Data and technology-driven improvements to electricity market design

Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy
in
Engineering and Public Policy

Luke A. Lavin

B.A., Physics and Anthropology, Amherst College

Carnegie Mellon University
Pittsburgh, PA

May, 2021

Acknowledgments

When I moved out to California I did not know if I would ever come back to life in the eastern time zone. Thanks to Carnegie Mellon University's Engineering and Public Policy department for making the return more rewarding and edifying than I could have imagined. Jay Apt has been the best advisor I could have wished for, allowing me to define my own research path over the past four years while providing prompt feedback and access to input from luminaries in the field whose work I had admired for years. Granger Morgan, Steve Rose, and Ben Hobbs have made up a dissertation committee with more collective forgotten knowledge on historical theory and practice of electricity market design than I could ever hope to obtain. Granger ensures my communication stays concise and policy-oriented. Steve has provided industry insight and ever-important high-quality data in addition to edits and his time. Ben graciously agreed to my entreaty to serve on the committee, and has drawn my attention to multiple theoretical and methodological debates on competitive electricity market design that laid a framework for what I observe in my work long before I was reading academic publications. I would not have learned the skills that made this research possible without the help of many students in EPP. Thanks are owed to Jeremy Keen, Brian Sergi, Gerad Freeman, Evan Sherwin, Priya Donti, Guannan He, Michael Craig, Mike Rath, Arthur Yip, Yamit Lavi, and Sean Smillie, among others, for making my code cleaner and wit sharper. Special thanks to Sinnott Murphy for the yeoman's work in processing and publishing on GADS data to make the research in Chapter 3 possible. In addition to Steve Rose, Roger Lueken and Seth Blumsack have patiently provided feedback on both research and life after CEIC and EPP. Thanks to Doug Sicker, Peter Adams, Vicki Finney, Adam Loucks, and Debbie Kuntz for their support of both me and commitment to what EPP is as a program. Thanks again go to Ben Hobbs for connecting us to Nik Zheng, my co-author and mathematics sounding board on the final publications in this dissertation. Nik worked tirelessly to address coding and other concerns during a pandemic where we could not collaborate in person as usual. I am confident he will provide even more insight in his own doctoral research. From work life before CMU thanks go to Marian Swain, Ana Mileva, Zach

Ming, Brian Conlon, Lucy McKenzie, Andy Satchwell, and Andrew Mills. It was essential to know what I did not know but could learn and contribute prior to starting doctoral research. These friends and mentors, among others, provided that insight. Thanks to my parents, Kevin and Mora, for endlessly encouraging my intellectual interests for as long as I can remember. My father always cited providing his children the best possible education as a primary motivator for his own long hours of work. I hope this completes my end of the deal. Thanks are owed as well to my brothers Joe and Tom for providing the color commentary. And to the Washington Nationals for winning the 2019 World Series. Finally, to Caitlin Lavin for being my co-pilot in all of this, especially through her own graduation, getting married in the midst of a pandemic, and the long hours in recent months. Words do not do justice to how essential your encouragement has been every step of the way from moving to Pittsburgh through graduation. This work was supported by the Carnegie Mellon Climate and Energy Decision Making Center (CEDM), formed through a cooperative agreement between the National Science Foundation and CMU (SES-0949710), and in part by the Richard King Mellon Foundation. Thanks to the funders as well as the Carnegie Mellon Electricity Industry Center for making this research possible.

Abstract

Wholesale electricity market and retail tariff design often uses anachronistic assumptions based on existing technology characteristics or historical computational and data limitations. This dissertation conducts four categories of analysis on how electricity market design can be modernized to increase efficiency and avert roadblocks to economy-wide deep decarbonization. First, electric utilities can capture most of the system benefit of customer-sited energy storage resource (ESR) adoption with critical peak pricing (CPP). CPP is proposed as a simple and Pareto-improving rate design for commercial and industrial customers with ESRs, similar to time-of-use rates for electric vehicle owners. Second, previous research quantifying correlated generator failures in the PJM Interconnection can be incorporated in both resource adequacy and scarcity pricing using an operating reserve demand curve (ORDC). Because correlated failures occur at very high and low temperatures when electricity demand is highest, there are substantial effects on target winter and summer planning reserve margins and increased social welfare from better accounting for generator failure probability when designing ORDCs. Third, as electricity markets evolve toward higher shares of variable, low marginal cost resources with ESRs new rules are needed to ensure resources' full, competitive participation. A bi-level model implemented on a realistic, high renewables nodal test system highlights strategies ESRs and hybrids can use to raise prices, particularly cross-product and decongestion strategies, and suggests offer uniformity over co-optimized temporal intervals as a remedy. Fourth, metrics like effective load carrying capability (ELCC) are increasingly common for quantifying the system-dependent contribution of variable generation and ESRs to resource adequacy. Extending these methods to a zonal evaluation of resource adequacy using data from the Midcontinent Independent System Operator (MISO) shows transmission and ESRs have a complementarity benefit in zonal resource adequacy that is not realized by variable nor conventional generators, suggesting the importance of increased zonal representation in planning with ESRs.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

Wholesale electricity market and retail tariff design often uses anachronistic assumptions based on existing technology characteristics or historical computational and data limitations. This dissertation conducts four categories of analysis across six chapters on how electricity market design can be modernized to increase efficiency and avert roadblocks to economy-wide deep decarbonization.

Chapter 2 develops an approach for optimizing the use of customer-sited energy storage resource (ESR) to increase electric utility avoided costs. Most of the system benefits of customer-sited ESR can be captured by the utility through peak pricing in a small subset of hours with minimal effect on the value of the ESR to the customer. However, these tariffs remain uncommon nationally, even for commercial and industrial customers. For policy simplicity, critical peak pricing can be thought of as appropriate for ESR-oriented retail rate design in a manner similar to the applicability of time-of-use tariffs for electric vehicle owners. More generally, ESRs may serve as an impetus for forward-thinking rate design modifications in coming years.

Chapters 3 and 4 draw on a model of generator failures developed by Sinnott Murphy [1] during his doctoral research to conduct two published analyses: one on resource adequacy implications in the PJM Interconnection as Chapter 3 [2], and one on better incorporating correlated generator failures in PJM's energy markets through dynamic operating reserve procurement as Chapter 4 [3]. Both papers apply insights from Dr. Murphy's research to an existing wholesale electricity market; the latter paper specifically proposes improvements to PJM's formulation of its operating reserve demand curves (ORDC) to better endogenize reserve procurement at the high and low temperatures when demand is highest and scarcity most commonly occurs.

Chapters 5 and 6 contribute insight for ongoing discussion at North American system operators on holistic competitive participation of energy storage resources (ESRs) with ability to set price for services of which they are technically capable in markets. In Chapter 5 we develop a modeling approach for identifying profit-maximizing bidding strategies for ESR- or hybrid-owning entities in a realistically parameterized multi-node, two-settlement electricity market with high penetration

of variable, low marginal cost resources. We identify cross-product and decongestion strategies by which ESRs and hybrids can commonly raise clearing prices above competitive levels. We propose an offer uniformity remedy that helps maintain competitive outcomes in delivering customers low-cost, reliable service under increasing ESR and hybrid penetrations.

Chapter 6 extends the model developed in Chapter 5 to focus on a specific, profitable strategy using resource hybridization to decongest generation pockets. Cases are developed to highlight the profitability of this strategy. Chapter 6 can be bypassed by readers of Chapter 5 without missing the general findings from the dissertation.

Chapter 7 quantifies concerns about incorporating ESRs in zonal resource adequacy (RA) constructs common in North American markets with centralized capacity auctions. RA modeling typically considers contributions of generators with limited or no representation of the transmission system. The ability of ESRs to substitute for transmission by time-shifting load makes it increasingly important to consider transmission limits in computing resource adequacy to accurately quantify the relative contribution of ESRs and generators to reliability metrics. We show how ESRs and transmission have a complementarity benefit in RA that does not apply for generators using a realistic zonal test system based on data from the Midcontinent Independent System Operator (MISO).

References

[1] S. Murphy, F. Sowell, and J. Apt, "A time-dependent model of generator failures and recoveries captures correlated events and quantifies temperature dependence," *Applied Energy*, vol. 253, p. 113513, 2019. doi:https://doi.org/10.1016/j.apenergy.2019.113513.

[2] S. Murphy, L. Lavin, and J. Apt, "Resource adequacy implications of temperature-dependent electric generator availability," *Applied Energy*, vol. 262, p. 114424, 2020. doi:https://doi.org/10.1016/j.apenergy.2019.114424.

[3] L. Lavin, S. Murphy, B. Sergi, and J. Apt, "Dynamic operating reserve procurement improves scarcity pricing in PJM," *Energy Policy*, vol. 147, p. 111857, 2020. doi:https://doi.org/10.1016/j.enpol.2020.111857.

Chapter 2 The Importance of Peak Pricing in Realizing System Benefits From Distributed Storage*

Abstract

A fundamental policy question for distributed energy resources (DER) is whether they create system benefits shared by all utility customers in addition to being profitable for the installing customer. This question has received considerable attention in "value of DER" and net metering reform proceedings for behind-the-meter solar photovoltaics in recent years. Commercial customer-sited lithium-ion batteries with a primary use case of demand charge management are forecast to greatly increase in the coming decade due to falling storage costs, making comparison of their customer and system benefits a timely topic in DER valuation. We conduct a national overview of the system benefits of commercial customer-sited storage and find system benefits will not be realized for many commercial customer-sited storage installations in the absence of incentives for storage dispatch during the top 50-100 annual hours that drive grid infrastructure investment. Regulatory implementation of default peak pricing during a small subset of annual hours for customer-sited storage can realize additional system benefits and offer Pareto improvement. Additional transparency in regulatory estimates of these system benefits helps catalyze longer-term visions for increased competition at the retail level using DERs.

* This paper is in review as L. Lavin and J. Apt "The Importance of Peak Pricing in Realizing System Benefits from Distributed Storage," *Energy Policy*. An earlier version received the 2019 Herbert L. Toor award for best EPP Part A qualifying paper.

Abbreviations and acronyms

5CP	5 coincident Peak hours in PJM
AMI	Advanced Metering Infrastructure
BTM	Behind the Meter
C&I	Commercial and Industrial
CBP	Capacity Bidding Program
ComEd	Commonwealth Edison, an Illinois-based electric distribution utility
CP	Coincident Peak
CPP	Critical Peak Pricing
CPR	Critical Peak Rebate
CPUC	California Public Utilities Commission
DER	Distributed Energy Resources
DLCO	Duquesne Light Company, a Pennsylvania-based electric distribution utility
DRAM	Demand Response Auction Mechanism
DRV	Demand Reduction Value
DRIPE	Demand Reduction Induced Price Effect
eGRID	Emissions and Generation Resource Integrated Database
EIA	United States Energy Information Administration
ERCOT	Electricity Reliability Council of Texas
FERC	Federal Energy Regulatory Commission
IOU	Investor-Owned Utility
LMP	Locational Marginal Price
LSE	Load-Serving Entity
LSRV	Locational System Relief Value
NARUC	National Association of Regulatory Utility Commissioners
NCP	Non-Coincident Peak
NEM	Net Energy Metering
NPV	Net Present Value
NREL	National Renewable Energy Laboratory
PCT	Participant Cost Test
PG&E	Pacific Gas and Electric, a California-based electric distribution utility
PURPA	Public Utilities Regulatory Policy Act
PV	Photovoltaic
REV	Reforming the Energy Vision
RIM	Ratepayer Impact Measure
RPS	Renewable Portfolio Standard
RTP	Real-Time Pricing
SGIP	Self-Generation Incentive Program
T&D	Transmission and Distribution
TOD	Time-of-day
TOU	Time-of-use
TRC	Total Resource Cost Test
VDER	Value of Distributed Energy Resources

2.1 Introduction

Rapidly falling costs of stationary battery storage, particularly lithium-ion chemistries [1], are driving increasingly aggressive forecasts for deployment at scale in the coming decade [2]. Private storage owners will tend to operate their storage assets to minimize their costs. That operation may either increase or decrease benefits to the system as a whole. Here we examine whether simple updates to customer rate structures can offer Pareto improvement for private customer and public system benefits.

We focus on behind-the-meter (BTM) battery storage, the deployment of which may be significant in the coming decade [2]. BTM storage can be further segmented into residential and non-residential segments; we focus on non-residential customers. Non-residential utility customers' electric tariffs typically have demand charges (a levy on their highest period of electricity consumption during a time period, often 15 minutes of every month). Non-residential customers also more commonly have time-of-use (TOU)[1] tariffs for their energy consumption than residential classes. For both demand and time-varying energy charges, the time-differentiation of billing creates an opportunity for storage to shift the customer's metered energy consumption and lower bills.

Previous literature focuses on the combination of storage cost, demand charges, and TOU bill management that make storage installation cost-effective from the customer perspective [3, 4]. In the parlance of distributed energy resource (DER) cost-effectiveness [5, 6], this literature evaluates whether distributed storage passes the Participant Cost Test (PCT), which compares the installing customer's ("participant") savings on their electricity bill to the cost the customer pays to install the DER subject to assumptions about future electricity bills and DER financing and lifetime. Passing the PCT under realistic assumptions is a key first screen for potential DER adoption, but a logical next step is to look at how the DER fares under a Total Resource Cost (TRC) or Ratepayer Impact Measure (RIM) Test. The TRC evaluates whether a DER creates benefits for society by comparing avoided costs against the costs of installing and maintaining the DER, while the RIM evaluates whether the DER creates benefits for the utility and/or non-participating utility customers on the on

[1]Both TOU and time-of-day (TOD) typically refer to "usage over broad blocks of hours (e.g., on-peak=6 hours for summer weekday afternoon; off-peak = all other hours in the summer months) where the price for each period is predetermined and constant" (REF: https://www.smartgrid.gov/recovery_act/time_based_rate_programs.html).

the system where the DER is installed by comparing participant bill savings against avoided costs.

The RIM is particularly relevant because it provides a window into how utilities and load-serving entities (LSEs) might respond to increased penetration of distributed storage on their system and, ultimately, it is those entities who propose the electricity rates on which the customer-sited cost-effectiveness of installation depends. Distributed solar photovoltaics (PV) installation has been generally cost-effective for customers when credited under net energy metering (NEM) at the full applicable retail rate. However, as installation has increased, studies have shown mixed results on whether full retail compensation accurately reflects the system value of distributed PV [7–9].

When full retail compensation is found to exceed a DER's system value, this can beget cost-shifting (or cross-subsidization): because a utility is guaranteed full cost recovery from its customers, rates are raised on non-participants to pay for the gap between DER compensation and its value to the system for participants who install that DER, though the financial details of this mechanism vary [10]. Utilities and state regulatory commissions, proffering this cost-shifting argument [11–13], have responded by moving away from full retail compensation or changing underlying rate structures to better reflect marginal cost of service for classes with significant distributed PV penetration in numerous jurisdictions in recent years. Because rate design can change in response to DER installation [14] and the key argument for doing so revolves around those resources doing poorly under the RIM, it is useful to evaluate distributed storage in the RIM framework before accepting that the combination of falling costs and current rate designs will lead to high levels of penetration. This paper seeks to fill that lacuna in the published literature on distributed storage.

We show that, in general, customer operation of storage leads to operation at times that do not relieve the peak system load and thus does not avoid system costs. This is because customer peaks that storage can reduce often occur in the early afternoon but system avoided cost peaks are in the late afternoon. In areas that install both storage and solar, large quantities of midday low marginal cost solar generation move system net load peaks and associated capacity-related avoided costs toward evening, further reducing the match between storage operation and system peak load. We find that dynamic rates are a simple and effective way to ensure that customer storage helps the grid. A full move to more complex real-time pricing may not be required to realize most system

avoided costs from customer-sited storage. Simpler critical peak pricing (CPP) rates targeted at the top 50-100 avoided cost hours can improve alignment of customer and system benefits for storage adoption.

2.2 Background

We include background on avoid costs and rate design to contextualize our research contribution.

2.2.1 Avoided Costs

Avoided costs are generally defined as a LSE's marginal cost of service, and are often used as a proxy for system value of a DER in the applicable service territory. Common monetized avoided cost components include energy, losses, ancillary services, capacity, and transmission and distribution (T&D) deferral, though many other components may be considered [15]. Some of these cost components are difficult to estimate and vary at a sub-jurisdictional level; for example, how to estimate distribution system value of DERs has been an area of debate [16] and focus of policy initiatives like New York's Value of DER (VDER) proceedings [17]. Additional jurisdiction-dependent components can include avoided policy compliance costs, such as Renewable Portfolio Standard (RPS) obligations. Avoided costs may also include societal components not directly monetized by the LSE; common societal avoided costs include greenhouse gases and other pollutants. Financial avoided costs like the market price response[2] [18,19] and avoided fuel price hedging costs may also be considered. Reliability and resilience benefits are another potential avoided cost; avoided customer outages may be particularly germane to storage if properly configured for islanded operation during outage events [20]. Our work considers only the main set of monetized avoided costs as a base case excluding financial and reliability/resilience related avoided costs, with jurisdiction-relevant monetized adders and consideration of societal avoided costs in sensitivity analysis. This focus allows for more direct evaluation of rate design alternatives but may not comprehensively include all avoided costs relevant to a particular jurisdiction.

[2]Sometimes also called Demand Reduction Induced Price Effect (DRIPE) in the context of energy efficiency programs, and the merit order effect in the context of wholesale markets.

2.2.2 Rate Design

Economics and policy literature have for decades considered the efficiency of electric utility rate structures balanced against other goals such as equity, transparency, stability, and ease or technical feasibility of implementation. This general framework for ratemaking is often said to follow the Bonbright Principles [21] and continues be applied in industry documents [22]. In the past decade diffusion of advanced metering infrastructure (AMI) has lowered technical barriers to implementation of more complicated and economically efficient rate designs; simultaneously, increasing adoption of DERs has created impetus for re-considering cost-causation in rates. [23] lays out the general parameters and parties of this debate over the future of rate design, centered around fixed cost recovery. More comprehensive reviews of alternative rate designs with a focus on future applications are available in [24, 25]. We note the importance of considering multiple objectives in rate design in interpreting our results.

2.3 Methods

We quantify the system benefits achieved by behind-the-meter C&I storage installation under realistic parameterizations of current rate design and avoided costs. Six different utility tariffs are each applied to 30 or more customer load shapes representative of common C&I customer types. BTM storage is then sized to 20% of the customer's peak load and its dispatch optimized as a linear program to minimize the customer bill. We compare customer bill savings to estimates of utility avoided costs from the optimized BTM storage dispatch and identify simple changes to rate structure that offer Pareto improvement by increasing system avoided costs without reducing customer bill savings.

We adopt a Ratepayer Impact Measure (RIM) framework comparing customer (participant) bill savings from a storage installation to bulk-system avoided costs from that same storage installation. Importantly, neither side of this comparison says anything about the cost of the installed storage device; we assume the combination of falling costs, subsidies, and rate design identified in region-specific adoption-focused market research [20, 26] and higher-level national evaluations [3, 27, 28] can

make customer-sited storage an attractive economic proposition in many US utility jurisdictions. Bill savings are reductions in the customer's annual utility bill and depend on a jurisdiction's rate design for the applicable customer. Avoided costs consist of avoided energy, generation capacity, losses, transmission and distribution deferral, again in the applicable jurisdiction, and are discussed in more detail below and in Appendix A.3. Typical non-residential utility customer bills are made up of customer or fixed charges ($/month), energy charges ($/kWh), and demand charges ($/kW). Applicable demand and energy charges may additionally vary by time period. For example, when energy charges vary according to predetermined time periods, this is known as a "time-of-use" (TOU) rate. When energy charges vary hourly with wholesale prices on the bulk electric grid, this is known as "real-time-pricing" (RTP) [29].

To reduce customer bills we size a four-hour duration battery to 20% of the applicable customer's peak load. While it is possible to optimize battery size as a function of battery costs, customer load, and applicable customer rate [30–32], this process adds considerable computational overhead, and [30] ultimately finds 3-5 hour duration batteries to be optimal at medium and high battery cost levels across a few of the most common representative Department of Energy (DOE) commercial load shapes [33] in four climate locations. For battery capacity we follow [4] in sizing to 20% of peak load in our base cases. Customer load shapes are explained further in the Section 2.4 Data and Appendix A.1, and are not assumed to be modified by the addition of any other DERs, though in practice co-siting batteries with solar PV to take advantage of the federal investment tax credit or state and local incentives may be common. We also assume load shapes are not modified by behavioral responses to battery installation.

Given a battery size, customer load shape, and applicable customer rate, the battery is dispatched to minimize the customer's bill. Dispatch is implemented as a cost-minimizing linear program in Pyomo with the open-source GLPK solver, minimizing the objective function of the customer's annual bill shown schematically below and detailed fully in Appendix B.

$$Min \sum_{t=1...8760} Load_t * \$Energy_t + MaxLoad * \$Demand + \$Customer \tag{2.1}$$

Where $Load_t$ is the customer load net of battery dispatch, $MaxLoad$ is the customer's peak load net of battery dispatch, and $Energy_t$, $\$Demand$, and $\$Customer$ are the customer's energy ($/kWh), demand ($/kW), and customer or fixed ($/kW-mo.) charges on the applicable rate schedule.

Battery dispatch is optimized on a monthly basis with perfect foresight in the base case. Twelve months of dispatch are summed for annual bill savings totals. Monthly optimization with hourly dispatch resolution is chosen for the optimization because demand charges are commonly levied on the customer's highest 15-60 minutes of electricity use in a month [34]. A perfect foresight assumption is common in literature and sets an upper bound on customer bill savings, but of course will not be achieved in practice; industry contacts suggest contracts for commercial storage more commonly guarantee 70-80% reduction in demand charges compared to perfect foresight [35]. In sensitivity analysis included in Appendix C we set a lower bound on customer bill savings with persistence forecast dispatch for two cases. Persistence forecast dispatch achieves 57-63% of bill savings for load shapes used in our base case. Finally, dispatch is implemented with consideration of battery degradation as future foregone revenue by setting the marginal benefit of use to $5/MWh for our peak shaving and bill management applications based on [36].

Avoided costs are calculated after the battery is sized and dispatched to minimize the customer's bill. Avoided costs are also annual with hourly granularity. To the best of our knowledge, the only US jurisdiction to historically release generic hourly avoided costs for its investor-owned utilities (IOUs) is the California Public Utilities Commission (CPUC) [37], though new estimates are available for New York as part of the Value of DER (VDER) proceedings [17, 38]. Other utilities or jurisdictions often file avoided costs for either energy efficiency programs or qualifying facility tariffs under the Public Utilities Regulatory Policy Act (PURPA), but are at either an annual or seasonal time-of-day granularity. Because of the short duration of battery storage charge and discharge cycles compared to availability of more common DERs like rooftop PV, combined heat and power, and energy efficiency, it is more important to evaluate avoided costs at (sub-)hourly granularity for batteries. Therefore, one recommendation coming from this work is for jurisdictions across the country to estimate time-dependent avoided costs hourly to better ascertain the cost-effectiveness of DERs. These estimates have additional benefits beyond batteries for other DERs, for example, [39] highlights

the importance of time-dependent valuation of energy efficiency measures.

In the absence of this information, we hew to the CPUC framework for allocating energy, losses, capacity, and T&D deferral avoided costs on an hourly basis for a recent year; details of this calculation for specific utilities included in Section 5 Results are in Appendix A.3. In the base cases of our results we assume inclusion of T&D deferral avoided costs at levels consistent with those used by the CPUC and developed by New York's IOUs as part of the VDER proceedings in their Demand Reduction Value (DRV, similar to transmission-level avoided costs) and Locational System Relief Value (LSRV, similar to distribution-level avoided costs) tariff components. However, we include sensitivities in Section 5.3 Avoided Cost Sensitivity without these avoided costs, which may depend on system-specific ability to avoid or defer upgrades to grid infrastructure due to near-term reductions in peak demand. For example, [9] finds only 1% of Philadelphia-based electric utility PECO's annual distribution capital expenditures to be growth-related, suggesting an upward limit on the ability of DERs to avoid or defer distribution costs strictly via load management outside of targeted planning processes. Generally, if transmission and distribution systems have significant excess capacity, the ability of DERs including storage to avoid or defer those costs may be more limited than assumed in the Section 5 Results base case.

Figure 2.1 is a schematic of our method.

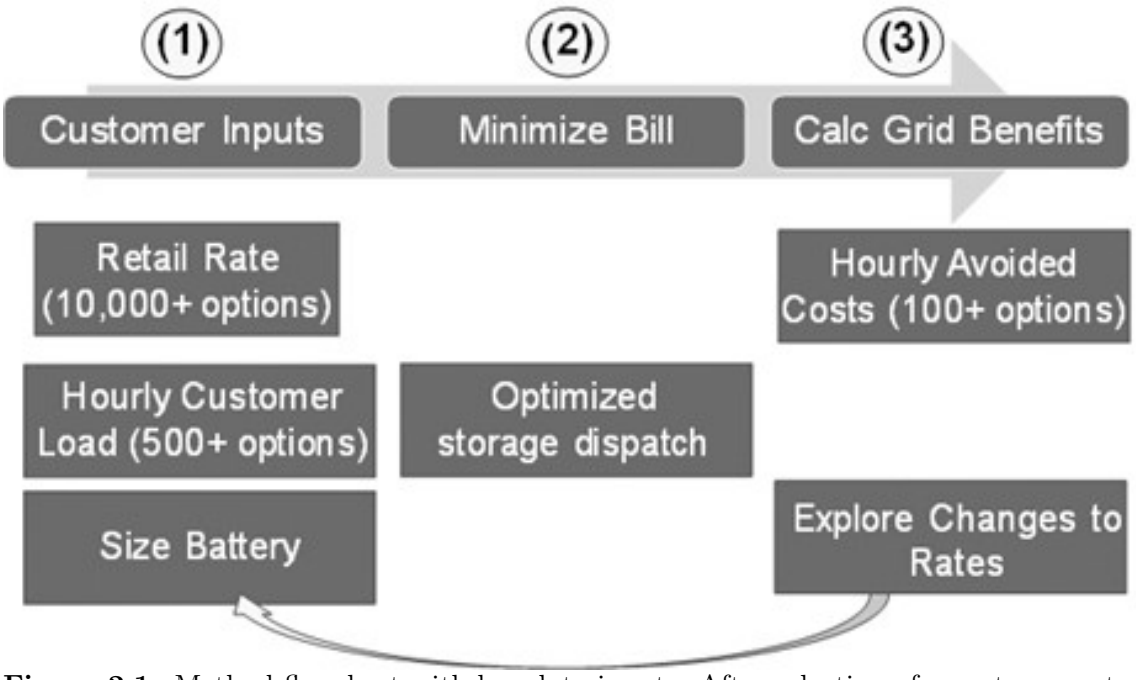

Figure 2.1: Method flowchart with key data inputs. After selection of a customer rate and load shape, a battery is sized to 20% of the customer's peak load and optimized with a linear program at monthly resolution to minimize the customer's bill. Bill savings and avoided costs can then be calculated. This process is then iterated as a new set of cases under changes to the customer's rate schedule.

2.4 Data

As shown in Figure 2.1, the three major input choices for conducting our analysis are the applicable customer rate, load shape, and estimate of utility avoided costs. Additionally, we size and parameterize battery dispatch. Data sources for these components are in Table 2.1, and described in more detail in Appendix A.

Because the main use case for customer-sited storage is reducing demand charges, we highlight the variation in demand charges by displaying in Figure 2.2 the rate with the highest demand charge by utility in each county across the United States. In counties with multiple utilities serving the county, we take a simple average of the utilities serving the county.

