based policy in the long-term since its knowledge of the wind profile is enhanced. We examine 4 relevant performance measures: the first two relate to filter accuracy, whereas the last two relate to IDS performance.

A well-conditioned KF is paramount: its mean and covariance are fed to the controller objective function. One way to verify the KF is to track the average reconstruction error (see Section 4.3). Additionally, the average filtered state entropy, defined as $\frac{1}{T}\sum_t ||\Sigma_{t|t}||_F$,

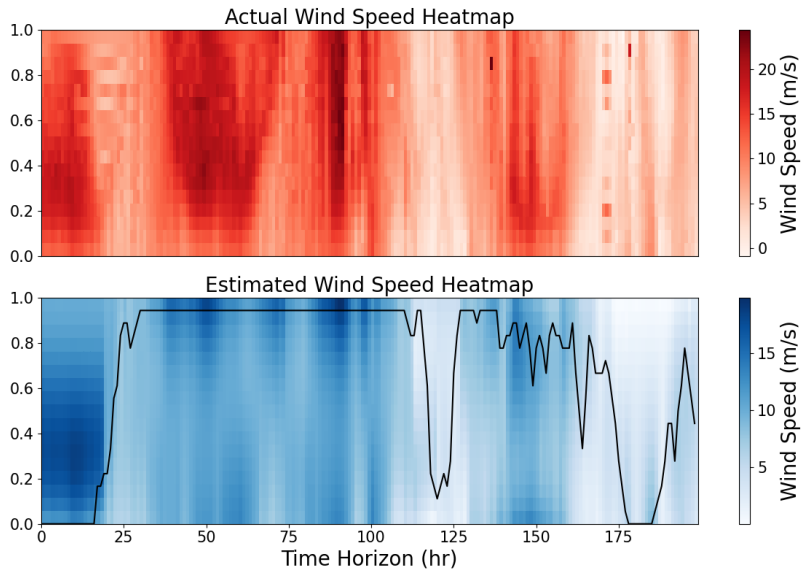


Figure 4.8: Comparison of actual wind-profile (top) and KF estimates (below) when using an IDS controller. The KF successfully tracks the main trend, except when the cost to explore is overwhelming (between 40 and 60 hours).

reveals the effectiveness of exploration in the policy. As a result, the information controller excels – without surprise – in this task and the IDS controller follows in second.

The IDS controller seeks to improve its knowledge of the maximum output distribution α_t. The average likelihood of the gross optimum, $\frac{1}{T}\sum_t \alpha_t(u_t^*)$, measures the efficiency of the controller's exploration component. Finally, the output average regret: $\frac{1}{T}\sum_t Regret(\pi_t, t)$ is the objective we ultimately seek to minimize. These last two performance measures illustrate the superiority of the IDS controller.

4.8 Conclusion and Further Work

This work focuses on identifying and leveraging a specific structure of the wind profile evolution to design an Information-Directed Sampling controller. The policy blends exploitation and exploration into a single objective function. We projected the wind profile dynamics onto a polynomial basis to maximize power output while ensuring the reconstruction loss stays acceptable. Under partial observability, the Kalman Filter's recursive equations yield the current and future system's state distribution given the next time step's altitude adjustment. One can then precisely estimate how exploring an altitude reveals information about the optimum power output. An online learning algorithm was also implemented to account for the variability of the wind profile conditions, modeled as heteroskedasticity of its variance.

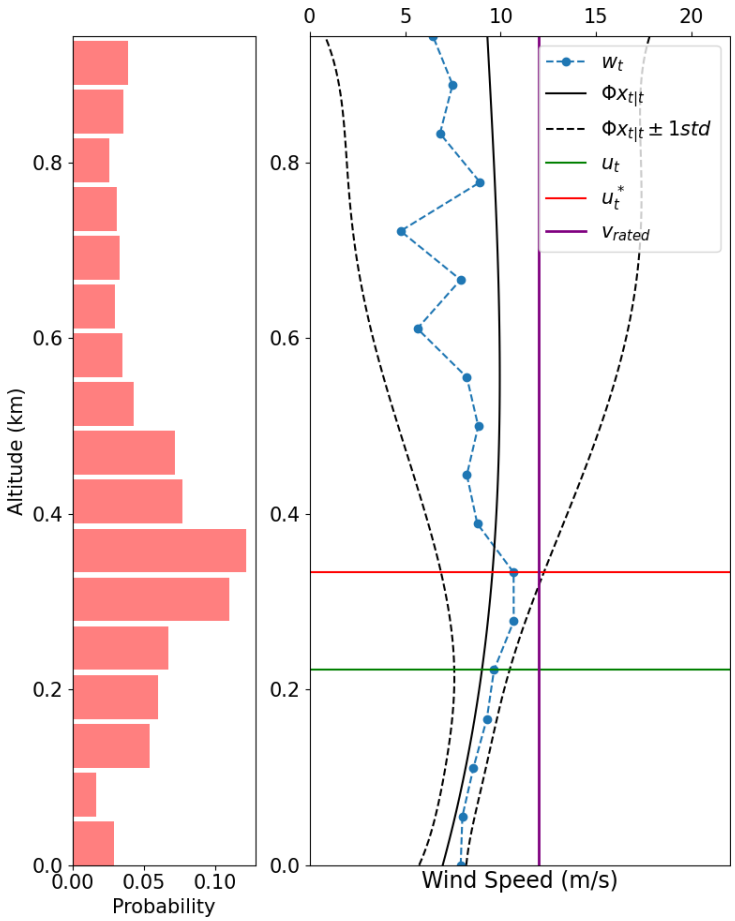


Figure 4.9: [Right] Inferred profile from the KF estimate $\Phi^\top x_{t|t}$ when using the IDS controller vs. the actual wind profile w_t. [Left] Probability distribution of the gross power optimum α_t. A properly controlled KF captures the general trend and outputs a correct distribution for the gross power.

We showed two important results based on real-world wind profile data: (i) The filter state estimates are reasonably accurate, even under sparse observations and with weak assumptions on the underlying evolution dynamics. (ii) Kalman Filtered-Information Directed Sampling improves net energy generation relative to other controllers.

Future work can include the following: (i) Quantify the performance loss for using a persistent model versus a more complex forecast model. (ii) Synthesize statistically accurate scenarios of wind speed profiles to rank controllers and evaluate the learning algorithm for the time-varying covariance matrix. (iii) Relax the KF hypotheses and consider alternative nonlinear filters. (iv) Leverage the stochastic process structure to derive tighter bounds on regret growth.

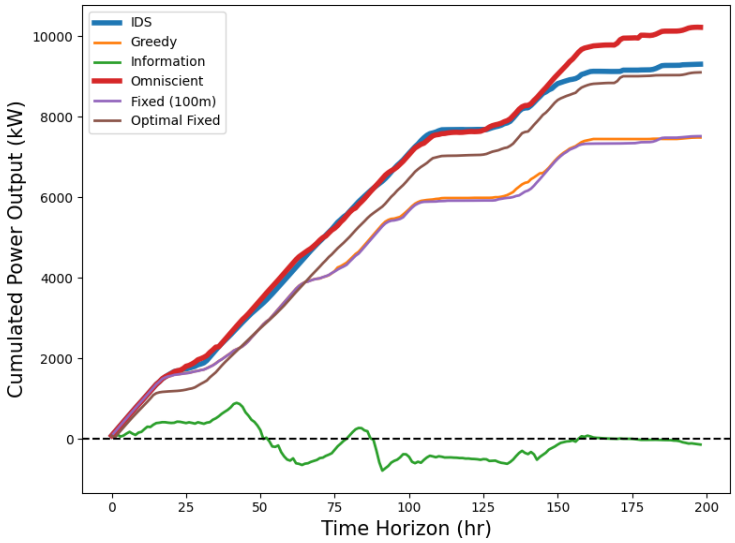


Figure 4.10: Cumulative power output for 6 different controllers applied to a sample rollout. IDS provides the greatest total energy output among all causal controllers.

	Selected Baselines		KF Policies		
	DP	100m	IDS	Greedy	Info.
Reconstruction Error (m/s)	2.67	4.91	2.64	3.18	**1.67**
Froebenious Norm State Entropy	5.80	9.48	5.70	6.65	**4.34**
Likelihood of Maximum		0.00	**0.16**	0.14	0.09
Output Average Regret (kW)		13.65	**4.62**	13.82	52.32

Table 4.2: Selected performance measures for different controllers. IDS presents a good tradeoff between observability of the profile (low reconstruction error, high likelihood of the maximum) and performance (least average regret)

Chapter 5

Mape_Maker: A Scenario Creator

We describe algorithms for creating probabilistic scenarios for renewables power production. Our approach allows for tailoring of forecast uncertainty, such that scenarios can be constructed to capture the situation where the underlying forecast methodology is more (or less) accurate than it has been historically. Such scenarios can be used in studies that extend into the future and may need to consider the possibility that forecast technology will improve. Our approach can also be used to generate alternative realizations of renewable energy production that are consistent with historical forecast accuracy, in effect serving as a method for creating families of realistic alternatives – which are often critical in simulation-based analysis methodologies. We illustrate the methods using real data for day-ahead wind forecasts.

Nomenclature

Observed variables

$x_t \in \mathbb{R}^n$ Timeseries of independent input data (e.g. actuals)
$y_t \in \mathbb{R}^n$ Timeseries of dependent input data (e.g. forecasts)
$\mathcal{X} \in \mathbb{R}^{n,2}$ Set of paired input data (actuals, forecasts) or (forecasts, actuals)
$x_t^{SID} \in \mathbb{R}^n$ Timeseries of simulation input data
$\mathcal{X}_{SID} \in \mathbb{R}^{n,2}$ Set of Simulation Input Data (SID) upon which the simulation is performed
$\tilde{r} \in \mathbb{R}$ target MARE (Mean Absolute Relative Error)

Random variables

$\tilde{\boldsymbol{\varepsilon}}_t \in \mathbb{R}^n$ Random vector of simulated errors
$\tilde{\boldsymbol{y}}_t \in \mathbb{R}^n$ Random vector of simulated values
$\tilde{\boldsymbol{u}}_t \in \mathbb{R}^n$ Random vector of uniform base process
$\boldsymbol{\mathcal{E}} \in \mathbb{R}^n$ Random variable of the error
$\tilde{\boldsymbol{\mathcal{E}}} \in \mathbb{R}^n$ Random variable of the simulated error
$\mathbf{X} \in \mathbb{R}^n$ Random variable of the input
$\mathbf{Z} \in \mathbb{R}^n$ Joint random variable : $Z = (\boldsymbol{\mathcal{E}}, \mathbf{X})$
f_X Marginal density function of the input data $\mathcal{X}$
f_ε Marginal density function of the error random variable
$f_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$ Conditional density function of the error given the input
$F_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$ Cumulative distribution function of the error given the input

Estimation

$a \in \mathbb{R}$ Percent of data used to estimate each conditional distribution
I_x^a Interval of $2a$ fraction of data around x in $\mathcal{X}$
$\bar{x}(x, a) \in \mathbb{R}$ Center of the interval I_x^a
$cap \in \mathbb{R}$ Capacity
$b(\cdot; \alpha, \beta, l, s)$ Density function of a beta for parameters : α, β, l, s
$\hat{\mathcal{S}}_x$ Set of estimated beta parameters of the conditional distributions $\boldsymbol{\mathcal{E}}|\mathbf{X} = x$ over $\mathcal{X}$
$\hat{m}(x) \in \mathbb{R}$ Expected value of the absolute estimated error given the input
$\tilde{m}(x, \tilde{r}, \omega) \in \mathbb{R}$ Expected value of the absolute simulated error given the input, a target MARE, and a weight function
$\omega_{\mathcal{X}}(\cdot)$ Weight function over $\mathcal{X}$
$\hat{F}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$ Cumulative distribution function of the estimated error given the input
$\hat{f}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$ Estimated conditional density function of the error given the input in $\mathcal{X}$

$m_{max}(x) \in \mathbb{R}$ Maximum value of the expected value of the absolute estimated error, given x
$r_{\hat{m}} \in \mathbb{R}$ Expected value of the mean absolute relative estimated error over $\mathcal{X}$
$\hat{r} \in \mathbb{R}$ Mean absolute relative error over $\mathcal{X}$ under the estimated conditional distributions
Simulation Distributions
$\tilde{F}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$ Cumulative distribution function of the simulated error given the input
$\tilde{f}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$ Simulation conditional density function of the error given the input in $\mathcal{X}_{SID}$
$\tilde{\mathcal{S}}_{x,m}$ Set of simulation beta parameters of the conditional distributions $\boldsymbol{\mathcal{E}}|\mathbf{X} = x$ over the SID
$\omega_{\mathcal{X}_{SID}}(\cdot)$ Weight function over $\mathcal{X}_{SID}$
P_{SID} Distribution plausibility score
Base Process
$\tilde{\boldsymbol{z}}_t \in \mathbb{R}^n$ Random vector of base process
ϕ Cdf of the standard normal distribution
$(\hat{z}_i)_{i\leq n} \in \mathbb{R}^n$ Estimated base process
$(a_i, b_i)_{i\leq p} \in \mathbb{R}^p$ ARMA parameters of the estimated base process
$\sigma_\delta \in \mathbb{R}$ Standard deviation of the error of the estimated base process
Simulation
$d \in \mathbb{R}$ Mean of the curvature of target input data

5.1 Introduction

Uncertainty associated with the forecasted production of renewable energy sources such as wind and solar mandates power systems analysis and management techniques that directly take stochastics into account. A growing literature describes methods for creating and evaluating probabilistic *scenarios*, which are forecasts of renewables power production with an attached probability. A representative sample of this literature can be found in [166, 167, 168, 169, 170, 171]. Here, we are interested in creating probabilistic scenarios for the situation when the underlying forecast methodology is modeled as being more (or less) accurate than it has been historically, e.g., to capture potential improvements in underlying numerical weather prediction models. Such scenarios can be used in studies that consider future power system conditions and configurations, and therefore may need to consider the possibility that forecast technology will improve. Our approach can also be used to construct alternative realizations of actual renewable energy production that are consistent with the accuracy of a particular forecasting methodology, in providing a method for creating families of realistic alternatives – which are often critical in simulation-based analysis methodologies. A general open-source software implementation of the methods described here – a package called *mape_maker* – is publicly available at `https://github.com/mape-maker/mape-maker`. While we focus our studies and exposition on renewables power production (specifically because the associated forecasting errors are significantly higher than for load), our methods and software operate on generic time series quantities of interest.

Forecasts of renewables power production are commonly available at periodic intervals,

with each forecast specifying a time series for a future system operating horizon, e.g., day-ahead for each hour of the next seven days. Corresponding time series for *ex post* production are available once the system operates through the forecasted operating horizon; we refer to the latter time series simply as *actuals*. Our methods and software are based on the availability of a historical (e.g, over the past year) set of such forecasts and associated actuals, which would generally be available to any system operator.

Given such a historical data set, we create a set of renewables power production scenarios that could reasonably correspond to a forecast (a current forecast or a forecast from the past). We refer to these scenarios as "actuals" to distinguish them from forecasts, although they are synthetic alternatives to the realized renewables power production time series. We can also create a set of forecasts that could reasonably correspond to a given time series of actuals. In other words, the process can be inverted. The correspondence between forecasts and actuals is based on analysis of historical forecast error distributions. Subsequently, the word "reasonably" is replaced with mathematical criteria concerning the error distribution, temporal correlation, and in the case of forecasts, curvature.

As a preview of the output of this capability, consider Figure 5.1. This figure provides a simple example where a set of 5 alternative synthetic scenarios for wind power production are constructed for several days in July of 2013 for the California Independent System Operator (CAISO) in the US. The results are based on day-ahead wind forecast error data obtained from CAISO for July 2013 through May 2015. The target fit error – specifically, the mean absolute percentage error or MAPE – is the quantity realized in the historical forecast error data. Because the scenarios are created for days in the past, we are able to show both the historical forecasts and realized actuals on the same plot as the synthetic actual scenarios.