Table 2.1: Data Overview

Category	Sources
Customer Rate	OpenEI utility rates database [40]; 40,000+ rates can be used by our tool based on OpenEI database; vetting and detailed info for chosen cases in Appendix A
Hourly Load Shapes	30 DOE buildings, 1998-2014 (17 years) of loads, 16 climate zones [41] 500 Duke (SC/NC) Commercial and Industrial Load Shapes (2012-2013) used as validation
Avoided Costs	California: 3 Utilities, 2018-2047, 16 climate zones [37]. Others: Build hourly avoided costs from LMP, Losses, Capacity, T&D, pollutants; detail in Appendix A

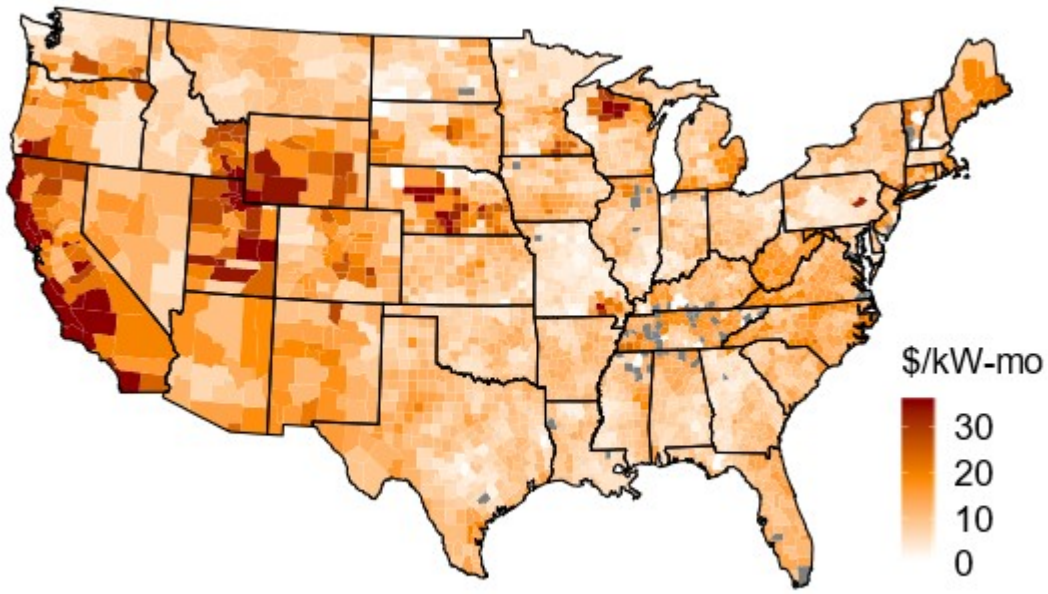

Figure 2.2: Average maximum commercial demand charge ($/kW-mo.) by county. Grey counties have no available data. Data from [40].

Figure 2.2 lends credence to the idea that, as noted by Manghani [3], Fisher and Apt [4], NREL [28], and others, deployment may happen outside traditional policy-forward jurisdictions due to heterogeneity in existing rate design as battery costs fall. For our cases, more than 40,000 utility-specific rates are pulled from OpenEI's Utility Rates Database [40] and shaped to hourly granularity using a set of scripts developed by the authors.

For customer load shape inputs, like [30] and other publications on the economics of commercial DERs [42], we use the 30 representative commercial customer load shapes developed by DOE in [33]. We verify that these DOE commercial load shapes are reasonable by comparison with metered load data in 2012-2013 for approximately 500 C&I customers in Duke Energy Carolinas' (NC/SC) service territory in Appendix A.

We estimate hourly avoided costs in a recent year for non-California jurisdictions where this level of detail is not available in avoided cost filings. Base case avoided cost estimates consist of marginal avoided energy, losses, capacity, and transmission and distribution (T&D) at hourly granularity for a utility's service territory. Hourly avoided CO_2 is also estimated at an Environmental Protection Agency (EPA) Emissions and Generation Resource Integrated Database (eGRID) subregion level using [43], monetized with the social cost of carbon used by [37],[3] and included as a sensitivity. Avoided CO_2 is not included in base cases because most utilities do not face CO_2 pricing as part of their cost of service. Further details of this estimation method are given in Appendix A.3.

Our method allows us to evaluate customer bill savings and avoided costs for any rate structure across the country given hourly customer load and avoided costs. However, three difficulties prevent us from realizing a fully representative study of commercial customers across the country

1. Representative hourly commercial customer load data is difficult to obtain.
2. Even when load data is obtained, the applicable rate for specific customers may be unknown, though we can make educated guesses.
3. For non-California jurisdictions we make generic assumptions to obtain hourly resolution avoided costs.

Additionally, the above criteria are not static; over time loads, rates, and avoided costs change. However, our main results showed enough consistency in the locations examined and with the sensitivity analyses performed to give confidence in the basic conclusions.

[3] \$63/ton CO_2 in nominal 2018\$

2.5 Results

2.5.1 Case Selection

We present results for six cases shown in Figure 2.3. The cases are meant to achieve a broad representation of the included continental United States temperatures (data from [41]) and the demand charges shown in Figure 2.2. Included utility service territories shaded in dark red in Figure 2.3A are Austin Energy (TX), Boston Edison (MA),[4] Commonwealth Edison (ComEd, IL), Dominion (VA), Duquesne Light Company (DLCO, PA), and Pacific Gas & Electric (PG&E, CA).

[4]The service territory shown on the map is NSTAR, which included Boston Edison Company, Cambridge Electric Light Company, Commonwealth Electric Company, and NSTAR Gas Company. NSTAR is now part of Eversource Energy, which includes additional utility service territories in Western Massachusetts, Connecticut, and New Hampshire. Cases throughout the text are referred to as Boston Edison to highlight that Boston-specific rate schedules and avoided costs are assigned for these customers.

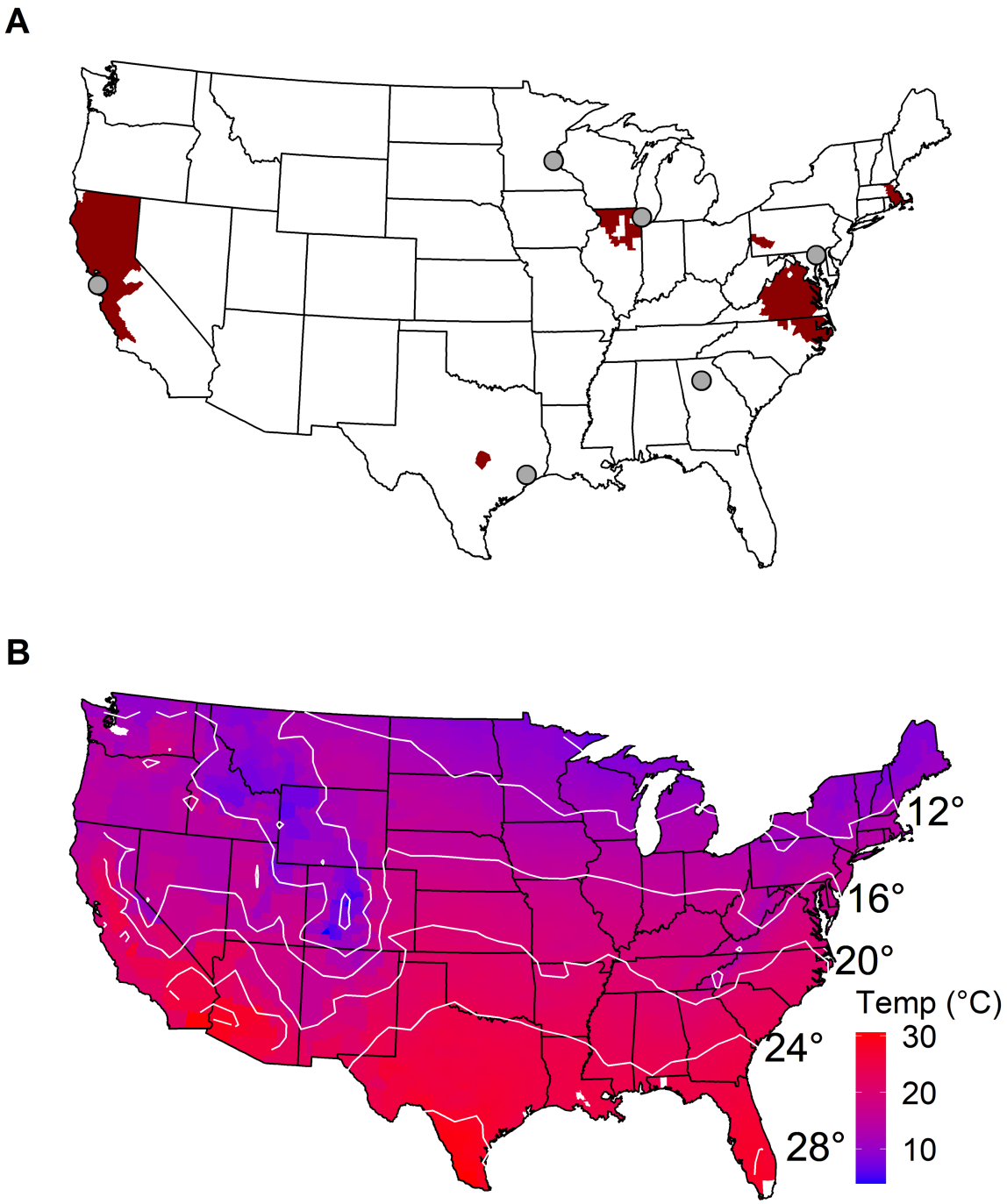

Figure 2.3: Six evaluated utility service territories (A) and annual average daily maximum temperature with isotherms (B). Gray dots (A) show the representative city for the climate load shape from DOE. Service territories are matched to the geographically closest dot for load shapes except Boston Edison (matched to Minneapolis for climate representation) and Dominion (matched to Atlanta for climate representation).

All results are presented in a RIM framework comparing percentage customer bill savings to system avoided costs from the customer-sited battery. We use percentages instead of absolute bill savings and avoided costs in results to better compare across customers with very different electricity consumption, though absolute savings are discussed in Section 2.5.4 Installation Likelihood. Violin plots (probability density plots of the data at different values) visualize a best-fit distribution and box-and-whisker plot assuming equal weight for each modeled customer load shape on the applicable utility tariff, commonly the 30 DOE load shapes for the applicable geography. The main result is presented Figure 2.4, and shows a critical peak price (CPP) tariff can increase avoided costs without decreasing customer bill savings, offering Pareto improvement. The policy implication is a full move to more complex real-time pricing may not be required to realize most system avoided costs from customer sited storage. Simpler critical peak pricing (CPP) rates targeted at the top 50-100 avoided cost hours realizes most system benefits and is more readily designed to ensure the storage adopting customer is not worse off.

Our main result is highly sensitive to which avoided cost components are considered; for this reason, a set of sensitivities on avoided costs are included in Figure 2.6. Other sensitivities are included in Appendix A and C.

2.5.2 Critical Peak Pricing

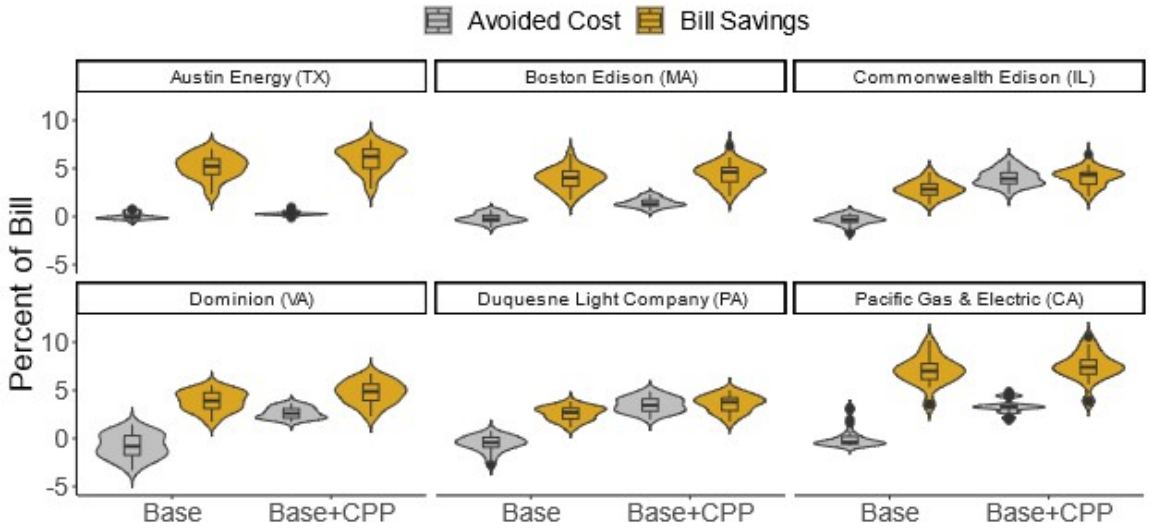

Figure 2.4: Critical Peak Pricing (CPP) can better match avoided costs and bill savings. Percent of Bill refers to the percent reduction in the customer's energy and demand bill before storage installation. Details of CPP are in Table 2.2. The distribution of values are shown as best-fit probability densities surrounding box plots.

Table 2.2: Figure 2.4 Case Parameterization.

Case Label	Load Shapes	Rate	Avoided Costs	Critical Peak Pricing
Base	Houston, Minneapolis, Chicago, Atlanta, Baltimore, San Francisco DOE shapes (n=30)	Austin Energy Commercial 10-300kW, Boston Edison Time-of-Use G-3, DLCO General Service >25kW, ComEd BES Large Load, Dominion GS-2 Demand, PG&E E-19 Medium TOU	Energy, losses, capacity, T&D	None
Base+CPP	Ibid	Ibid	Ibid	$0.50/kWh adder for optimal 15 four hour annual calls

Figure 2.4 suggests a well-targeted critical peak price (or rebate)[5] helps ensure storage dispatch during hours with the highest avoided costs. The re-dispatch of storage to achieve these avoided

[5]A critical peak price is levied on a customer's entire consumption during a set of hours. A rebate is a payment made to a customer associated with reducing load from a pre-defined baseline during those same hours. In principle, the two are economically equivalent if appropriately designed; in practice, the political economy of implementation and behavioral response may differ. We use CPP throughout the rest of the paper.

costs need not reduce customer bill savings from installation; in fact, additional battery revenue from the CPP hours outweighs foregone revenue from re-dispatch in most cases.

Figure 2.4 shows the highest incremental CPP avoided costs in the service territories in PJM ISO (ComEd, DLCO, Dominion) primarily because load serving entities in PJM allocate generating capacity costs to customers on the basis of five annual coincident peak hours (5CP, Appendix A.3 Avoided Costs). Bill savings are highest in PG&E due to higher, more granular demand charges (see Figure 2.2) and additional temporal resolution in time-of-use energy charges on PG&E's default commercial tariff. The CPP has the least effect in Austin Energy (TX) because the applicable wholesale market for the Electricity Reliability Council of Texas (ERCOT) does not have a capacity payment for the CPP to target and avoid.

We next investigate why a tariff such as the CPP is needed to achieve higher avoided costs from customer-sited storage in many jurisdictions. The primary reason the base cases achieve much higher customer bill savings than avoided costs is that many commercial customers have load shapes where monthly customer peak demand (which often determines their demand charge) does not align well with avoided costs. Results for the six jurisdictions are presented in Figure 2.5.

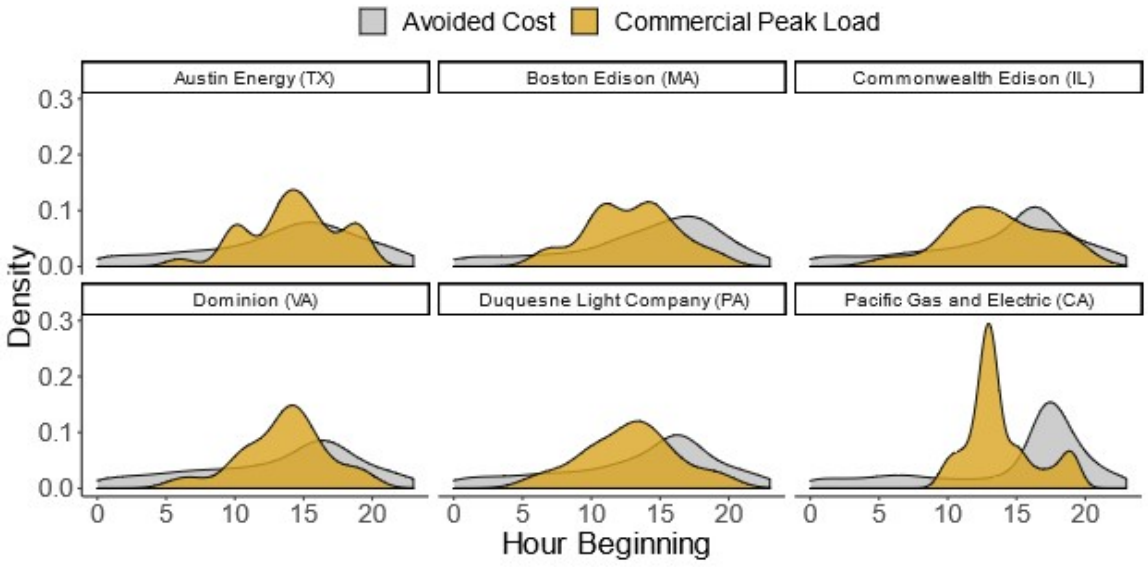

Figure 2.5: Density of annual hourly average avoided costs and peak hour by commercial load shape. Avoided cost peaks tend to lag commercial customer load peaks in most jurisdictions; the effect is most pronounced in California where increasing solar penetration has pushed avoided cost peaks toward the evening. Appendix A.3.

If customer peaks occur in the early afternoon but system avoided cost peaks are in the late afternoon (Figure 2.5), both these sets of hours may fall during the afternoon peak periods of even "smart" non-residential rates with coincident peak demand charges or more granular pre-defined TOU peak periods [44]. Smart non-residential rate design for storage, therefore, may require an additional level of granularity provided by CPP calls specifically targeted at peak avoided cost hours.

Further, as data for PG&E in Figure 2.5 suggests, increasing penetrations of variable renewable energy (VRE) have additional implications for correlation between customer loads and avoided costs. Large quantities of midday low marginal cost solar generation move system net load peaks and associated capacity-related avoided costs toward evening [45]. As a result, whatever misalignment currently exists between storage dispatch for minimizing customer bills and system avoided costs under current rates may be exacerbated by VRE, and could undermine presumed co-benefits of storage and VRE deployment if rate design does not adapt.

2.5.3 Avoided Cost Sensitivity

Our result on the value of CPP is sensitive to the question of which system costs can be avoided by battery dispatch. The base case in Figure 2.4 considered marginal energy, losses, generation capacity, and T&D costs avoidable. Figure 2.6 shows the incremental value of these avoided costs under CPP. An additional sensitivity analysis adds avoided marginal CO_2 emissions.

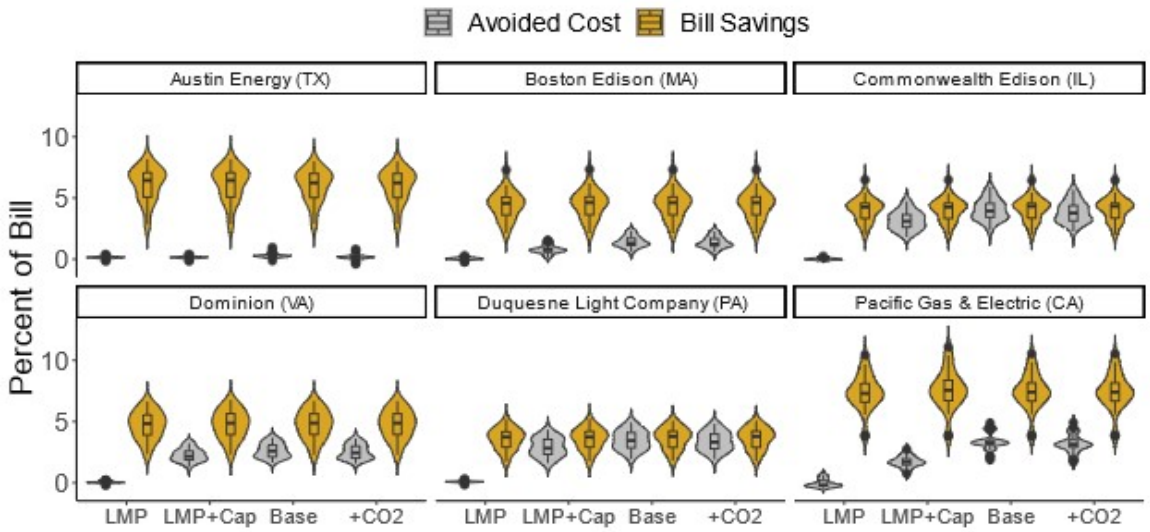

Figure 2.6: Inclusion of peak-driven avoided investment costs in generation capacity and T&D infrastructure are necessary for realizing the system benefits of CPP from storage. Including non-monetized marginal environmental benefits like avoided CO_2 has little further effect. As for Figure 2.4, the distribution of values are shown as probability densities surrounding box plots. Details of cases are in Table 2.3.

Table 2.3: Figure 2.6 Case Parameterization.

Case Label	Load Shapes	Rate	Avoided Costs	Critical Peak Pricing
LMP	Houston, Minneapolis, Chicago, Atlanta, Baltimore, San Francisco DOE shapes (n=30)	Austin Energy Commercial 10-300kW, Boston Edison Time-of-Use G-3, DLCO General Service >25kW, ComEd BES Large Load, Dominion GS-2 Demand, PG&E E-19 Medium TOU	Energy, losses	$0.50/kWh adder for optimal 15 four hour annual calls
LMP+Cap	Ibid	Ibid	Energy, losses, generation capacity	Ibid
Base	Ibid	Ibid	Energy, losses, generation capacity, T&D	Ibid
$+CO_2$	Ibid	Ibid	Energy, losses, generation capacity, T&D, CO_2	Ibid

Peak-driven investments at the bulk system, transmission, and distribution levels concentrated in just a few dozen hours create the greatest opportunity for CPP to avoid costs often not reflected in current rates. Avoided energy costs alone do not readily justify CPP, so jurisdictions should be careful to consider which investment costs are avoidable. Many jurisdictions are located within wholesale

electricity markets with well-defined capacity costs at the bulk system level, but avoidable T&D costs on a utility's sub transmission or distribution system are much less well-defined. Quantifying these avoided costs has been a subject of considerable attention in proceedings like New York's Value of DER (VDER) proceedings, resulting in tariffs with a utility (DRV) and sub-utility locational (LSRV) component broadly reflecting T&D avoided cost magnitude estimates and tariff dispatch criteria for DERs to accrue these avoided costs [17]. Figure 2.6 suggests distributed storage deployment should serve as an impetus for more utilities to consider developing similar estimates.

Additional results highlighted by Figure 2.6 include that consideration of avoided CO_2 costs[6] using a marginal emissions factor (MEF) framework has little effect on the difference between avoided costs and bill savings. Avoided CO_2 emissions when dispatching storage to minimize customer bills are small in the evaluated territories for current electric grid MEFs; this confirms more detailed research in [46, 47]. However, even the most detailed and accurate marginal emissions factors assess only dispatch on the current electric grid, though adjustments can be made to better forecast future grid generation mixes [48]. Also, investment effects are exogenous to the MEF approach. Bistline and Young [49] show that for storage technologies like the distributed batteries this paper considers, investment-related emissions effects considered in capacity expansion models are much larger than dispatch-related emissions effects considered by MEF or dispatch models under a wide range of future scenarios, with increased storage penetration generally yielding increased emissions reductions as cost declines allow or policy requires more VRE on the grid. The MEF approach may therefore significantly underestimate the magnitude of emissions reductions associated with distributed storage deployment, though this does not affect our main set of conclusions about rate design.

2.5.4 Installation Likelihood

Considerable other literature has focused on cost-effectiveness for the customer or third-party storage owner and forecast increasing penetrations of distributed battery storage. For this reason, we take the policy relevance of increasing penetrations of distributed battery storage as a given and focus on cost-effectiveness from the utility perspective in previous parts of Section 2.5 Results. To

[6]Climate avoided costs are monetized avoided CO_2 based on a \$63/ton CO_2 damages, drawing on [37]

enable comparisons across utilities and customers Figure 2.4-Figure 2.6 weight all customer load shapes equally and are normalized by customer bills. However, certain types of commercial customers in certain utility territories are clearly more likely to install distributed storage than others. Some of the reason for a customer being more likely to install battery storage will be financial, which we highlight in Figure 2.7.

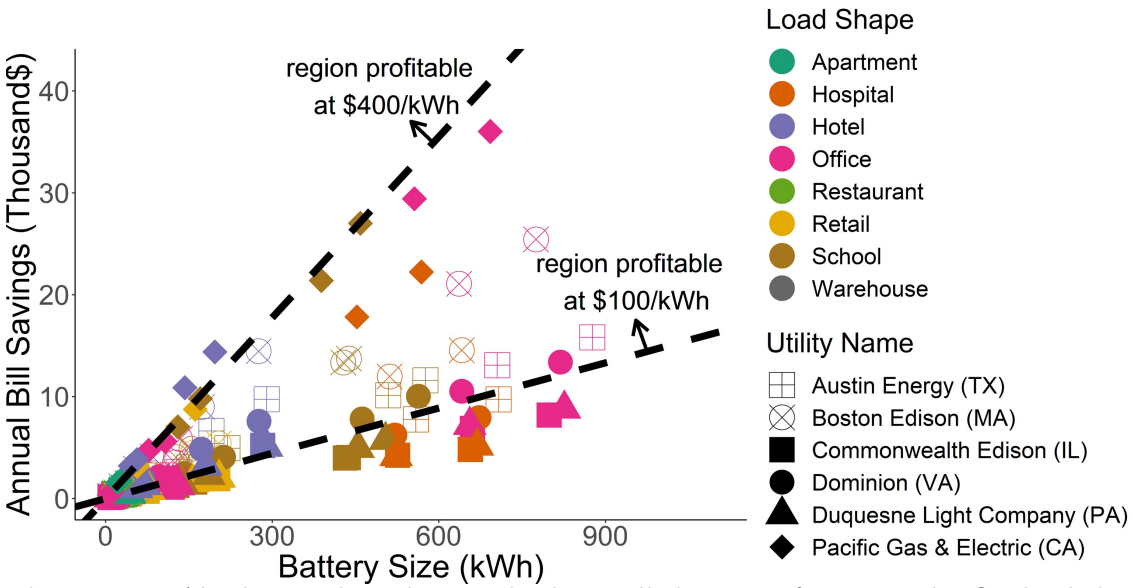

Figure 2.7: Absolute and per battery kWh installed savings for grouped DOE load shapes across the six modeled utilities. The 15 DOE cases are grouped into the 8 plotted groups as described in Appendix C.6. Black dashed lines show break-even storage installed costs in \$/kWh assuming a 10 year battery lifetime and 10% discount rate, with annual payments occurring at the end of each calendar year. Points in region above line have bill savings exceeding that levelized installed cost ("profitable" excluding operating costs). Bill savings are without CPP and assume perfect foresight. All customers have four-hour duration batteries sized to 20% of peak load.

Generally we can infer lower load factors, higher demand charges, and economies of scale in battery installation costs mean customers make customers with higher savings per battery kWh installed as well as higher total bill savings more likely to install, all else equal. Figure 2.7 shows that among modelled DOE load shapes the customers with both high total bill savings and high per kWh installed bill savings are on PG&E's electric rates, particularly the Large Office and Secondary School load shapes. These customers are characterized by the highest demand charges (PG&E's) among the six modelled utilities, diurnal peak loads on regular weekday operating hours, and comparatively large overall loads among the modelled load shapes. The dashed black lines in Figure 2.7 suggest

retail bill savings alone can support a $400/kWh breakeven upfront installed cost for the most lucrative modelled customers assuming a 10 year battery lifetime, constant annual bill savings, and a 10% discount rate. This is a simplified view of the customer or third-party storage owner's financial decision that does not consider the full stack of costs (e.g., operations and maintenance), policy (e.g., tax credits), financing options, available revenue streams (e.g., enrollment in a demand response program), and non-monetized benefits (e.g., avoided customer outages). The simplified view highlights consideration of other aspects of the installation decision can help further target rate design at those customers most likely to install. For example, the utility might assume a customer with large critical loads like a hospital will place a higher value on avoided outages and be more likely to install. Or the utility might assume that all customers will place a higher value on avoided outages after recently experiencing extended mandatory interruptions in utility service, as occurred during the Public Safety Power Shutoffs affecting many California IOU customers in 2019 and 2020.

2.6 Conclusions and Policy Implications

We find that dynamic rates are a simple and effective way to ensure that customer storage helps the grid through generation capacity and T&D avoided. A full move to more complex real-time pricing may not be required to realize most system avoided costs from customer-sited storage. Simpler critical peak pricing (CPP) rates targeted at the top 50-100 avoided cost hours can improve alignment of customer and system benefits for storage adoption and offer Pareto improvement. Therefore, utilities and regulators should consider these rate options as part of storage-oriented policy.

CPP rates are currently offered to commercial customers on at least an optional or pilot basis in some jurisdictions, but are not generally common (Figure 2.8). For example, non-residential storage installations receiving SGIP incentives in California are often on CPP rates, or other event-based programs targeted at peak hours like the capacity bidding program (CBP) and demand response auction mechanism (DRAM) [26]. In New York, Consolidated Edison is piloting a rider (Q) to their standard commercial rate allowing for an as-used daily demand period "that aligns demand charges

with when network peaks occur" [38, 50]. All New York IOUs now offer a Value Stack Tariff, with the locational LSRV component of this tariff accrued based on dispatch during minimally 10 annual 1-4 hour calls announced with at least 21 hours advanced notice [17]. As California generally and the densely populated parts of New York have among the highest demand charges (Figure 2.2) and have additional incentives for storage deployment, options for rates that fit the criteria outlined in this work are notable. However, options for CPP, critical peak rebates (CPR), and/or real-time pricing (RTP) rates remain generally uncommon (Figure 2.8), and are much less common than simpler TOU rates, leaving considerable room for ambition in rate design beyond current policy.

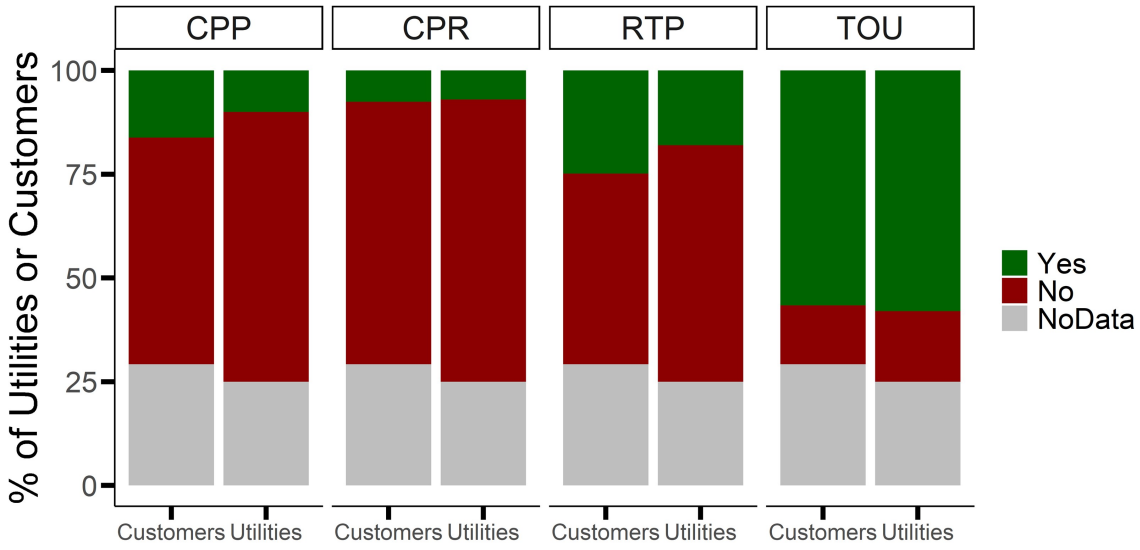

Figure 2.8: Availability of dynamic rate options for commercial customers among the top 100 IOUs by commercial customer count in the United States in 2019. Data from 2019 EIA 861 [51]. 25 of the top 100 IOUs (29% of customers) do not report data on this form in 2019.

Finally, we make three points about the broader effects of changes to rate design to better accommodate battery storage.

First, if distributed storage encourages a move toward default advanced rate designs for C&I customers, this has broader effects for C&I customers without storage. Econometric studies often find the broader effects of advanced rate design to be positive at the class level [52, 53]. Therefore, we follow Linvill and Lazar [34] in recommending the time has come for these advanced rate designs in C&I customer classes.

Second, we considered only a limited set of revenue streams potentially available to distributed storage. We noted earlier the existence of a potentially sizable and salient but difficult to quantify avoided customer outage value stream, as well as some non-monetized benefits. Additional monetizable revenue streams may be available through aggregation of customer-sited storage, particularly with the implementation of FERC Orders 841 and 2222 [54, 55] respectively enabling broader storage and DER aggregation participation in wholesale markets. However, wholesale market revenue streams are available to storage sited at the distribution or bulk system level, customer-sited aggregation has transaction costs, and economies of scale may be realized by larger battery installations that can access these revenue streams.

Last, there are regulatory barriers to implementation of some of the incentives and rate designs suggested by this work. Two tensions are the incentive structure of a regulated monopoly that earns rate-of-return on its capital investments, and additional flexibility afforded to distribution utilities in setting both interconnection rules and rate design beyond the scope of our analysis. Regarding utility incentives for encouraging third-party or customer-owned assets, remedies are extensively explored in literature on revenue decoupling and performance-based ratemaking. Slow, costly interconnection processes for customer-sited storage and increased fixed or sufficiently onerous standby charges are potential disincentives for storage adoption and the value of dynamic rate design. Here we simply note these approaches, which have been tried in some jurisdictions seeing distributed solar PV adoption, are generally unpopular, undermine customer choice, and may not be justified by cost-causation. More dynamic rate designs can realize increased benefits for customers and utilities from customer-sited storage installation in the near-term, and have potentially increasing co-benefits beyond storage in a smarter, more transactive grid.

References

[1] Lazard, "Levelized Cost of Storage Analysis 2018," November, 2018. https://www.lazard.com/media/450774/lazards-levelized-cost-of-storage-version-40-vfinal.pdf.

[2] D. Finn-Foley, "The Next Five Years in Energy Storage According to 500 Energy Professionals Live attendee poll results from Greentech Media's 2017 Energy Storage Summit Crowdsourced

Insights Panel," 2018. https://www.greentechmedia.com/articles/read/
the-next-five-years-in-energy-storage-according-to-500-energy-professionals.

[3] R. Manghani, "The Economics of Commercial Energy Storage in the U.S.: The Outlook for Demand Charge Management," vol. 4, no. 8, pp. 1–4, 2016.
https://www.woodmac.com/our-expertise/focus/Power--Renewables/
The-Economics-of-Commercial-Energy-Storage-in-the-U.S.
-The-Outlook-for-Demand-Charge-Management/.

[4] M. J. Fisher and J. Apt, "Emissions and Economics of Behind-the-Meter Electricity Storage," *Environmental Science & Technology*, vol. 51, pp. 1094–1101, Feb 2017.
doi:http://dx.doi.org/10.1021/acs.est.6b03536.

[5] CPUC, "California Standard Practice Manual Economic Analysis of Demand-Side Programs," October, 2001. http://www.cpuc.ca.gov/WorkArea/DownloadAsset.aspx?id=7741.

[6] EPA, "Understanding Cost-Effectiveness of Energy Efficiency Programs : Best Practices , Technical Methods , and Emerging," November, 2008.
https://19january2017snapshot.epa.gov/sites/production/files/2015-08/documents/
understanding_cost-effectiveness_of_energy_efficiency_programs_best_practices_
technical_methods_and_emerging_issues_for_policy-makers.pdf.

[7] L. Hansen, V. Lacy, and D. Glick, "A review of solar PV benefit and cost studies," *Rocky Mountain Institute Electricity Innovation Lab*, no. September, p. 1-59, 2013.
https://rmi.org/insight/a-review-of-solar-pv-benefit-and-cost-studies/.

[8] P. Vaishnav, N. Horner, and I. L. Azevedo, "Was it worthwhile? Where have the benefits of rooftop solar photovoltaic generation exceeded the cost?," *Environmental Research Letters*, vol. 12, no. 9, 2017. doi:http://dx.doi.org/10.1088/1748-9326/aa815e

[9] J. F. Keen and J. Apt, "How much capacity deferral value can targeted solar deployment create in Pennsylvania?," *Energy Policy*, vol. 134, p. 110902, September, 2019.
doi:https://doi.org/10.1016/j.enpol.2019.110902

[10] A. Satchwell, A. Mills, G. Barbose, R. Wiser, P. Cappers, and N. Darghouth, "Financial Impacts of Net-Metered PV on Utilities and Ratepayers : A Scoping Study of Two Prototypical U . S . Utilities," September, 2014.
https://emp.lbl.gov/publications/financial-impacts-net-metered-pv.

[11] SAIC, "2013 Updated Solar PV Value Report: Arizona Public Service," May, 2013.
https://www.azsolarcenter.org/images/docs/reports/SolarValueStudy-SAIC-2013-05.pdf

[12] Xcel Energy Services, "Costs and Benefits of Distributed Solar Generation Public Service Company of Colorado System Study Report in Response to Colorado Public Utilities Commission Decision No. C09-1223," 2013.
http://www.eei.org/issuesandpolicy/generation/NetMetering/Documents/
CostsandBenefitsofDistributedSolarGenerationonthePublicServiceCompanyofColoradoSystemXcelEnergy.
pdf.

[13] Edison Electric Institute, "Solar Energy and Net Metering," August 2015, 2016. http://www.eei.org/issuesandpolicy/generation/NetMetering/Documents/StraightTalkAboutNetMetering.pdf.

[14] N. R. Darghouth, R. Wiser, G. Barbose, and A. Mills, "Net Metering and Market Feedback Loops: Exploring the Impact of Retail Rate Design on Distributed PV Deployment," LBNL-183185, 2015. https://emp.lbl.gov/publications/net-metering-and-market-feedback.

[15] J. Lazar and K. Colburn, "Recognizing the Full Value of Energy Efficiency (What's Under the Feel-Good Frosting of the World's Most Valuable Layer Cake of Benefits)," September, 2013. https://www.puc.nh.gov/electric/15-137%20non%20docket%20info/RAP_LazarColburn_LayerCakePaper_2013_Sept_9%20(1).pdf.

[16] S. F. Tierney, "The Value of DER to D: The Role of Distributed Energy Resources in Supporting Local Electric Distribution System Reliability," 2016. https://www.cpuc.ca.gov/uploadedFiles/CPUC_Public_Website/Content/About_Us/Organization/Divisions/Policy_and_Planning/Thought_Leaders_Events/Tierney%20White%20Paper%20-%20Value%20of%20DER%20to%20D%20-%203-30-2016%20FINAL.pdf.

[17] NYPSC, "Order Regarding Value Stack Compensation," 2019. https://www.nyserda.ny.gov/-/media/NYSun/files/Updated-Value-Stack-Order-2019-04-18.pdf.

[18] M. T. Craig, P. Jaramillo, B. M. Hodge, N. J. Williams, and E. Severnini, "A retrospective analysis of the market price response to distributed photovoltaic generation in California," *Energy Policy*, vol. 121, pp. 394–403, February, 2018. doig:https://doi.org/10.1016/j.enpol.2018.05.061

[19] P. Chernick and C. Neme, "The Value of Demand Reduction Induced Price Effects (DRIPE)," 2015. https://www.raponline.org/knowledge-center/the-value-of-demand-reduction-induced-price-effects-dripe/.

[20] R. Hledik, "The Economic Potential for Energy Storage in Nevada Public Utilities Commission of Nevada," 2018. http://files.brattle.com/files/14618_economic_potential_for_storage_in_nevada_-_final.pdf.

[21] J. C. Bonbright, "Principles of Public Utility Rates," 1961. https://www.raponline.org/knowledge-center/principles-of-public-utility-rates/.

[22] NARUC, "Distributed Energy Resources Rate Design and Compensation," November, 2016. https://pubs.naruc.org/pub/19FDF48B-AA57-5160-DBA1-BE2E9C2F7EA0.

[23] L. Wood, J. Howat, R. Cavanagh, S. Borenstein, J. Deason, and L. Schwartz, "Recovery of Utility Fixed Costs: Utility, Consumer, Environmental and Economist Perspectives," no. 5, pp. 1–86, 2016. https://emp.lbl.gov/publications/recovery-utility-fixed-costs-utility.

[24] A. Chitkara, D. Cross-Call, B. Li, and J. Sherwood, "A Review of Alternative Rate Designs," May, 2016. https://rmi.org/insight/review-alternative-rate-designs/.

[25] J. Lazar and W. Gonzalez, "Smart Rate Design For a Smart Future," July, 2015. https://www.raponline.org/knowledge-center/smart-rate-design-for-a-smart-future/.

[26] Itron and E3, "2017 SGIP Advanced Energy Storage Impact," 2018. http://www.cpuc.ca.gov/uploadedFiles/CPUC_Public_Website/Content/Utilities_and_Industries/Energy/Energy_Programs/Demand_Side_Management/Customer_Gen_and_Storage/2017_SGIP_AES_Impact_Evaluation.pdf.

[27] Lazard, "Lazard Levelized Cost of Storage Version 5.0," tech. rep., 2019. https://www.lazard.com/media/451087/lazards-levelized-cost-of-storage-version-50-vf.pdf.

[28] NREL, "Identifying Potential Markets for Behind-the-Meter Battery Energy Storage : A Survey of U.S . Demand Charges," pp. 1–7, 2017. https://www.nrel.gov/docs/fy17osti/68963.pdf.

[29] C. Eid, E. Koliou, M. Valles, J. Reneses, and R. Hakvoort, "Time-based pricing and electricity demand response: Existing barriers and next steps," *Utilities Policy*, vol. 40, pp. 15–25, 2016. doi:https://doi.org/10.1016/j.jup.2016.04.001.

[30] D. Wu, M. Kintner-Meyer, T. Yang, and P. Balducci, "Economic analysis and optimal sizing for behind-the-meter battery storage," *IEEE Power and Energy Society General Meeting*, vol. 2016-Nov, 2016. doi:https://doi.org/10.1109/PESGM.2016.7741210.

[31] D. Wu, C. Jin, P. Balducci, and M. Kintner-Meyer, "An energy storage assessment: Using optimal control strategies to capture multiple services," *IEEE Power and Energy Society General Meeting*, vol. 2015-Sept, 2015. doi:https://doi.org/10.1109/PESGM.2015.7285820.

[32] D. Wu, M. Kintner-Meyer, T. Yang, and P. Balducci, "Analytical sizing methods for behind-the-meter battery storage," *Journal of Energy Storage*, vol. 12, pp. 297–304, 2017. doi:https://doi.org/10.1016/j.est.2017.04.009.

[33] M. Deru, K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski, M. Rosenberg, M. Yazdanian, J. Huang, and D. Crawley, "U.S. Department of Energy commercial reference building models of the national building stock," *Publications (E)*, pp. 1 – 118, February 2011. http://digitalscholarship.unlv.edu/renew_pubs/44.

[34] C. Linvill and J. Lazar, "Smart non-residential rate design: Aligning rates with system value," *Electricity Journal*, vol. 31, no. 8, pp. 1–8, 2018. doi:https://doi.org/10.1016/j.tej.2018.09.011.

[35] M. J. Fisher, "Personal Communication," 2018.

[36] G. He, Q. Chen, P. Moutis, S. Kar, and J. F. Whitacre, "An intertemporal decision framework for electrochemical energy storage management," *Nature Energy*, vol. 3, no. 5, pp. 404–412, 2018. doi:https://doi.org/10.1038/s41560-018-0129-9.

[37] California Public Utilities Commission, "Avoided Cost Calculator User Manual," August, 2016. http://www.cpuc.ca.gov/General.aspx?id=5267.

[38] NYSERDA, "Standby Rate + Con Ed Rider Q Fact Sheet," pp. 2–6, 2018. https://www.nyserda.ny.gov/-/media/Files/Programs/Energy-Storage/Rider-Q.pdf.

[39] N. Mims, T. Eckman, and C. Goldman, "Time-varying value of electric energy efficiency," June, 2017. https://emp.lbl.gov/publications/time-varying-value-electric-energy.

[40] "Utility Rate Database | Open Energy Information," 2018. https://openei.org/wiki/Utility_Rate_Database.

[41] NLDAS, "North America Land Data Assimilation System (NLDAS) Daily Air Temperatures and Heat Index (1979-2011) Request," 2018. https://wonder.cdc.gov/nasa-nldas.html.

[42] J. F. Keen and J. Apt, "Are high penetrations of commercial cogeneration good for society?," *Environmental Research Letters*, vol. 11, no. 12, 2016. doi:https://doi.org/10.1088/1748-9326/11/12/124014.

[43] I. Azevedo, N. Horner, K. Siler-Evans, and P. Vaishnav, "Electricity Marginal Factor Estimates," 2017. http://cedmcenter.org.

[44] C. Linvill, J. Lazar, M. Dupuy, J. Shipley, and D. Brutkoski, "Smart Non-Residential Rate Design," pp. 1–56, December, 2017. https://www.raponline.org/knowledge-center/smart-non-residential-rate-design/.

[45] J. Lazar, "Teaching the "Duck" to Fly," no. January, 2014. http://www.raponline.org/document/download/id/6977.

[46] E. S. Hittinger and I. M. Azevedo, "Bulk energy storage increases united states electricity system emissions," *Environmental Science and Technology*, vol. 49, no. 5, pp. 3203–3210, 2015. doi:https://doi.org/10.1021/es505027p.

[47] O. Babacan, A. Abdulla, R. Hanna, J. Kleissl, and D. G. Victor, "Unintended Effects of Residential Energy Storage on Emissions from the Electric Power System," *Environmental Science & Technology*, vol. 52, p. acs.est.8b03834, 2018. doi:https://doi.org/10.1021/acs.est.8b03834.

[48] T. A. Deetjen and I. L. Azevedo, "Reduced-Order Dispatch Model for Simulating Marginal Emissions Factors for the United States Power Sector," *Environmental Science and Technology*, vol. 53, no. 17, pp. 10506–10513, 2019. doi:https://doi.org/10.1021/acs.est.9b02500.

[49] J. E. Bistline and D. T. Young, "Emissions impacts of future battery storage deployment on regional power systems," *Applied Energy*, vol. 264, p. 114678, February, 2020. doi:https://doi.org/10.1016/j.apenergy.2020.114678.

[50] G. C. Sayre and D. X. Burman, "New York State Energy Storage Roadmap," no. 518, 2018. https://www.nyserda.ny.gov/AllPrograms/Programs/EnergyStorage/AchievingNYEnergyGoals/TheNewYorkStateEnergyStorageRoadmap.

[51] U.S. Energy Information Administration, "EIA 861," 2019. https://www.eia.gov/electricity/data/eia861/.

[52] A. Faruqui, S. Sergici, and C. Warner, "Arcturus 2.0: A Meta-Analysis of Time-Varying Rates for Electricity," *The Electricity Journal*, pp. 1–29, 2017. doi:https://doi.org/10.1016/j.tej.2017.11.003.

[53] A. Faruqui, R. Hledik, and J. Tsoukalis, "The Power of Dynamic Pricing," *The Electricity Journal*, February, 2009. doi:https://doi.org/10.1016/j.tej.2009.02.011.

[54] FERC, "Electric Storage Participation in Markets Operated by Regional Transmission Organizations and Independent System Operators," no. 1, 2018. https://www.ferc.gov/whats-new/comm-meet/2018/021518/E-1.pdf.

[55] F. Register, "Federal Register / Vol . 85 , No . 204 / Wednesday , October 21 , 2020 / Rules and Regulations," vol. 85, no. 204, pp. 67094–67158, 2020. https://www.govinfo.gov/content/pkg/FR-2020-10-21/pdf/2020-23159.pdf.

[56] S. Borenstein and J. B. Bushnell, "Do Two Electricity Pricing Wrongs Make a Right? Cost Recovery, Externalities, and Efficiency," *Nber Working Paper Series*, vol. 29, no. 34, pp. 4471–4480, 2018. https://www.nber.org/papers/w24756.

[57] R. Hledik, J. Zahniser-Word, and J. Cohen, "Storage-oriented rate design : Stacked benefits or the next death spiral?," *The Electricity Journal*, vol. 31, no. 8, pp. 23–27, 2018. doi:https://doi.org/10.1016/j.tej.2018.09.012.

[58] M. A. Cohen, P. A. Kauzmann, and D. S. Callaway, *Effects of distributed PV generation on California's distribution system, part 2: Economic analysis*, vol. 128. Feb., 2016. doi:https://doi.org/10.1016/j.solener.2016.01.004.

[59] K. Patel, Z. Ming, L. Lavin, G. D. Moor, B. Horii, and S. Price, "The benefits and costs of net energy metering in New York," *Energy and Environmental Economics*, 2015. https://www.ethree.com/wp-content/uploads/2017/01/E3-NY-Legislative-NEM-Study-Report-121115-FINAL-SENT.pdf.

[60] Synapse Energy Economics, Resource Insight, Les Deman Consulting, North Side Energy, and Sustainable Energy Advantage, "Avoided Energy Supply Components in New England: 2018 Report," *AESC 2018 Study Group*, 2018.

[61] E. . E. Economics, "Cost-Benefit Analysis of Plug-in Electric Vehicle Adoption in the AEP Ohio Service Territory," 2017. http://www.synapse-energy.com/sites/default/files/AESC-2018-17-080-June-Release.pdf.

[62] W. House, "Social Cost of Greenhouse Gases," 2013. https://obamawhitehouse.archives.gov/omb/oira/social-cost-of-carbon.

[63] K. Ricke, L. Drouet, K. Caldeira, and M. Tavoni, "Country-level social cost of carbon," *Nature Climate Change*, vol. 8, no. 10, pp. 895–900, 2018. doi:https://doi.org/10.1038/s41558-018-0282-y.

[64] S. P. Holland, E. T. Mansur, N. Muller, and A. J. Yates, "Decompositions and Policy Consequences of an Extraordinary Decline in Air Pollution from Electricity Generation," 2018. http://www.nber.org/papers/w25339.

A Data Appendix

A.1 Customer Load Shapes

Comparison to Actual Load Data in Duke Energy Carolinas

As a preliminary validation of the use of the DOE commercial load shapes in for the territories, we compare results for the 2013 Baltimore DOE shapes against approximately 500 Duke Energy Carolinas C&I customer load shapes from 2013.

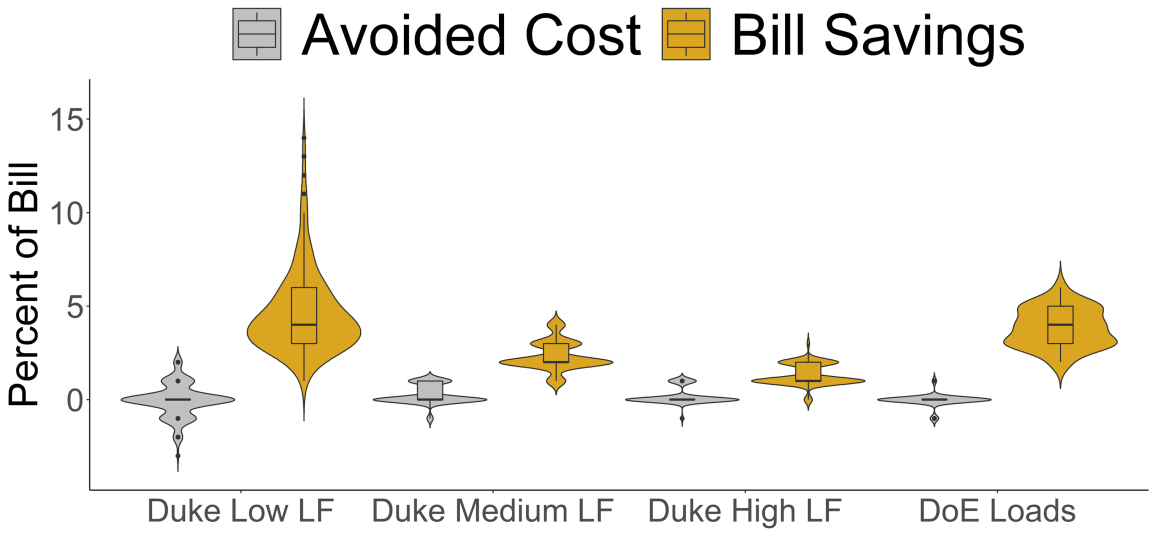

Figure A-1: Duke Energy Carolinas load shape comparison with DOE commercial loads.

Table A-1: Duke Energy Carolinas comparison inputs.

Case Label	Load Shapes	Rate	Avoided Costs
Duke Low LF	Duke Energy Carolinas Load Factor <0.5	Dominion GS-2 Demand	Dominion 2013
Duke Medium LF	Duke Energy Carolinas 0.5<Load Factor<0.65	Dominion GS-2 Demand	Dominion 2013
Duke High LF	Duke Energy Carolinas Load Factor >0.65	Dominion GS-2 Demand	Dominion 2013
DoE Loads	Baltimore DoE shapes	Dominion GS-2 Demand	Dominion 2013

Similar results between the Duke 2013 load shapes with the lowest load factors and the DOE loads suggest the DOE shapes are broadly representative of the customers most likely to install storage. A customer's load factor is the ratio of their average to peak demand; a lower load factor suggests the customer pays a larger portion of their bill as demand charges and have more scope to install storage and reduce their demand charge.

Main Case Parameterization

Details about the DOE load shapes [33] used for the six cases in the main text are provided below.

Table A-2: Parameterization of main six cases.

Case Utility	Representative City Shapes	Climate Zone[a]	Year	Weight
Austin Energy (TX)	Houston DOE (n=30)	2A	2006	All 30 shapes weighted equally
Boston Edison (MA)	Minneapolis DOE (n=30)	6A	2006	All 30 shapes weighted equally
Commonwealth Edison (IL)	Chicago DOE (n=30)	5A	2006	All 30 shapes weighted equally
Dominion (VA)	Atlanta DOE (n=30)	3A	2006	All 30 shapes weighted equally
Duquesne Light Company (PA)	Baltimore DOE (n=30)	4A	2006	All 30 shapes weighted equally
Pacific Gas & Electric	San Francisco DOE (n=30)	3C	2003	All 30 shapes weighted equally

[a] See https://www.energy.gov/eere/buildings/commercial-reference-buildings

In the main text the shapes are listed by their representative city. The cities are representative of 16 modeled climate zone profiles in the United States. These 16 zones are further subdivided into 8 general climate regions. We include one shape from each of the five most common climate zones, matched to a utility for which we could develop avoided cost data. An additional shape is included in climate zone 3 because of differentiation between the west and east coast. The climate regions not included (1,7,8) are the hottest and coldest zones of the United States, respectively, and are less common. The representative cities for those zones are 1-Miami (FL), 7-Duluth (MN), and 8-Fairbanks (AK).

Years for the shapes are selected to match the weekday/weekend profile of the year for which we

develop avoided costs. For example, 2006 and 2017 are both common years beginning on a Sunday.

Shapes are weighted equally in all cases to achieve a broader representation of customer shapes. In practice, certain shapes may make up a larger portion of a utility's load than others. Additionally, some types of customers may be more likely to adopt storage for behavioral or cost-benefit reasons.

Further information on DOE Load Shapes

Log average hourly loads for each of the 30 shapes across the six areas used in the main cases are shown below for time-of-day and monthly resolution.

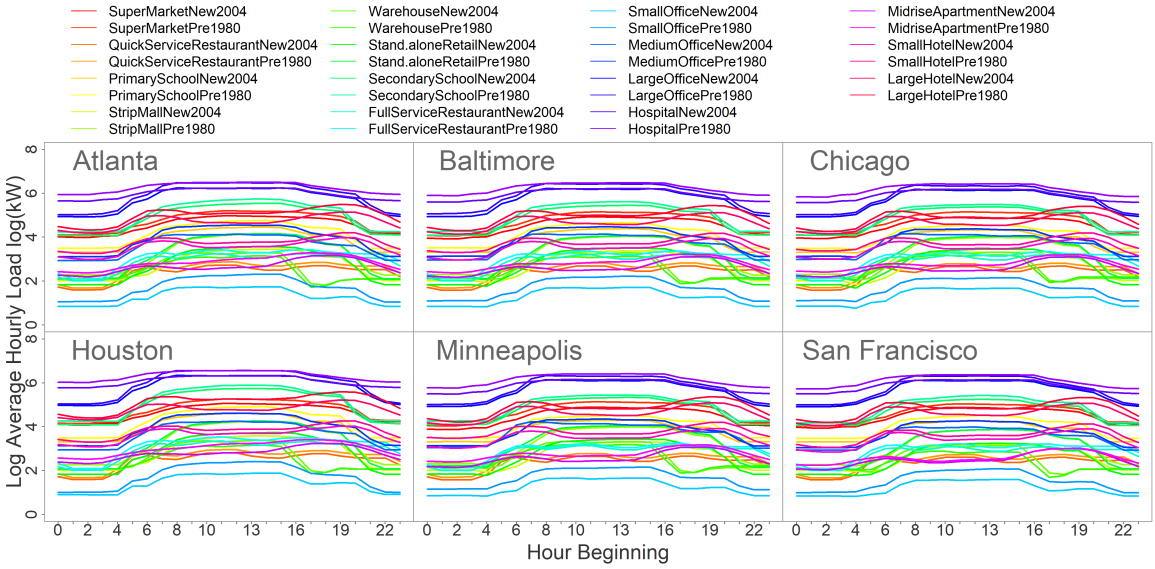

Figure A-2: Time of Day Shapes.