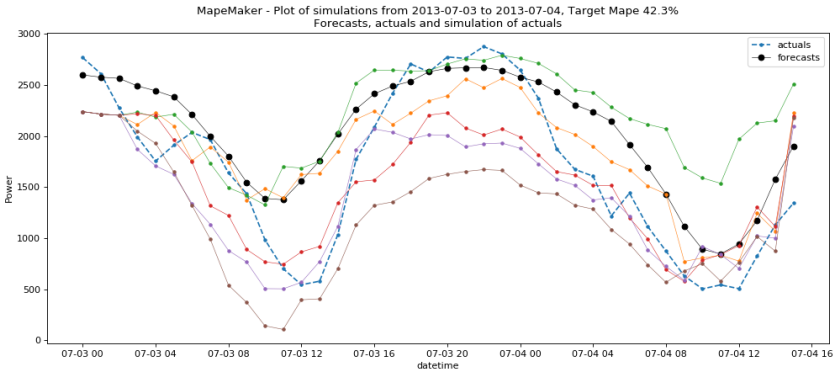


Figure 5.1: Illustration of 5 scenarios of wind production in CAISO for several days in July 2013, each representing a synthetic realization. The historical forecasts and realized actuals are also shown.

Our method represents, to the best of our knowledge, a first-of-kind capability for constructing either synthetic forecasts or realizations of renewables power production that are

consistent with arbitrary forecast precision. Such a capability enables a range of novel analyses for power systems operations models that have to date – while of significant interest and potential impact to both practitioners and researchers – not been possible. In general, we view our methods as having the potential to significantly enhance the utility of production cost models (PCMs), which are the key analytic tool used by industry for power systems planning analysis.

Related Literature

The present work builds on prior research efforts conducted by the authors to construct probabilistic scenarios for wind power production [172], solar power production [171], and energy demand [173]. These works in turn are either related to or build on approaches for constructing probabilistic scenarios (largely in renewables power production), either to provide situational awareness regarding uncertainty to power system operators or to serve as input to advanced stochastic power systems operations optimization models.

One widespread approach to constructing probabilistic wind power scenarios involves fitting models using historically observed wind power production characteristics. For example, Morales *et al.* propose a methodology based on a time series analysis of historical wind power production, while also maintaining spatial correlation across distinct wind farms [174]. Such approaches are limited, as they do not incorporate information concerning short-term forecasts of resource availability, which provide the best information about near-term conditions. Early attempts to construct probabilistic wind power production scenarios from forecasts were straightforward. For example, Wang *et al.* assume wind power production is normally distributed around a forecasted quantity [175]. Pinson *et al.* proposed a greatly improved method based on multivariate Gaussian random variables, estimated with the covariance matrix of prediction errors [176]. The methods outlined in [172] and [171] are designed to improve upon the Pinson *et al.* method, focusing on developing non-parametric forecast error distributions and controlled methods for generating low-probability scenarios. In [177], a time series based auto regressive moving average model is simulated to be transformed into actual wind speeds by distribution transformation. The wind speed scenarios are converted to wind power scenarios using a power curve. Unlike many methods and ours, their method does not consider an external forecast.

Associated with the above body of literature is research devoted to developing methods and metrics to assess the quality of the resulting scenarios. The vast majority of these approaches strictly consider static attributes of scenarios, in contrast to evaluating the scenarios in terms of their benefit in a decision context (e.g., power system operations). Well-known and widely studied examples include the energy score, the Brier Score, and Minimum Spanning Tree (MST) rank histograms [169]. More recent work in this area, intended to improve discrimination ability - such as the Variogram score, is reported in [178, 170]. In this paper, we introduce new criteria for scenarios and evaluate our methods using them.

Existing research on constructing probabilistic scenarios for renewables power production does not consider the ability to either (1) vary forecast errors in a controlled manner or (2)

rigorously generate synthetic realizations that are consistent with a given forecast error target. There is also no literature concerned with creating forecast scenarios that plausibly correspond to observed power. This paper aims to fill these gaps.

Measures of Forecast Error

Let $(x_i)_i \in \mathbb{R}^n$ and $(y_i)_i \in \mathbb{R}^n$ denote two time series. For simplicity, we subsequently refer to these time series as x_t and y_t. We then define the following functions:

$$\begin{array}{rcccc} RE & : & \mathbb{R}^* \times \mathbb{R} & \to & \mathbb{R} \\ & & x, y & \mapsto & \frac{y-x}{x} \end{array} \quad \text{(Relative Error)}$$

$$\begin{array}{rcccc} MARE & : & \mathbb{R}^{*n} \times \mathbb{R}^n & \to & \mathbb{R}_+ \\ & & x_t, y_t & \mapsto & \sum_{i=1}^n \frac{|RE(x_i,y_i)|}{n} \end{array} \quad \text{(Mean Absolute Relative Error)}$$

The MAPE (Mean Absolute Percentage Error) is simply the MARE (Mean Absolute Relative Error) given as a percentage. Our software library communicates with users in terms of MAPE, but in our discussions here it is convenient to use MARE and sometimes MAE (Mean Absolute Error) variants.

While MAPE is a very popular way of characterizing forecast accuracy for renewables production, it is well-known to have a number of undesirable properties (see, e.g., [179]). One undesirable property is that x values of zero must be ignored in the calculation. We have organized our methods in such a way as to avoid division by zero. Most of the development here is based on converting the MAPE target to an absolute error conditional on the value of x, so it would be a relatively straightforward extension to convert our algorithms to use some measure of accuracy other than the MAPE.

Notation Scheme

We use x_t as the time series of input data that will be treated as the independent and y_t as the time series of input data that will be modeled as dependent. These names make sense as our goal is ultimately to obtain a simulation time series $\tilde{y}_t$, given an input time series x_t^{SID}. Note that which of these pairs is the forecast and which is the actual depends on the user's choice, i.e., what is being simulated.

Denote $\mathcal{X} = \{(x_{t,i}, y_{t,i}), 1 \leq i \leq n\}$, the set of each paired occurrence. We use x_t^{SID}, $\mathcal{X}_{SID}$ (resp.), to denote the Simulation Input Data (SID) time series, set (resp.), upon which the simulation of $\tilde{y}$ will be performed.

While the time series notation (x_t, y_t, x_t^{SID}) is going to come in handy when discussing the base process procedure (see Section 5.4), the set notation ($\mathcal{X}$, $\mathcal{X}_{SID}$) is going to be useful to discuss about the conditional distributions and obtain useful guarantees on the mean absolute relative error (MARE) of our simulations (see Section 5.2).

We use bold upper case font to denote random variables and bold lower case font to denote random vectors. We will focus on generating a plausible distribution of simulated errors $\tilde{\boldsymbol{\varepsilon}}_t$ in the following. The title of the paper and the name of our software library derives from the requirement that simulated values $\tilde{\mathbf{y}}_t = x_t^{SID} + \tilde{\boldsymbol{\varepsilon}}_t$ must result in a MARE close enough to the target MARE. We formalize this constraint as

$$\mathbb{E}[MARE(x_t^{SID}, x_t^{SID} + \tilde{\boldsymbol{\varepsilon}}_t)] = \tilde{r},$$

where $\tilde{r}$ is the target MARE, the target MAPE divided by 100%.

Plausibility Objectives

A main theme underlying this work that we will use to justify some of our design choices involves what we refer to as *plausibility objectives*. For any requested MARE, the distribution of errors computed should be as close as possible to the original error distribution while satisfying the target MARE. If a user were to select the estimated MARE as the requested one, one would naturally expect the distribution of errors drawn from the simulated distributions to be somehow "close" to the estimated distribution. For example, if the system of forecasts is producing a wide range of errors at very low forecasted power output, then even if the forecast technology is improving, one would expect it to still produce a relatively wider range of errors at low power regardless of the requested MAPE. We formalize these objectives as follows in Definition 5.1.1.

Definition 5.1.1. A scenario set is said to be *plausible* if:

1. the shape of the error distribution for the scenarios is close to the shape of the empirical distribution of errors, i.e, its distribution plausibility score is close to 1 (as defined in later in Section 5.3);

2. the computed autocorrelation coefficients for the set are close the empirical values; and

3. the computed curvature for the set is close to the empirical value, especially when the scenarios are forecasts (because we observe that forecasts typically have lower curvature than actuals.)

We discuss ways to evaluate the plausibility of a scenario set and provide illustrations in Section 5.5.

5.2 Modeling the Joint Distribution of $(\boldsymbol{\mathcal{E}}, \mathbf{X})$

The symbol $\boldsymbol{\mathcal{E}}$ represents the random variable of the error between the input x and the corresponding dependent input y. In this section, we model the fact that the error depends on the input x and does not have the property of 'white noise'. Therefore, we must account

for the fact that x is also a random variable that we will note $\mathbf{X}$. To model their joint distribution, we define the random variable : $\mathbf{Z} = (\boldsymbol{\mathcal{E}}, \mathbf{X})$.

Here, $\mathbf{Z}$ denotes a random variable with values in $(-\infty, +\infty) \times (0, +\infty)$ – or, if the production capacity *cap* is known by the forecaster, values in $[-cap, cap] \times [0, cap]$. We denote by $f_{\mathbf{Z}}$ the density of $\mathbf{Z}$, and denote by $f_{\boldsymbol{\mathcal{E}}}$ and $f_{\mathbf{X}}$ the marginals of $f_{\mathbf{Z}}$. Then,

$$f_{\boldsymbol{\mathcal{E}}}(\varepsilon) = \int_{-\infty}^{\infty} f_{\mathbf{Z}}(\varepsilon, x)dx, \qquad f_{\mathbf{X}}(x) = \int_{-\infty}^{\infty} f_{\mathbf{Z}}(\varepsilon, x)d\varepsilon$$

We also define the conditional density of $\boldsymbol{\mathcal{E}}$ given $\mathbf{X} = x$ as:

$$f_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}(\varepsilon) = \frac{f(\varepsilon, x)}{f_{\mathbf{X}}(x)}$$

Modeling the conditional distribution of errors is important as it can vary significantly with the value of input data. For example, when the forecasts and the actuals are both low, the errors will be biased because the power produced cannot be below zero (in settings where power consumed by wind farms is both forecast and reported, the lower bound could be negative, but we use zero without loss of generality). Symmetrically, close to the maximum capacity, *cap*, errors are bounded by the fact that power cannot exceed maximum production capacity.

In this context, we introduce the functional $m(x)$ to denote the expected value of the absolute error of the distribution conditioned on x, defined as:

$$m(x) = \mathbb{E}[|\boldsymbol{\mathcal{E}}| \, | \mathbf{X} = x] = \int_{\varepsilon=-\infty}^{\infty} |\varepsilon| f_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}(\varepsilon)d\varepsilon$$

We then introduce r to denote the mean absolute relative error, defined as:

$$r = \mathbb{E}[\mathbb{E}[\frac{|\boldsymbol{\mathcal{E}}| \, | \mathbf{X}}{\mathbf{X}}]] = \mathbb{E}[\frac{m(\mathbf{X})}{\mathbf{X}}]$$

In Figure 5.2, we provide an illustrative visualization of the distribution of the relative error *RE* with respect to the actuals. We note that because the actuals are correlated with the forecasts, the figure would be very similar if the forecasts were used in instead. The data is for CAISO wind power data, ranging from July 1, 2013 to June 30, 2015. We will use this dataset for illustration throughout the paper, and refer to it informally as the *CAISO Wind* data set. These data are available in the mape_maker software distribution; the file is `wind_total_forecast_actual_070113_063015.csv`.

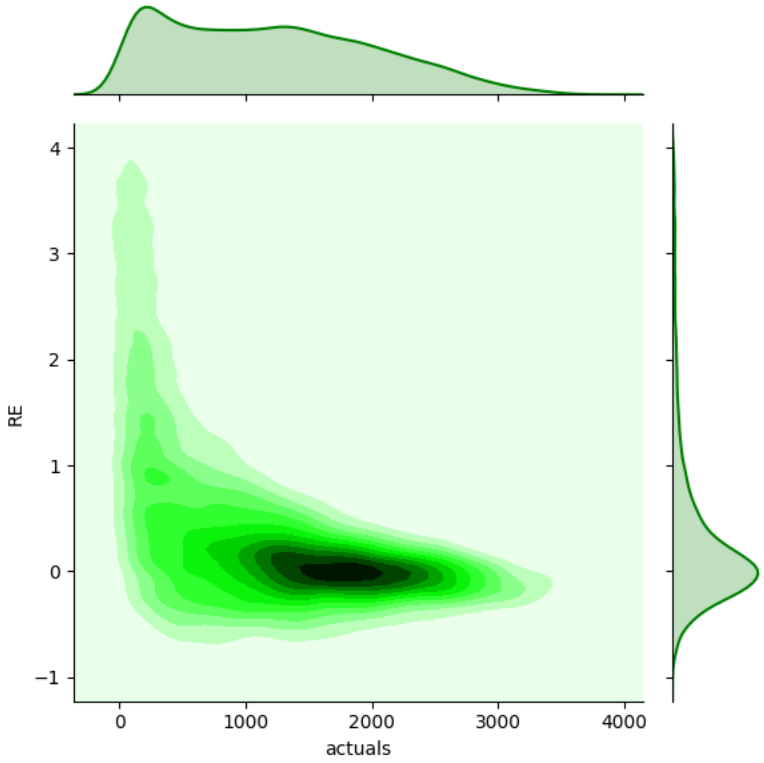


Figure 5.2: Empirical joint distribution of $(\frac{\boldsymbol{\mathcal{E}}}{\mathbf{X}}, \mathbf{X})$ - CAISO Wind Power

Estimating the Conditional Distribution of $\boldsymbol{\mathcal{E}}|\mathbf{X}$, $\hat{f}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$

In [180], it was shown that an appropriate probability density function for the wind power forecast error is the beta distribution. Besides, its variable kurtosis and its bounds make it more suitable than the Normal distribution to our end. Finally, its simplicity (only two shape parameters) is useful for the estimation.
Given the notation $x \in \mathcal{X}$, we use the beta distribution on $[l, s+l]$ to model $f_{\boldsymbol{\mathcal{E}}|X=x}$. In addition to the l and s that we will refer to as location parameters, a beta distribution requires two additional parameters – α and β, i.e., the shape parameters. We then define

$$f_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}(\varepsilon) = beta(\varepsilon;\ (\alpha, \beta, l, s)) = \frac{(\frac{\varepsilon - l}{s})^{\alpha-1}(1 - \frac{\varepsilon - l}{s})^{\beta-1}}{B(\alpha, \beta)}$$

with

$$B(\alpha, \beta) = \int_{\varepsilon=l}^{l+s} (\frac{\varepsilon - l}{s})^{\alpha-1}(1 - \frac{\varepsilon - l}{s})^{\beta-1} d\varepsilon$$

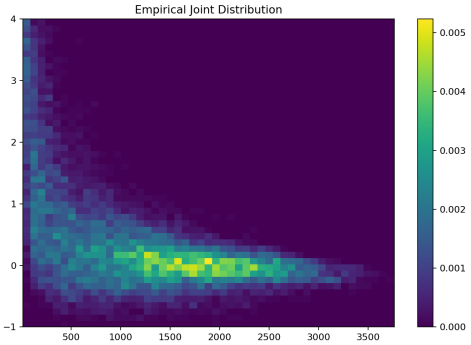


Figure 5.3: Empirical joint distribution of CAISO

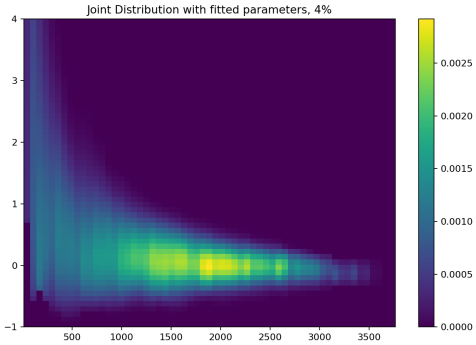


Figure 5.4: Simulated joint distribution with $a = 4\%$ of CAISO

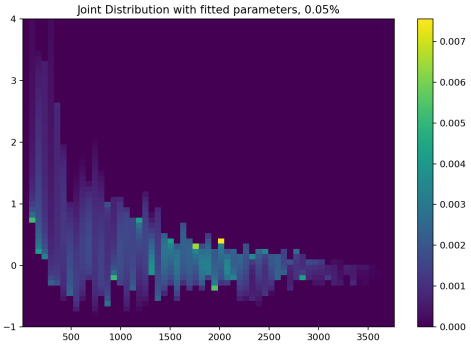


Figure 5.5: Simulated joint distribution with $a = 0.5\%$

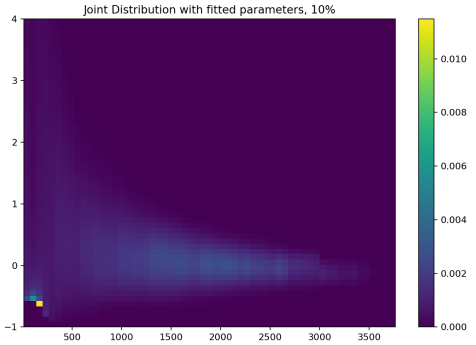


Figure 5.6: Simulated joint distribution with $a = 10\%$

Intervals for Conditional Estimation

We estimate the parameters of the conditional density for each x of the input dataset, $\mathcal{X}$, using a a fraction a (e.g., 0.05; see Section 5.2) of data before and after each x. Let G_X denote the empirical cumulative distribution function.
Then, let $I_x^a = [G_X^{-1}(G_X(x) - a),\ G_X^{-1}(G_X(x) + a)]$. Thus I_x^a is centered on

$$\bar{x}(x;\ a) = \frac{G_X^{-1}(G_X(x) - a) + G_X^{-1}(G_X(x) + a)}{2}$$

with $2a$ fraction of the data. We fit the shape parameters of a beta distribution on the sample $E_{I_x^a} = \{\varepsilon_i,\ 1 \leq i \leq n, x_i \in I_x^a, \}$. Note that for production values near zero and near the capacity, there could be as few as a fraction of the values used.

To compute an estimate for a particular value x', our method uses the interval I_x^a for which $\bar{x}(x; a)$ is closest to x' and finds the corresponding set $E_{I_x^a}$ to compute the parameters

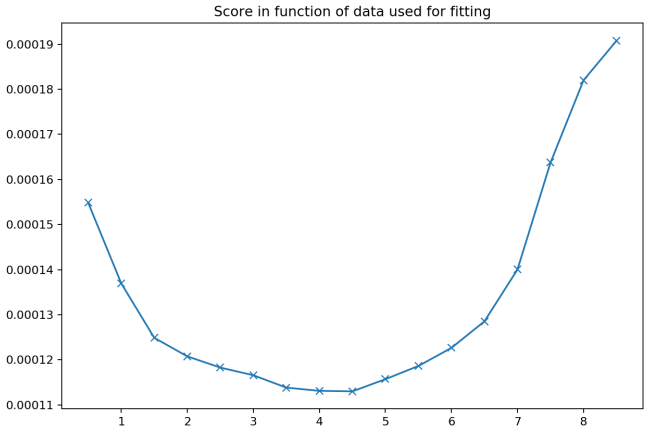


Figure 5.7: D^2 score for a% of data selected. We choose 4 percent.

for x'. For every $x' \in \mathcal{X}$ that is not close to zero or *cap*, the closest $\bar{x}(x;a)$ to x' will often be just $\bar{x}(x';a)$. However, for very small or large values of x' and when $\mathcal{X}_{SID} \not\subset \mathcal{X}$, the use of the interval with the closest mean is most appropriate.

We will now describe how our method fits the parameters of the beta distributions. Because every estimated quantity will depend on a, we drop a as a subscript or function parameter for notational simplicity. Figures 5.3-5.6 and finally 5.7 highlight how a can be selected.

Fixing l, s and Estimating α, β

Constraints on the Location Parameters

An informed choice of the location parameters will avoid simulating errors leading to y values lower than 0 or greater then the *cap* of the dataset. We now define the function y_{max}, which returns the maximum possible simulated value at x according to a conditional distribution $f_{\boldsymbol{\varepsilon}|\mathbf{X}=x}$. Because the inverse of the corresponding cumulative distribution function (CDF) evaluated at one, $F^{-1}_{\boldsymbol{\varepsilon}|\mathbf{X}=x}(1) = l + s$, is the maximum of the error simulated; $F^{-1}_{\boldsymbol{\varepsilon}|\mathbf{X}=x}(0) = l$ is the minimum; and because we want to avoid simulating values above the cap or below zero we have

$$\begin{aligned} y_{max}(x) &= x + F^{-1}_{\boldsymbol{\varepsilon}|\mathbf{X}=x}(1) \\ &= x + s + l \\ &\leq cap. \end{aligned}$$

Similarly,

$$\begin{aligned} y_{min}(x) &= x + F^{-1}_{\boldsymbol{\varepsilon}|\mathbf{X}=x}(0) \\ &= x + l \\ &\geq 0. \end{aligned}$$

These two conditions give

$$\begin{aligned} l &\geq -x \\ s &\leq cap - x - l. \end{aligned}$$

Thus, we can define the estimators of the location parameters for each x as:

$$\hat{l}(x) = \begin{cases} -x & \text{if } \min\left(\varepsilon_i,\ x_i \in I_x\right) \leq -x \\ \min\left(\varepsilon_i,\ x_i \in I_x\right) & \text{else} \end{cases}$$

$$\hat{s}(x) = \begin{cases} cap - x - \hat{l}(x) & \text{if } \max\left(\varepsilon_i,\ x_i \in I_x\right) \geq cap - x \\ \max\left(\varepsilon_i,\ x_i \in I_x\right) - \hat{l}(x) & \text{else} \end{cases}$$

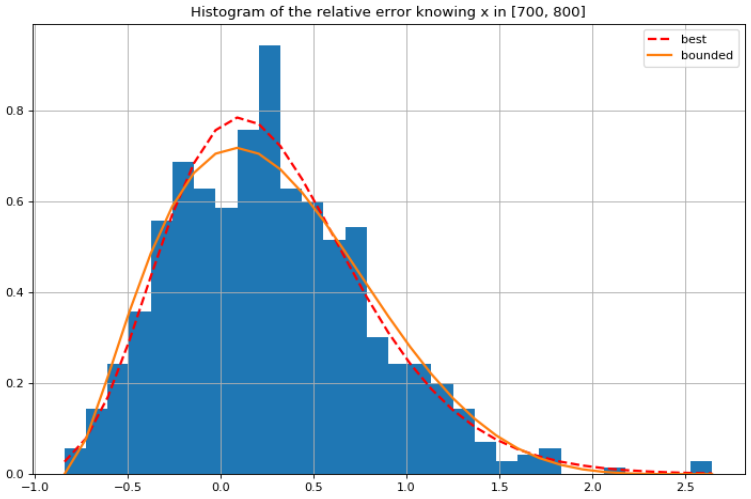


Figure 5.8: Approximation of the conditional density $f_{\boldsymbol{\varepsilon}|x=750}$ by a beta distribution- CAISO

Choosing the Shape Parameters by the Method of Moments

The mean and variance of a beta(α, β, l, s) distribution are:

$$\begin{aligned} \mu &= \frac{s\alpha}{\beta + \alpha} + l \\ V &= \frac{1}{s^2} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{aligned}$$

We can now choose shape parameters by solving these two equations for α and β

$$\hat{\mu}(x) = \frac{\hat{s}(x)\alpha}{\beta + \alpha} + \hat{l}(x)$$
$$\hat{V}(x) = \frac{1}{\hat{s}(x)^2} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

to obtain $\hat{\alpha}(x)$ and $\hat{\beta}(x)$.

For any $x \in \mathcal{X} \cup \mathcal{X}_{SID}$ assign

$$\hat{\mathcal{S}}_x = (\hat{\alpha}(x), \hat{\beta}(x), \hat{l}(x), \hat{s}(x))$$

Selecting a

We now develop an empirical way to select the best a. If a is small, the sample on which to fit the distribution will be small since I_x^a is small. Fitting a distribution on very little data is of course dangerous. On the other hand, if a is large, then the sample is too large to provide us with an estimation of the conditional density. In the extreme, where $a = 1$, every conditional density will be equal to the density of the relative error.

We select a using a least squares fit based on a discrepancy score between the empirical distribution function and the one obtained by estimating each conditional distributions with $2a$ of the data. Let g be the empirical joint density of (X, ε). Let $\hat{f}$ be the joint density of $(\mathbf{X}, \boldsymbol{\mathcal{E}})$ taken as $\hat{f}_a(x, \varepsilon) = \hat{f}_{\mathbf{X}}(x) * \hat{f}^a_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}(\varepsilon)$. We choose a to minimize the deviation between the real density and the simulated density:

$$D^2(a) = \int_x \int_\varepsilon (g(x, \varepsilon) - \hat{f}_a(x, \varepsilon))^2 d\varepsilon dx.$$

5.3 Adjusting Conditional Densities to Fit a Target MARE

In order to satisfy the plausibility of objectives, we need to adjust the conditional densities so as to retain their shape and the relative contribution to the error across values of x.

- While $\boldsymbol{\mathcal{E}}$ denotes the random variable describing the error with properties that can be estimated from $\mathcal{X}$, $\tilde{\boldsymbol{\mathcal{E}}}$ denotes the simulated error as a random variable defined by a distribution that we will develop with desired properties for the simulation over $\mathcal{X}_{SID}$.
- In the previous section, we defined the population density, $f_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$, and the estimated density, $\hat{f}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$. In this section, we introduce a third conditional distribution: the simulation density $\tilde{f}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$.

The simulation density distribution of $\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{X}$ will be chosen to verify the plausibility objectives while ensuring that the expected relative absolute error of the simulated random variable $\tilde{\boldsymbol{\mathcal{E}}}$ matches the desired MARE $\tilde{r}$:

$$\mathbb{E}_{\tilde{\boldsymbol{\mathcal{E}}}}[\frac{1}{n_{SID}}\sum_{x\in\mathcal{X}_{SID}}\frac{|\tilde{\boldsymbol{\mathcal{E}}}|\,|\mathbf{X}=x}{x}] = \tilde{r}.$$

Adjusting the Location Parameters to Fit a Target MAE

We want to adjust each conditional distribution such that the global distribution of $\tilde{\boldsymbol{\mathcal{E}}}$ satisfies the target MARE and maintains the same shape parameters as the original distributions. Toward this objective, we compute the mean absolute error of a beta distribution when α and β are fixed. First, let $l < 0$ and $s + l > 0$. Then, let $b(\cdot;\alpha,\beta,l,s)$ be an arbitrary beta density function with parameters (α,β,l,s) for which we define a mean absolute error function of l and s given values for α and β as

$$\nu(l,s;\alpha,\beta) = \int_{\varepsilon=l}^{s+l}|\varepsilon|b(\varepsilon;\alpha,\beta,l,s)d\varepsilon.$$

We then observe:

$$\lim_{s\to 0}\nu(l,s;\alpha,\beta) = 0, \quad \forall l<0$$

$$\nu(l,s;\alpha,\beta) \underset{s\to\infty}{\sim} \frac{s\alpha}{\alpha+\beta}$$

Since ν is continuous (it is a sum of continuous functions), the intermediate value theorem applies which means that $\nu(l,s;\alpha,\beta)$ can achieve any value and in particular, the value needed in order to hit the specified error target.

Thus, once we are given α, β, and a target value for the absolute error at a particular value of x, we need to find the intersection between a hyperplane defined by the target and the surface defined by $\nu(l,s;\alpha,\beta)$ to establish values for $\tilde{l}$ and $\tilde{s}$. For $x\in\mathbb{R}_+$ we will want to choose the solution that minimizes the distance to the estimated values $\hat{l}(x)$ and $\hat{s}(x)$ while hitting a target mean absolute error $m(x)$ and without changing the shape parameters.

This could be accomplished using

$$\begin{aligned}(\tilde{l}(x),\tilde{s}(x)) = \arg\min_{l,s} \quad & (l-\hat{l}(x))^2 + (s-\hat{s}(x))^2 \\ \text{s.t.} \quad & l\in\mathbb{R},\ s\in\mathbb{R}_+ \\ & 0\geq l\geq -x \\ & 0\leq s\leq cap - x - l \\ & \nu(l,s;\hat{\alpha}(x),\hat{\beta}(x)) = m(x)\end{aligned} \tag{5.1}$$

but requiring no deviation from $m(x)$ can cause computational issues. Thus, for expediency, we instead use in our experiments the following formulation that yields very similar results:

$$\begin{aligned}(\tilde{l}(x), \tilde{s}(x)) = \arg\min_{l,s} \quad & \Big(\nu(l, s; \hat{\alpha}(x), \hat{\beta}(x)) - m(x)\Big)^2 + \left(\frac{l - \hat{l}(x)}{o(x)}\right)^2 + \left(\frac{s - \hat{s}(x)}{o(x)}\right)^2 \\ \text{s.t.} \quad & 0 \geq l \geq -x \\ & 0 \leq s \leq cap - x.\end{aligned} \tag{5.2}$$

In this formulation, we employ $o(x)$ as a heuristic to scale the deviation for l and s to be less important than the deviation for m using, for example, $o(x) = \max(|\hat{l}(x)|, |\hat{s}(x)|, m(x))$.

Since there are bound constraints on l and s (see section 5.2), ν cannot hit every target $m(x)$. We compute a maximum target function that can be hit as:

$$m_{max}(x) = \max_{l \in (-x, 0], s \in [0, cap - x)} \nu(l, s; \hat{\alpha}(x), \hat{\beta}(x)).$$

The target function m must then be bounded for every x by:

$$m(x) \leq m_{max}(x). \tag{5.3}$$

Given a mean absolute error target function m satisfying inequality (5.3) we obtain for any x, a beta distribution of parameters $\tilde{\mathcal{S}}_{x,m} = (\hat{\alpha}(x), \hat{\beta}(x), \tilde{l}(x), \tilde{s}(x))$ that satisfies the mean absolute error target and that is the closest possible to the estimated distribution. We now proceed to allocate an error target to each $x \in \mathcal{X}_{SID}$ that we will call $\tilde{m}$ and that depends on the target MARE and on a weight function.

The following figures show the iterative steps to adjusting a given beta distribution to match a given target (m=0.43) as represented on figure 5.9. Figure 5.10, show how the location and scale parameters of the beta distribution would adjust to match 3 different target absolute relative error. Finally, figure 5.11 plots the resulting distributions for those 3 target AREs.

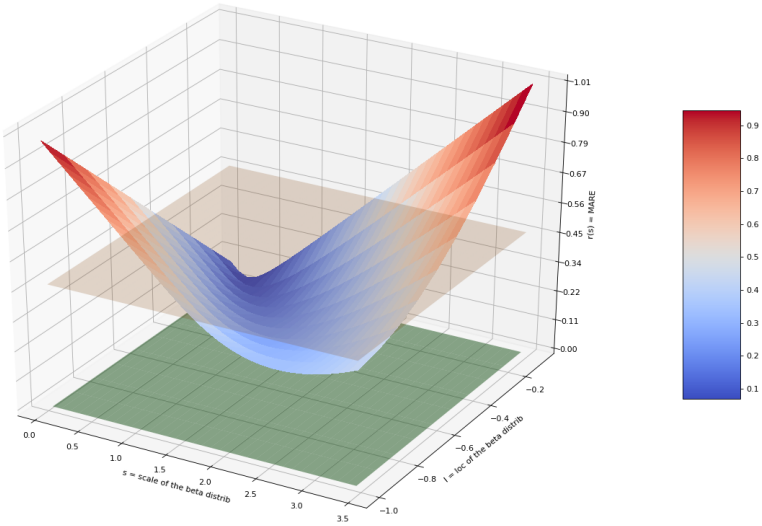


Figure 5.9: $\nu(l, s; \alpha, \beta)$, the green hyperplane is 0, and in rose is the target m=0.43

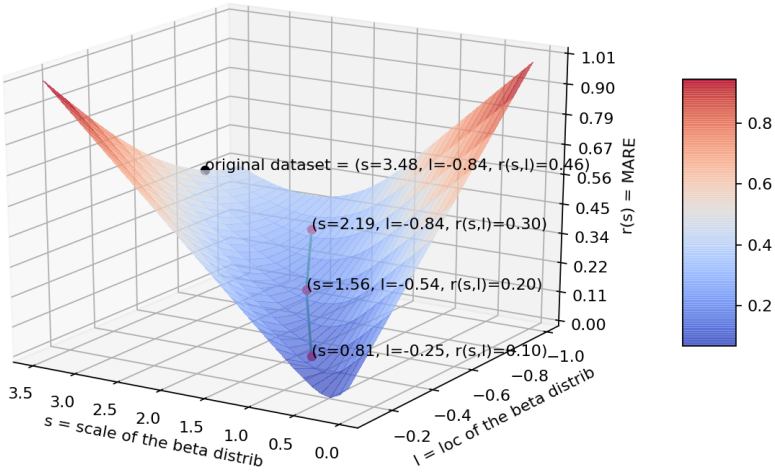


Figure 5.10: Closest roots for different target ARE

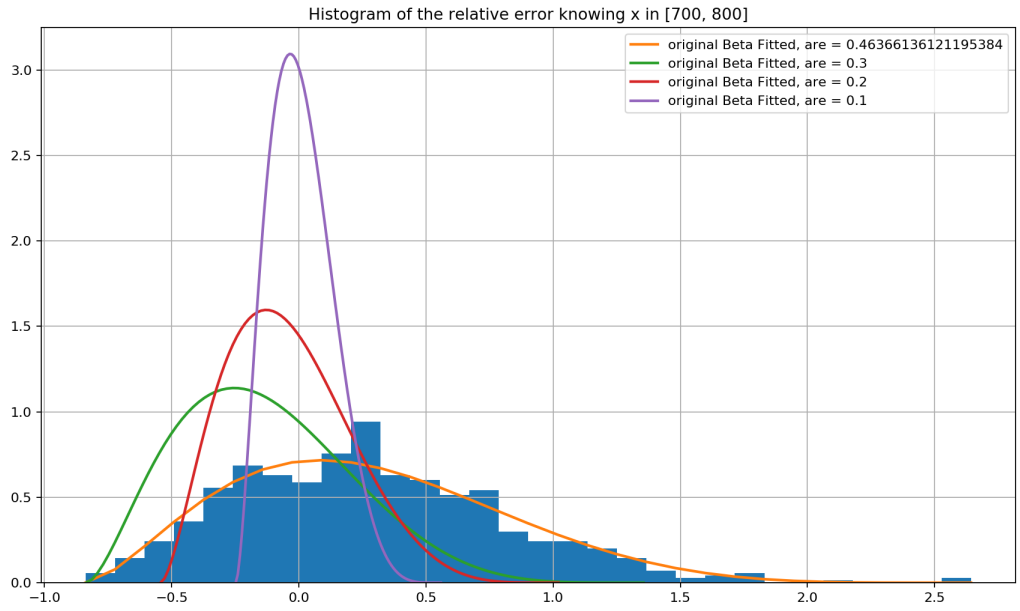


Figure 5.11: shape of the distributions matching the target ares

Changing the Conditional Distributions

Weight Functions

Let's define $\Omega_{\mathcal{X}_{SID}}$ as the set of functions $\omega_{\mathcal{X}_{SID}}$ defined on $\mathcal{X}_{SID}$ such that

$$\frac{1}{n_{SID}} \sum_{x \in \mathcal{X}_{SID}} \omega_{\mathcal{X}_{SID}}(x) = 1.$$

We call them *weight functions*. Weight functions will be used to assign a target MAE to obtain from each of the conditional distributions $\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{X} = x$, for all $x \in \mathcal{X}_{SID}$. It can also be seen as the function that weights the contribution of the Absolute Error of each conditional distribution to the Mean Absolute Relative Error of the simulation.