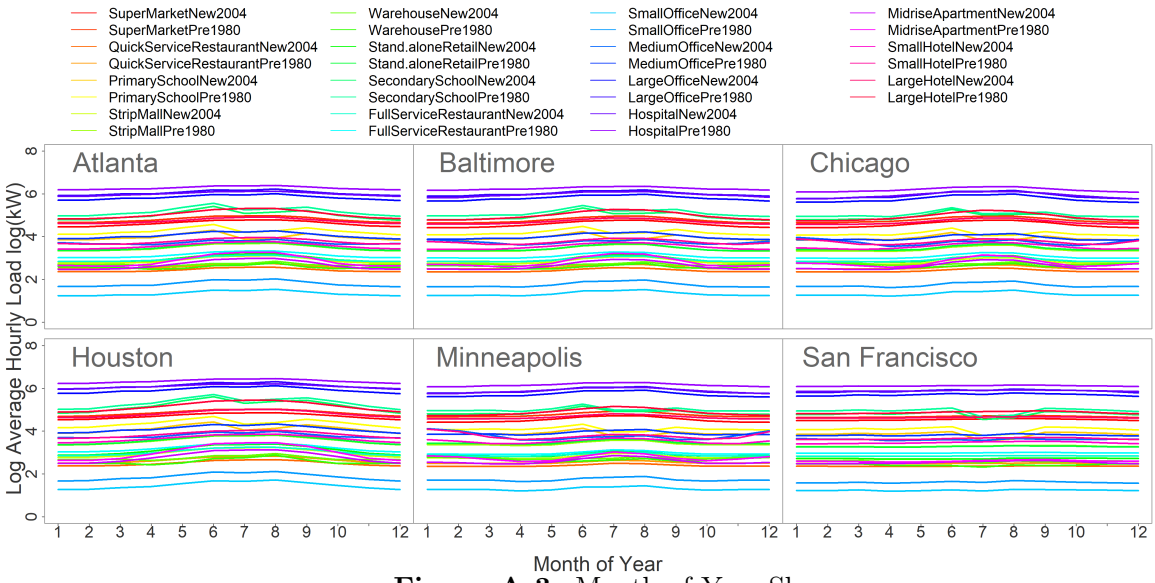

Figure A-3: Month of Year Shapes.

A.2 Customer Rates

Rate Choice: Southern California Edison Sensitivity

In the main text all customers in a utility's service territory are assigned the same commercial rate schedule. In practice, multiple rate schedules may be available to commercial customers; for example, customers with larger loads (e.g., hospital) may be on a different schedule than small commercial loads (e.g., restaurant).

To help ascertain the importance of the assigned rate schedule, we conduct a sensitivity using multiple types of rate schedules available to commercial customers in the Southern California Edison (SCE, CA) service territory. SCE is chosen because of the multiplicity of rate schedules available to commercial customers in the service territory of a large Californian utility.

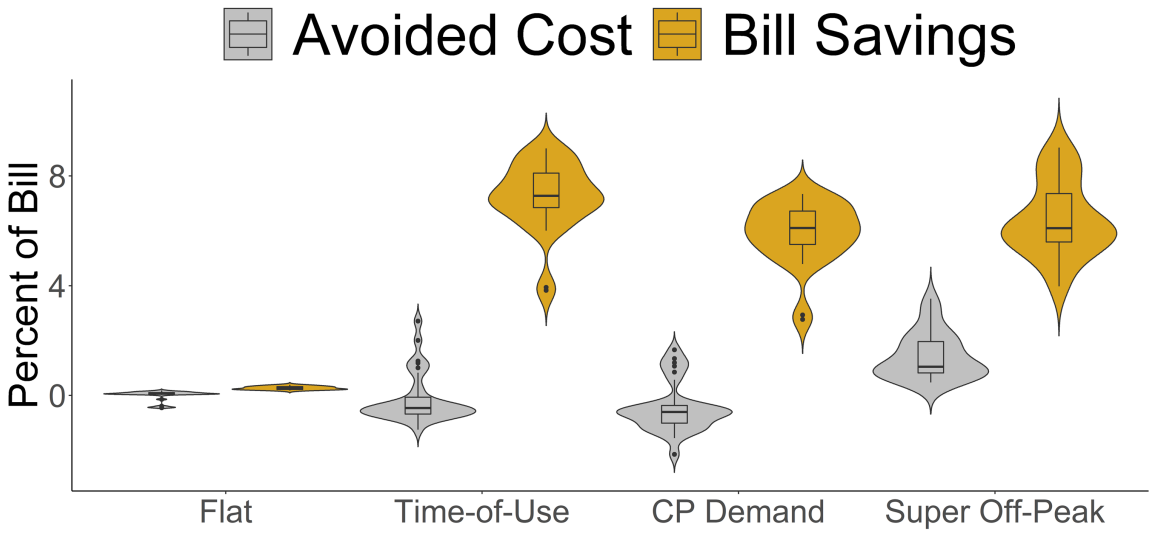

Figure A-4: Southern California Edison commercial rate analysis.

Table A-3: SCE Case Labeling.

Case Label	Load Shapes	Rate	Avoided Costs
Flat	Los Angeles DOE	SCE GS-1	SCE Climate Zone 6 2018
Time-of-Use	Los Angeles DOE	SCE TOU-8 Option B	SCE Climate Zone 6 2018
CP Demand	Los Angeles DOE	SCE GS-2 Demand	SCE Climate Zone 6 2018
Super Off-Peak	Los Angeles DOE	SCE TOU-GS-3-SOP	SCE Climate Zone 6 2018

We see that in terms of bill savings, the only outlier among the tested rate is the flat rate. This is because the flat rate, meant for small commercial customers, has no time-differentiation of energy charges nor a demand charge. There is no incentive for customers to install storage on such a rate, and these types of rates are not considered in the main text. The other rates achieve similar bill savings regardless of the details of how their time-dependent energy and demand charges are levied, the most notable difference is that the super-off peak rate achieves higher avoided costs because the super-off peak energy rate encourages battery charging at times when avoided costs are lowest. Overall, results suggest there are not significant differences in bill savings given rate schedules of similar levels of complexity.

Main Case Parameterization

Table A-4: Electricity Tariff Schedules.

Case Utility	Tariff ID	Energy Charge ($/kWh)	Demand Charge ($/kW)	Customer Charge ($/mo.)
Austin Energy (TX)	Commercial - Secondary Voltage 10kW-300kW - Inside	$.058	$12.49 NCP	$27.5
Boston Edison (MA)	Boston Edison Company General Service Time of Use G-3 (B3, B6)	$.16109	$17.96/$23.93 Winter NCP/ Summer NCP	$237.07
Commonwealth Edison (IL)	BES-Large Load Delivery Class (Secondary)	$.07927/.07679 Winter/Summer	$7.05 NCP	$121.24
Dominion (VA)	GS-2 Demand	$.0798-$.0363 (declining block)	$9.908/$11.346 Winter NCP/ Summer NCP	$21.17
Duquesne Light Company (PA)	General Service Medium More Than or Equal to 25 kW	$.0903	$7.09 NCP	$54
Pacific Gas & Electric (CA)	E-19 Medium General Demand TOU (Secondary)	$.08717/.10165/.14726 Off-peak/part-peak summer & peak winter/peak summer Weekdays	$17.33/+$.13/+$5.23 +$18.74 NCP/winter peak/ summer part peak/ summer peak Weekdays	$599.59

Tariff IDs are as listed in OpenEI's Utility Rates Database [40]. Customer charges vary considerably based on the selected rate within a service territory, rates meant for larger consumption customers have larger fixed charges. Because customer charges for the above-selected rate schedules may thus not represent the customer charge faced by a customer, they are excluded when calculating the percent reduction in bill in the main text figures.

A.3 Avoided Costs

Hourly average avoided costs in PG&E Climate Zone 3A (San Francisco) for 2018 from the CPUC avoided cost calculator are shown below [37]. We assume a 2018 resource balance year.

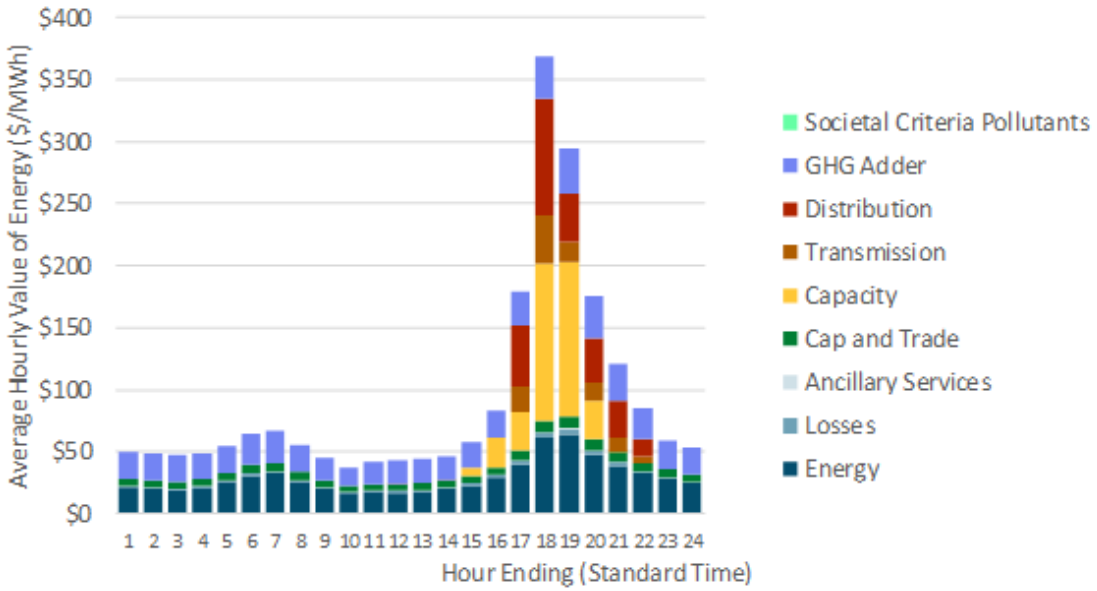

Figure A-5: CPUC Avoided Cost Calculator PG&E Climate Zone 3A 2018 avoided costs.

We similarly stack hourly average avoided costs for all jurisdictions below without differentiation of avoided cost components. Component estimates are explained in subsequent sections.

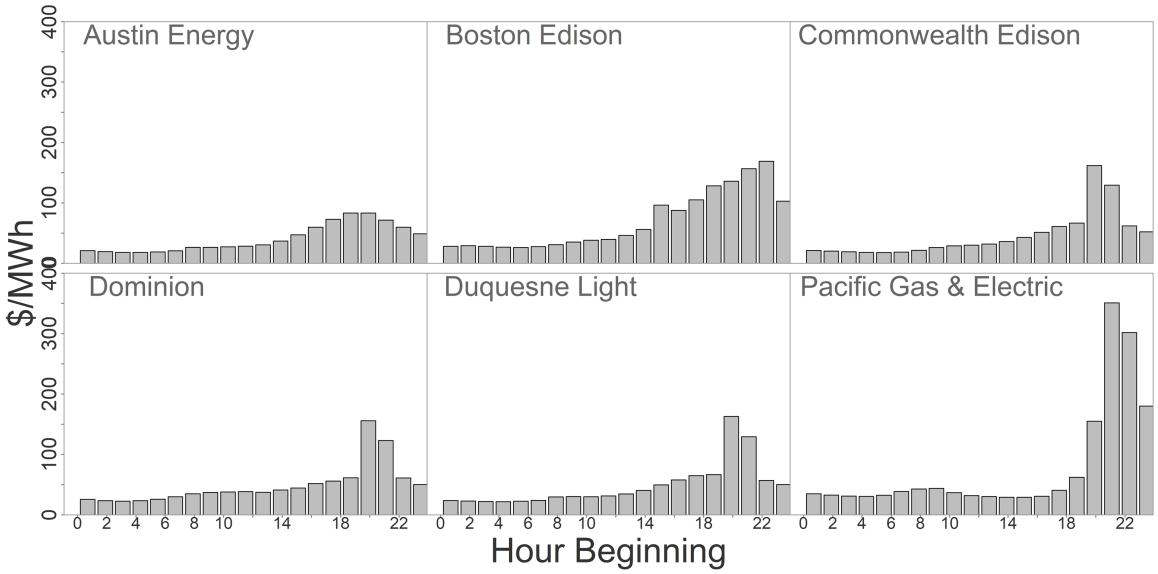

Figure A-6: Hourly average avoided costs for six main cases.

Energy and Losses

Table A-5: LMP Nodes for energy avoided costs and losses.

Case Utility	LMP Node	ISO	Distribution Losses Adder	Year
Austin Energy (TX)	LZ_SOUTH	ERCOT	5%	2017
Boston Edison (MA)	.ZNEMASSBOST	ISO-NE	5%	2017
Commonwealth Edison (IL)	COMED	PJM	5%	2017
Dominion (VA)	DOM ANACONDA115 KV	PJM	5%	2017
Duquesne Light Company (PA)	AES DUQ 13 KV MAIN	PJM	5%	2017
Pacific Gas & Electric	TH_NP15_GEN-APND	CAISO	CPUC Avoided Cost Calculator	2017

Energy costs are approximately by locational marginal prices (LMPs) in nodal competitive wholesale markets in many parts of the United States. LMPs account for losses at the bulk system level, but most losses occur below the resolution of ISO pricing nodes as resistive losses on distribution feeders. We add 5% to LMPs to account for this effect of losses, except in the California case, where we default to the values in the CPUC Avoided Cost Calculator [37]. Borenstein and Bushnell [56] find average US distribution losses to be 6.2%, though marginal losses are higher at 8.9% due to the reality that marginal MWhs are served at higher load than the average MWh, with considerable variation across the country and distribution feeders.

Shapes for the LMPs (excluding distribution losses) for the six main cases are shown below as an annual hourly price duration curve, with an inset adding resolution on the top 100 hours.

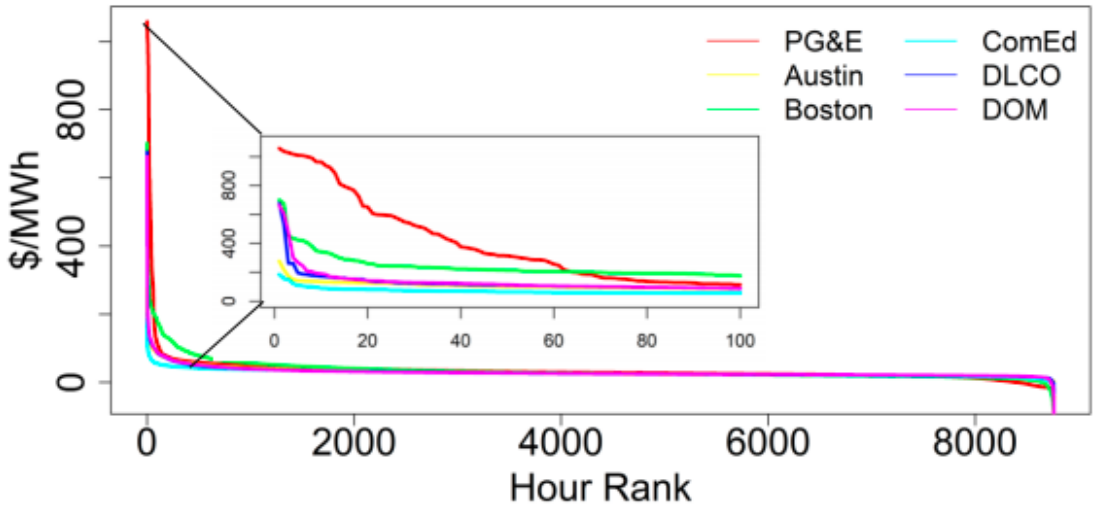

Figure A-7: Price Duration Curve for LMPs at nodes for six main cases.

Generation Capacity, Transmission and Distribution

Generation capacity values in markets with capacity remuneration mechanisms and the allocation method for these costs are listed below.

Table A-6: Generation capacity avoided cost assumptions.

Case Utility	Capacity Zone	ISO	Capacity Price	Year	Allocation Method
Austin Energy (TX)	N/A	ERCOT	N/A	2017	N/A
Boston Edison (MA)	NEMA-BOSTON	ISO-NE	$493.15/MW-day	2017	PCAF
Commonwealth Edison (IL)	COMED	PJM	$153.61/MW-day	2017-18	5CP
Dominion (VA)	DOM	PJM	$153.61/MW-day	2017-18	5CP
Duquesne Light Company (PA)	DLCO	PJM	$153.61/MW-day	2017-18	5CP
Pacific Gas & Electric (CA)	CAISO (bilateral)	CAISO	CPUC Avoided Cost Calculator 2018 Resource Balance Year	2018	CPUC Avoided Cost Calculator

Marginal transmission and distribution costs and the allocation method for these costs are listed in the table below.

Table A-7: Transmission and Distribution avoided cost assumptions.

Case Utility	Value	Allocation Method	Allocator Shape
Austin Energy (TX)	$80/kW-year	PCAF	LZ_SOUTH 2017 Load
Boston Edison (MA)	$80/kW-year	PCAF	LD.ALEWIFE 13.8
Commonwealth Edison (IL)	$80/kW-year	PCAF	COMED 2017 Zonal Load
Dominion (VA)	$80/kW-year	PCAF	DOM 2017 Zonal Load
Duquesne Light Company (PA)	$80/kW-year	PCAF	DUQ 2017 Zonal Load
Pacific Gas & Electric (CA)	CPUC Avoided Cost Calculator ($130/kW-year)	CPUC Avoided Cost Calculator (uses PCAF)	CPUC Avoided Cost Calculator

The magnitude of avoidable generation capacity costs is based on capacity market clearing prices in PJM and ISO-NE for the relevant year. ERCOT does not have a capacity remuneration mechanism. California does not have a centralized capacity market, load-serving entities contract for capacity bilaterally to meet obligations. We follow the CPUC Avoided Cost Calculator in setting the cost of the marginal capacity contract to the net cost of new entry (net CONE), and assume capacity is needed on the margin in the near-term by setting the resource balance year to 2018. Generation capacity costs are allocated using the year's five coincident peak (5CP) days in PJM that are used to allocate these costs to customers. 5CP days are calculated after the summer season, so in practice customers and load-serving entities must make educated guesses about which days will count toward the 5CP allocation. Our approach implicitly assumes 15 CPP calls will be enough to accurately target the 5CP days.

On actual systems the magnitude of avoidable or deferrable T&D costs vary by feeder [57] [58]. We choose a central non-California estimate of $80/kW-year based on review of [17], [20], [37], [59] [60]. Allocations of this cost, which determine whether storage installations can avoid it, are as described below.

Avoided T&D costs and generation capacity in ISO-NE are allocated using the peak capacity factor allocation (PCAF) method. The PCAF method was developed by Pacific Gas & Electric in 1993 for their General Rate Case, and has been in use since for these applications [37], [61]. The PCAF method sets a load threshold above which avoided grid infrastructure costs are allocated, generally based on either the standard deviation in loading on the feeder or a percent of annual hours. We base our threshold on the standard deviation in load on the assigned allocation shape

(e.g., the LD.ALEWIFE 13.8 bus for Boston Edison) as compared to the peak load. The threshold calculation for a year y of hourly loads is thus

$$Threshold_y = max(Load_y) - \sigma^{Load} \tag{A.1}$$

Allocation weights are then calculated as

$$PCAF_{y,h} = \frac{max(Load_{y,h} - Threshold_{y,h}, 0)}{\sum_{h=1...8760} max(Load_{y,h} - Threshold_y, 0)} \tag{A.2}$$

where y is the year for which the PCAF is calculated, and h is the hour of the year.

Using these methods, the generation capacity and T&D avoided costs are allocated on an hourly basis. Below, we show the annual distribution of the hours to which the generation capacity and T&D avoided costs are assigned.

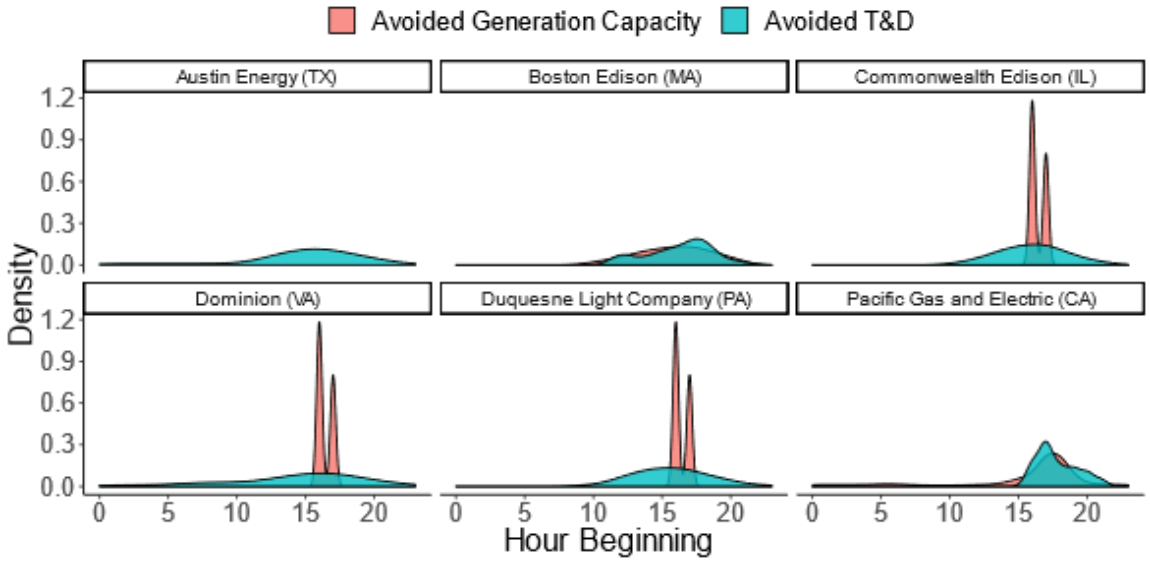

Figure A-8: Avoided Capacity, T&D hourly distribution.

This shows the effect of the 5CP allocation method for generating capacity avoided costs in PJM on the figures in the main text. Because the 5CP allocates avoided capacity in just five hours, these costs are more easily targeted and avoided by an optimized CPP rate.

Carbon Dioxide Sensitivity

We conduct a sensitivity on climate-related damages associated with emissions of CO_2. Marginal emissions factors for CO_2 are taken from Azevedo et al. [43] for the five utilities outside California. These emissions factors are shown below:

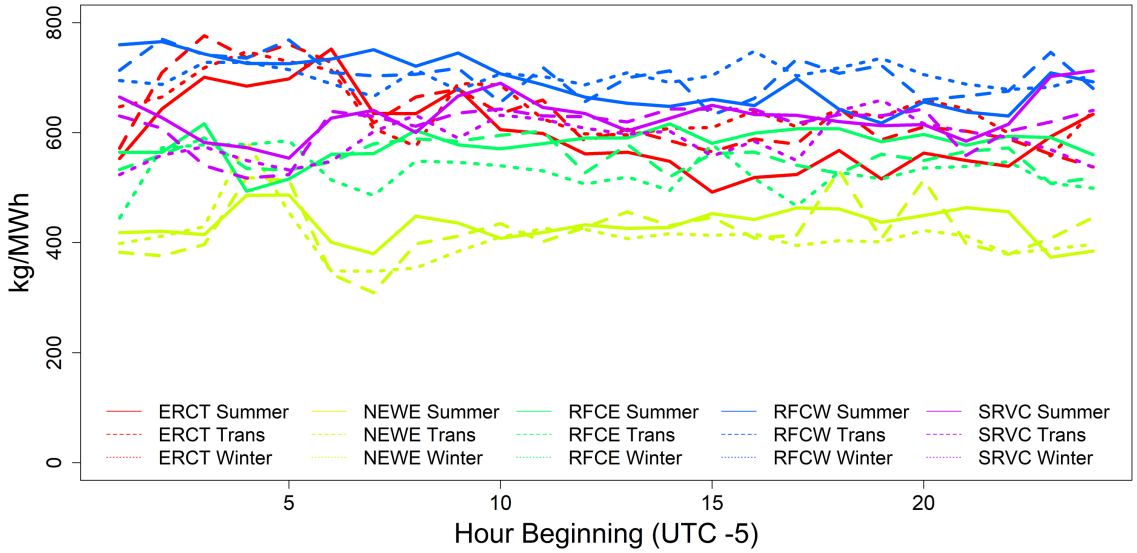

Figure A-9: Hourly marginal emissions factors for non-California cases.

In California, we use the CPUC Avoided Cost Calculator emissions factors. In all jurisdictions we apply a \$63/ton (2018\$) carbon price to monetize the emissions, this value is used to be consistent with the CPUC Avoided Cost Calculator and is broadly consistent with social cost of carbon (SCC) estimates [62], though some more recent estimates develop a higher SCC [63]. Because inclusion of CO_2 has little effect on avoided costs, we do not conduct further sensitivity to the SCC.

Other

Additional avoided costs are included in the California case for avoided cap and trade allowances, ancillary services, and societal criteria pollutants. Cap and trade is California state policy, and the other two components are small (<\$1/MWh).

Three other important categories of avoided costs are sometimes considered, but not included in this study. They are

1. Other avoided environmental damages

2. Financial avoided costs

3. Avoided customer outages

Other environmental damages include health damages associated with SO_x, NO_x, and other emissions harmful to human health. Though these damages may be considerable [64], they are not generally part of a utility's internalized cost of service, and would likely move in a similar direction to the CO_2 sensitivity we do conduct, as in Fisher and Apt [4].

Financial avoided costs include the market price effect, fuel price hedging, and other risks [15]. The market price effect, or demand reduction induced price effect (DRIPE) results from a reduction in inframarginal rents earned by still-operating generators in competitive wholesale markets with a market clearing price based on the marginal generator bid. If DERs decrease market prices, increase in consumer surplus associated with this effect can be sizable [18]. Because storage, unlike efficiency or behind-the-meter solar PV, increases metered customer load in some hours and decreases it in others, it is less clear what the direction of this effect would be. Additionally, the market price effect is arguably a short-run transfer from producers to consumers, rather than an avoided cost, though to the extent it corrects existing market inefficiencies it can be a durable benefit [15]. Fuel price hedging is meant to defray risks of gas price volatility, some of which could arguably be avoided by decreasing reliance on natural gas as the marginal generation fuel due to DERs.

Finally, avoided customer outages are a potentially sizable customer benefit of storage installation if appropriately configured to provide this service [20]. We do not consider this benefit partially because it is difficult to estimate and likely varies across customers, but also because it does not have direct impact on a utility's avoided costs because served by storage during an outage is by definition not load that could be served by the utility. However, this benefit may have important implications for customer-sited storage adoption, particularly in jurisdictions like parts of the California IOU service territories during the 2019 Public Safety Power Shutoffs where customers have recently endured long duration outages.

Sensitivity to CPP Rate

An example case shown below for a San Francisco climate zone Primary School on PG&E's E-19 rate demonstrates the general effect of CPP implementation on battery dispatch and avoided costs. As in the main text, CPP calls for this figure are 4 hours in duration, 15 times per year, optimized by the tool to be during the 15 periods with highest avoided costs.