Target Function Generator

We also define the following functional that we call target function generator.

$$\begin{array}{rccccc} \tilde{m} & : & \mathcal{X}_{SID} \times \mathbb{R}_+ \times \Omega_{\mathcal{X}_{SID}} & \to & \mathbb{R}_+ & \text{(Target function generator)} \\ & & x, \tilde{r}, \omega & \mapsto & \tilde{r} x \omega(x),\ x > 0 & \end{array}$$

For a fixed $\tilde{r}$ and ω, $\tilde{m}(.\,,\,\tilde{r},\omega)$ is a target function for $\tilde{r}$ over the SID. Since the target function will be used to directly adjust the conditional distribution, it must respect the inequality (5.3). Finally, we say that a target MARE $\tilde{r}$ is *feasible* for a given $\omega \in \Omega_{\mathcal{X}_{SID}}$ if

$$\forall x \in \mathcal{X},\ \tilde{m}(x,\tilde{r},\omega) \leq m_{max}(x)$$

Zero Power Input

We recall that the zero input does not count in the computation of the MARE. However, we want the distribution of the simulated errors to be drawn from the estimated distribution. In other words :

$$\forall \tilde{r} \in \mathbb{R}_+, \quad \tilde{l}(0) = \hat{l}(0) \ \textit{and}\ \tilde{s}(0) = \hat{s}(0)$$

We assign

$$\tilde{m}(0) = \hat{m}(0)$$

To avoid big discontinuities in the parameters of the beta distributions, we could take as $\tilde{l}(0) = \lim_{x\to 0} \tilde{\ell}(x)$, $\tilde{s}(0) = \lim_{x\to 0} \tilde{s}(x)$,

Convergence to the Requested MARE

Using the function $\tilde{m}$ to assign a target MAE for each SID input will allow us to hit the target MARE using the simulation distributions. Indeed, let us define the random variable $\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{X}$ with density $\tilde{f}_{\tilde{\boldsymbol{\mathcal{E}}}|\mathbf{X}=x}(\varepsilon) = b(\varepsilon, \tilde{\mathcal{S}}_{x,\tilde{m}})$, $\varepsilon \in (-cap, cap)$. If we establish the distribution parameters as described in Section 5.3 and solve Program (5.1) with $m(x) = \tilde{m}(x;\,\tilde{r},\omega_{\mathcal{X}_{SID}})$ we have,

$$\int_{\varepsilon=-\infty}^{\infty} |\varepsilon| b(\varepsilon, \tilde{\mathcal{S}}_{x,\tilde{m}}) d\varepsilon = \tilde{m}(x;\,\tilde{r},\omega_{\mathcal{X}_{SID}}), \quad \forall x \in \mathcal{X}_{SID}.$$

Then, the expected MARE with the errors drawn from these distributions and with the inputs in the $\mathcal{X}_{SID}$ is :

$$\begin{aligned}
\mathbb{E}_{\tilde{\boldsymbol{\mathcal{E}}}}\Big[\frac{1}{n_{SID}} \sum_{x\in\mathcal{X}_{SID}} \frac{|\tilde{\boldsymbol{\mathcal{E}}}|\ |\mathbf{X}=x}{x}\Big] &= \frac{1}{n_{SID}} \sum_{x\in\mathcal{X}_{SID}} \frac{\mathbb{E}_{\tilde{\boldsymbol{\mathcal{E}}}}[|\tilde{\boldsymbol{\mathcal{E}}}|\ |\mathbf{X}=x]}{x} \\
&= \frac{1}{n_{SID}} \sum_{x\in\mathcal{X}_{SID}} \frac{\tilde{m}(x;\,\tilde{r},\omega_{\mathcal{X}_{SID}})}{x} \\
&= \frac{\tilde{r}}{n_{SID}} \sum_{x\in\mathcal{X}_{SID}} \omega_{\mathcal{X}_{SID}}(x) \\
&= \tilde{r}
\end{aligned}$$

This is true with any weight function for which $\frac{1}{n_{SID}} \sum_{x\in\mathcal{X}_{SID}} \omega_{\mathcal{X}_{SID}}(x) = 1$. We now proceed to describe the construction of a sensible weight function.

Weight Function for $\mathcal{X}_{SID} = \mathcal{X}$

We recall the plausibility objectives: we want the distribution of the simulated errors to be as close as possible to the population distribution. In particular, suppose that we want to do a simulation with a target MARE that happens to be the same as the MARE of the original data and further suppose that we want to simulate using values from the entire data set (i.e., $\mathcal{X}_{SID} = \mathcal{X}$). Then we expect the simulated conditional distributions to be equal to the estimated conditional distributions. In other words,

$$\mathcal{X}_{SID} = \mathcal{X}, \quad \tilde{r} = r_{\hat{m}} \implies \forall x \in \mathcal{X}, \tilde{l}(x) = \hat{l}(x) \text{ and } \tilde{s}(x) = \hat{s}(x).$$

Solving Program (5.2) defined in subsection 5.3, leads to $\tilde{l}(x) \approx \hat{l}(x)$ and $\tilde{s}(x) \approx \hat{s}(x), \forall x \in \mathcal{X}$.

Let us define the following $\hat{\omega}_{\mathcal{X}}$ function,

$$\forall x \in \mathcal{X},\ \hat{\omega}_{\mathcal{X}}(x) := \frac{\hat{m}(x)}{x r_{\hat{m}}} = \frac{\int_{\varepsilon=-\infty}^{\infty} |\varepsilon| \hat{f}_{\boldsymbol{\varepsilon}|\mathbf{X}=x}(\varepsilon) d\varepsilon}{x r_{\hat{m}}}$$

By definition, we have $\frac{1}{n}\sum_{x \in \mathcal{X}} \hat{\omega}_{\mathcal{X}}(x) = 1$. It is a weight function.

The choice of this weight function is natural when $\mathcal{X}_{SID} = \mathcal{X}$ because it is the ratio of the expected absolute relative error simulated at x over the mean absolute relative error when the errors are distributed according to the estimated joint distribution. However, choosing it when $\mathcal{X}_{SID} \neq \mathcal{X}$ would satisfy our requirement for plausibility but it would prevent us from hitting the requested MARE.

Figure 5.12 illustrates that for the full CAISO wind dataset, the weight function presents a hyperbolic shape. The low values account for the biggest part of the MAPE.

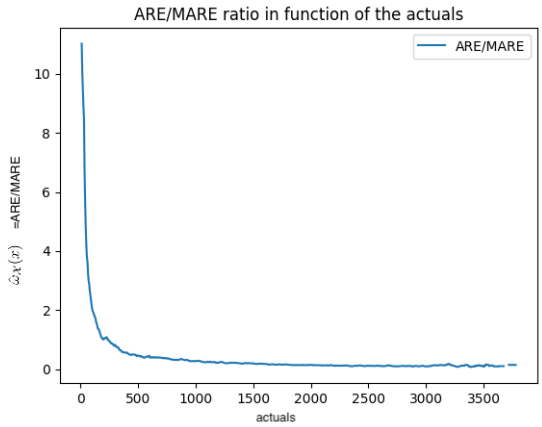


Figure 5.12: $\hat{\omega}_{\mathcal{X}}(x) = \frac{\hat{m}(x)}{x\hat{r}}$ ratio for the CAISO wind dataset.

Weight Function for $\mathcal{X}_{SID}$ and an Arbitrary $\tilde{r}$

Let us define the following real number that we call the distribution plausibility score :

$$P_{\mathcal{X}_{SID}} = \frac{1}{n_{SID}} \sum_{x \in \mathcal{X}_{SID}} \hat{\omega}_{\mathcal{X}}(x)$$

When $\mathcal{X}_{SID} \neq \mathcal{X}$, the distribution of the SID is different from the distribution of the input dataset. Thus we do not necessarily have $P_{\mathcal{X}_{SID}} = 1$. A goal of our method is to meet the requested MARE, at least in expectation, no matter the $\mathcal{X}_{SID}$. If $P_{\mathcal{X}_{SID}}$ is greater than 1, it means that the distribution of $\mathcal{X}_{SID}$ has more data in the range where the weight function takes high values. This means that if use $\hat{\omega}_{\mathcal{X}}$, we are going to simulate too many errors with high values. While it makes some physical sense, we are nonetheless going to simulate a greater MARE than expected. Symmetrically, if $P_{\mathcal{X}_{SID}}$ is smaller than one, we are going to retrieve a lower MAPE than expected. This is illustrated in Figure 5.13 where the density for the $\mathcal{X}_{SID}$ between December 2013 and March 2014 indicates more values at lower power than for the entire dataset, $\mathcal{X}$. If we simply used $\hat{\omega}_{\mathcal{X}}(x)$, for $x \in \mathcal{X}_{SID}$, then meeting the target AREs for each x would result in a MARE much greater than specified. In other words, since the ARE/MARE ratio is very high for the low power input, and since these inputs are over represented under the distribution of the December 2013 - March 2014 SID, we are going to simulate too many errors with a high target of mean absolute error. To meet the target MARE, a re-scaled weight function must therefore be computed.

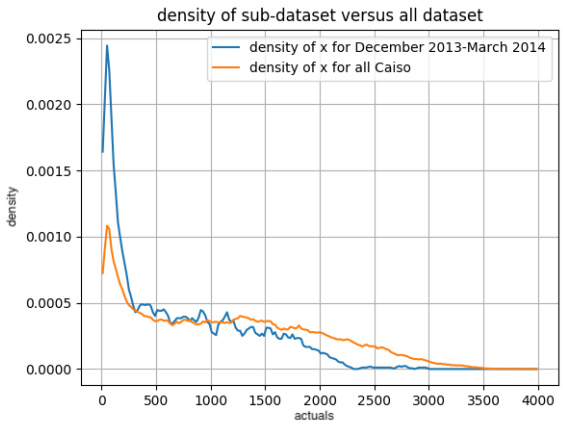


Figure 5.13: Comparison test density versus all dataset density

Let us define the following SID weight function :

$$\forall x \in \mathcal{X}_{SID},\ \tilde{\omega}_{\mathcal{X}_{SID}}(x) := \frac{\hat{\omega}_{\mathcal{X}}(x)}{P_{\mathcal{X}_{SID}}}$$

With the re-scaled factor, we have $\frac{1}{n_{SID}} \sum_{x \in \mathcal{X}_{SID}} \tilde{\omega}_{\mathcal{X}_{SID}}(x) = 1$ so $\tilde{\omega}_{\mathcal{X}_{SID}} \in \Omega_{\mathcal{X}_{SID}}$.

Finally, for a given feasible $\tilde{r} \in \mathbb{R}_+$, we compute a $\tilde{\omega}_{\mathcal{X}_{SID}}$ which allocates the absolute errors across $\mathcal{X}_{SID}$ based on the allocation across $\mathcal{X}$. With these two parameters we can compute $\tilde{m}(x;\, \tilde{r}, \tilde{\omega}_{\mathcal{X}_{SID}})$, $x \in \mathcal{X}_{SID}$. According to Section 5.3, defining $\tilde{\boldsymbol{\mathcal{E}}}$ from this target function, will get us $\mathbb{E}_{\tilde{\boldsymbol{\mathcal{E}}}}[\frac{1}{n_{SID}} \sum_{x \in \mathcal{X}_{SID}} \frac{|\tilde{\boldsymbol{\mathcal{E}}}|\,|\mathbf{X}=x}{x}] = \tilde{r}$.

We can also get the feasibility region for the target MARE. For a given $\tilde{r}$ to be a feasible target MARE, it must satisfy $\forall x \in \mathcal{X}, \tilde{m}(x, \tilde{r}, \tilde{\omega}_{\mathcal{X}_{SID}}) \leq m_{max}(x)$. We can deduce a feasibility region for the target MARE:

$$\tilde{r} \leq \tilde{\mathcal{R}}_{\mathcal{X}_{SID}} = P_{\mathcal{X}_{SID}} \min\big(\frac{m_{max}(s)}{s\hat{\omega}_{\mathcal{X}}(s)}, s \in \mathcal{X}_{SID}\big)$$

Simulating Without autocorrelation

Simulation of forecast errors without autocorrelation is now straightforward. We generate a vector that is identically and independently distributed (i.i.d.) following a Uniform on $[0, 1]$, $\tilde{\mathbf{u}}_t = (\tilde{\mathbf{u}}_{t,i})_{i \leq n_{SID}}$. Then,

$$\tilde{\boldsymbol{\varepsilon}}_{t,i} = \tilde{F}^{-1}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x^{SID}_{t,i}}(\tilde{\mathbf{u}}_{t,i}), \quad \forall i \leq n_{SID}$$

and

$$\tilde{\mathbf{y}}_{t,i} = x_{t,i} + \tilde{\boldsymbol{\varepsilon}}_{t,i}, \quad \forall i \leq n_{SID},$$

so

$$\mathbb{E}[MARE(x^{SID}_t, \tilde{\mathbf{y}}_t)] = \tilde{r}.$$

While we are hitting the target MARE in expectation, the entire autocorrelation of the simulated errors relies solely on the autocorrelation of the input. In the case where the errors are not depending on the input, i.e. $\tilde{F}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x^{SID}_{t,i}} = \tilde{F}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x^{SID}_{t,0}}$, $\forall i \leq n_{SID}$ - which is the case for the middle power range for the CAISO wind data - then our simulations would have a null autocorrelation function (except for lag 0). It would resemble a white noise process, which is not plausible according to our criteria. Therefore, we need to implement a base process that will replace $(\tilde{\mathbf{u}}_t)_{t \leq n_{SID}}$ and will generate the needed autocorrelation to satisfy the second point of the plausibility criteria.

5.4 Ensuring auto-correlation plausibility

Inferring a Base Process

Simulating forecast errors that have the estimated beta distribution as well as the appropriate temporal autocorrelation requires the estimation and use of a *base process*. Instead of simply passing i.i.d. uniform pseudo-random numbers to the inverse of the beta distributions, we create a base process in a way that is similar to the methods used in [177]. Our method

uses a base process to generate a pseudo-random vector $\tilde{\mathbf{u}}_t = (\tilde{\mathbf{u}}_{t,i})_{i \leq n_{SID}}$ that is marginally Uniform on [0,1] and that depend on the past p lags of the base process and the past q lags of errors over the base process. Then, as seen previously in section 5.3, we simulate the forecast errors via the transformation $\hat{F}^{-1}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x_t}(\tilde{\mathbf{u}}_t)$.

In order to obtain a distribution that enables creation of a vector of Uniform pseudo random numbers with the appropriate characterisics, we use techniques inspired by the ARTA fit method (see [181]). The method makes use of the CDF for the standard normal distribution, which we denote using ϕ. The CDF of the conditional distribution $\boldsymbol{\mathcal{E}}|\mathbf{X}$, which is a beta distribution fit using $\mathcal{X}$ is denoted by $\hat{F}_{\boldsymbol{\mathcal{E}}|\mathbf{X}}$. Let us define the following time series $\hat{z}_t$, as the base process time series of the dataset:

$$\forall i \leq n, \quad \hat{z}_{t,i} = \phi^{-1}(\hat{F}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x_{t,i}}(\varepsilon_{t,i}))$$

Its empirical distribution is close to a standard Gaussian. Indeed, in Section 5.2 we are estimating the conditional distribution so that $\boldsymbol{\mathcal{E}}|\mathbf{X} = x$ has distribution that is approximated by $\hat{f}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}$, thus, $\hat{F}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x}(\boldsymbol{\mathcal{E}}) \dot{\sim} \mathcal{U}[0,1]$ and $\hat{\mathbf{z}}_t \dot{\sim} \mathcal{N}(0,1)$. We fit on this base process an ARMA process. The standard definition of an ARMA process of order p and q uses $(a_i)_{i \leq p}$ and $(b_i)_{i \leq q}$ as coefficients so we temporarily reuse those symbols in this section.

Definition 5.4.1. $\{\mathbf{z}_i\}$ is a base process if

- $\{\mathbf{z}_i\}$ follows an ARMA process of order p and q :

$$\mathbf{z}_i = \sum_{h=1}^{p} a_h \mathbf{z}_{i-h} + \sum_{h=1}^{q} b_h \delta_{i-h} + \delta_i$$

  Where $\{\delta_t\}$ are i.i.d. Gaussian random variables with mean 0 and variance σ_δ^2.

- $Var[\mathbf{z}_i] = 1$, $E[\mathbf{z}_i] = 0$, so that for all i, $\quad \mathbf{z}_i \sim N(0,1)$.

We run a grid search over multiple (p,q) and we select the ARMA model to minimize the BIC criterion. The $(a_i)_{i \leq p}$ and $(b_i)_{i \leq q}$ found during the process define a function that enables us to generate base processes which will create the autocorrelation that we are looking for. We further specify σ_δ so that we get $Var[\tilde{\mathbf{z}}_t] = 1$.

Then, we can simulate the error directly by

$$\tilde{\boldsymbol{\varepsilon}}_{t,i} = \tilde{F}^{-1}_{\boldsymbol{\mathcal{E}}|\mathbf{X}=x_{t,i}}(\phi(\tilde{\mathbf{z}}_{t,i})), \quad \forall i \leq n_{SID}$$

and also get the result for the expected MARE established in Section 5.3.

By using an ARMA process as a base process we get the desired autocorrelation without the assumption of Gaussian errors that accompanies direct use an ARMA process [181]. As an aside, we note that the textbook use of i.i.d. uniforms as described in Section 5.3 can be seen as a degenerate case where the base process is of order $p = 0$ and $q = 0$.

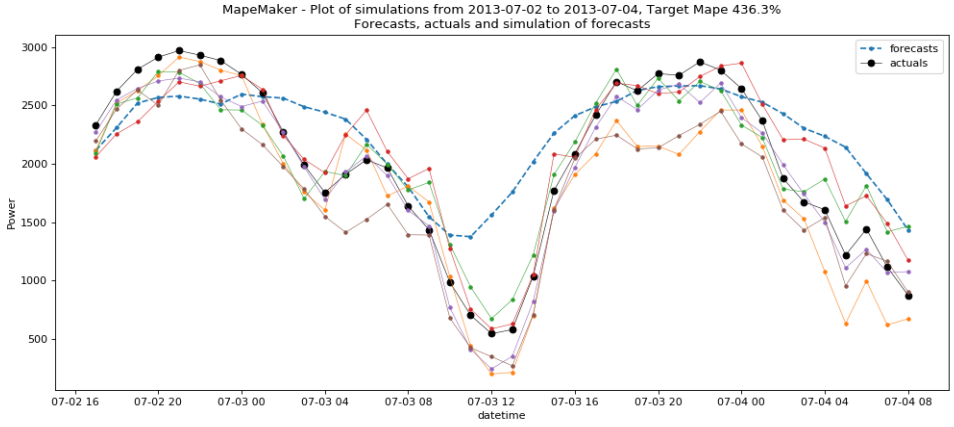
Enforcing Curvature

Let $(y_i)_i \in \mathbb{R}^n$ denote a simulation output time series. We define the *curvature* at a point i in $(y_{t,i})_i$ as

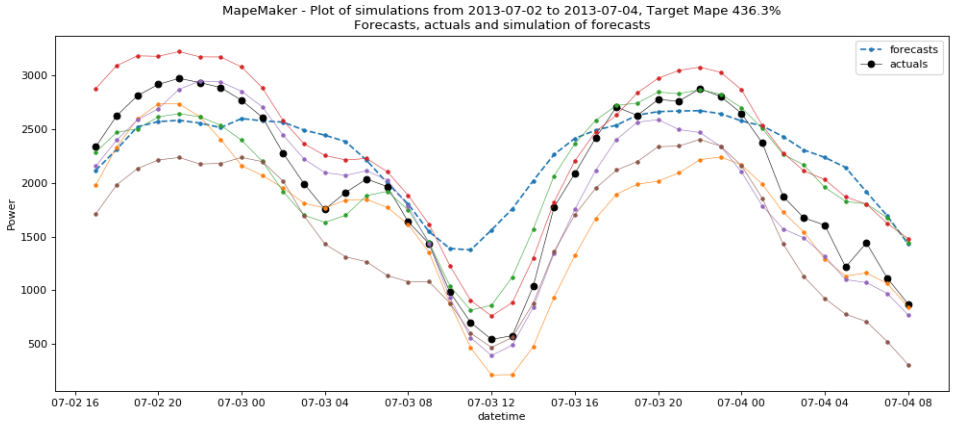
$$s_i = y_{i+2} - 2y_{i+1} + y_i \quad \forall i < n-2,$$

i.e., a second difference.

Methods described in Section 5.4 successfully model temporal correlation between the errors while satisfying a target MARE. However, some scenarios might not "look right" because of their lack of smooth curvature. This is specified as the third point of the plausibility criteria. This is especially unsatisfying in the case of forecasted renewables power production, which are much less sharp and erratic when compared to actual quantities.



(a) Without curvature adjustment.



(b) With curvature adjustment

Figure 5.14: One set of forecast scenarios with, and without, curvature adjustment

We now concretely illustrate this issue in Figure 5.14 with analysis of the CAISO wind power production data introduced previously. In Figure 5.14(a), we show baseline scenarios resulting from our proposed methods, i.e., without adjustment for curvature. As is clearly observed, the simulated forecasts in this case closely mirror the actuals – and not the one "true" forecast. In contrast, we show in Figure 5.14(b) a closely related set simulated forecasts – obtained by the procedure we now describe – that instead exhibit significantly smoother and more realistic-looking curvature. Ultimately, the need for such adjustment depends entirely on the application.

In order to adjust the curvature of a forecast while still achieving a target MARE, one approach is to *a posteriori* adjust a time series that already satisfies a target MARE such that specific curvature characteristics are imposed. We now formalize this general approach.

We introduce a minimization problem in which we penalize deviations from both a target second difference and the simulated forecast error. Per earlier analysis, we can simulate $(\tilde{\varepsilon}_i)_i$ using an ARMA base process. Then, define $d \in \mathbb{R}_+$ as the target second difference, and we let the user decides on the penalization weights: W_s and W_ε in $\mathbb{R}_+$.

We then let $(y_i)_i$ denote the solution of the following mathematical program:

$$\begin{aligned} \min_{y} \quad & \sum_{i=3}^{n} W_s \Big(|y_i - 2y_{i-1} + y_{i-2}| - d \Big)^2 + W_\varepsilon \Big(y_i - x_i - \tilde{\varepsilon}_i \Big)^2 \\ \text{s.t.} \quad & y \in [0, cap]^n \end{aligned} \tag{5.4}$$

For practical computation, we now transform this mathematical program so that the objective function is quadratic and constraints are linear – such that widely available mathematical programming solvers can be leveraged. The transformation yields the following equivalent mixed-integer linear program (MILP), with $3n$ additional variables, n equality constraints, and $3n$ inequality constraints ($6n$ if we consider that the three real vectors are negatively bounded by 0):

$$\begin{aligned} \min_{y,\, \lambda^+,\, \lambda^-,\, b} \quad & \sum_{i=2}^{n} W_s \Big(\lambda_i^+ + \lambda_i^- - d \Big)^2 + W_\varepsilon \Big(y_i - x_i - \varepsilon_i \Big)^2 \\ \text{s.t.} \quad & y \in \mathbb{R}_+^n,\ \lambda^+ \in \mathbb{R}_+^n,\ \lambda^- \in \mathbb{R}_+^n,\ b \in \{0,1\}^n \\ & y_i \leq cap \\ & \lambda_i^+ - \lambda_i^- = y_i - 2y_{i-1} + y_{i-2}, \quad \forall i \leq n \\ & \lambda_i^+ \leq b_i d_{max} \\ & \lambda_i^- \leq (1 - b_i) d_{max} \end{aligned} \tag{5.5}$$

where d_{max} denotes a large constant; a safe value is four times the capacity.

To verify equivalence of the two mathematical programs, we note that if $y_i - 2y_{i-1} + y_{i-2} \geq 0$, then because $b_i \in \{0,1\}$, λ_i^- is equal to 0 with the two last equations. Then $\lambda_i^+ = y_i - 2y_{i-1} + y_{i-2}$ and $\lambda_i^+ + \lambda_i^- = y_i - 2y_{i-1} + y_{i-2}$. We use the same reasoning when $y_i - 2y_{i-1} + y_{i-2} < 0$.

Numerous open source and commercial solvers are available for such a mathematical program. However, solution time does generally increase with n. In many applications the restrictions on curvature are motivated by aesthetic or heuristic considerations. Thus, it can be reasonable to specify a "loose" optimality gap to avoid excessive computation time.

5.5 Evaluation

In this article, we proposed three types of simulations all based on the conditional estimation and adjustment of section (5.3). We then added that we could safely model a base process to be one step closer to the autocorrelation of the observed dependent variable, and that -eventually- by correcting the curvature we could have smooth forecast simulations. We denote those three procedures:

- A) "IID" as the base process which is i.i.d.,
- B) "ARMA" as the base process which follows an ARMA process,
- C) "ARMA + curvature" as the base process which follows an ARMA process and we correct the curvature *a posteriori*,

Behavior of the simulated MAREs as M Grows

To illustrate the behavior of the simulated scenarios as the number of scenarios created, M, grows we conducted experiments using the CAISO wind dataset and created scenarios for three days. Figure 5.15 shows that for this example, the achieved MARE is close to the target MARE as soon as there are about 5 scenarios.

Behavior of the curvature and autocorrelation function of the simulations

Let p be the maximum lag of autocorrelation we wish to assess.

Let us define the value of the autocorrelation function of ε_t at lag j,

$$\hat{\rho}(\varepsilon_t, j) = \frac{1}{(n-j)\sigma^2} \sum_{i=0}^{n-j} \varepsilon_{t,i+j}\varepsilon_{t,i}$$

From there, we can define a score, the squared root of the sum of the squared deviation of the autocorrelation of the simulation with the observed errors :

$$S_{auto\_correlation}(\varepsilon_t, \tilde{\varepsilon}_t) = \sqrt{\sum_{j=1}^{p} (\hat{\rho}(\varepsilon_t, j) - \hat{\rho}(\tilde{\varepsilon}_t, j))^2}$$

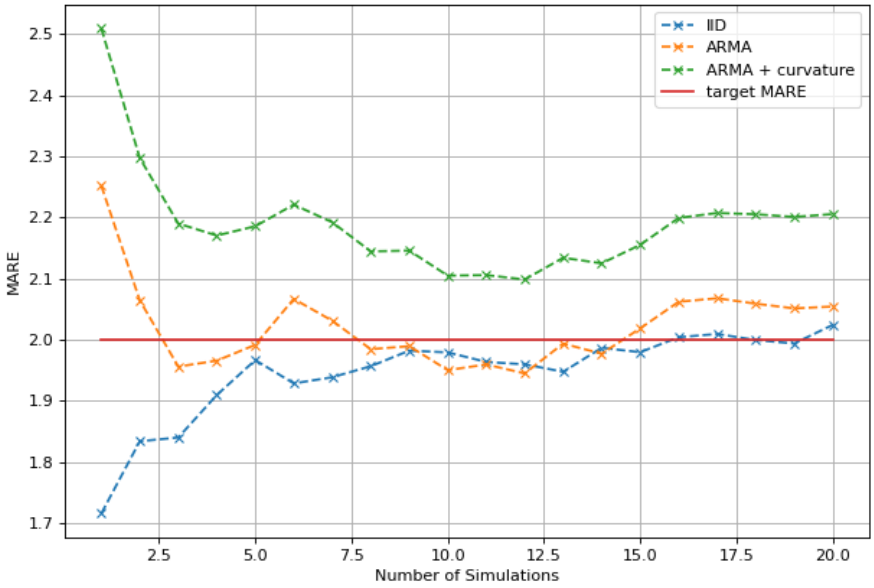


Figure 5.15: MARE score as a function of the number of scenarios created by simulation.

We ran $M = 20$ simulations over the CAISO wind dataset and created scenarios for a target MAPE of 200% and for data ranging from 2013-07-01 to 2015-06-30.

	MARE	$S_{auto\_correlation}$	Curvature
Observed	3.77162	0	71.658
Targeted	2	0	71.658
IID	1.896 ± 0.142	1.949 ± 0.025	370.306 ± 2.091
ARMA	1.914 ± 0.215	0.011 ± 0.004	169.071 ± 1.332
ARMA + curvature	1.888 ± 0.298	0.011 ± 0.022	85.531 ± 19.625

Table 5.1: The first two lines indicate the values observed in the input dataset and the target values for the scenarios. The rest of the table gives the mean $\pm$ the standard deviation for 20 scenarios simulated with respect to three procedures.

First we confirm, that each procedure results in a close MARE with respect to the targeted MARE. The scenarios that use curvature correction result in a value that is closest to the target, which makes sense because Program 5.5 corrects for the MARE after the scenarios are created. However, the other scenarios are reasonably close. We also notice that for base processes that are not i.i.d., the autocorrelation score is quite good almost regardless of the

number of scenarios. Finally, we can notice that for each value (MARE, autocorrelation score and Curvature) the methods are ordered as expected.

Putting it all Together

In this section we summarize the process to deliver a simulation with correct targets.

Procedures for Estimation

First, as shown in Algorithm 2, we preprocess the data and estimate the conditional distributions using the methods explained in Section 5.2. This results in a set of beta distribution parameters for each input from the whole dataset, called $\hat{\mathcal{S}}$. To estimate the parameters we recall that the user should specify a data fraction for the sampling (e.g., 0.05). (The software provides an option to produce a curve for the scores described in Section 5.2.)

Algorithm 2 Estimating the beta distributions

Input: x_t, y_t, a\{Input time series and percent of data\}
Output: $\hat{\mathcal{S}}$

procedure COMPUTING_ESTIMATION_PARAMETERS(x_t, y_t, a)
  $\mathcal{X} \leftarrow sort(x_t)$
  for $x \in \mathcal{X}$ **do**\{Applying the methodology explained in Section 5.2\}
    Compute the interval of estimation I_x^a and sample E_x^a .
    $\bar{x}(x,a) \leftarrow \bar{I_x^a}$
    $\hat{l}(\bar{x}(x,a)), \hat{s}(\bar{x}(x,a)) \leftarrow Bounds(E_x^a, x)$\{See section 5.2\}
    $\hat{\alpha}(\bar{x}(x,a)), \hat{\beta}(\bar{x}(x,a)) \leftarrow Moments(mean(E_x^a), std(E_x^a))$\{See section 5.2\}
  end for
  for $x \in \mathcal{X}$ **do**\{Take the closest computed point of estimation\}
    $x' \leftarrow argmin|\bar{x}(x',a) - x|$
    $\hat{\mathcal{S}}_x \leftarrow (\hat{\alpha}, \hat{\beta}, \hat{l}, \hat{s})(\bar{x}(x',a))$
  end for
  return $\hat{\mathcal{S}}$
end procedure=0

Next, as shown in Algorithm 3, we estimate the partitioning of the mean absolute percent errors according to the input and we encode this information in the weight function. An important feature of this procedure is the computation of $r_{\hat{m}}$ which is the expected mean absolute relative error from the conditional distributions (which may be close in value to, but is different from, $\hat{r}$.) This procedure is explained in section 5.3.