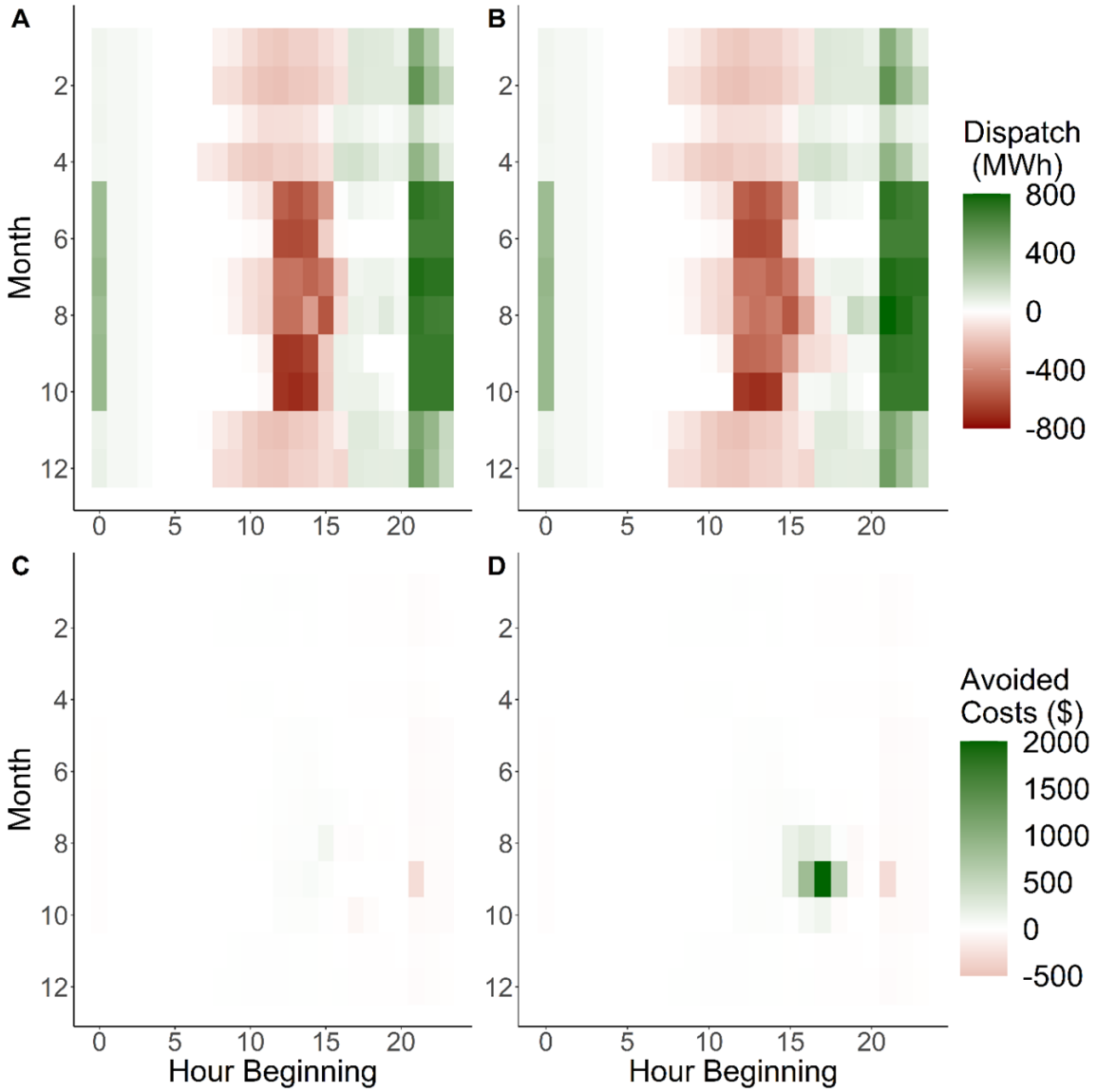

Figure A-10: Detailed storage dispatch (A,B) and avoided costs (C,D) for Primary School on PG&E E-19 rate without (A,C) and with (B,D) peak pricing. Positive storage dispatch is charging battery, negative storage dispatch is discharging battery.

Re-dispatch is small in MWh terms, but has large effects on the avoided costs because those handful of hours during the 15 four-hour CPP calls are so valuable.

We also consider the sensitivity of the cases to the level the CPP rate adder is set at. In the main text, the CPP is implemented as a \$.50/kWh adder during the CPP calls. This level is chosen for some consistency with existing rates in California plus adjustment for relatively lower rates in most other jurisdictions. For example, the CPP is \$.862/kWh on PG&E's Peak Day Pricing rate.[7] However, we note our optimization is very sensitive to inclusion of the CPP price; a very low CPP adder achieves most benefits from battery re-dispatch.

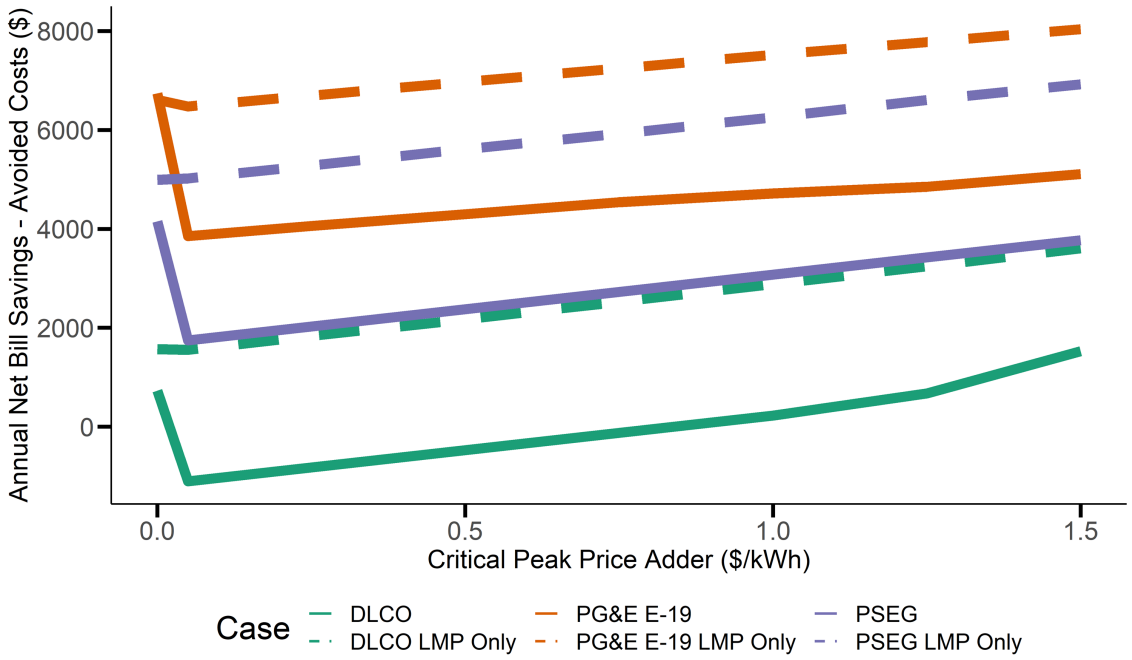

Figure A-11: Difference between bill savings and avoided cost for example customer on CPP rate. PSEG is Public Service Electric and Gas, a New Jersey utility. LMP cases consider only avoided energy and losses.

In practice, CPP rates will have to overcome transaction and informational costs, and the setting of a higher CPP rate may allow for decreases in off-peak prices resulting in the same overall bill for customers who do not respond at all to a CPP rate.

[7] https://www.pge.com/en_US/business/rate-plans/rate-plans/peak-day-pricing/peak-day-pricing.page

A.4 Correlation of Avoided Costs and Load Shapes

The DOE load shapes are weather-normalized, while avoided cost data is based on a specific year (often 2017) of marginal energy, capacity, T&D, and CO_2 costs. This creates concern about whether the weather-normalization process is underestimating customer loads on realized peak days (e.g., in 2017) in the avoided cost data. As a check on robustness to this concern, we again compare with metered 2013 data for approximately 500 C&I customer loads in Duke Energy Carolinas. We check whether this load data is any more correlated with loads and LMPs in the nearest PJM zone (Dominion) for which there is hourly wholesale pricing data than are the DOE shapes (Baltimore) assigned to our Dominion case. Results do not suggest meaningful differences in the correlation between loads and prices associated with using metered data vs. the weather-normalized shapes.

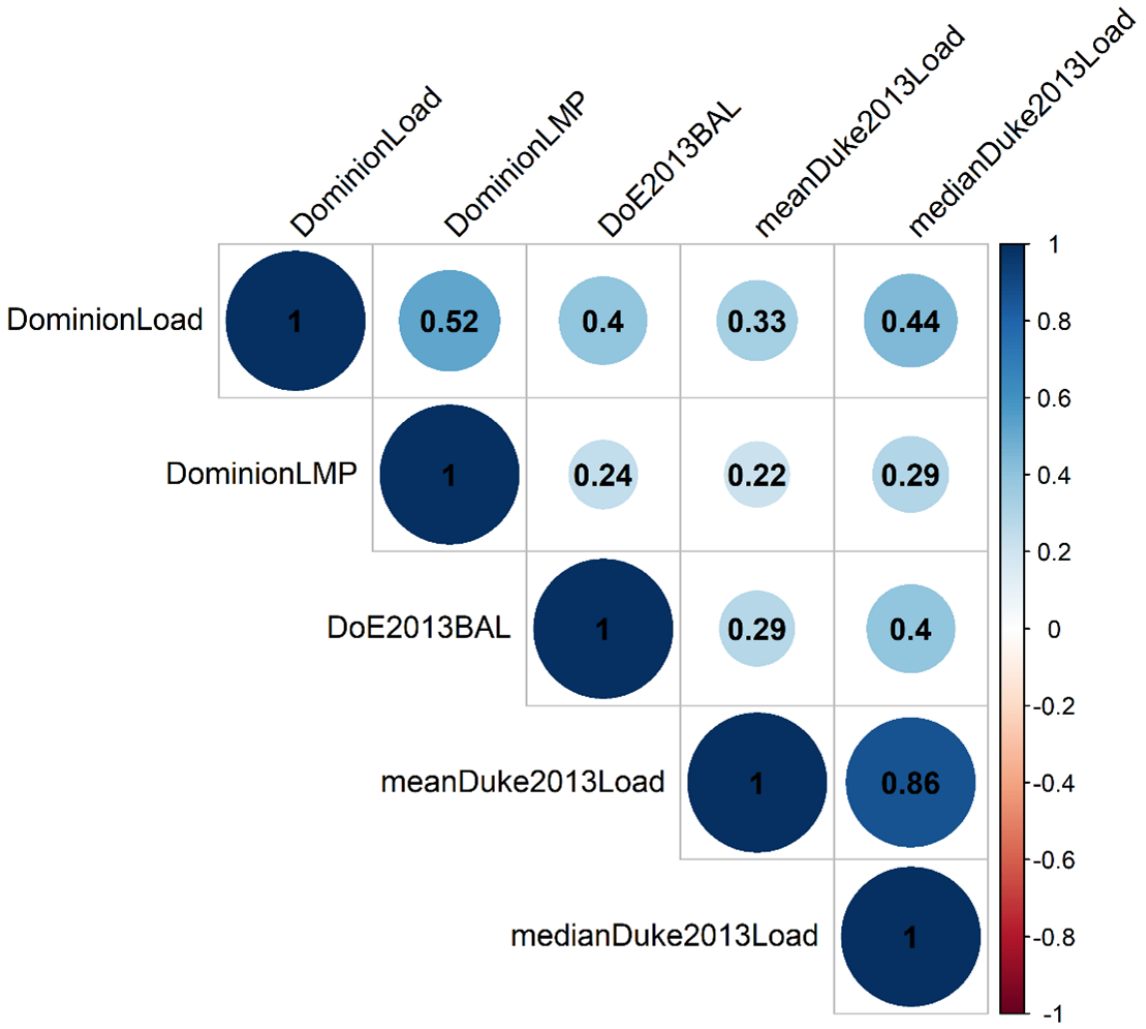

Figure A-12: Correlation plot comparing metered and DOE customer loads to Dominion zonal loads and prices in 2013. Duke Energy Carolinas metered loads are weighted both by the average and median customer load for comparison.

B Optimization Formulation and Computational Methods Appendix

B.1 Optimization Formulation

Constants and Parameters

Constant or Parameter	Definition	Base Case Value	Units
η	Round-trip efficiency of battery	85	%
Deg	Marginal benefit of usage (or levelized cost of degradation)	\$5 [36]	MWh^{-1}
$E_{i,t}$	Cost of electricity in period i at time t	Varies by customer tariff	\$/kWh
C	Customer charge	Varies by customer tariff	\$/mo.
D	Cost of peak demand across all t	Varies by customer tariff	\$/kW
D_j	Cost of peak demand in time period j	Varies by customer tariff	\$/mo.
NL_t	Native load at time t	Varies by customer tariff	kWh
NML	Native maximum customer load	Varies by customer tariff	kW
P	Rated power of battery	$0.2 * NML$	kW
DU	Duration of battery	4	hours
SOC^{max}	maximum state of charge of battery	$P * DU$	kWh
SOC^{min}	minimum state of charge of battery	$0.2 * P * DU$	kWh
SOC^0	initial SOC of battery entering first time period	SOC^{min}	kWh

Variables

Variable	Definition	Units
C_t	Battery charging at time t	kW
D_t	Battery discharging at time t	kW
SOC_t	State of charge of battery at time t	kWh
L_t	Customer net load at time t	kWh
ML	Customer net max load across all t	kW
ML_j	Customer net max load in time period j	kW

Sets

Set	Definition
$t \in T$	Hour of the year. Optimization is run monthly and added together to obtain annual totals
$j \in J$	Set of demand charge periods; are a subset of the hours of the year
$i \in I$	Set of energy charge periods; are a subset of the hours of the year

Objective Function

$$Min \sum_{t=1...8760} L_t * E_{i,t} + ML_j * D_j + ML * D + C + (C_t + D_t) * Deg \tag{B.1}$$

Constraints

$$L_t = NL_t + C_t - D_t \tag{B.2}$$

$$SOC_t = SOC_{t-1} + C_t * \sqrt{\eta} - \frac{D_t}{\sqrt{\eta}} \tag{B.3}$$

$$ML \geq L_t, \forall t \tag{B.4}$$

$$ML_j \geq L_{t,j}, \forall t, j \in J \tag{B.5}$$

$$0 \leq C_t \leq P \tag{B.6}$$

$$0 \leq D_t \leq P \tag{B.7}$$

$$SOC^{min} \leq SOC_t \leq SOC^{max} \tag{B.8}$$

Further Explanation

C_t and D_t are the hourly battery charge and discharge decision variables; other variables attain their value as per the constraints based on these decision variables. We do not constrain the battery to exclusively be either charging or discharging, as the formulation of the problem makes this constraint dominated in all hours (i.e., the optimal solution will never include charging and discharging in the

same hour). Constraints (B.4) and (B.5) linearize non-coincident and period-based demand charges, allowing formulation of the dispatch as a linear program. They result in an equivalent solution to implementing a non-linear $max()$ statement for positive loads and demand charges. Defining SOC^0 as SOC^{min} ensures the initial and final battery state of charge will be the same without additional constraints because the objective function minimizes the customer's bill, so a battery without any forward-looking degradation costs for subsequent time periods will always end in its minimum state of charge.

B.2 Computational Methods

The above optimization is implemented as a linear program in Pyomo. It is solved using GLPK[8], an open-source solver.

C Results Sensitivities Appendix

C.1 Perfect Foresight Sensitivity

Our battery dispatch optimization has monthly perfect foresight as implemented in the main cases. Because customer peak loads cannot be perfectly forecast, these results set an upper bound on bill savings. We use a more naïve persistence forecast of customer load to set a lower bound on bill savings for the PG&E and Dominion cases. Persistence forecast cases assume that customer loads will be exactly equivalent to the previous day's loads, except for adjustment when a weekday follows a weekend day or holiday (and vice-versa), in those cases the forecast load is equivalent to the previous day of the same day type. Persistence sets a lower-bound as, in practice, a better forecast (e.g., including forecast temperature variables) could improve on peak load forecasting, though never doing quite as well as perfect foresight. Results are in Figure C-1.

[8]https://www.gnu.org/software/glpk/

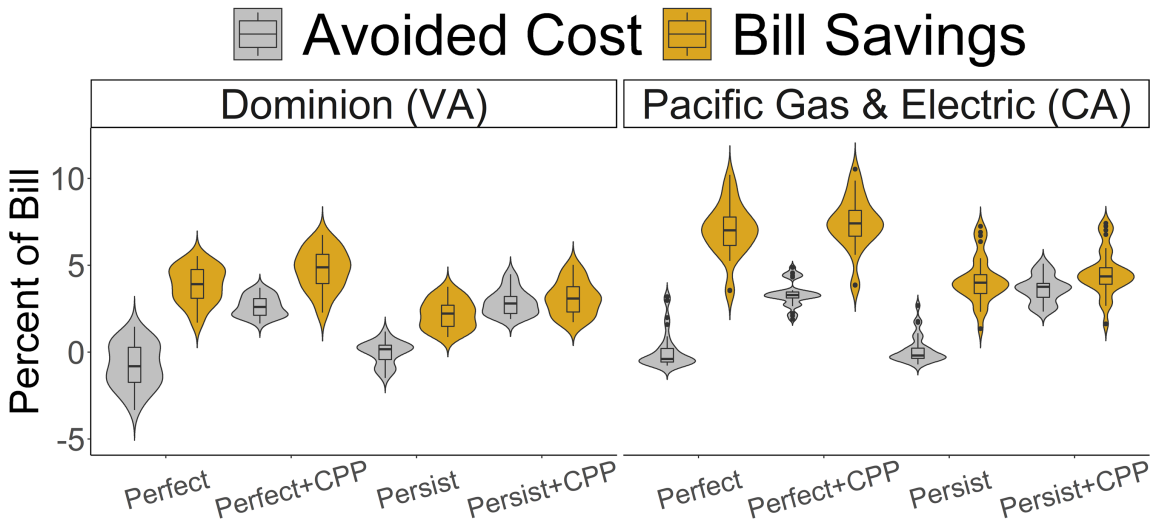

Figure C-1: Persistence sensitivity.

Table C-1: Figure C-1 Case Parameterization.

Case Label	Load Shapes	Rate	Avoided Costs	Critical Peak Pricing	Optimization Dispatch Logic
Perfect	Baltimore, San Francisco, DOE shapes (n=30)	Dominion GS-2 Demand, PG&E E-19 Medium TOU	Energy, losses, capacity, T&D	None	Perfect knowledge of customer peak load hours
Perfect+CPP	Ibid	Ibid	Ibid	$0.50/kWh adder for optimal 15 four-hour annual calls	Ibid
Persist	Ibid	Ibid	Ibid	None	Expect customer load to be the same as previous day's load
Persist+CPP	Ibid	Ibid	Ibid	$0.50/kWh adder for optimal 15 four-hour annual calls	Ibid

Median bill savings in persistence cases are 57-63% of perfect foresight bill savings across the two utilities and for cases with and without CPP. These are the anticipated lower and upper bounds of savings. Median avoided costs are largely unaffected, though the distribution changes in some cases with the re-dispatch of the customer's battery. The relatively unchanged median across CPP cases and increase in avoided costs versus non-CPP cases shows a persistence forecast can still do well in targeting battery dispatch to the very highest system avoided cost hours during which critical peaks occur, assuming the utility properly targets its critical peaks and sends the appropriate price signal in advance. Therefore, while relaxing the perfect forecast assumption significantly reduces bill savings, negatively affecting whether a customer will realize sufficient savings to make installing a battery financially worthwhile [4], it has less effect on the savings accrued by the utility under smarter non-residential rate designs.

C.2 Persistence Comparison of DOE and Actual Duke Load Shapes

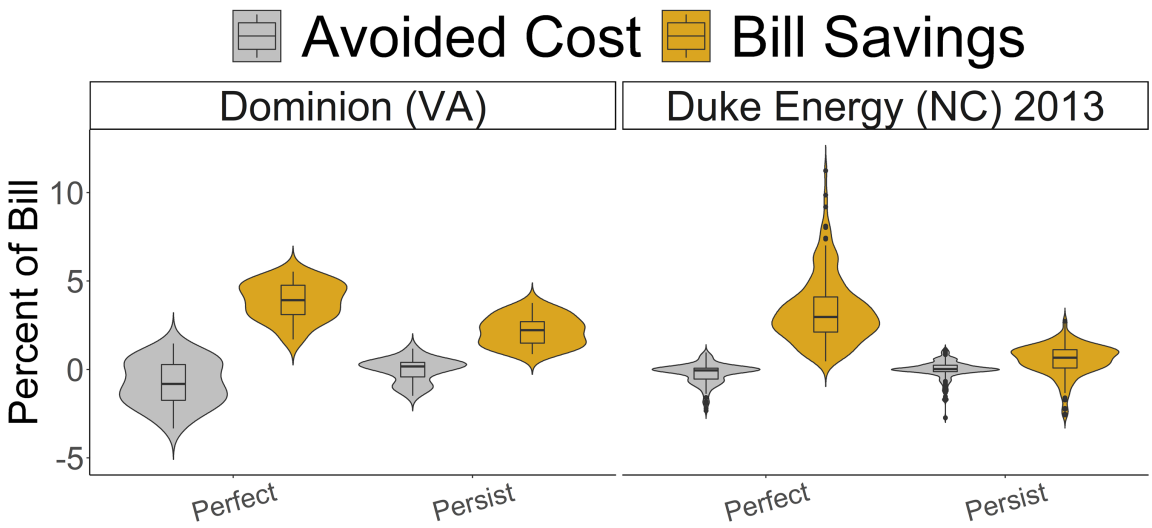

Figure C-2: Persistence sensitivity using Duke Energy Carolinas 2013 load shapes.

As in Appendix A for the main results, Figure C-2 compares the main results, which use the DOE representative commercial load shapes, with metered load data for approximately 500 C&I customers in Duke Energy Carolinas' service territory in 2013. As is the case for the main persistence sensitivity, bill savings decline substantially under a persistence rather than perfect foresight forecast of customer loads, while median avoided costs are mostly constant, though with a shift in the distribution. However, while median persistence forecast bill savings in are 57% of perfect foresight forecast bill savings in the Dominion cases (left panel of Fig. C-1), they are only 22% of perfect foresight forecast bill savings in the Duke Energy Carolinas cases. This suggests actual metered customer loads are peakier than the representative DOE shapes, and therefore more affected by imperfect forecasts of when customer peak loads will occur. As in the main persistence sensitivity, this affects the customer's financial proposition for installation more than the utility's avoided cost-based proposition.

C.3 Demand Charge Temporal Resolution

All our cases are run at hourly resolution, so demand charges are also levied at hourly resolution on a customer's peak monthly load hour. In practice, many customers' demand charges are levied on shorter time increments of 30 or 15 minutes, though 60 minute resolution also exists [34]. We plan to run sensitivities to the resolution of the demand charge before submitting for publication, but do not anticipate large changes to results.

C.4 Battery Sizing

All customers have four-hour duration batteries sized to 20% of their annual peak load in the main cases. Customers may size batteries differently than this assumption, both in terms of duration and capacity. Reasons for these sizing parameters are given in the main text. We may conduct additional sensitivities to battery sizing as time allows, but because we are explicitly not evaluating adoption or the cost-effectiveness of installation from the customer's perspective, we think the sizing of the battery within reasonable bounds is unlikely to matter in our RIM framework.

C.5 Optimization Degeneracy of Time-of-Use Periods

Some customers in our main cases, such as those on the PG&E E-19 rate, face energy charges with pre-defined time-of-use periods. TOU periods are often longer than the duration of the battery, so the solution to the battery dispatch optimization may be degenerate regarding when the battery charges or discharges during a given TOU period with no binding demand-related constraint. An example of how our optimization responds to this situation is shown below for dispatch on the PG&E E-19 rate:

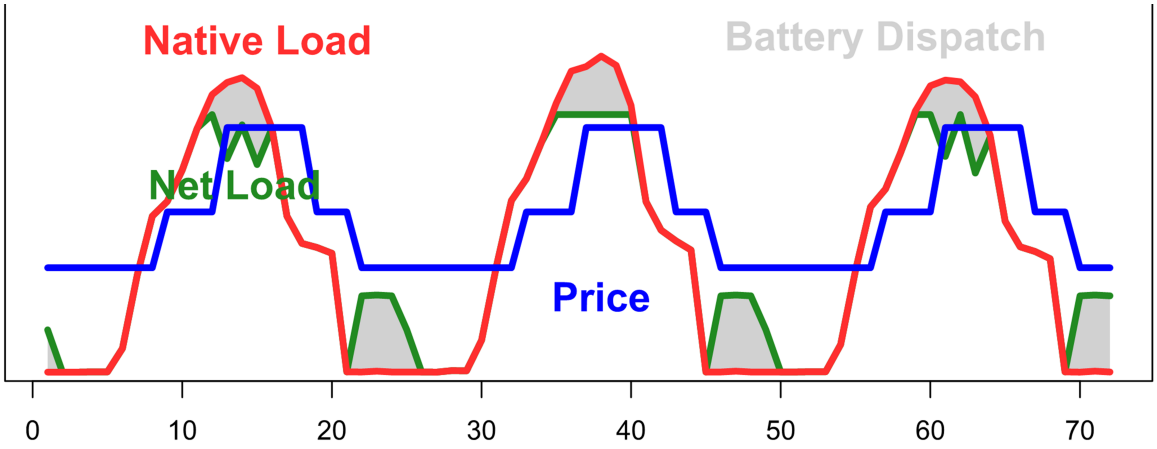

Figure C-3: Three days of customer-sited storage dispatch on a TOU rate.

In the absence of adding a penalty to the objective function for deviation from an ideal state of charge for the battery, we consider the results above to be the most sensible outcome. The battery charges as soon as the lowest price TOU period begins, even though charging any time during this period would be equivalent form a customer bill perspective. The reason to prefer earlier charging is, in practice, lack of forecast of future customer load will slightly preference charging the battery earlier to guard against an unforeseen load spike (or, if the battery were enabled for it, avoidance of a customer outage).

C.6 Installation Likelihood Assumptions

All cases in Figure 2.7 assume a 10-year battery lifetime over which to accrual bill savings, 10% real discount rate, bill savings equivalent in each of the ten years, and ten payments for the upfront battery costs made as an annuity at the end of each year based on the discount rate. The purpose of this figure is explicitly not to conduct a Participant Cost Test, so we show only a few breakeven upfront installed costs for comparison and do not try to estimate actual installation costs faced by the customer or a third-party battery owner. For the same reason we conduct no further sensitivities on battery lifetime and discount rate.

To make Figure 2.7 more legible while still plotting all 30 DOE load shapes across the six modeled utilities we group similar types of buildings into 8 labels, as described in Table C-1.

Table C-2: DOE Load Shape labels in Figure 2.7.

Shape Name in Figure 2.7	Included DOE Load Shapes
Apartment	MidriseApartmentNew2004, MidriseApartmentPre1980
Hospital	HospitalNew2004, HospitalPre1980
Hotel	SmallHotelNew2004, SmallHotelPre1980, LargeHotelNew2004, LargeHotelPre1980
Office	SmallOfficeNew2004, SmallOfficePre1980, MediumOfficeNew2004, MediumOfficePre1980, LargeOfficeNew2004, LargeOfficePre1980
Restaurant	QuickServiceRestaurantNew2004, QuickServiceRestaurantPre1980, FullServiceRestaurantNew2004, FullServiceRestaurantPre1980
Retail	SuperMarketNew2004, SuperMarketPre1980, StripMallNew2004, StripMallPre1980, Stand-aloneRetailNew2004, Stand-aloneRetailPre1980
School	PrimarySchoolNew2004, PrimarySchoolPre1980, SecondarySchoolNew2004, SecondarySchoolPre1980
Warehouse	WarehouseNew2004, WarehousePre1980

Chapter 3 Resource Adequacy Implications of Temperature-Dependent Electric Generator Availability*

Abstract

Current grid resource adequacy modeling assumes generator failures are both independent and invariant to ambient conditions. We evaluate the resource adequacy policy implications of correlated generator failures in the PJM Interconnection by making use of observed temperature-dependent forced outage rates. Correlated failures pose substantial resource adequacy risk, increasing PJM's required reserve margin from 15.9% to 22.9% in the 2018/2019 delivery year. However, PJM actually procured a 26.6% reserve margin in this delivery year, translating to excess capacity payments of $315 million and an implied value of lost load of approximately $700,000/MWh, a figure two orders of magnitude greater than typically used in operational contexts. Capacity requirements vary by month, with more than 95% of loss-of-load risk accruing in July. Setting monthly capacity targets could reduce annual PJM procurement by approximately 16%. We examine the resource adequacy implications of the ongoing replacement of nuclear and coal in PJM with combined-cycle gas generators, finding moderate benefits: approximately 2% reduction in capacity requirements. We identify modest resource adequacy risks from potential future climate scenarios, modeled as temperature increases of 1 and 2 degrees Celsius relative to our study period. Holding loads fixed, these scenarios increase capacity requirements by approximately 0.5% and 1.5%, respectively.