The next phase, shown in Algorithm 4, is estimation of the underlying base_process that generates autocorrelation in the time series of the errors. This is done by using the CDF B of the beta distribution whose parameters have been inferred in step 1. Then we operate a grid search over the p and q parameters to select the order of the model that minimizes the

Algorithm 3 Estimating the weight function

Input: $\hat{\mathcal{S}}_x$, $\mathcal{X}$
Output: $\omega_{\mathcal{X}}$
0: **procedure** COMPUTING_ESTIMATED_WEIGHT_FUNCTION($\hat{\mathcal{S}}_x$, $\mathcal{X}$)
0: $r_{\hat{m}} \leftarrow 0$
0: **for** $x \in \mathcal{X}$ **do**{Applying the methodology explained in Section 5.3}
0: $\hat{m}(x) \leftarrow \int_{\varepsilon=-\infty}^{\infty} |\varepsilon| beta(\varepsilon; \hat{\mathcal{S}}(x)) d\varepsilon$.
0: $m_{max}(x) \leftarrow \max \nu(l, s, \hat{\alpha}(x), \hat{\beta}(x))$ {See constraints on target function (5.3)}
0: $\omega_{\mathcal{X}}(x) \leftarrow \frac{\hat{m}(x)}{x}$
0: $r_{\hat{m}} \leftarrow r_{\hat{m}} + \hat{m}(x)$
0: **end for**
0: $r_{\hat{m}} \leftarrow \frac{r_{\hat{m}}}{|\mathcal{X}|}$
0: $\omega_{\mathcal{X}} \leftarrow \frac{\omega_{\mathcal{X}}}{r_{\hat{m}}}$
0: **return** $\omega_{\mathcal{X}}$
0: **end procedure**=0

BIC criterion. We save the coefficients. Recalling that we want the marginal of Z to follow a standard normal, we set the variance of the errors of the base process so that $Var[\mathbf{Z}_t] = 1$. This procedure is explained in Section 5.4.

Procedures to Deliver the Target MARE

First, as shown in Algorithm 5, given a target MARE $\tilde{r}$, and a $\mathcal{X}_{SID}$ we verify that $\tilde{r}$ is feasible. If it is, we aim at targeting a mean absolute error for each conditional distribution with input in the $\mathcal{X}_{SID}$. For this, we compute a target function using the estimated weight function (see section 5.3).

Second, as shown in Algorithm 6, according to a target function $\tilde{m}$, we assign adjusted parameters for each conditional distribution whose input is in the $\mathcal{X}_{SID}$. We adjust the location parameters from the estimated values while keeping the shape parameters. See section 5.3.

Procedure to Simulate the Output

Using methods summarized in Algorithm 7, we simulate a base process sample of length $|\mathcal{X}_{SID}|$ and use the simulated conditional distributions to obtain conditioned errors. We directly get the simulation by summing the errors and the input data. Finally, if the user asks for it, we optimize the curvature *a posteriori*, see section 5.4.

Algorithm 4 Fitting the Base Process ARMA process

Input: x_t, ε_t, $\hat{\mathcal{S}}_x$
Output: $(a_i)_{i\leq p}$, $(b_i)_{i\leq q}$, σ_δ
0: **procedure** FIT_ARMA_PROCESS(x_t, ε_t, $\hat{\mathcal{S}}_x$)
0: **for** $i \in [1, len(x)]$ **do**{Estimating the base process see Section 5.4}
0: $\hat{z}_{t,i} \leftarrow \phi^{-1}(B(\varepsilon_{t,i}, \mathcal{S}_{\mathcal{X}}(x_i)))$
0: **end for**
0: $BIC \leftarrow +\infty$
0: $p,\ q \leftarrow 0,\ 0$
0: **for** $p', q' \in [0,5]^2$ **do**{Grid Searching}
0: $tempBIC \leftarrow BIC(ARMA(\hat{z}_t, (p', 0, q'))$
0: **if** $tempBIC < BIC$ **then**
0: $BIC \leftarrow tempBIC$
0: $p,\ q \leftarrow p',\ q'$
0: **end if**
0: **end for**
0: $(a_i)_{i\leq p},\ (b_i)_{i\leq q} \leftarrow ARMA(\hat{z}_t, (p, 0, q))$
0: $\sigma_\delta \leftarrow \arg\min_\sigma(std(ARMA((a_i)_{i\leq p},\ (b_i)_{i\leq q}, \sigma) - 1)^2$
0: **return** $(a_i)_{i\leq p},\ (b_i)_{i\leq p},\ \sigma_\delta$
0: **end procedure**=0

Algorithm 5 Inferring a target function for the SID

Input: $\hat{\mathcal{S}}$, $\mathcal{X}_{SID}, \tilde{r}, \hat{\omega}_{\mathcal{X}}$
Output: $\tilde{m}$
0: **procedure** COMPUTING_SIMULATION_TARGET_FUNCTION($\hat{\mathcal{S}}$, $\mathcal{X}_{SID}, \tilde{r}, \hat{\omega}_{\mathcal{X}}$)
0: $P_{\mathcal{X}_{SID}} \leftarrow 0$ {Computing the Distribution plausibility score}
0: **for** $s \in \mathcal{X}_{SID}$ **do**
0: $P_{\mathcal{X}_{SID}} \leftarrow P_{\mathcal{X}_{SID}} + \frac{\hat{\omega}_{\mathcal{X}}(s)}{|\mathcal{X}_{SID}|}$.
0: **end for**
0: $\tilde{r}_{max} \leftarrow P_{\mathcal{X}_{SID}} \min\left(\frac{m_{max}(s)}{s\hat{\omega}_{\mathcal{X}}(s)}, s \in \mathcal{X}_{SID}\right)$
0: **if** $\tilde{r} > \tilde{r}_{max}$ **then**
0: **Report Error**
0: **end if**
0: $\tilde{\omega}_{\mathcal{X}_{SID}} \leftarrow \frac{\hat{\omega}_{\mathcal{X}}}{P_{\mathcal{X}_{SID}}}$
0: **for** $s \in \mathcal{X}_{SID}$ **do**{Applying the function as explained in Section 5.3}
0: $\tilde{m}(s) \leftarrow \tilde{r}s\tilde{\omega}_{\mathcal{X}_{SID}}(s)$
0: **end for**
0: **return** $\tilde{m}$
0: **end procedure**=0

Algorithm 6 Inferring the simulation beta distributions

Input: $\tilde{m}$, $\mathcal{S}$, $\mathcal{X}_{SID}$
Output: $\tilde{\mathcal{S}}_{\tilde{m}}$
0: **procedure** ADJUSTING_SIMULATION_PARAMETERS($\tilde{m}$, $\mathcal{S}$, $\mathcal{X}_{SID}$)
0: **for** $\delta \in \mathcal{X}_{SID}$ **do**{Applying the methodology explained in section}
0: $x \leftarrow closest(\delta, \mathcal{X})$ {not necessarily $\mathcal{X}_{SID} \subset \mathcal{X}$}
0: $\tilde{\alpha}(\delta), \tilde{\beta}(\delta) \leftarrow \hat{\alpha}(x), \hat{\beta}(x)$
0: $\tilde{l}(\delta), \tilde{s}(\delta) \leftarrow Program_1(\tilde{\alpha}(\delta), \tilde{\beta}(\delta), \delta, \tilde{m}(\delta))$ {See equation 5.2}
0: $\tilde{\mathcal{S}}_{\delta,\tilde{m}} = (\tilde{\alpha}(\delta), \tilde{\beta}(\delta), \tilde{l}(\delta), \tilde{s}(\delta))$
0: **end for**
0: **return** $\tilde{\mathcal{S}}_{\tilde{m}}$
0: **end procedure**=0

Algorithm 7 Simulating a sample of output

Input: $\tilde{m}$, $\mathcal{S}$, x_t^{SID}, $(a_i)_{i\leq p}$, $(b_i)_{i\leq p}$, σ_δ, $\tilde{\mathcal{S}}_{\mathcal{X}_{SID}}$, which implies $\tilde{F}^{-1}_{\boldsymbol{\varepsilon}|\mathbf{X}=x_t}$
Output: $(\tilde{y}_{t,i})_{i\leq n_{SID}}$
0: **procedure** COMPUTING_ESTIMATION_PARAMETERS($\tilde{m}, \mathcal{S}, \mathcal{X}_{SID}, (a_i)_{i\leq p}$, $(b_i)_{i\leq p}$, σ_δ)
0: $(\tilde{z}_{t,i})_{i\leq n_{SID}} \leftarrow createArmaSample((a_i)_{i\leq p}$, $(b_i)_{i\leq p}$, σ_δ, $n_{SID})$
0: **for** $i \in [1, n_{SID}]$ **do**
0: $\tilde{\varepsilon}_{t,i} = \tilde{F}^{-1}_{\boldsymbol{\varepsilon}|\mathbf{X}=x_{t,i}^{SID}}(\phi(\tilde{z}_{t,i}))$
0: $\tilde{y}_{t,i} = x_{t,i}^{SID} + \tilde{\varepsilon}_{t,i}$
0: **end for**
0: **if** Curvature is $True$ **then**
0: $(\tilde{y}_{t,i})_{i\leq n_{SID}} \leftarrow Optimization_1(\tilde{\varepsilon}_t, d, x_t^{SID}, cap)$ {See Program 5.5}
0: **end if**
0: **return** $(\tilde{y}_{t,i})_{i\leq n_{SID}}$
0: **end procedure**=0

5.6 Conclusions and Future Directions

We have described methods for creating scenarios that make use of a history of forecast errors. The methods are summarized in algorithm form in the Appendix and the corresponding software is available for download and use. Although we used wind data from CAISO in our illustrations, the method can be used for any situation where there is a history of forecasts and actuals. In particular, the software has been used to create scenarios for load, solar, and wind for the rts-gmlc data `https://github.com/GridMod/RTS-GMLC`.

The use of solar requires pre- and post-processing of the input data to work well. Instead of power values, the forecasts and actuals should be presented as fractions of capacity and with the value of *cap* set to one during the day and zero at night. This is because solar power is always zero at night and because the concept of "low power" changes during the day.

This work enables new research in areas such as:

1. Rigorous quantification of the potential benefits associated with improvements to renewables power production forecasts, to determine if the costs associated with procurement of improved numerical weather prediction (NWP) models and associated hardware to provide boundary / initial conditions (e.g., meteorological towers) translate to improved system reliability and reduced operations costs.

2. Analyses of specific events in which forecast error was significantly higher than typically observed, e.g., during extreme weather conditions. A recent real-world motivating case stems from significant wind forecast errors observed by the Midcontinent Independent System Operator (MISO) in the US during polar vortices.

3. Bootstrapping to produce realistic alternative synthetic actuals for renewables power production, to mitigate issues associated with availability of limited (e.g., a year) historical data when conducting power system operations simulations; very large numbers of samples are required for rigorous reliability and cost-benefit analyses.

Future research includes consideration of error measures other than the MAPE. On the purely software front, we are working to parallelize computations. The software and the methods described here are intended to be an addition to the kit of tools available for dealing with uncertainty in power generation planning and operations.

Acknowledgement

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes. We are grateful for the referees' comments, which improved the paper.

Chapter 6

Conclusion

This dissertation focuses on integrating renewable energy and battery systems into the power grid and addresses the various uncertainties associated with this integration. Through four distinct research projects, this work has contributed to understanding and resolving some challenges in this field.

Battery Storage in Wholesale Markets (Chapter 2): this chapter provided a framework for optimizing the siting and sizing of a battery in a nodal wholesale market. We developed a mixed-integer linear programming model that captures the price-maker behavior of a battery system. This work underscores the strategic importance of battery storage in electricity markets, especially in managing congestion and capitalizing on locational marginal pricing disparities.

Modeling and State Estimation of Lithium-Sulfur Batteries (Chapter 3): this chapter deals with the problem of system identification and state estimation for Li-S batteries. We integrated a piecewise affine model within a moving horizon estimation framework. This addresses the challenges posed by the inherent complexity and constraints of traditional DAE-based models, paving the way for more accurate and efficient battery management strategies. Control of Airborne Wind Energy Systems (Chapter 4): this chapter optimizes the operation of an AWE through an Information-Directed Sampling control strategy. By inherently balancing the exploration of wind conditions and exploitation of immediate power output, this work contributes to maximizing the efficiency and effectiveness of AWEs in a dynamically changing environment.

Probabilistic Scenarios for Renewable Power Production (Chapter 5): this chapter developed a method to create probabilistic scenarios reflecting various levels of forecast accuracy. This work generates a variety of plausible scenarios matching the variability in renewable power production. This chapter's contribution is critical for stochastic economic dispatch, simulation-based analysis and planning in renewable energy systems.

Adding renewable energy to the power grid involves complex technical challenges. This work brings together optimization, control theory, and energy economics, showcasing the multifaceted approach needed in power systems engineering today. As the global shift towards sustainable energy continues, the methods and insights from this dissertation offer

useful guidance in dealing with these complex issues.

Perspective on future work

Each chapter of this dissertation offers opportunities for further refinement and improvement. For instance, the battery storage project does not account for uncertainties in both the bid/ask curves and the overall power flows. An efficient decomposition of the problem using methods like the Alternating Direction Method of Multipliers (ADMM) or Benders decomposition could lead to a more tractable formulation. Future work could consider the use of Gaussian processes to avoid the exhaustive search. Additionally, exploring multi-market participation is essential, considering that battery storage systems primarily derive their revenue from these markets. Our novel way to estimate DAE states for lithium-sulfur batteries offer promising prospects for industrial applications for more general battery chemistry. First, applying few-shot, implicit learning to dynamically refine the DAE approximation model - accounting for shuttle effects and overall cell degradation - is now feasible with fine-tuning techniques developed for Large Language Models. Second, it's vital to demonstrate the advantages of moving horizon estimation in industrial contexts, particularly in terms of computational efficiency, when compared to the existing equivalent circuit model (ECM) and extended Kalman filter (EKF) methods. The use of tools such as JAX or PyTorch for implicit differentiation opens new possibilities in this domain. Online learning is also vital for the control of Airborne Wind Energy Systems (AWES), where employing surrogate power functions could facilitate the analysis and demonstrate the sub-linear regret of the Information-Directed Sampling strategy. Lastly, our current scenario generation method is univariate; transitioning to a multivariate, system-wide scenario generation approach is essential for stochastic unit commitment. Employing techniques like Gaussian copulas could effectively create base processes that are auto-correlated both temporally and spatially.

Bibliography

[1] Carl Sulzberger. "Thomas Edison's 1882 Pearl Street Generating Station". In: *IEEE Global History Network* (2010).

[2] Jada Garofalo. "An Electricity Regulation Primer—The History of Electricity Regulation in the United States". In: *Available at SSRN 3974880* (2021).

[3] Deeptha Mathavan. "From Dabhol to Ratnagiri: The electricity act of 2003 and reform of India's power sector". In: *Colum. J. Transnat'l L.* 47 (2008), p. 387.

[4] Electricité de France. *Statuts d'EDF*. `https://www.edf.fr/groupe-edf/espaces-dedies/investisseurs/statuts-d-edf`. Accessed: [Your Access Date]. 2023.

[5] Judith A McGaw. *More work for mother: The ironies of household technology from the open hearth to the microwave*. 1984.

[6] Julie A Cohn. *The grid: biography of an American technology*. MIT press, 2017.

[7] HyungSeon Oh. "Analytical solution to swing equations in power grids". In: *Plos one* 14.11 (2019), e0225097.

[8] Hadi Saadat et al. *Power system analysis*. Vol. 2. McGraw-hill, 1999.

[9] United States. Bonneville Power Administration. *Columbia River power for the people: A history of policies of the Bonneville Power Administration*. US Department of Energy, Bonneville Power Administration, 1981.

[10] ABB. *Energy Efficiency in the Power Grid*. `http://large.stanford.edu/courses/2014/ph240/singh2/docs/TDEnergyEff.pdf`. Accessed: [Your Access Date]. 2014.

[11] [First Name(s)] Weiss and [First Name(s)] Hagerty. *The Coming Electrification of the North American Economy*. Report. Prepared for WIRES. The Brattle Group, Mar. 2019.