* This paper was published as S. Murphy, L. Lavin, and J. Apt, "Resource Adequacy Implications of Temperature-Dependent Electric Generator Availability," *Applied Energy*, 2020. 262: 114424. doi:https://doi.org/10.1016/j.apenergy.2019.114424

Abbreviations and acronyms

CC	Combined cycle natural gas generator type
CDD	Cooling degree days
DR	Demand response
EFDH	Equivalent forced derating hours
EFOF	Equivalent forced outage factor
FOH	Forced outage hours
EFORd	Equivalent demand forced outage rate, a measure of the probability a generating unit will that not be available due to forced outages or forced deratings when there is demand on the unit to generate
ELCC	Effective load carrying capability, a measure of the additional load the power system can serve at equivalent reliability due to the addition of a generator
EUE	Expected unserved energy, the aggregate quantity of load expected to be unmet during the delivery year due to generation insufficiency
FEF	Forecast error factor
FERC	Federal Energy Regulatory Commission
GADS	Generating Availability Data System
HDD	Heating degree days
HE	Hour Ending
LOLE	Loss of load expectation, the number of occurrences of generation insufficiency per year for a given forecast peak load and forecast generator fleet
MPP	Most probable peak, in a week
NERC	North American Electric Reliability Corporation
NU	Nuclear generator type
PH	Period hours
PJM	The PJM Interconnection, the largest system operator by installed generation capacity and load in North America
RECAP	Renewable Energy Capacity Planning Model
RRS	PJM's annual Reserve Requirement Study, which determines the reserve margin required to achieve 0.1 LOLE reliability target
ST	Steam turbine generator type
VOLL	Value of lost load, the valuation of a unit of unmet load
XEFORd	Equivalent demand forced outage rate but excluding causes of outages that are outside management control (for example, loss of a transmission line)

3.1 Introduction

Significant attention has been given to the question of how much generation capacity is required to ensure power system reliability. In the U.S., system operators[1] have formalized probabilistic approaches for determining capacity requirements in their balancing areas, termed resource adequacy

[1]We use this term to refer to both Independent System Operators (ISOs) and Regional Transmission Organizations (RTOs).

modeling. These methods are designed to comply with resource adequacy standards established by the North American Electric Reliability Corporation (NERC) and approved by the U.S. Federal Energy Regulatory Commission (FERC) [1]. Probabilistic resource adequacy methods have existed since the mid-20th century. For example, [2] uses probability theory to calculate loss of load expectation for small power systems using pooled forced outage rates and assuming independent generator failures. Hall et. al [3] develops a method of calculating cumulative generator availability probabilities that assumes exponentially distributed available and unavailable residence times for individual generators. This capacity availability model is combined with a Markov model of load states in [4] to calculate cumulative capacity reserve state probabilities. Authoritative volumes describing established practice had been written by the end of the century [5, 6]. Common to these methods is the representation of generators using Markov models whose probability of unscheduled failure can be represented by an availability statistic such as the forced outage rate [7]. The earliest methods for calculating resource adequacy requirements were analytical, with simulation methods developed subsequently [8].

After the creation of the PJM market, increasing data and computational power, variable renewable energy penetrations, and experience with market mechanisms for capacity procurement have expanded the scope of resource adequacy modeling. Felder [9] identifies common-cause failures and temperature as important for resource adequacy, but is able to analyze only test cases in the absence of system level data available in our study. Dragoon and Dvortsov [10] proposes a heuristic method for calculating marginal adequacy contributions of resources; we describe here quantitative methods based on observed correlated failure statistics to calculate marginal resource adequacy contributions. In more recent years research has focused on improving quantification of variable renewable energy (predominately wind and solar photovoltaics) contributions to resource adequacy [11, 12], with additional attention paid to methods for crediting these resources in capacity markets [13, 14] and whether these resources create a need to better incorporate intertemporal resource flexibility in resource adequacy modeling [15]. Finally, capacity markets and market design around resource adequacy are an active area of research. Attention has focused better quantifying the economic cost of shedding firm load due to generation insufficiency [16]; variance in these estimates

is important in setting reliability criteria [17]. The relative merits of energy-only constructs with scarcity pricing [18] versus longer time horizon fixed payments in capacity markets [19, 20] is an ongoing debate in market design. Within the forward capacity market framework, research has focused on articulating prices and quantities on a demand curve for a capacity product [21, 22], with a sloped demand curve [23] and performance incentives [24, 25] identified as important attributes. The present paper contributes to these ongoing market design debates by quantifying a key source of uncertainty in articulating a demand curve for capacity.

In recent years, many system operators have capacity well in excess of their NERC target. The 2018 NERC Summer Reliability Assessment reports that most assessment areas have anticipated planning reserve margins (hereafter, simply reserve margins) well in excess of their NERC reference summer target [26]. Similarly, the PJM Interconnection LLC (PJM), the largest system operator by installed generation capacity and load in North America, forecasts a reserve margin of approximately 25-29% over the period of 2019-2028, while its annual Reserve Requirement Study calls for no more than 16% over that period [27].[2]

Higher procurement improves reliability and therefore may be justified by a low cost of maintaining excess supply or in response to systematic under-forecasting of high-demand or large-outage events. In PJM's case, several recent weather events have stressed the grid, including Hurricane Sandy (2012), the Polar Vortex and subsequent cold weather of January 2014, and cold snaps in subsequent winters. That such events could pose reliability challenges despite high reserve margins suggests that current approaches to resource adequacy modeling are not capturing important reliability risks, such as correlated generator failures. Better understanding of the determinants of generator availability, and incorporation of these considerations into resource adequacy modeling, is the focus of this paper and could improve system operators' ability to determine capacity requirements.

Current approaches to resource adequacy modeling in North America, including in PJM, assume that each generator in the power system has a constant failure probability that captures its outage

[2] A 25% reserve margin indicates capacity procurement 25% greater than the forecast peak load for the planning period of interest. Reserve margins are typically reported in terms of installed (nameplate) capacity, which does not account for expected reductions in generator availability under peak load conditions (e.g. deflating by the forced outage rate for thermal generators). The resource adequacy modeling process determines the minimum reserve margin necessary for achieving a desired risk of generation insufficiency (e.g. a loss-of-load expectation of one generation insufficiency event in 10 years, termed 0.1 LOLE).

risk during all hours of the planning year [27]. This implies that generator failures are unconditionally independent of one another and invariant to environmental conditions. Recent research examined the independence assumption using four years of Generating Availability Data System (GADS) data for the eight North American reliability regions [28]. The analysis demonstrated that large generation outage events occur much more often than is consistent with the constant failure probability assumption—i.e., generator failures are correlated. The current practice of assuming unconditional independence lowers the computed probability of large-outage events at the system level [29]. If elevated generator failure probabilities occur during times of high load, current resource adequacy modeling practice could underestimate power system capacity requirements.

To identify a plausible mechanism to explain observed correlated failures, subsequent research explicitly modeled generator transition probabilities. Based on 23 years of PJM GADS data, [30] used logistic regression to demonstrate significant correlations between transition probabilities, load, and ambient temperature. At both very cold and very hot temperatures, PJM's fleet is less available than on average. For example, nuclear generators are less available at high temperatures, likely due to cooling water constraints, while natural gas generators are less available at low temperatures, in part due to fuel unavailability. The regression specification captured about 45% of weekly average variation in unavailable capacity from unscheduled generator failures on the holdout dataset. It also picked up nearly all correlated failure events observed in PJM over the data period, with the exception of Hurricane Sandy, which was not an extreme-temperature event. Given that extreme temperatures tend to be associated with high loads, and that temperatures are highly spatially correlated at the scale of an individual balancing area such as PJM, it follows that temperature dependence could pose significant resource adequacy risks for PJM.

Here we examine the resource adequacy implications of temperature-dependent generator availability, a potential source of risk that is not currently considered by PJM [27]. To do this we use the temperature dependence of the six types of generators identified in [30] and a modified version of the Renewable Energy Capacity Planning Model (RECAP) [31], an open-source resource adequacy tool originally developed by Energy+Environmental Economics (E3) in collaboration with the California

Independent System Operator.[3] Like the resource adequacy planning tools used by PJM, RECAP computes loss of load expectation (LOLE), a measure of generation insufficiency for a power system, given a parameterized fleet of generators and a load forecast. We modify RECAP to allow each generator's forced outage rate to depend on ambient temperature, rather than fixing it at an average value for every hour within a given month or year as is common industry practice.

We use this modification to compute capacity requirements for PJM under two scenarios: 1) current practice (representing assumptions of unconditional independence and constant failure probabilities); and 2) when allowing generator availability to depend on temperature. By comparing these two sets of results, we can identify the magnitude of latent resource adequacy risk introduced by overly optimistic generator failure assumptions (no correlated temperature-dependent failures). Given that extreme temperatures are seasonal, we also use RECAP to set monthly capacity targets to achieve the same reliability metric and quantify the reductions in capacity that could be obtained as opposed to the current annual procurement approach. Finally, we explore how various bounding changes to the resource mix and future temperature increases under climate change scenarios may affect resource adequacy in PJM, given the temperature dependence found in [30]. To our knowledge, the resource adequacy implications of temperature dependence of conventional generator resources has not been previously studied, nor is it accounted for in the resource adequacy modeling practices of any system operator in North America.

In brief, we find that relaxing the assumptions of unconditionally independent generator failures and constant generator failure probabilities uncovers substantial latent resource adequacy risk: a 6% increase in annual procurement (corresponding to a 7 percentage point increase in the reserve margin) is required to maintain the 0.1 LOLE reliability target in the 2018/2019 delivery year. This is due to reduced generator availability during peak-load hours, driven by high summer temperatures. When accounting for temperature dependence of generator availability, monthly capacity procurement targets would substantially reduce annual average reserve procurement in PJM with negligible impact

[3]E3 includes the following disclaimer with the user guide to the public version of the model: "RECAP is open-source and not a commercial software package, thus users are welcome to modify or tailor any part of the RECAP model to their needs. RECAP has a complicated set of inputs and model settings and E3 takes no responsibility for the validity of results produced by third parties using this model, which requires some intuition and knowledge of the methodology to use successfully." We change the underlying RECAP Python source code to enable input of hourly generator forced outages dependent on ambient conditions.

on LOLE. This is because spring and fall months experience mild temperatures, leading to both lower loads and increased generator availability.

The remainder of the paper is organized as follows. Section 2 reviews how resource adequacy modeling is currently conducted in PJM. Section 3 describes the RECAP model. Section 4 describes data development, modeling, and RECAP parameterization. Section 5 presents results. Section 6 concludes.

3.2 Overview of resource adequacy modeling in PJM

Resource adequacy is concerned with ensuring sufficient forward procurement of generation capacity to serve firm load within the operational parameters of the grid. Where capacity markets exist, as in PJM, capacity obligations for load-serving entities in the balancing area may be used to parameterize the demand curve for the capacity market [22, 32]. Capacity markets are intended to help provide generators with revenue sufficiency and stability in support of long-term resource adequacy [33, 34]. Each year PJM conducts a resource adequacy planning analysis that seeks to determine the level of capacity required to limit the frequency of loss-of-load events to once in 10 years, termed 0.1 loss of load expectation (0.1 LOLE) [1].[4] Each analysis forecasts capacity requirements for the next 11 delivery years, which begin June 1 and end the following May 31. The modeling process considers historical generator forced outage rates, generator maintenance requirements, load forecast error, peak load variability, interzonal transmission constraints, and emergency transfer capacity from neighboring synchronous balancing authorities [27, 36]. The reserve margin is the quantity of capacity required to meet the reliability criterion, reported as a percentage above the forecast 1-in-2 peak load, corresponding to the median forecast peak load [27, 37].

PJM forecasts the 1-in-2 peak load for the delivery year using an econometric model in conjunction with historical temperature profiles [27, 37–39]. Load uncertainty is modeled weekly based on at least seven years of historical data, with each week having a normalized daily peak based on its modeled

[4]The LOLE metric used by PJM does not consider either the duration or magnitude of the load-shed event. Some U.S. system operators instead interpret "once in ten years" to mean 24 total hours of expected load shed per 10 years; some international systems consider load-shed magnitude by using a quantity of expected unserved energy as the reliability metric [35].

mean and standard deviation. These normalized daily peak values are used to sort the 52 weeks of each historical load year in descending order by weekly peak by season. After magnitude-ordering the weeks, the normalized weekday peak-load observations for each calendar week are used to calculate a mean and standard deviation, defining a Gaussian distribution for that week, assumed to apply to each weekday of the week. This information, along with the forecast peak load and a forecast error factor, is used to calculate the most probable peak load for each week of the delivery year. The most probable peak for week i (MPP_i) is calculated as:

$$MPP_i = \mu_i + 1.163 * \psi_i \tag{3.1}$$

where μ_i is the mean of the Gaussian distribution for week i, $\psi_i = \sqrt{\sigma_i^2 + FEF^2}$ is the standard deviation after including the forecast error factor (FEF), σ_i is the standard deviation of the Gaussian distribution for week i, and 1.163 is a multiplier that relates the expected value of the maximum of a sample drawn from a Gaussian distribution to the mean and standard deviation of that distribution, based on the number of draws [40].[5] Uncertainty in each weekly peak load is assumed to follow a Gaussian distribution centered at μ_i with standard deviation ψ_i. For purposes of calculating LOLE, PJM represents each week's Gaussian as a probability mass function using 21 equally spaced points spanning +/- 4.2 standard deviations. As described in detail below, the goal of PJM's procedure is to determine whether sufficient generation capacity will be available to meet each of the 21 points in each week of the delivery year. Points for which this is not true accrue LOLE.

With the 21 possible peak load values for each week of the delivery year established, PJM next forecasts the performance of its generator fleet. For each generator expected to serve the balancing area during the delivery year, PJM uses the most recent five years of GADS data to calculate the forced outage rate.[6] Each generator is modeled as a two-state homogeneous Markov model, with available and unavailable states, where the forced outage rate is assumed to represent the (constant) probability of the generator being unavailable when needed by the system. To determine the probability distribution of available capacity for the power system, all possible combinations

[5]Here, 1.163 is the first Gaussian order statistic for n=5 draws, corresponding to five weekdays in each week.
[6]Generators with fewer than 5 years of historical data are supplemented by class-average data.

of generator states are enumerated [27, 38]. The probability of a given state is computed as the product of the relevant generator-level probabilities. This forecast distribution of available capacity is then compared to the forecast peak load distribution in each weekday (approximately 260 days), described previously, to compute the total LOLE for the delivery year (Equation 3.2):

$$LOLE = \sum_{i=1}^{260} LOLE_i = \sum_{i=1}^{260} \sum_{j=1}^{21} LOLP_j = \sum_{i=1}^{260} \sum_{j=1}^{21} P(D_j) * P(G < D_j) \tag{3.2}$$

where $LOLE_i$ is the aggregate loss of load expectation for weekday i of the delivery year, $LOLP_j$ is the loss of load probability (LOLP) at the load value corresponding to the j^{th} position in the 21-point Gaussian representing weekday i's MPP, $P(D_j)$ is the probability of the j^{th} load value occurring, and $P(G < D_j)$ is the probability of available generation (i.e., generation not on forced, maintenance, or planned outage) being insufficient to meet the j^{th} load value.[7] The forecast peak load is then scaled until aggregate loss of load expectation (Equation 3.2) precisely satisfies the reliability criterion (0.1 LOLE), and then the final reserve margin for the delivery year is calculated.

As discussed in detail below, our contribution is to evaluate the resource adequacy implications of relaxing the assumption of an invariant G, based on previous research demonstrating that the distribution of available capacity depends strongly on temperature [30].

3.3 Overview of the RECAP model

RECAP [31] was developed by Energy+Environmental Economics (E3) in collaboration with the California Independent System Operator to improve resource adequacy valuation of unconventional capacity resources like wind and solar. RECAP is similar in scope and capabilities to PJM's own resource adequacy tools and has been used by system operators and utilities. We briefly describe the main functionality of the open-source version of the model used in this paper and our modifications.

RECAP calculates the reserve margin required to achieve a desired reliability metric, which we set

[7] As PJM accounts for ambient deratings during the summer and planned outages in its resource adequacy analysis, it would be more precise to use G_i to represent the distribution of available capacity in Equation 3.2. However, we refrain in order to emphasize that PJM currently assumes generator availability from unscheduled failures is constant throughout the delivery year.

to 0.1 LOLE for consistency with PJM's target. While PJM's modeling approach considers only the peak load hour of each weekday, RECAP models all hours of the delivery year. By default, RECAP divides the delivery year into 576 time-slice bins: month (12) x hour (24) x weekday/weekend (2).[8] The user may define additional bins by load percentile within the month x hour x weekday/weekend time-slices to improve the model's ability to represent peak loads; we use three such bins (0-80%, 80-90%, and 90+%), yielding a total of 1,728 bins. Using historical data on hourly load and wind and solar capacity factors in each time-slice bin, RECAP produces an empirical distribution of net loads (gross load less contemporaneous variable renewable generation) as the convolution of load, wind, and solar in each time-slice. RECAP then fits a user-selected probability distribution to the net load data in each bin to model load uncertainty; we employ Gaussian distributions for consistency with PJM.

Consistent with PJM's resource analysis approach, RECAP uses the forced outage rate for each conventional generator to compute a probability distribution of available capacity for the power system in each time-slice bin. By default, RECAP can consider annual or monthly forced outage rates; if annual values are used, the distribution of available capacity is invariant across time-slice bins just as G is invariant across all weekdays of the delivery year in Equation 3.2. We modify the RECAP source code to instead consider hourly ambient temperature when generating this probability distribution. In this way we relax the incorrect assumption that available capacity of conventional generation is invariant to temperature in PJM [30].

RECAP computes the LOLP in each time-slice bin through convolution of the corresponding net load and available generation probability density functions. Results are aggregated across any additional user-defined load bins, weighting appropriately to obtain the final annual LOLE result.[9] By default, RECAP sums up resulting LOLP over all hours of the delivery year; however, to better match PJM's definition of LOLE, we modify RECAP to consider only the peak LOLP hour in each day.

[8] While PJM does not include weekends in its resource adequacy analysis, there is little discrepancy introduced by this difference: we find that weekends represent less than 0.1% of total LOLE.

[9] For example, if three load bins were used, with splitting occurring at the 80th and 90th percentiles, the three LOLP values for each month x hour x weekday/weekend time-slice bin would be weighted by 80%, 10%, and 10%.

3.4 Data development, modeling and parameterization

We parameterize RECAP with PJM's conventional generator fleet, hourly wind and solar generation profiles, hourly temperature data, and normalized historical PJM hourly load data. We discuss the development of each set of inputs below.

3.4.1 Conventional generator fleet

We use the PJM GADS database [41] to define currently operating conventional generators serving the PJM balancing area, which includes all generators other than wind, solar, and behind-the-meter resources. From PJM GADS we obtain the full operating history of each such generator for the period 1995-2018Q1, totaling 1.94 million events records affecting 1,845 generators. Reporting to PJM GADS is mandatory for all conventional resources participating on PJM's markets, so we have complete coverage of these resource types. In RECAP we parameterize conventional generators by their nameplate capacity and generator type, both of which are reported in PJM GADS, and their forced outage rates, which we calculate from PJM GADS.

We develop both temperature-dependent and unconditional forced outage rates for each generator as described in [30]. In brief, temperature-dependent hourly generator transition probabilities are calculated using logistic regression. Separate models are fit to hour-over-hour transitions when a generator can experience a failure (i.e., the generator begins in the available state) versus when it can experience a recovery (i.e., the generator begins in the outage state). In both cases, the model specification includes linear and quadratic terms for both cold and hot temperatures (delineated using a threshold of 18.3 degrees Celsius), as well as a linear load term, though we apply backward elimination to remove statistically insignificant covariates. Models are fit using all available data. To reduce bias, we require that generators have at least one full year of data reporting and at least 10 failures and recoveries per final model covariate to be retained [42]. Using this procedure, we retain 78% of the capacity that has ever reported to PJM GADS.

To obtain temperature-dependent forced outage rates, we evaluate each retained generator's fitted models at a temperature of interest, along with the median load value at that temperature.[10]

[10]Median load values are determined by linking historical PJM metered load data to contemporaneous historical

This yields a transition probability matrix governing failures and recoveries for each generator at that temperature. Eigendecomposition of this matrix allows us to obtain the ergodic probability of the generator being unavailable at that temperature. We repeat this procedure over a temperature range of -30 degrees Celsius to 40 degrees Celsius for each generator in 5-degree increments and then compute the capacity-weighted average forced outage rate at each temperature by generator type. See Procedure 4 of [30] for more information. We apply these generator-type averages to all generators reporting to PJM GADS that were in service as of 2018 Q1, even generators with too little data to be retained in the modeling procedure described in the previous paragraph.

We calculate unconditional forced outage rates using the following equation [43]:

$$EFOF_i = (FOH_i + EFDH_i)/PH_i \tag{3.3}$$

where $EFOF_i$ is the equivalent forced outage factor for generator i, FOH_i (forced outage hours) is the count of hours where generator i experiences a forced outage, $EFDH_i$ (equivalent forced derating hours) is the count of hours where generator i experiences a forced derating, reported on a full-outage-equivalent basis, and PH_i (period hours) is the total number of hours of data reporting for generator i. $EFOF$ provides an approximation of the forced outage rate statistic used by PJM that allows us to circumvent data reporting limitations in PJM GADS prior to 2004. While resource adequacy studies in PJM are conducted using the five most recent years of GADS data, here we use all historical data for each generator, consistent with our approach for computing temperature-dependent forced outage rates. As with the temperature-dependent forced outage rates, we then compute averages by generator type. We use the unconditional forced outage rates to establish a baseline representing current PJM practice. This allows us to quantify latent resource adequacy risk from temperature-dependent generator availability.

Temperature-dependent distributions of available capacity for PJM's conventional generator fleet are shown in Figure 3.1. The distributions of available capacity at both cold and hot temperatures are shifted to the left of the distributions at moderate temperatures, indicating less available capacity on

temperature data and then taking the median value over all hours with temperatures within a 10-degree "neighborhood" of the temperature of interest. For example, the median load at 20 degrees Celsius is calculated based on all historical loads occurring between a temperature of 10 degrees Celsius and 30 degrees Celsius.

average; they are also shorter and wider than the distributions at moderate temperatures, indicating greater risk of large deviations from the average.

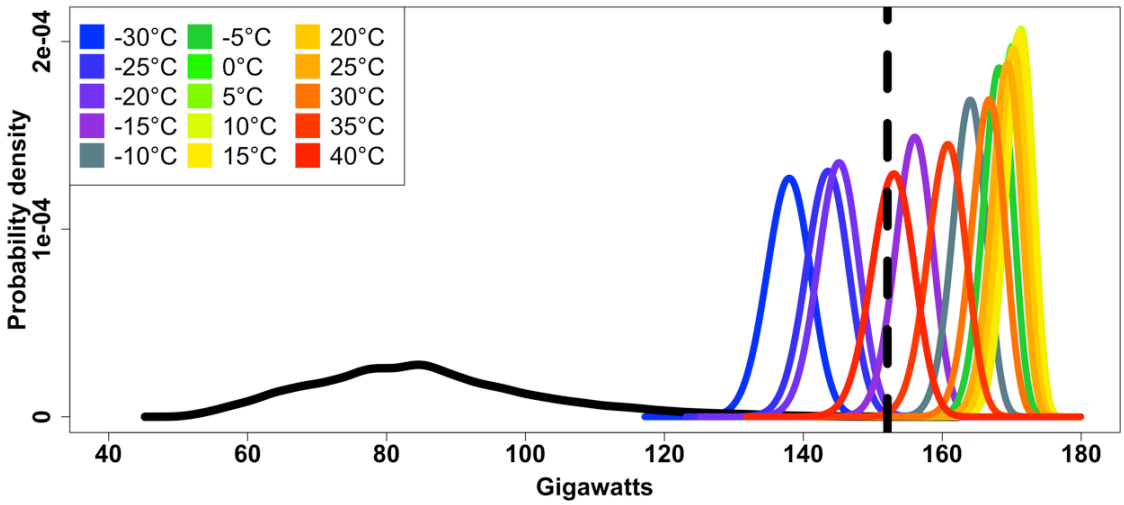

Figure 3.1: Distributions of available capacity as a function of temperature for PJM's conventional generator fleet (narrow curves, various colors). Note reduced availability at both cold and hot temperatures. Wide black curve is the distribution of the 1-in-2 (median) load forecast. Dashed black line is the peak value from the median load forecast.

A comparison of available capacity for selected hours in July for the temperature-dependent and unconditional forced outage rate scenarios is shown in Figure 3.2. The temperature-dependent afternoon distribution is shifted to the left of, and has a higher variance than, the unconditional distribution due to the high temperatures that prevail. As described in Section 3.2, current resource adequacy modeling practice does not account for this shift in the available capacity distribution.

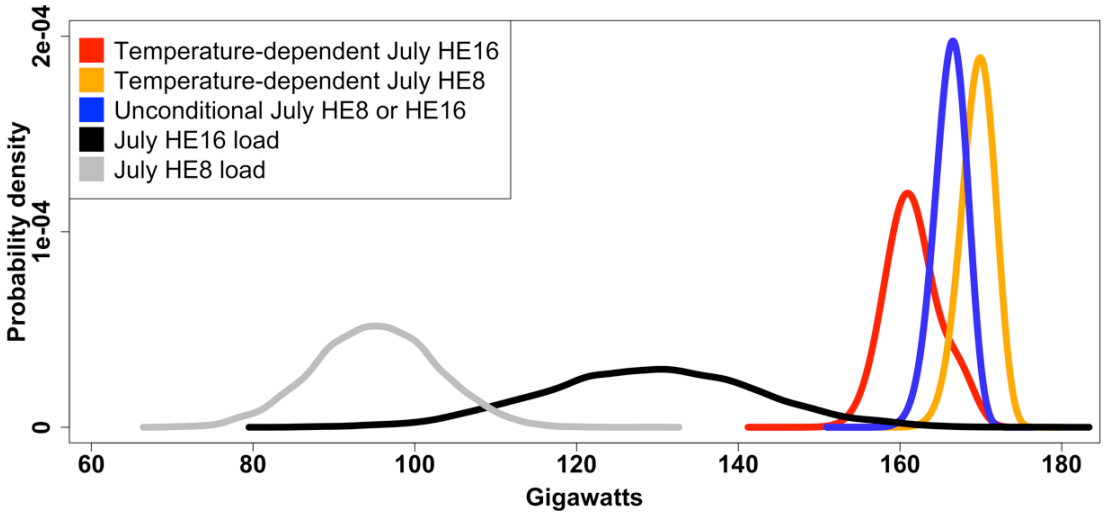

Figure 3.2: Illustration of how temperature-dependent forced outage rates can identify latent resource adequacy risk. Distributions of available capacity for selected July hours using temperature-dependent forced outage rates are shown in red and orange; the distribution of available capacity when using unconditional forced outage rates is shown in blue. Black and gray curves indicate load distributions for corresponding July hours. HE indicates hour-ending. Current resource adequacy modeling practice does not recognize that the distributions of available capacity will differ between morning and afternoon due to different prevailing temperatures, and consequently understates the risk of not being able to serve load.

3.4.2 Wind generation

To model wind generation and its coincidence with peak load, RECAP requires an hourly normalized historical generation profile and total quantity of installed wind capacity for the modeled delivery year. Because of the relatively recent vintage and sparse public hourly generation data for most wind plants in PJM, we choose to model wind generation rather than use empirical profiles.

To develop normalized wind generation profiles, we first identify online wind plants serving PJM as of December 2017 using Energy Information Administration (EIA) Form 923 (EIA-923) [44]. Plants are subset by their plant identification number recorded in EIA-923. Identified plants are then matched against the U.S. Wind Turbine Database [45] to obtain information about physical location, generator online date, turbine hub height and turbine technology, as well as to verify installed capacity. We exclude all plants with less than 4 MW capacity and those that could not be matched, leaving 89 wind plant locations in PJM representing 8.368 GW of installed capacity, more

than 97% of the total capacity recorded in EIA-923. The 89 retained wind plant locations have an average installed capacity of 94.0±72.2 MW, with a median of 80 MW and a maximum of 305.8 MW.