[12] Trieu Mai et al. *Electrification Futures Study: Scenarios of Electric Technology Adoption and Power Consumption for the United States*. NREL Report NREL/TP-6A20-71500. Golden, CO: National Renewable Energy Laboratory (NREL), 2018.

[13] Joshua D. Rhodes. *The old, dirty, creaky US electric grid would cost $5 trillion to replace. Where should infrastructure spending go?* `https://theconversation.com/the-old-dirty-creaky-us-electric-grid-would-cost-5-trillion-to-replace-where-should-infrastructure-spending-go-68290`. Accessed: [Your Access Date]. Mar. 2017. (Visited on 03/16/2017).

[14] U.S. Department of Energy. *Building a Better Grid Initiative*. Accessed: [Insert Date of Access]. 2021. URL: `https://www.energy.gov/gdo/building-better-grid-initiative`.

[15] Theodora Konstantinou and Konstantina Gkritza. "Examining the barriers to electric truck adoption as a system: A Grey-DEMATEL approach". In: *Transportation Research Interdisciplinary Perspectives* 17 (2023), p. 100746.

[16] Anna M Brockway and Laurel N Dunn. "Weathering adaptation: Grid infrastructure planning in a changing climate". In: *Climate Risk Management* 30 (2020), p. 100256.

[17] Sofia Taylor and Line A Roald. "A framework for risk assessment and optimal line upgrade selection to mitigate wildfire risk". In: *Electric Power Systems Research* 213 (2022), p. 108592.

[18] Noah Rhodes and Line Roald. "Co-optimization of power line de-energization and restoration under high wildfire ignition risk". In: *arXiv preprint arXiv:2204.02507* (2022).

[19] Guanglun Zhang et al. "Texas electric power crisis of 2021 warns of a new blackout mechanism". In: *CSEE journal of Power and Energy Systems* 8.1 (2022), pp. 1–9.

[20] Melissa Dumas, Binita Kc, and Colin I Cunliff. *Extreme weather and climate vulnerabilities of the electric grid: A summary of environmental sensitivity quantification methods*. Tech. rep. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2019.

[21] Alexandre Moreira et al. "Climate-aware generation and transmission expansion planning: A three-stage robust optimization approach". In: *European Journal of Operational Research* 295.3 (2021), pp. 1099–1118.

[22] Nicholas Abi-Samra. *Power Grid Resiliency for Adverse Conditions*. Artech House, 2017.

[23] HC Righolt and FG Rieck. "Energy chain and efficiency in urban traffic for ICE and EV". In: *2013 World Electric Vehicle Symposium and Exhibition (EVS27)*. IEEE. 2013, pp. 1–7.

[24] Yanyan Xu et al. "Planning for Electric Vehicles Coupled with Urban Mobility". In: *arXiv preprint arXiv:2303.15578* (2023).

[25] Teng Zeng et al. "Inducing human behavior to maximize operation performance at PEV charging station". In: *IEEE Transactions on Smart Grid* 12.4 (2021), pp. 3353–3363.

[26] International Energy Agency. *The Future of Heat Pumps*. https://www.iea.org/reports/the-future-of-heat-pumps. Accessed: [Insert Date of Access]. 2023.

[27] Tesla, Inc. *Master Plan Part 3*. https://www.tesla.com/ns_videos/Tesla-Master-Plan-Part-3.pdf. Accessed: [Insert Date of Access]. 2023.

[28] Sandia National Laboratories. *Energy Storage Handbook*. https://www.sandia.gov/ess-ssl/wp-content/uploads/2020/12/ESHB_Ch12_Thermal_Ho.pdf. Accessed: [Insert Date of Access]. 2020.

[29] Johannes Hyvönen, Tero Koivunen, and Sanna Syri. "Possible bottlenecks in clean energy transitions: Overview and modelled effects–Case Finland". In: *Journal of Cleaner Production* (2023), p. 137317.

[30] U.S. Department of Energy. *Next Generation Grid Technologies Report*. https://www.energy.gov/oe/articles/next-generation-grid-technologies-report-download?utm_medium=email&utm_source=govdelivery. Accessed: [Insert Date of Access]. 2023.

[31] Julio Romero Aguero et al. "Modernizing the grid: Challenges and opportunities for a sustainable future". In: *IEEE Power and Energy Magazine* 15.3 (2017), pp. 74–83.

[32] Meiken Hansen and Bettina Hauge. "Prosumers and smart grid technologies in Denmark: developing user competences in smart grid households". In: *Energy Efficiency* 10.5 (2017), pp. 1215–1234.

[33] Feng Li, Ilhan Kocar, and Antoine Lesage-Landry. "Mitigating Equipment Overloads due to Electric Vehicle Charging Using Customer Incentives". In: *2023 IEEE Power Energy Society General Meeting (PESGM)*. 2023, pp. 1–5. DOI: 10.1109/PESGM52003.2023.10252605.

[34] Feng Li, Ilhan Kocar, and Antoine Lesage-Landry. "Mitigating equipment overloads due to electric vehicle charging using customer incentives". In: *2023 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE. 2023, pp. 1–5.

[35] Alexandra Von Meier. *Electric power systems: a conceptual introduction*. John Wiley & Sons, 2006.

[36] Shifei Yuan et al. "Stability analysis for li-ion battery model parameters and state of charge estimation by measurement uncertainty consideration". In: *Energies* 8.8 (2015), pp. 7729–7751.

[37] Nima Lotfi et al. "Reduced-order electrochemical model-based SOC observer with output model uncertainty estimation". In: *IEEE Transactions on Control Systems Technology* 25.4 (2016), pp. 1217–1230.

[38] Manh-Kien Tran et al. "Comparative study of equivalent circuit models performance in four common lithium-ion batteries: LFP, NMC, LMO, NCA". In: *Batteries* 7.3 (2021), p. 51.

[39] Paolo Attilio Pegoraro et al. "Compensation of systematic measurement errors in a PMU-based monitoring system for electric distribution grids". In: *IEEE Transactions on Instrumentation and Measurement* 68.10 (2019), pp. 3871–3882.

[40] Emma Stewart. "Micro-synchrophasors for distribution systems". In: *Lawrence Berkeley National Laboratory* (2015), pp. 2016–09.

[41] Siddharth Bhela, Vassilis Kekatos, and Sriharsha Veeramachaneni. "Enhancing observability in distribution grids using smart meter data". In: *IEEE Transactions on Smart Grid* 9.6 (2017), pp. 5953–5961.

[42] Xinfan Lin. "Theoretical analysis of battery SOC estimation errors under sensor bias and variance". In: *IEEE Transactions on Industrial Electronics* 65.9 (2018), pp. 7138–7148.

[43] Morteza Aien, Ali Hajebrahimi, and Mahmud Fotuhi-Firuzabad. "A comprehensive review on uncertainty modeling techniques in power system studies". In: *Renewable and Sustainable Energy Reviews* 57 (2016), pp. 1077–1089.

[44] Ricardo Bessa et al. "Handling renewable energy variability and uncertainty in power system operation". In: *Advances in Energy Systems: The Large-scale Renewable Energy Integration Challenge* (2019), pp. 1–26.

[45] Xiang Li et al. "Are renewables really that expensive? The impact of uncertainty on the cost of the energy transition". In: *Computer Aided Chemical Engineering*. Vol. 46. Elsevier, 2019, pp. 1753–1758.

[46] Shan Jin et al. "Modeling and solving a large-scale generation expansion planning problem under uncertainty". In: *Energy Systems* 2 (2011), pp. 209–242.

[47] Hao Tu et al. "Integrating physics-based modeling with machine learning for lithium-ion batteries". In: *Applied Energy* 329 (2023), p. 120289.

[48] Hermann W Dommel and William F Tinney. "Optimal power flow solutions". In: *IEEE Transactions on power apparatus and systems* 10 (1968), pp. 1866–1876.

[49] Keith Moffat, Mohini Bariya, and Alexandra Von Meier. "Unsupervised impedance and topology estimation of distribution networks—Limitations and tools". In: *IEEE Transactions on Smart Grid* 11.1 (2019), pp. 846–856.

[50] Mathilde Badoual. "Renewable Energy and Flexibility Integration on the Electricity Market". PhD thesis. University of California, Berkeley, 2021.

[51] Yujie Wang et al. "A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems". In: *Renewable and Sustainable Energy Reviews* 131 (2020), p. 110015.

[52] Yih-Fang Huang et al. "State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid". In: *IEEE Signal Processing Magazine* 29.5 (2012), pp. 33–43.

[53] Antoine Lesage-Landry and Joshua A Taylor. "Setpoint tracking with partially observed loads". In: *IEEE Transactions on Power Systems* 33.5 (2018), pp. 5615–5627.

[54] Panqanamala Ramana Kumar and Pravin Varaiya. *Stochastic systems: Estimation, identification, and adaptive control*. SIAM, 2015.

[55] Nesrine Amor, Ghulam Rasool, and Nidhal C Bouaynaya. "Constrained state estimation-a review". In: *arXiv preprint arXiv:1807.03463* (2018).

[56] Greg Welch, Gary Bishop, et al. "An introduction to the Kalman filter". In: (1995).

[57] Rached Dhaouadi, Ned Mohan, and Lars Norum. "Design and implementation of an extended Kalman filter for the state estimation of a permanent magnet synchronous motor". In: *IEEE Transactions on Power Electronics* 6.3 (1991), pp. 491–497.

[58] Rambabu Kandepu, Bjarne Foss, and Lars Imsland. "Applying the unscented Kalman filter for nonlinear state estimation". In: *Journal of process control* 18.7-8 (2008), pp. 753–768.

[59] Geir Evensen. "The ensemble Kalman filter: Theoretical formulation and practical implementation". In: *Ocean dynamics* 53 (2003), pp. 343–367.

[60] Petar M Djuric et al. "Particle filtering". In: *IEEE signal processing magazine* 20.5 (2003), pp. 19–38.

[61] Frank Allgöwer et al. "Nonlinear predictive control and moving horizon estimation—an introductory overview". In: *Advances in control: Highlights of ECC'99* (1999), pp. 391–449.

[62] Qing-Guo Wang. "Control design for transient performance". In: *IEEE/CAA Journal of Automatica Sinica* 10.1 (2023), pp. 3–7.

[63] Torkel Glad and Lennart Ljung. *Control theory*. CRC press, 2018.

[64] Harry L Trentelman, Anton A Stoorvogel, and Malo Hautus. *Control theory for linear systems*. Springer Science & Business Media, 2012.

[65] THS Abdelaziz and M Valášek. "Pole-placement for SISO linear systems by state-derivative feedback". In: *IEE Proceedings-Control Theory and Applications* 151.4 (2004), pp. 377–385.

[66] Pierre OM Scokaert and James B Rawlings. "Constrained linear quadratic regulation". In: *IEEE Transactions on automatic control* 43.8 (1998), pp. 1163–1169.

[67] S Joe Qin and Thomas A Badgwell. "An overview of industrial model predictive control technology". In: *AIche symposium series*. Vol. 93. 316. New York, NY: American Institute of Chemical Engineers, 1971-c2002. 1997, pp. 232–256.

[68] Manfred Morari and Jay H Lee. "Model predictive control: past, present and future". In: *Computers & chemical engineering* 23.4-5 (1999), pp. 667–682.

[69] Richard Bellman. "Dynamic programming". In: *Science* 153.3731 (1966), pp. 34–37.

[70] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. "Reinforcement learning: A survey". In: *Journal of artificial intelligence research* 4 (1996), pp. 237–285.

[71] Kai Arulkumaran et al. "Deep reinforcement learning: A brief survey". In: *IEEE Signal Processing Magazine* 34.6 (2017), pp. 26–38.

[72] Yashen Lin et al. *Research roadmap on grid-forming inverters*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2020.

[73] Amirnaser Yazdani and Reza Iravani. *Voltage-sourced converters in power systems: modeling, control, and applications*. John Wiley & Sons, 2010.

[74] Karel De Brabandere et al. "A voltage and frequency droop control method for parallel inverters". In: *IEEE Transactions on power electronics* 22.4 (2007), pp. 1107–1115.

[75] Athanasios Vasilakis et al. "The evolution of research in microgrids control". In: *IEEE Open Access Journal of Power and Energy* 7 (2020), pp. 331–343.

[76] Yash P Bhatt and Mihir C Shah. "Design, analysis and simulation of synchronous reference frame based phase lock loop for grid connected inverter". In: *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*. IEEE. 2016, pp. 1–5.

[77] Jorge Elizondo Martinez. "Microrid operation strategy for improved recovery and inertial response after large disturbances". PhD thesis. Massachusetts Institute of Technology, 2016.

[78] Yousef Khayat et al. "Decentralized optimal frequency control in autonomous microgrids". In: *IEEE Transactions on Power Systems* 34.3 (2018), pp. 2345–2353.

[79] Guannan Lou et al. "Distributed MPC-based secondary voltage control scheme for autonomous droop-controlled microgrids". In: *IEEE transactions on sustainable energy* 8.2 (2016), pp. 792–804.

[80] Pengwei Du et al. "New ancillary service market for ERCOT". In: *IEEE Access* 8 (2020), pp. 178391–178401.

[81] Yuanyuan Shi et al. "Optimal battery control under cycle aging mechanisms in pay for performance settings". In: *IEEE Transactions on Automatic Control* 64.6 (2018), pp. 2324–2339.

[82] Bolong Cheng and Warren B Powell. "Co-optimizing battery storage for the frequency regulation and energy arbitrage using multi-scale dynamic programming". In: *IEEE Transactions on Smart Grid* 9.3 (2016), pp. 1997–2005.

[83] Scott J Moura, Nalin A Chaturvedi, and Miroslav Krstic. "PDE estimation techniques for advanced battery management systems—Part I: SOC estimation". In: *2012 American Control Conference (ACC)*. IEEE. 2012, pp. 559–565.

[84] Chitra Dangwal et al. "Pack Level State-of-Power Prediction for Heterogeneous Cells". In: *2022 American Control Conference (ACC)*. IEEE. 2022, pp. 1066–1073.

[85] Menglong Hao et al. "Efficient thermal management of Li-ion batteries with a passive interfacial thermal regulator based on a shape memory alloy". In: *Nature Energy* 3.10 (2018), pp. 899–906.

[86] Dylan Kato and Scott J Moura. "1d pde model for thermal dynamics in fluid-cooled battery packs: Numerical methods and sensor placement". In: *2021 American Control Conference (ACC)*. IEEE. 2021, pp. 3102–3107.

[87] Satadru Dey, Hector E Perez, and Scott J Moura. "Model-based battery thermal fault diagnostics: Algorithms, analysis, and experiments". In: *IEEE Transactions on Control Systems Technology* 27.2 (2017), pp. 576–587.

[88] Juan Arteaga and Hamidreza Zareipour. "A price-maker/price-taker model for the operation of battery storage systems in electricity markets". In: *IEEE Transactions on Smart Grid* 10.6 (2019), pp. 6912–6920.

[89] Juan M Morales et al. *Integrating renewables in electricity markets: operational problems*. Vol. 205. Springer Science & Business Media, 2013.

[90] Korosh Vatanparvar and Ratnesh Sharma. "Battery optimal approach to demand charge reduction in behind-the-meter energy management systems". In: *2018 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE. 2018, pp. 1–5.

[91] Jose Daniel Lara, Daniel E Olivares, and Claudio A Canizares. "Robust energy management of isolated microgrids". In: *IEEE Systems Journal* 13.1 (2018), pp. 680–691.

[92] Preet Gill et al. "State-Of-Health Estimation Pipeline for Li-ion Battery Packs with Heterogeneous Cells". In: *2022 American Control Conference (ACC)*. IEEE. 2022, pp. 1080–1086.

[93] Dong Zhang et al. "Remaining useful life estimation of lithium-ion batteries based on thermal dynamics". In: *2017 American Control Conference (ACC)*. IEEE. 2017, pp. 4042–4047.

[94] Bernard Knueven et al. *State-of-the-Art Techniques for Large-Scale Stochastic Unit Commitment*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2021.

[95] Eric Gimon. *How Market Rules Are Holding Back Energy Storage*. 2019. URL: https://www.greentechmedia.com/articles/read/energy-storage-wholesale-market-rules.

[96] *Global Grid-Scale Battery Market Size, Market Share, Application Analysis, Regional Outlook, Growth Trends, Key Players, Competitive Strategies and Forecasts, 2019 To 2027*. Acute Market Reports, 2020.

[97] *Technology Roadmap - Energy Storage*. 2014. URL: https://www.iea.org/reports/technology-roadmap-energy-storage.

[98] Paul L Joskow. "Challenges for wholesale electricity markets with intermittent renewable generation at scale: the US experience". In: *Oxford Review of Economic Policy* 35.2 (2019), pp. 291–331.

[99] Paul Denholm (NREL) Thomas Bowen Ilya Chernyakhovskiy. *Grid-Scale Battery Storage: Frequently Asked Questions*. 2017. URL: https://www.nrel.gov/docs/fy19osti/74426.pdf.

[100] KC Divya and Jacob Østergaard. "Battery energy storage technology for power systems—An overview". In: *Electric power systems research* 79.4 (2009), pp. 511–520.

[101] Federica Cucchiella, Idiano D'adamo, and Massimo Gastaldi. "Photovoltaic energy systems with battery storage for residential areas: an economic analysis". In: *Journal of Cleaner Production* 131 (2016), pp. 460–474.

[102] A Nottrott, Jan Kleissl, and Byron Washom. "Energy dispatch schedule optimization and cost benefit analysis for grid-connected, photovoltaic-battery storage systems". In: *Renewable Energy* 55 (2013), pp. 230–240.

[103] XY Wang, D Mahinda Vilathgamuwa, and San Shing Choi. "Determination of battery storage capacity in energy buffer for wind farm". In: *IEEE Transactions on Energy Conversion* 23.3 (2008), pp. 868–878.

[104] A Rajendra Prasad and E Natarajan. "Optimization of integrated photovoltaic–wind power generation systems with battery storage". In: *Energy* 31.12 (2006), pp. 1943–1954.

[105] Marija Petkovic. *Tesla big battery at Hornsdale earns record revenue in September 2019*. 2019. URL: https://reneweconomy.com.au/tesla-big-battery-at-hornsdale-earns-record-revenue-in-september-2019/.

[106] Arjan S Sidhu, Michael G Pollitt, and Karim L Anaya. "A social cost benefit analysis of grid-scale electrical energy storage projects: A case study". In: *Applied energy* 212 (2018), pp. 881–894.

[107] Sophie Vorrath Giles Parkinson. *The stunning numbers behind success of Tesla big battery*. 2018. URL: https://reneweconomy.com.au/the-stunning-numbers-behind-success-of-tesla-big-battery-63917/.

[108] Yuhua Du, Rishabh Jain, and Srdjan M Lukic. "Effects of battery degradation on economic viability of energy storage systems participating in regulation markets". In: *2017 IEEE Power & Energy Society General Meeting*. IEEE. 2017, pp. 1–5.

[109] AEMO. *Hornsdale Wind Farm 2 FCAS Trial*. 2018.

[110] *Noen Hornsdale Wind Farm FCAS Trial Report*. Neoen, 2018.

[111] Rodolfo Dufo-López. "Optimisation of size and control of grid-connected storage under real time electricity pricing conditions". In: *Applied Energy* 140 (2015), pp. 395–408.

[112] Mostafa Kazemi et al. "Operation scheduling of battery storage systems in joint energy and ancillary services markets". In: *IEEE Transactions on Sustainable Energy* 8.4 (2017), pp. 1726–1735.

[113] Guannan He et al. "Optimal bidding strategy of battery storage in power markets considering performance-based regulation and battery cycle life". In: *IEEE Transactions on Smart Grid* 7.5 (2015), pp. 2359–2367.

[114] Mauricio BC Salles et al. "Potential arbitrage revenue of energy storage systems in PJM". In: *Energies* 10.8 (2017), p. 1100.

[115] Raymond H Byrne et al. "Opportunities for energy storage in CAISO: Day-ahead and real-time market arbitrage". In: *2018 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*. IEEE. 2018, pp. 63–68.

[116] Hamed Mohsenian-Rad. "Coordinated price-maker operation of large energy storage units in nodal energy markets". In: *IEEE Transactions on Power Systems* 31.1 (2015), pp. 786–797.

[117] Jonathan William Mather. "Market Design and Analysis for Uncertain, Flexible, and Decentralized Power Systems". PhD thesis. UC Berkeley, 2018.

[118] Steven A Gabriel et al. *Complementarity modeling in energy markets*. Vol. 180. Springer Science & Business Media, 2012.

[119] Trevor Alvey et al. "A security-constrained bid-clearing system for the New Zealand wholesale electricity market". In: *IEEE Transactions on power systems* 13.2 (1998), pp. 340–346.

[120] URL: `https://github.com/GuillaumeGoujard/LMP_NZ`.

[121] David Young, Stephen Poletti, and Oliver Browne. "Can agent-based models forecast spot prices in electricity markets? Evidence from the New Zealand electricity market". In: *Energy Economics* 45 (2014), pp. 419–434.

[122] Ran Fu, Timothy W Remo, and Robert M Margolis. *2018 US utility-scale photovoltaics-plus-energy storage system costs benchmark*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2018.

[123] Kandler Smith et al. "Life prediction model for grid-connected Li-ion battery energy storage system". In: *2017 American Control Conference (ACC)*. IEEE. 2017, pp. 4062–4068.

[124] Caitlin D Parke et al. "Progress on continuum modeling of lithium–sulfur batteries". In: *Sustainable Energy & Fuels* 5.23 (2021), pp. 5946–5966.

[125] Sizhe Wang et al. "Insight into MoS2–MoN Heterostructure to Accelerate Polysulfide Conversion toward High-Energy-Density Lithium–Sulfur Batteries". In: *Advanced Energy Materials* 11.11 (2021), p. 2003314.

[126] Ting-Zheng Hou et al. "The formation of strong-couple interactions between nitrogen-doped graphene and sulfur/lithium (poly) sulfides in lithium-sulfur batteries". In: *2D Materials* 2.1 (2015), p. 014011.

[127] Karthikeyan Kumaresan, Yuriy Mikhaylik, and Ralph E White. "A mathematical model for a lithium–sulfur cell". In: *Journal of the electrochemical society* 155.8 (2008), A576.

[128] Monica Marinescu, Teng Zhang, and Gregory J Offer. "A zero dimensional model of lithium–sulfur batteries during charge and discharge". In: *Physical Chemistry Chemical Physics* 18.1 (2016), pp. 584–593.

[129] Abbas Fotouhi et al. "Lithium-sulfur battery technology readiness and applications—a review". In: *Energies* 10.12 (2017), p. 1937.

[130] Chu Xu et al. "Online state estimation for a physics-based Lithium-Sulfur battery model". In: *Journal of Power Sources* 489 (2021), p. 229495.

[131] Zhijia Huang et al. "On Electrochemical Model-based State Estimation for Lithium-Sulfur Batteries". In: *ongoing review* (2022).

[132] Gérard Biau and Erwan Scornet. "A random forest guided tour". In: *Test* 25 (2016), pp. 197–227.

[133] Giancarlo Ferrari-Trecate, Domenico Mignone, and Manfred Morari. "Moving horizon estimation for hybrid systems". In: *IEEE transactions on automatic control* 47.10 (2002), pp. 1663–1676.

[134] Marcello Torchio et al. "Design of Piecewise Affine and Linear Time-Varying Model Predictive Control Strategies for Advanced Battery Management Systems". In: *Journal of The Electrochemical Society* 164.4 (Mar. 2017), A949. DOI: 10.1149/2.0201706jes. URL: https://dx.doi.org/10.1149/2.0201706jes.

[135] C Dangwal et al. "Global Sensitivity Analysis of 0-D Lithium Sulfur Electrochemical Model". In: *IFAC World Congress 2023*. 2023.

[136] Leo Breiman et al. "Cart". In: *Classification and regression trees* (1984).

[137] Alberto Bemporad and Manfred Morari. "Control of systems integrating logic, dynamics, and constraints". In: *Automatica* 35.3 (1999), pp. 407–427.

[138] Mickey Francis. *Renewables Became the Second-Most Prevalent U.S. Electricity Source in 2020*. URL: https://www.eia.gov/todayinenergy/detail.php?id=48896#.

[139] Moritz Diehl. "Airborne wind energy: Basic concepts and physical foundations". In: *Airborne wind energy*. Springer, 2013, pp. 3–22.

[140] Cristina L Archer and Ken Caldeira. "Global assessment of high-altitude wind power". In: *Energies* 2.2 (2009), pp. 307–319.

[141] Lorenzo Fagiano, Mario Milanese, and Dario Piga. "High-altitude wind power generation". In: *IEEE Transactions on Energy Conversion* 25.1 (2009), pp. 168–180.

[142] Cristina L Archer, Luca Delle Monache, and Daran L Rife. "Airborne wind energy: Optimal locations and variability". In: *Renewable Energy* 64 (2014), pp. 180–186.

[143] Antonello Cherubini et al. "Airborne Wind Energy Systems: A review of the technologies". In: *Renewable and Sustainable Energy Reviews* 51 (2015), pp. 1461–1476.

[144] Alireza Bafandeh and Chris Vermillion. "Real-time altitude optimization of airborne wind energy systems using Lyapunov-based switched extremum seeking control". In: *2016 American Control Conference (ACC)*. IEEE. 2016, pp. 4990–4995.

[145] Chris Vermillion et al. "Model-based plant design and hierarchical control of a prototype lighter-than-air wind energy system, with experimental flight test results". In: *IEEE Transactions on Control Systems Technology* 22.2 (2013), pp. 531–542.

[146] Cécile Penland and Prashant D Sardeshmukh. "The optimal growth of tropical sea surface temperature anomalies". In: *Journal of climate* 8.8 (1995), pp. 1999–2024.

[147] Cecile Penland. "Random forcing and forecasting using principal oscillation pattern analysis". In: *Monthly Weather Review* 117.10 (1989), pp. 2165–2185.

[148] Hermann Mena and Lena Pfurtscheller. "An efficient SPDE approach for El Niño". In: *Applied Mathematics and Computation* 352 (2019), pp. 146–156.

[149] Emil B Iversen et al. "Short-term probabilistic forecasting of wind speed using stochastic differential equations". In: *International Journal of Forecasting* 32.3 (2016), pp. 981–990.

[150] Kenneth Scerri, Michael Dewar, and Visakan Kadirkamanathan. "Estimation and model selection for an IDE-based spatio-temporal model". In: *IEEE Transactions on Signal Processing* 57.2 (2008), pp. 482–492.

[151] Daniel Russo and Benjamin Van Roy. "Learning to optimize via information-directed sampling". In: *Operations Research* 66.1 (2018), pp. 230–252. ISSN: 15265463. DOI: 10.1287/opre.2017.1663. arXiv: 1403.5556.

[152] Shamir Bin-Karim, Alireza Bafandeh, and Christopher Vermillion. "Spatio-temporal optimization through model predictive control: A case study in airborne wind energy". In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 2016, pp. 4239–4244.

[153] Alireza Bafandeh. "Hierarchical Control Strategies for Spatiotemporally Varying Systems with Application to Airborne Wind Energy". PhD thesis. The University of North Carolina at Charlotte, 2018.

[154] Ali Baheri and Christopher Vermillion. "Altitude optimization of airborne wind energy systems: A Bayesian optimization approach". In: *2017 American Control Conference (ACC)*. IEEE. 2017, pp. 1365–1370.

[155] Ali Baheri et al. "Real-time control using Bayesian optimization: A case study in airborne wind energy systems". In: *Control Engineering Practice* 69 (2017), pp. 131–140.

[156] Laurel N Dunn et al. "On Wind Speed Sensor Configurations and Altitude Control in Airborne Wind Energy Systems". In: *2019 American Control Conference (ACC)*. IEEE. 2019, pp. 2197–2202.

[157] Chao Qin, Diego Klabjan, and Daniel Russo. "Improving the expected improvement algorithm". In: *arXiv preprint arXiv:1705.10033* (2017).

[158] Patrick Keyantuo et al. "A Vector Auto-Regression Based Forecast of Wind Speeds in Airborne Wind Energy Systems". In: *2021 IEEE Conference on Control Technology and Applications (CCTA)*. 2021.

[159] Laurel N Dunn. *Data-Driven Decision Analysis in Electric Power Systems*. University of California, Berkeley, 2020.

[160] Cristina L. Archer. *Wind profiler at Cape Henlopen*. URL: https://www.ceoe.udel.edu/%20our-people/profiles/carcher/fsmw.

[161] Suwanchai Sangsuk-Iam and Thomas E Bullock. "Analysis of discrete-time Kalman filtering under incorrect noise covariances". In: *IEEE Transactions on Automatic Control* 35.12 (1990), pp. 1304–1309.

[162] H Heffes. "The effect of erroneous models on the Kalman filter response". In: *IEEE Transactions on Automatic Control* 11.3 (1966), pp. 541–543.

[163] Robert F Engle, Olivier Ledoit, and Michael Wolf. "Large dynamic covariance matrices". In: *Journal of Business & Economic Statistics* 37.2 (2019), pp. 363–375.

[164] Stuart Gibson and Brett Ninness. "Robust maximum-likelihood estimation of multivariable dynamic systems". In: *Automatica* 41.10 (2005), pp. 1667–1682.

[165] Simo Särkkä. *Bayesian filtering and smoothing*. 3. Cambridge University Press, 2013.

[166] Y. Dvorkin et al. "Comparison of scenario reduction techniques for the stochastic unit commitment". In: *2014 IEEE PES General Meeting — Conference Exposition*. July 2014, pp. 1–5. DOI: 10.1109/PESGM.2014.6939042.

[167] P. Pinson et al. "Properties of Quantile and Interval Forecasts of Wind Generation and their Evaluation". In: *Proceedings of the European Wind Energy Conference & Exhibition*. http://www.ewea.org. Athens, 2006. URL: http://www2.imm.dtu.dk/pubdb/p.php?4250.

[168] P Pinson et al. "From probabilistic forecasts to statistical scenarios of short-term wind power production". In: *Wind Energy* 12 (1 2008), pp. 51–62.

[169] Pierre Pinson and Robin Girard. "Evaluating the quality of scenarios of short-term wind power generation". In: *Applied Energy* 96 (2012), pp. 12–20.

[170] Didem Sari et al. "Statistical metrics for assessing the quality of wind power scenarios for stochastic unit commitment". In: *Wind Energy* 19.5 (2016), pp. 873–893. ISSN: 1099-1824. DOI: 10.1002/we.1872. URL: http://dx.doi.org/10.1002/we.1872.

[171] David L. Woodruff et al. "Constructing Probabilistic Scenarios for Wide-Area Solar Power Generation". In: *Solar Energy* 160.15 (2018), pp. 153–167.

[172] Andrea Staid et al. "Generating short-term probabilistic wind power scenarios via nonparametric forecast error density estimators". In: *Wind Energy* 20.12 (2017), pp. 1911–1925.

[173] Yonghan Feng et al. "Toward Scalable Stochastic Unit Commitment. Part 1: Load Scenario Generation". In: *Energy Systems* 6.3 (2015), pp. 309–329.

[174] Juan M. Morales, Roberto Minguez, and Antonio J. Conejo. "A methodology to generate statistically dependent wind speed scenarios". In: *Applied Energy* 87.3 (2010), pp. 843–855.

[175] Jianhui Wang, Mohammad Shahidehpour, and Zuyi Li. "Security-constrained unit commitment with volatile wind power generation". In: *IEEE Transactions on Power Systems* 23.3 (2008), pp. 1319–1327.

[176] Pierre Pinson et al. "From probabilistic forecasts to statistical scenarios of short-term wind power production". In: *Wind Energy* 12.1 (2009), pp. 51–62.

[177] Kailash Chand Sharma, Prerna Jain, and Rohit Bhakar. "Wind power scenario generation and reduction in stochastic programming framework". In: *Electric Power Components and Systems* 41.3 (2013), pp. 271–285.

[178] Pierre Pinson and Julija Tastu. "Discrimination ability of the Energy score". In: *Technical Report* (2013). URL: `http://orbit.dtu.dk/fedora/objects/orbit:122326/datastreams/file_b919613a-9043-4240-bb6c-160c88270881/content`.

[179] Stephan Kolassa and Roland Martin. "Percentage Errors Can Ruin Your Day (and Rolling the Dice Shows How)". In: *Foresight: The International Journal of Applied Forecasting* 23 (2011), pp. 21–27. URL: `https://EconPapers.repec.org/RePEc:for:ijafaa:y:2011:i:23:p:21-27`.

[180] Hans Bludszuweit, José Antonio Domínguez-Navarro, and Andrés Llombart. "Statistical analysis of wind power forecast error". In: *IEEE Transactions on Power Systems* 23.3 (2008), pp. 983–991.

[181] Bahar Biller and Barry L. Nelson. "Fitting Time-Series Input Processes for Simulation". In: *Operations Research* 53 (2005), pp. 549–559.