To simulate hourly output, we match each wind site to the closest available 100-meter hub height wind speeds from the National Renewable Energy Laboratory's (NREL) Wind Integration National Dataset Toolkit [46], which contains 5-minute wind speeds for 2007-2012 at an assumed 100-meter turbine hub height. Matched 100-meter wind speeds are scaled to individual generator hub heights using Equation 2 of [47], reproduced for our use case as Equation 3.5 below. Wind speeds are then converted to power output using an appropriate power curve given the vintage and type of installation [46]. To account for degradation in wind turbine performance over time, we apply a historical degradation factor as a function of generator age in 2018 using Figure 39 of Lawrence Berkeley National Laboratory's 2017 Wind Technologies Market Report [48]. Equations specifying and a table describing this process are included below.

$$NormWindGen_{t,g} = WindSpeed_{t,g,hh} * PowerCurve_g * Degradation_g * Derate \tag{3.4}$$

$$WindSpeed_{hh} = WindSpeed_{100} * \frac{ln(hh)/.03}{ln(100)/.03} \tag{3.5}$$

Table 3.1: Wind Plant Generation.

Variable (Units)	Description
$WindSpeed_{t,g,hh}$ (m/s)	Wind speed at time t for location of plant g at its hub height hh of plant's turbines, scaled from 100-meter hub height Wind Speed per equation 3.5.
$PowerCurve_g$ $(MW - s/m)$	Wind speed at time t for location of plant g at its hub height hh of plant's turbines, scaled from 100-meter hub height Wind Speed per equation 3.5.
$Degradation_g$ (%)	Wind speed at time t for location of plant g at its hub height hh of plant's turbines, scaled from 100-meter hub height Wind Speed per equation 3.5.
$Derate$ (%)	Wind speed at time t for location of plant g at its hub height hh of plant's turbines, scaled from 100-meter hub height Wind Speed per equation 3.5.
hh	Hub height of generator g. Used to scale 100-meter hub height windspeeds from NREL dataset. 3.5.

Our resulting simulated wind generation output for 2007-2012 has a 39% capacity factor compared to a reported capacity factor of 32.6% for installations in the Great Lakes region and 44.4% for installations in the interior region of the USA (Figure 41 of [48]). Because most PJM wind capacity is located in Illinois, Indiana, and Ohio, closely mapping to the Great Lakes region, we derate our simulated hourly profiles to match the Great Lakes Region capacity factor.[11] While RECAP can calculate the capacity value of each wind generation site in PJM individually, for study scope and model runtime reasons we aggregate generation into a single shape for input into RECAP. The 8.368 GW of installed wind capacity with this generation profile has a marginal effective load carry capability (ELCC) of 16.9% in the base case, reflecting its lower availability during peak summer load hours with non-zero LOLE.

3.4.3 Solar generation

As with wind generation, RECAP takes two key inputs for solar generation: hourly normalized generation profile(s) and a total quantity of installed capacity corresponding to each profile for the modeled delivery year. Given our focus on conventional resources and the relatively small quantity of utility-scale solar capacity in PJM, we model a single aggregate solar generation profile.

To develop the hourly normalized solar profile for PJM, we obtain solar radiation and weather data from NREL's National Solar Radiation Database [50, 51] and convert it to alternating current (AC) generation output using Pacific Northwest National Laboratory's (PNNL) GridLab-D solar panel and inverter objects [52]. GridLab-D uses the same solar modeling as NREL's System Advisory Model [53], a widely used engineering-economic tool.

Key assumptions for our PJM solar shape are given in Table 4.2. Panel tilt, efficiency, and other variables convert solar irradiance to a DC power output, which is then converted to AC power output based on assumptions about inverter efficiency and inverter loading ratio. The full set of equations for this conversion are complex, so we refer to [51] for further information. Hourly modeled profiles are adjusted to account for daylight savings time and leap years. The resulting annual average AC capacity factor for this PJM shape is 19.3%.

[11] Reasons for simulated generation exceeding actual performance may include curtailment, turbine down time for maintenance, wake effects, and icing on blades [49].

Table 3.2: Solar Generation Key Assumptions.

Parameter	Assumption
Solar panel tilt	Fixed, 30 degrees
Inverter loading ratio	1.2
Inverter Efficiency	96%
Panel Efficiency	17%
emphPower Factor	Constant, 1.0

Using EIA Form 923 [44] we identify approximately 2 GW_{AC} of installed utility-scale solar capacity online in PJM as of December 2017. This includes 1.3 GW_{AC} of capacity in Maryland, Virginia, Delaware, Ohio, Pennsylvania, Indiana, Illinois, and Kentucky, and an additional 0.7 GW_{AC} in North Carolina that may be deliverable to PJM. We assume behind-the-meter solar generation is accounted for in the input load profiles. At this level of installation in the base case the utility scale solar has a marginal ELCC of 51.6%, reflecting its higher availability during peak summer load hours with non-zero LOLE.

3.4.4 Temperature data

We use temperature data for two purposes: 1) to specify a load regression and obtain a time trend for PJM loads (Section 3.4.5); and 2) to estimate the ambient temperature in PJM for the 2018 delivery year so that we can model generator availability in each time-slice bin in the temperature-dependent scenario. We obtain historical temperature data from the National Oceanic and Atmospheric Administration [52]. We select weather stations corresponding to Chicago, Cleveland, Philadelphia and Washington, D.C. for 2006-2017, the time period of our historical load data. We process each station's raw observations into hourly time series and then average the four series to obtain our input temperature series for PJM. Over this period, only 25 hourly observations were missing across the four weather stations. We interpolate each weather station's missing observations by forward propagation of its nearest non-missing observation. The final temperature time series is used in conjunction with the temperature-dependent forced outage rates described in Section 3.4.1 to determine unavailable capacity from unscheduled events for PJM's conventional fleet in each time-slice bin in the temperature-dependent scenario. The temperature

time series is not used in the current practice scenario because current resource adequacy modeling practice assumes that generator outage risk is invariant to ambient conditions (Equation 3.3).

3.4.5 Load Forecast

We obtain hourly historical metered load data from PJM by zone for 2006-2017 and aggregate to a PJM-wide load shape for input into RECAP. We select January 1^{st}, 2006 as a starting point for historical loads because the years immediately prior saw significant expansion of PJM's footprint. Since 2006, three zones have been added to PJM: American Transmission Systems Inc. (ATSI, 2011), Duke Energy Ohio and Kentucky (DEOK, 2012), and East Kentucky Power Cooperative (EKPC, 2013). To account for these missing zonal loads, we develop a correlation matrix between ATSI, DEOK, EKPC, and the zones that were present since January 1^{st}, 2006. We then match ATSI, DEOK, and EKPC to the zone with which each is most highly correlated. ATSI and DEOK most closely match Dayton Power and Light, while EKPC most closely matches American Electric Power. The correlation of each of the three zones with their matched zone is always greater than 0.95. We then fill in the unobserved loads in ATSI, DEOK, and EKPC using the corresponding load in the matched zone, scaling by the ratio of the average loads.

With the unobserved loads incorporated, we next estimate the time trend in the historical PJM load data so that we can make the historical values comparable to loads in our delivery year. We model daily loads using the following linear regression specification:

$$\begin{aligned} Load_t = Chi\_HDD_t + Chi\_CDD_t + Cle\_HDD_t + Cle\_CDD_t + DC\_HDD_t \\ +DC\_CDD_t + Phi\_HDD_t + Phi\_CDD_t + Weekday_t + T_t \end{aligned} \tag{3.6}$$

where $Load_t$ is the sum of the hourly PJM loads in day t, Chi, Cle, DC, and Phi are abbreviations of the four cities representing geographic diversity in PJM's footprint for which we obtain temperature data (Chicago, Cleveland, Washington, D.C., and Philadelphia), CDD_t is cooling degree days in day t, HDD_t is heating degree days in day t, Weekday is a Boolean variable indicating whether day t is a weekday, and T is a linear time trend. More detail is in Table 3.3. After fitting this model we use

the coefficient on T to de-trend the historical loads to the 2018 delivery year for input into RECAP.

Table 3.3: Load Regression Specification.

Variable (Units)	Description
Chi_HDD_t (°C)	Average hourly temperature above 18.3°C on day t at Chicago weather station (O'Hare International Airport, Weather Bureau Army Navy (WBAN) Code 94846). 0 if average temperature below 18.3°C.
Chi_CDD_t (°C)	Average hourly temperature below 18.3°C on day t at Chicago weather station (O'Hare International Airport, Weather Bureau Army Navy (WBAN) Code 94846). 0 if average temperature above 18.3°C.
Cle_HDD_t (°C)	Average hourly temperature above 18.3°C on day t at Cleveland weather station (Cleveland-Hopkins International Airport, WBAN Code 14280). 0 if average temperature below 18.3°C.
Cle_CDD_t (°C)	Average hourly temperature below 18.3°C on day t at Cleveland weather station (Cleveland-Hopkins International Airport, WBAN Code 14280). 0 if average temperature above 18.3°C.
DC_HDD_t (°C)	Average hourly temperature above 18.3°C on day t at Washington, D.C. weather station (Ronald Reagan International Airport, WBAN Code 13734). 0 if average temperature below 18.3°C.
DC_CDD_t (°C)	Average hourly temperature below 18.3°C on day t at Washington, D.C. weather station (Ronald Reagan International Airport, WBAN Code 13734). 0 if average temperature above 18.3°C.
Phi_HDD_t (°C)	Average hourly temperature above 18.3°C on day t at Philadelphia weather station (Philadelphia International Airport, WBAN Code 19739). 0 if average temperature below 18.3°C.
Phi_CDD_t (°C)	Average hourly temperature below 18.3°C on day t at Philadelphia weather station (Philadelphia International Airport, WBAN Code 19739). 0 if average temperature above 18.3°C.
$Weekday_t(Boolean)$	1 if day t occurs during a non-holiday weekday, 0 otherwise.
$T_t(GW_{avg})$	Daily time trend in load t on day T after 12/31/2005, the day on which our load data timeseries begins (i.e., T_t=1 for hours on 1/1/2006, T_t=366 for hours on 1/1/2007, etc.). Helps account for changing electricity consumption due to reasons other than weather and daytype (e.g., due to energy efficiency, demographics) in delivery year compared to historical years.

In addition to the historical loads, RECAP takes in the forecast peak load and forecast total annual load for the delivery year. The 1-in-2 (median) unrestricted forecast peak load for our delivery year is 152.1 GW and the forecast total annual load is 806.7 TWh (92.1 GW_{avg}) [54].

3.4.6 RECAP parameterization

Finally we parameterize RECAP with the developed generator fleet and load forecast for the modeled delivery year. We summarize our base parameterization in Table 3.4, matching PJM's parameterization reported in the 2018 Reserve Requirement Study (RRS) to the extent feasible [27].

Table 3.4: Parameterizing the RECAP model.

Metric	Value
Conventional generation[ab]	207.5 GW installed capacity
Solar generation	2.0 GW installed capacity
Wind generation	8.4 GW installed capacity
Demand response[c]	0 GW
Emergency imports[d]	0 GW
Footprint	All current PJM zones
Zonal disaggregation	None
Scheduled outages	Average historical requirements by generator type and month
Peak load forecast[e]	152.1 GW
Load bins[f]	0-80, 80-90, 90-100
Reliability metric	0.1 LOLE
Capacity addition resource[g]	Combined-cycle gas generator

[a] We enforce a 2.5 GW reduction to thermal generator capacity June-August. This roughly corresponds to the 2.5 GW of thermal capacity that PJM puts on planned outage during weeks 6-15 of the delivery year to account for capability reductions driven by high temperatures and humidity.

[b] This includes all non-retired conventional resources reporting to PJM GADS as of 2018 Q1, both resources located in the PJM footprint as well as external resources that have obtained firm transmission and available transfer capability into PJM [27]. For comparison, 209 GW of installed capacity was eligible to be offered into the 2018/2019 Base Residual Auction [55].

[c] We do not include demand response (DR) due to the difficulty of parameterizing it in RECAP; PJM similarly omits it from their resource adequacy analysis.

[d] Represents emergency capacity from neighboring power systems. PJM assumes 3.5 GW of emergency import capacity [27].

[e] Corresponds to the forecast peak load used by PJM to model the 2018/19 delivery year [27].

[f] Percentiles of the load used to disaggregate each hour x month x weekday/weekend time-slice bin.

[g] This is the resource assumed to be built if the load forecast results in greater than 0.1 LOLE for the existing fleet.

Our parameterization does not account for all modeling details that are included in PJM's resource adequacy analysis. In particular, we do not include transmission constraints nor emergency imports from neighboring power systems. We also do not include demand response (DR) due to the difficulty of modeling duration-limited resources in RECAP. However, PJM similarly omits DR from the RRS, so no inconsistency is introduced. We also present a table of the temperature-dependent and unconditional forced outage rates employed in the modeling of conventional generator availability (Table 3.5) [30].

Table 3.5: Temperature-dependent and unconditional forced outage rates for conventional generator types. Values drawn from [30]. CC is combined cycle gas, CT is simple cycle gas, DS is diesel, HD is hydroelectric and pumped storage, NU is nuclear, ST is steam turbine (coal).

	Temperature (°C)	**CC**	**CT**	**DS**	**HD**	**NU**	**ST**
Temperature -dependent forced outage rates	-15	14.9%	19.9%	21.2%	7.0 %	1.9%	13.3%
	-10	8.1%	9.9%	17.0%	4.3%	1.8%	11.2%
	-5	4.8%	5.1%	13.7%	3.2%	1.7%	9.9%
	0	3.3%	3.1%	11.6%	2.7%	1.8%	9.1%
	5	2.7%	2.4%	10.6%	2.6%	1.8%	8.6%
	10	2.5%	2.2%	10.2%	2.6%	1.9%	8.3%
	15	2.8%	2.4%	10.4%	2.7%	2.1%	8.4%
	20	3.5%	2.7%	13.6%	2.7%	2.7%	8.6%
	25	3.5%	3.1%	13.5%	2.5%	3.7%	9.4%
	30	4.1%	3.9%	14.3%	2.9%	6.6%	11.4%
	35	7.2%	6.6%	17.5%	8.2%	12.4%	14.0%
Unconditional forced outage rates	All	3.3%	2.8%	10.9%	2.4%	2.6%	9.4%

3.5 Results and discussion

Several RECAP runs were conducted in order to examine the resource adequacy implications of temperature-dependent generator availability. First capacity requirements to achieve the 0.1 LOLE reliability target were compared under unconditional and temperature-dependent forced outage rates to quantify the magnitude of latent resource adequacy risk from current resource adequacy modeling practice in PJM. We then look at capacity requirements for achieving reliability targets ranging from 0.02 LOLE to 0.18 LOLE and compare those values to PJM's actual procurement in the 2018/2019 delivery year. Next, we examine potential reductions in capacity requirements when setting monthly targets, rather than a single annual target, to achieve 0.1 LOLE. We then quantify the value of lost load (VOLL) implied by the same range of reliability targets to examine the cost of achieving these levels of resource adequacy. Given the economic pressure on nuclear and coal generators in PJM, we consider the resource adequacy implications of their retirement. Finally, given that all PJM generator types exhibit reduced availability during hot temperatures, we examine the resource adequacy implications of plausible future temperature scenarios for PJM's conventional fleet.

3.5.1 Aggregate effect of accounting for temperature dependence

Our first step is determining capacity requirements for achieving the 0.1 LOLE reliability target when treating generator forced outage rates as invariant to temperature. In this scenario, 174.5 GW of conventional generation capacity, corresponding to a 15.9% reserve margin, is required to cover the forecast 1-in-2 peak load.[12][13] For comparison, PJM reports that a 16.2% reserve margin achieves 0.1 LOLE for the 2018/2019 delivery year [27].[14] The close agreement between these two results gives us confidence that we have captured the primary drivers of LOLE in PJM despite not accounting for emergency imports or zonal transmission constraints, nor using PJM's internal modeling tools.

When using temperature-dependent forced outage rates, capacity requirements to achieve 0.1 LOLE increase by 10.6 GW to 185.1 GW, corresponding to a 22.9% reserve margin. This difference represents substantial latent resource adequacy risk from the simplifying assumption that conventional generator availability is invariant to temperature. However, even the higher 185.1 GW figure is less than the 192.6 GW of total capacity (26.6% reserve margin) PJM procured for the 2018/2019 delivery year [55,57]. Using a capacity price of \$184/MW-day (\$67.2/kW-year), this translates to \$315 million in additional capacity payments beyond what is required to achieve 0.1 LOLE when accounting for temperature dependence in generator availability.[15] While greater reliability is desirable, a demand curve for forward capacity procurement in markets like PJM's is meant to balance the incremental costs and benefits of capacity, and should be articulated as accurately as possible to achieve this aim. The method presented here offers a means of more precisely quantifying capacity requirements for achieving a target LOLE and therefore of improving the articulation of the demand curve for forward capacity procurement. We note that the economic benefits reported here do not consider dynamic impacts. In particular, reduced capacity procurement would likely increase the prevalence

[12] We report installed capacity values rather than unforced capacity values. Unforced capacity deflates installed capacity by the generator's forced outage rate and is used by PJM to determine capacity payments for generators that clear the capacity market.

[13] We focus on conventional generation requirements (i.e., capacity requirements net of wind and solar resources) when reporting quantities of capacity. However we include the capacity value of wind and solar resources when reporting the corresponding reserve margin.

[14] PJM finds that the reserve margin would increase by 1.9 percentage points if no emergency imports were available; its base-case analysis assumes 3.5 GW of import capability [56].

[15] \$184/MW-day is the average of PJM's 2018/2019 final zonal capacity market clearing prices, weighted by zonal capacity obligations [36]. PJM capacity prices have fluctuated in recent years but generally clear in the range of \$100-200/MW-day; other centralized capacity markets have experienced similar or more price volatility [58].

of scarcity pricing events, increasing energy and operating reserves costs and thereby reducing the realized level of benefits.

Next, we determine the quantity of capacity required to achieve a range of reliability targets, from 0.02 LOLE (one loss-of-load event per 50 years) to 0.18 LOLE (nearly one loss-of-load event per 5 years) (Figure 3.3). Achieving a lower target requires that additional capacity be procured; progressively larger increments of capacity are required as the reliability target becomes more stringent. The difference in capacity requirements between the unconditional and temperature-dependent forced outage rates scenarios is reasonably constant, but increases as the reliability target becomes more stringent. PJM's procurement of 192.6 GW is nearly 16 GW more than what our model finds is necessary to achieve the 0.1 LOLE reliability target when assuming unconditional generator availability, and is sufficient to achieve 0.02 LOLE even accounting for temperature dependence.

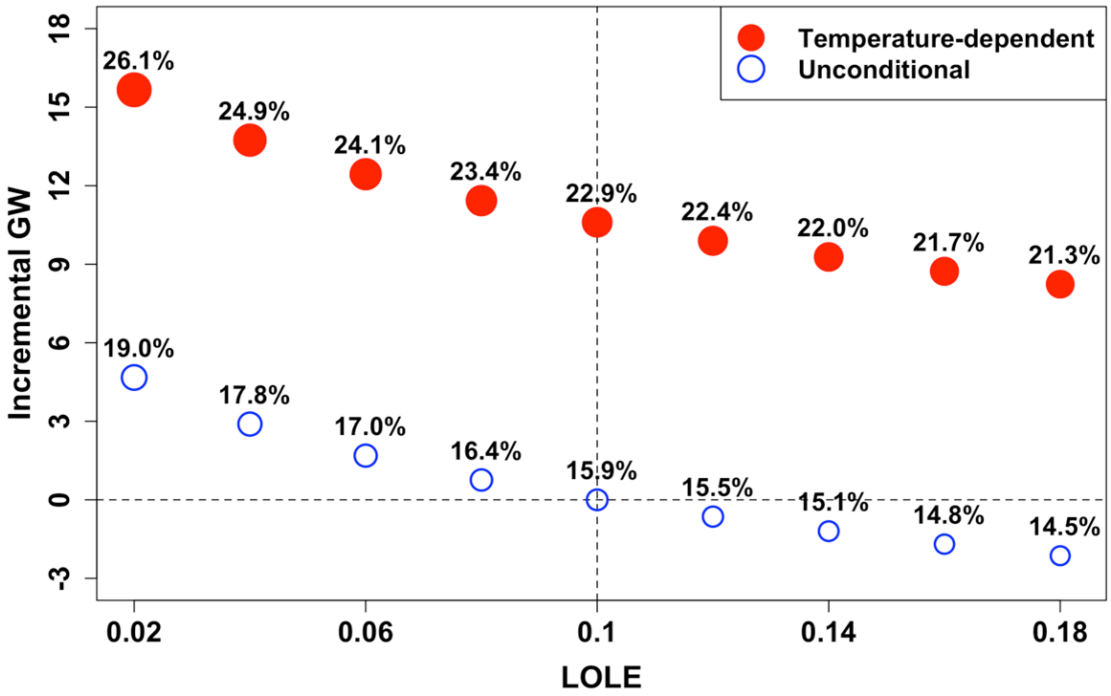

Figure 3.3: Incremental generation capacity required to achieve various LOLE targets under unconditional (hollow blue circles) and temperature-dependent (solid red circles) forced outage rates. Values are calculated with respect to the 0.1 LOLE reliability target under unconditional forced outage rates. Circle size indicates corresponding reserve margin; reserve margins are also indicated above each dot. For a fixed reliability target, the distance between the hollow blue circle and the red circle indicates the magnitude of latent resource adequacy risk in PJM during the 2018/2019 delivery year when incorrectly treating forced outages as invariant to temperature. Note that 0.02 LOLE in the unconditional forced outage rates scenario is a less stringent reliability target than 0.18 LOLE in the temperature-dependent forced outage rates scenario.

To examine the cost implications of current resource adequacy standards, we compute the implied incremental VOLL for each reliability target included in the previous analysis (Figure 3.4). These values offer a means of determining a desirable level of resource adequacy, balancing the cost of procuring additional capacity with the benefit of reduced load-shed. We compute each implied VOLL by calculating the change in capacity procurement required to achieve the reliability target of interest, along with the corresponding change in EUE. These values can be combined with an assumed cost of incremental capacity procurement to obtain an implied VOLL. We calculate the VOLL for both 0.08 LOLE and 0.12 LOLE using 0.10 LOLE as the baseline, and proceed "outward"

from there.[16] We again use a capacity value of $184/MW-day ($67.2M/GW-year).

Implied VOLLs for these reliability targets range from $100,000/MWh to $700,000/MWh, approximately two orders of magnitude higher than price caps employed in operational contexts in U.S. wholesale electricity markets [59]. As PJM's current procurement is at the lower (more reliable) end of our explored LOLE range, its implicit VOLL corresponds to the upper end of our estimated VOLL range. However, even at the 0.1 LOLE reliability target the VOLL is $200,000/MWh. This result adds to previous studies that have argued that 0.1 LOLE may overvalue unserved energy and lead to a higher reserve margin than is economically optimal, e.g. [60]. Misalignment between valuation of reserves in planning and operational contexts has operational consequences, and may help to explain why PJM has rarely triggered its operational scarcity pricing, even during recent extreme weather events [61]. By this same logic, enhancements to scarcity pricing on operational timescales, including ongoing efforts at PJM [62], may increase energy market revenues, lessen capacity market clearing prices, and therefore reduce the implied VOLL as well as economic benefits of capacity market reforms.

[16] For example, the VOLL for 0.06 LOLE is computed from a starting point of 0.08 LOLE, while the VOLL for 0.14 LOLE is computed from a starting point of 0.12 LOLE. The VOLL corresponding to 0.1 LOLE is calculated as the arithmetic mean of the VOLLs for 0.08 LOLE and 0.12 LOLE.

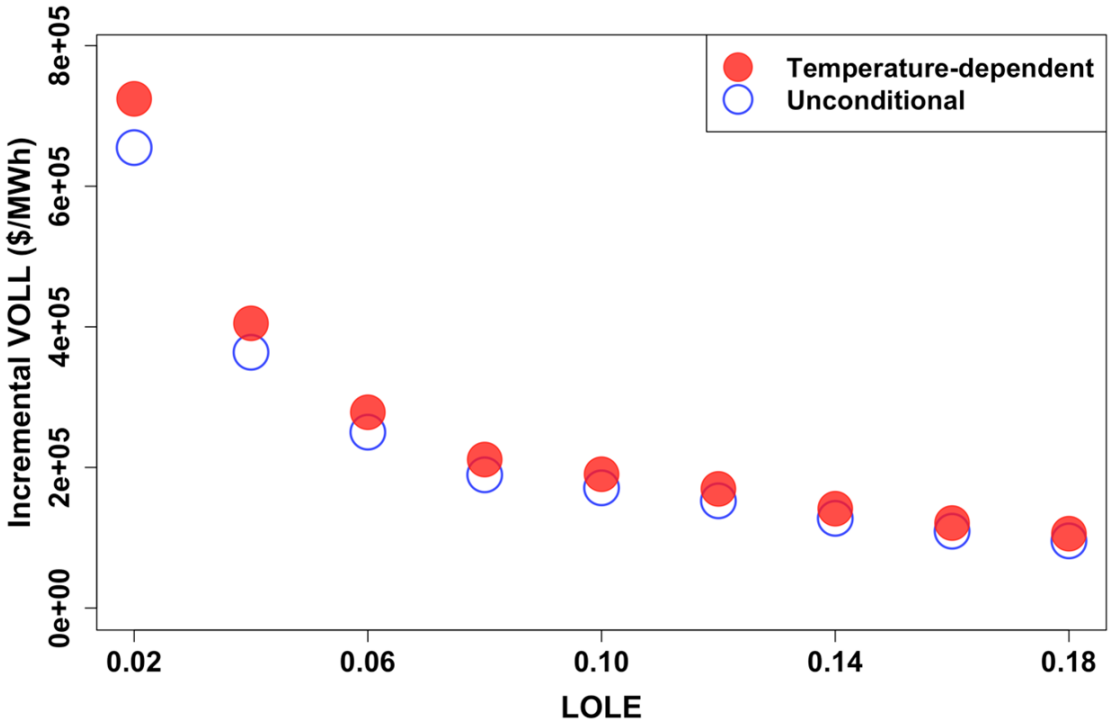

Figure 3.4: Implied incremental VOLL, in $/MWh, as a function of the reliability target, under the temperature-dependent (solid red circles) and unconditional (hollow blue circles) forced outage rate scenarios. Values calculated using $184/MW-day cost of incremental capacity. With the exception of 0.1 LOLE, the VOLL at each reliability target is calculated with respect to the immediately adjacent reliability target in the direction of 0.1 LOLE; for example, the VOLL at 0.06 LOLE is calculated using 0.08 LOLE as the baseline. The VOLL corresponding to 0.1 LOLE is calculated as the arithmetic mean of the VOLLs for 0.08 LOLE and 0.12 LOLE. The substantial increase in VOLL as the reliability target is tightened is driven both by lower incremental EUE reductions and higher incremental capacity procurement requirements.

3.5.2 Resource adequacy implication of monthly capacity targets

Accounting for temperature dependence in generator availability significantly increases capacity procurement requirements over PJM's computed reserve margin (but as previously noted not above their procured reserves). Given the strong seasonality of extreme temperatures, we next consider whether and to what extent capacity requirements could be reduced through monthly procurement targets rather than the current PJM practice of annual procurement.

To conduct this analysis, we begin by looking at accumulated LOLE in each calendar month

under an annual procurement approach, using both unconditional and temperature-dependent forced outage rates. In any month accounting for at least 0.01% of the total annual LOLE (i.e., 0.00001 LOLE given a 0.1 LOLE annual target), we retain the annual capacity procurement level as that month's requirement. For the remaining months—i.e., those with zero, or effectively zero, LOLE—we allow RECAP to reduce capacity procurement until each achieves the 0.01% LOLE contribution limit.[17] In this way, we can quantify opportunities for reducing capacity procurement in mild, low-load months without increasing overall LOLE in the delivery year.

In the unconditional forced outage rates scenario, only July and August have more than 0.01% of annual LOLE, so we allow RECAP to decrease capacity procurement until the threshold value is achieved in the other 10 months. This results in an average monthly procurement of 148.5 GW, a 15% reduction in capacity requirements on an annual basis. The result in the temperature-dependent forced outage rates scenario is similar, with again only July and August having more than 0.01% of annual LOLE. Here, decreasing procurement in the other 10 months to our threshold results in an average monthly procurement of 156.2 GW, a 16% reduction in capacity requirements on an annual basis. This capacity requirement is 10% below annual base-case procurement in even the unconditional forced outage rates scenario, demonstrating the extent to which LOLE is concentrated in a small subset of months, resulting in over-procurement in other months when procuring on an annual basis. A barplot showing monthly procurement requirements for achieving 0.1 LOLE under both forced outage rate scenarios is Figure 3.5.

[17] This is an arbitrary but reasonable threshold intended to ensure that total LOLE in the delivery year is not significantly increased while working within the confines of RECAP to estimate monthly capacity requirements. In the worst case (where only one month accrues more than 0.01% of total LOLE), this threshold would increase total LOLE in the delivery year by 1.1%.

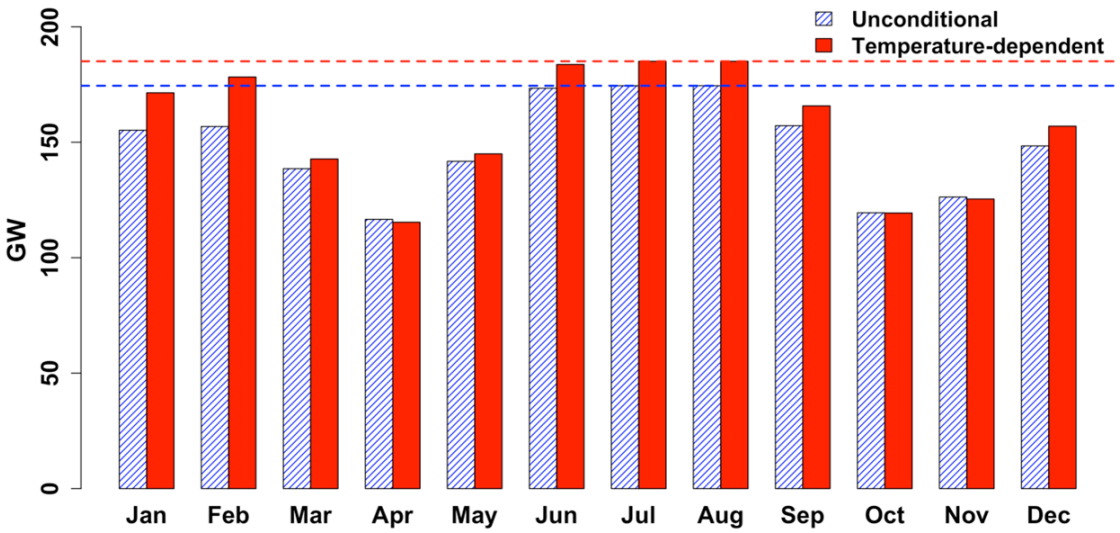

Figure 3.5: Monthly capacity requirements for achieving 0.1 LOLE under unconditional (blue hatching lines) and temperature-dependent (red) forced outage rates. Annual capacity requirements in each scenario indicated by dashed horizontal lines and correspond to the height of the respective July bars. Monthly procurement levels for April, October and November are slightly lower in the temperature-dependent forced outage rates scenario than in the unconditional forced outage rates scenario due to mild prevailing temperatures.

The 16% reduction obtained when procuring to achieve 0.1 LOLE on a monthly basis in the temperature-dependent scenario is a lower bound on the reduction that would be obtained by PJM if it continues to procure sufficient capacity to achieve 0.02 LOLE (corresponding to PJM's 26.6% realized reserve margin). This is because capacity requirements increase non-linearly in LOLE reductions. While we do not attempt to quantify the market implications of monthly procurement for capacity prices, economic fundamentals suggest that to the extent monthly procurement better enables inherently time-varying resources like wind, solar, and seasonally available demand-responsive loads to participate in the capacity market, monthly procurement should increase efficiency and reduce costs in the aggregate. However, we caveat that extreme temperatures are not the only cause of correlated failures, although they explain most correlated generator failures. Further analysis on additional drivers of correlated failures should be conducted prior to implementing seasonal or monthly procurement.

3.5.3 Resource adequacy implications of future generation resource scenarios

The results presented thus far have considered capacity requirements for PJM's existing conventional generator fleet. Given the economic pressure on nuclear and coal generators as a result of low natural gas prices [63–65], we next consider illustrative bounding scenarios on future fleet composition. In particular, we consider the resource adequacy implications of replacing the following with combined-cycle gas generators: 1) all existing nuclear generators; 2) all existing coal generators; and 3) all existing nuclear and coal generators. In each case, the replacement combined-cycle gas generator is made to be the same size as the retired coal or nuclear generator and has the same forced outage rates as PJM's existing combined-cycle gas generators.[18] Any of these fleet composition changes could increase or decrease capacity requirements, depending on how the change affects the system-level distribution of available capacity, particularly during peak-load hours.

In the unconditional forced outage rates scenario, replacing existing nuclear generators with combined-cycle gas generators results in a very slight increase in capacity requirements, while replacing existing coal generators reduces capacity requirements by approximately 5 GW. This is because combined-cycle gas generators' unconditional forced outage rates are slightly higher than those of nuclear generators but much lower than those of coal generators. In the temperature-dependent forced outage rates scenario, replacing existing nuclear generators reduces capacity requirements by approximately 2 GW, while replacing existing coal generators again reduces capacity requirements by approximately 5 GW. This is because combined-cycle gas generators' temperature-dependent forced outage rates are lower than those of nuclear and coal generators during the hot, high-load hours that contribute overwhelmingly to LOLE. Results are summarized in Table 3.6.

[18] Assuming gas generators are of equal nameplate capacity as the nuclear and coal generators they replace, rather than the more realistic scenario of a larger number of smaller generators, yields a lower bound on the reliability benefits achieved by the replacement scenario because it maximizes the variance in the system-level distribution of available capacity.

Table 3.6: Resource adequacy implications of fleet composition change scenarios under unconditional and temperature-dependent forced outage rates. Values reflect annual procurement. CC is combined cycle gas, NU is nuclear, ST is steam turbine (coal).

Case	Unconditional	Temperature-dependent
Base case	174.5 GW	185.1 GW
Replace NU with CC	174.8 GW	183.2 GW
Replace ST with CC	169.1 GW	179.5 GW
Replace NU and ST with CC	169.5 GW	178.9 GW

The inconsistent resource adequacy implication of retiring nuclear generators under the two scenarios illustrates the importance of robust modeling of generator availability during high-load hours. As high-load hours occur in conjunction with very hot and very cold temperatures in PJM, ignoring the effects of extreme temperatures on generator availability underestimates capacity procurement requirements. We caution that if winter resource adequacy risks, particularly natural gas supply deliverability risks, are insufficiently captured by the logistic regression model of [30] then nuclear generators may provide reliability value not accounted for here. Increases in gas-fired generation capacity could exacerbate fuel supply constraints during cold weather events, particularly affecting the availability of gas generators employing interruptible contracts [66]. In contrast, nuclear generators do not experience reduced availability during cold weather events [30].

3.5.4 Resource adequacy implications of future temperature scenarios

Given that all PJM generator types exhibit reduced availability during hot temperatures [19], we next examine the resource adequacy risks of increased temperatures for PJM's existing fleet. Previous research examining the implications of climate change for power system reliability has focused on understanding changes in heatwave frequency influencing peak loads [67], changing seasonal load shapes [68], as well as generator-level effects, such as increased deratings needed to comply with thermal pollution regulations [69,70], with some work aggregating the implications for power systems planning [71]. Here we consider the resource adequacy implications of temperature increases as they translate to reduced generator availability for PJM's conventional fleet. We examine two simplified scenarios: 1) where the temperature in each hour increases by 1 degree Celsius; and 2) where the

temperature in each hour increases by 2 degrees Celsius.[19] This approach is reasonable given that nearly all resource adequacy risk in PJM accrues in the summer, particularly during the hottest hours of July. These temperature scenarios increase capacity requirements by approximately 1 and 3 GW in the temperature-dependent forced outage rates scenario.[20] Increased temperatures are assumed to have no effect on generator availability in the unconditional forced outage rates scenario. Results are summarized in Table 3.7.

Table 3.7: Resource adequacy implications of temperature increases under unconditional and temperature-dependent forced outage rates. Values reflect annual procurement.

Case	Unconditional	Temperature-dependent
Base case	174.5 GW	185.1 GW
+1 degree Celsius	174.5 GW[a]	186.1 GW
+2 degrees Celsius	174.5 GW[a]	188.0 GW

[a] Increased temperatures are assumed to have no effect on generator availability in the unconditional forced outage rates scenario, so these values are identical to the base case.

3.5.5 Resource adequacy implications of future temperature scenarios in conjunction with future generation resource scenarios

Finally, we consider capacity requirements under increased temperatures in conjunction with replacement of all nuclear and/or coal generators by combined-cycle gas generators. This combines the three generator replacement scenarios from Section 3.5.3 with the +2 degrees Celsius case from Section 3.5.4. In the unconditional forced outage rates scenario, increased temperatures are again assumed to have no effect on generator availability, so each fleet replacement case matches the corresponding result from Table 3.6. In the temperature-dependent forced outage rates scenario, the capacity reductions from generator replacement offset the capacity increases required from increased temperatures to various extents; the net effect ranges from a reduction of 2 GW when replacing

[19] These temperature increases are with respect to the average temperatures occurring during 2006-2017, the period of our historical temperature data. Global temperatures during this time were already approximately 1 degree Celsius above the pre-industrial reference point [72]. Under business-as-usual emissions trajectories, an additional 1 and 2 degrees of temperature increase could be realized in approximately 2050 and 2100, respectively, though outcomes are highly sensitive to future emissions scenarios [73].

[20] We do not account for load increases as a result of temperature increases. This would increase capacity requirements in both scenarios, and creates an important interaction for consideration in power systems planning [71].

all coal and nuclear generators, to an increase of 1.5 GW when replacing only nuclear generators. Results are summarized in Table 3.8. While these resource adequacy risks appear modest, they warrant consideration in power systems planning. The method demonstrated here offers a tractable means of doing so.

Table 3.8: Resource adequacy implications of temperature increases under unconditional and temperature-dependent forced outage rates. Values reflect annual procurement.

Case	Unconditional	Temperature-dependent
Base case (current temperature, current fleet)	174.5 GW	185.1 GW
+2 degrees Celsius with no fleet change	174.5 GW[a]	188.0 GW
+2 degree Celsius with NU to CC	174.8 GW[a]	186.6 GW
+2 degree Celsius with ST to CC	169.1 GW[a]	184.3 GW
+2 degree Celsius with NU and ST to CC	169.5 GW[a]	183.0 GW

[a] Increased temperatures are assumed to have no effect on generator availability in the unconditional forced outage rates scenario, so these values are identical to the corresponding fleet change element of Table 3.6.

3.6 Discussion

The results presented here demonstrate the importance of incorporating temperature dependence of conventional generator availability into resource adequacy modeling in PJM. While we have endeavored to parameterize our model to be consistent with PJM practice, this analysis is not a substitute for PJM's own resource adequacy modeling process. For example, we do not account for transmission constraints nor emergency imports from neighboring systems due to a lack of publicly available data, but both were important during the Polar Vortex of 2014 [74]. Were the necessary data made available, a combined transmission and generation adequacy analysis for PJM and adjacent power systems could likely further improve the practice of resource adequacy modeling. In addition, we make the following caveats about our efforts to quantify the risks of correlated generator failures in PJM. First, the regression specification employed may not capture the full extent of winter resource adequacy risk, particularly the risk of interruptions to non-firm gas transportation. That effect becomes even more important under the various retirement scenarios in which coal and/or nuclear generators are replaced by additional gas capacity. In addition, we note that the

temperature-dependent forced outage rates are calculated on as much as 23 years of PJM GADS data. PJM's Capacity Performance program, which enacted penalties for resources that fail to perform when called, was implemented only after the Polar Vortex of 2014, and has led to significant investment in generator maintenance, dual-fuel capabilities, and firm gas supply contracts. Thus, it is possible that the relationship between extreme temperatures and forced outages has weakened in recent years. However, in analysis not presented here, we have found remarkably consistent temperature dependence of generator availability across vintages of coal generators in PJM. We also caution that extreme temperatures are not the only driver of correlated failures. For example, PJM is occasionally affected by hurricanes during fall months, which are typically accompanied by moderate temperatures. Despite these caveats, this analysis demonstrates the importance of considering ambient conditions in resource adequacy modeling.

3.7 Conclusions and Policy Implications

We have examined the implications of temperature-dependent generator availability for resource adequacy in the PJM Interconnection. By combining PJM's Generating Availability Data System (GADS) database with an open-source resource adequacy modeling tool and publicly available datasets, we demonstrate that the increased failure probabilities previously shown to affect PJM's conventional generator fleet at extreme temperatures [30] pose significant reliability risks not considered in standard resource adequacy modeling. Temperature dependence increases PJM's required reserve margin from 15.9% to 22.9% to achieve the 0.1 loss of load expectation (LOLE) reliability target in the 2018/2019 delivery year. However, PJM actually procured a 26.6% reserve margin for this delivery year, sufficient to achieve 0.02 LOLE and resulting in capacity payments of $315 million more than for a 22.9% reserve margin. Our estimate of the economic benefits that could be realized from reduced capacity procurement does not consider dynamic impacts, such as increases in energy and operating reserve costs associated with any increase in the prevalence of scarcity pricing events.

While greater procurement of generation capacity reduces the probability of loss-of-load events,

the incremental reliability benefits should be weighed against the incremental costs of procuring that capacity. We estimate that the value of lost load (VOLL) implicit in PJM's selected level of capacity procurement during the 2018/2019 delivery year is approximately $700,000/MWh, two orders of magnitude above values commonly used in operational contexts. This excess procurement helps explain why PJM has rarely triggered scarcity pricing, even during recent extreme weather events [61], and creates doubt as to whether current initiatives on scarcity pricing reform [62] will have meaningful effects. However, increased scarcity prices could reduce capacity market clearing prices, reduce implied VOLL, and therefore reduce economic benefits associated with resource adequacy reforms.

To the extent PJM understands that extreme temperatures increase generator failures and that this effect cannot be accounted for in its current resource adequacy modeling approach, the decision to procure additional capacity is prudent. The key contribution of this work is to demonstrate a tractable method for incorporating correlated failures into resource adequacy modeling, such that their reliability implications can be quantified directly, rather than accounted for heuristically.

Given the seasonality of temperatures, we examine the potential benefits of monthly, rather than annual, capacity procurement for PJM. We find that monthly capacity targets could reduce annual capacity procurement by approximately 16% without increasing LOLE. While we do not estimate the economic benefits of a monthly capacity market, we note this potential reduction in capacity requirements is substantial. Further, a monthly or seasonal structure should allow for improved efficiency in the supply offers of inherently seasonal resources, such as solar, wind, and some demand response. However, some of the benefits associated with increased efficiency may be offset by increased risk premiums for firms, given decreased certainty around annual capacity revenue. [75] finds annual capacity markets reduce resource risk asymmetrically, advantaging resources with a higher share of operating costs (e.g., most natural gas-fired generation), suggesting a more granular capacity market may differentially increase resource risk premiums and affect longer-term procurement in PJM.

Because economic pressure is exerted on nuclear and coal generators by low wholesale electricity prices due to inexpensive natural gas, we examine the resource adequacy implications of their retirement and replacement by combined-cycle gas generators. Assuming the new combined-cycle

gas generators have forced outage rates equal to PJM's existing combined-cycle fleet, the system experiences moderate resource adequacy benefits of approximately 1% and 3%, respectively. This result is driven by the lower forced outage rates of combined-cycle gas generators during the hot, high-load hours that contribute overwhelmingly to LOLE in PJM. The inconsistent resource adequacy implication of replacing nuclear generators with combined-cycle gas generators under our two scenarios illustrates the importance of robust modeling of generator availability during high-load hours.

Since all PJM generator types exhibit reduced availability during hot temperatures, we examine the resource adequacy implications of future climate change scenarios, operationalized as temperature increases of 1 and 2 degrees Celsius. Holding other factors constant, these scenarios result in increased capacity requirements of approximately 0.5% and 1.5%, respectively. Combining the 2-degrees Celsius temperature increase with the retirement cases results in a 1% increase and 0.5% decrease in capacity requirements when replacing nuclear and coal generators, respectively. While these resource adequacy risks appear to be mild in PJM, it is only by explicitly modeling generator availability as a function of temperature that the magnitude of the risk can be determined.

Planning models such as RECAP (Renewable Energy Capacity Planning Model) and PJM's own resource adequacy tools do not perform full sequential modeling of the system over time. Thus, even when accounting for temperature-dependent generator availability, resource adequacy modeling may miss important reliability risks. Future work will consider temperature-dependent generator availability on operational timescales using a security-constrained unit commitment and economic dispatch model. Such an approach can account for the risk of lost load from generator outages during sustained extreme weather events, incorporate demand response resources, quantify operational flexibility needs, and allow system operators to better understand the value of procuring operating reserves, complementing this analysis.

References

[1] ReliabilityFirst Corporation, "BAL-502-RFC-03: Planning resource adequacy analysis, assessment and documentation," 2017. https://www.nerc.com/pa/Stand/Reliability%20Standards/BAL-502-RF-03.pdf.

[2] G. Calabrese, "Generating reserve capacity determined by the probability method," *Transactions of the American Institute of Electrical Engineers*, vol. 66, no. 1, pp. 1439–1450, 1947. doi:https://doi.org/10.1109/T-AIEE.1947.5059596.

[3] J. D. Hall, R. J. Ringlee, and A. J. Wood, "Frequency and duration methods for power system reliability calculations: Part I–generation system model," *IEEE Transactions on Power Apparatus and Systems*, vol. 87, no. 9, pp. 1787–1797, 1968. doi:https://doi.org/10.1109/TPAS.1968.291986.

[4] R. J. Ringlee and A. J. Wood, "Frequency and duration methods for power system reliability calculations: Part II–demand model and capacity reserve model," *IEEE Transactions on Power Apparatus and Systems*, vol. 88, no. 4, pp. 375–388, 1969. doi:https://doi.org/10.1109/TPAS.1969.292458.

[5] R. Billinton and R. N. Allan, *Reliability evaluation of power systems.* Plenum, 2 ed., 1996.

[6] R. Billinton and R. N. Allan, *Reliability assessment of large electric power systems.* Kluwer Academic Publishers, 1 ed., 1988.

[7] M. P. Bhavaraju, J. A. Hynds, and G. A. Nunan, "A method for estimating equivalent forced outage rates of multistate peaking units," *IEEE Transactions on Power Apparatus and Systems*, no. 6, pp. 2067–2075, 1978. doi:https://doi.org/10.1109/TPAS.1978.354710

[8] R. Billinton and W. Li, *Reliability assessment of electric power systems using Monte Carlo methods.* Springer Science & Business Media, 1 ed., 1994.

[9] F. A. Felder, "Incorporating Resource Dynamics to Determine Generation Adequacy Levels in Restructured Bulk Power Systems," *KIEE International Transactions on Power Engineering*, pp. 100–105. 2004.

[10] K. Dragoon and V. Dvortsov, "Z-method for power system resource adequacy applications," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 982–988, 2006. doi:https://doi.org/10.1109/TPWRS.2006.873417.

[11] J. Kwon, Z. Zhou, T. Levin, and A. Botterud, "Resource Adequacy in Electricity Markets with Renewable Energy," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 773–781, 2019. doi:https://doi.org/10.1109/tpwrs.2019.2930934

[12] B. A. Frew, W. J. Cole, Y. Sun, T. T. Mai, and J. Richards, "8760-Based Method for Representing Variable Generation Capacity Value in Capacity Expansion Models: Preprint," 2017. https://www.osti.gov/scitech/biblio/1375305.

[13] M. Milligan and K. Porter, "The Capacity Value of Wind in the United States: Methods and Implementation," *Electricity Journal*, vol. 19, no. 2, pp. 91–99, 2006. doi:https://doi.org/10.1016/j.tej.2005.12.010.

[14] C. Bothwell and B. F. Hobbs, "Crediting Renewables in Electricity Capacity Markets: The Effects of Alternative Definitions upon Market Efficiency," *The Energy Journal On-Line*, pp. 1–20, June, 2016. https://EconPapers.repec.org/RePEc:aen:journl:ej38-si1-bothwell.

[15] E. Lannoye, D. Flynn, and M. O'Malley, "Evaluation of power system flexibility," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 922–931, 2012. doi:https://doi.org/10.1109/TPWRS.2011.2177280.

[16] K. Carden, N. Wintermantel, and J. Pfeifenberger, "The Economics of Resource Adequacy Planning : Why Reserve Margins Are Not Just About Keeping the Lights On," 2011. http://www.nrri.org/pubs/electricity/NRRI_resource_adequacy_planning_april11-09.pdf.

[17] G. Slipac, M. Zeljko, and D. Šljivac, "Importance of reliability criterion in power system expansion planning," *Energies*, vol. 12, no. 9, 2019. doi:https://doi.org/10.3390/en12091714.

[18] W. W. Hogan, "On an "Energy Only" Electricity Market Design for Resource Adequacy," *Economics of Energy & Environmental Policy*, pp. 1–39, 2005. https://ferc.gov/EventCalendar/Files/20060207132019-Hogan_Energy_Only_092305.pdf.

[19] J. Bowring, "Capacity Markets in PJM," *Economics of Energy and Environmental Policy*, vol. 2, no. 2, pp. 47–64, 2013. doi:https://doi.org/10.5547/2160-5890.2.2.3.

[20] K. Spees, S. A. Newell, and J. P. Pfeifenberger, "Capacity Markets - Lessons Learned from the First Decade," *Economics of Energy and Environmental Policy*, vol. 2, no. 2, pp. 1–26, 2013. doi:https://doi.org/10.5547/2160-5890.2.2.1.

[21] J. P. Pfeifenberger, S. A. Newell, K. Spees, A. Murray, and I. Karkatsouli, "Third Triennial Review of PJM's Variable Resource Requirement Curve," 2014. https://www.pjm.com/-/media/library/reports-notices/reliability-pricing-model/20140515-brattle-2014-pjm-vrr-curve-report.ashx?la=en.

[22] S. A. Newell, D. L. Oates, J. P. Pfeifenberger, K. Spees, J. M. Hagerty, J. I. Pedtke, M. Witkin, and E. Shorin, "Fourth Review of PJM's Variable Resource Requirement Curve," 2018. https://www.pjm.com/-/media/library/reports-notices/reliability-pricing-model/20180425-pjm-2018-variable-resource-requirement-curve-study.ashx?la=en.

[23] B. F. Hobbs, M. C. Hu, J. G. Iñón, S. E. Stoft, and M. P. Bhavaraju, "A dynamic analysis of a demand curve-based capacity market proposal: The PJM reliability pricing model," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 3–14, 2007. doi:https://doi.org/10.1109/TPWRS.2006.887954.

[24] P. A. Ruiz and G. Gross, "Short-term resource adequacy in electricity market design," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 916–926, 2008. doi:https://doi.org/10.1109/TPWRS.2008.926094

[25] P. Mastropietro, P. Rodilla, and C. Batlle, "Performance incentives in capacity mechanisms: Conceptual considerations and empirical evidence," *Economics of Energy and Environmental Policy*, vol. 6, no. 1, pp. 149–163, 2017. doi:https://doi.org/10.5547/2160-5890.6.1.pmas.

[26] North American Electric Reliability Corporation, "2018 Summer Reliability Assessment," 2018. https://www.nerc.com/pa/RAPA/ra/Reliability%5CAssessments%5C%0ADL/NERC_SRA_05252018_Final.pdf.

[27] PJM, "2018 PJM Reserve Requirement Study," 2018. https://www.pjm.com/-/media/committees-groups/subcommittees/raas/20181004/20181004-pjm-reserve-requirement-study-draft-2018.ashx.

[28] S. Murphy, J. Apt, J. Moura, and F. Sowell, "Resource adequacy risks to the bulk power system in North America," *Applied Energy*, vol. 212, no. June 2017, pp. 1360–1376, 2018. doi:https://doi.org/10.1016/j.apenergy.2017.12.097.

[29] D. P. Gaver, F. E. Montmeat, and A. D. Patton, "Power system reliability I: Measures of reliability and methods of calculation," *IEEE Transactions on Power Apparatus and Systems*, vol. 83, pp. 727–737, Jul 1964. doi:https://doi.org/10.1109/TPAS.1964.4766068

[30] S. Murphy, F. Sowell, and J. Apt, "A time-dependent model of generator failures and recoveries captures correlated events and quantifies temperature dependence," *Applied Energy*, vol. 253, no. July, p. 113513, 2019. doi:https://doi.org/10.1016/j.apenergy.2019.113513

[31] Energy+Environmental Economics Inc., "Renewable energy capacity planning model." https://www.ethree.com/tools/recap-renewable-energy-capacity-planning-model/.

[32] Monitoring Analytics, "State of the Market Report for PJM," 2017. https://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2017.shtml.

[33] P. Cramton, A. Ockenfels, and S. Stoft, "Capacity market fundamentals," *Economics of Energy & Environmental Policy*, vol. 2, no. 2, pp. 27–46, 2013.

[34] P. Cramton and S. Stoft, "Forward reliability markets: Less risk, less market power, more efficiency," *Utilities Policy*, vol. 16, no. 3, pp. 194–201, 2008. doi:https://doi.org/10.1016/j.jup.2008.01.007

[35] J. P. Pfeifenberger, K. Spees, K. Carden, and N. Wintermantel, "Resource adequacy requirements: Reliability and economic implications," tech. rep., 2013. https://brattlefiles.blob.core.windows.net/files/7636_resource_adequacy_requirements_-_reliability_and_economic_requirements.pdf.

[36] PJM Interconnection, "Final zonal UCAP obligations, capacity prices and CTR credit rates," 2018. https://www.pjm.com/-/media/markets-ops/rpm/rpm-auction-info/2018-2019-%0Afinal-zonal-ucap-obligations-capacity-prices-ctr-credit-rates.ashx?la=en.

[37] PJM Interconnection, "PJM Manual 20: PJM resource adequacy analysis," 2019. https://www.pjm.com/~/media/documents/manuals/m20.ashx.

[38] PJM Interconnection, "PJM generating adequacy analysis: Technical methods," 2003. https://www.pjm.com/~/media/etools/oasis/references/whitepaper-sections-12.ashx.

[39] PJM Interconnection, "PJM load forecasting model whitepaper," 2016. https://www.pjm.com/~/media/library/reports-notices/load-forecast/2016-load-forecastwhitepaper.ashx.

[40] E. S. Pearson and H. O. Hartley, *Biometrika tables for statisticians*. Cambridge (United Kingdom): Cambridge University Press, 1966.

[41] I. Integ Enterprise Consulting, "PowerGADS 3.0 user manual," 2015. https://www.pjm.com/markets-and-operations/etools/egads.aspx.

[42] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstem, "A simulation study of the number of events per variable in logistic regression analysis," *Journal of Clinical Epidemiology*, vol. 49, no. 12, pp. 1373–1379, 1996. doi:https://doi.org/10.1016/s0895-4356(96)00236-3.

[43] IEEE Power Engineering Society, *Standard Definitions for Use in Reporting Electric Generating Unit Reliability, Availability, and Productivity*, vol. 2006. 2007. doi:https://doi.org/10.1109/IEEESTD.2007.335902.

[44] U.S. Energy Information Administration, "Form 923, final 2017 data," 2018. https://www.eia.gov/electricity/data/eia923/.

[45] B. D. Hoen, J. E. Diffendorfer, J. T. Rand, L. A. Kramer, C. P. Garrity, and H. E. Hunt, "United States Wind Turbine Database," 2018. https://eerscmap.usgs.gov/uswtdb.

[46] C. Draxl, A. Clifton, B. M. Hodge, and J. McCaa, "The Wind Integration National Dataset (WIND) toolkit," *Applied Energy*, vol. 151, pp. 355–366, 2015. https://www.nrel.gov/docs/fy15osti/64691.pdf.

[47] W. Katzenstein, E. Fertig, and J. Apt, "The variability of interconnected wind plants," *Energy Policy*, vol. 38, no. 8, pp. 4400–4410, 2010. doi:https://doi.org/10.1016/j.enpol.2010.03.069.

[48] R. H. Wiser and M. Bolinger, "2017 Wind Technologies Market Report," 2018. https://emp.lbl.gov/wind-technologies-market-report.

[49] J. Rand, "Personal communication," tech. rep., nov 2018.

[50] National Renewable Energy Laboratory, "NSRDB Data Viewer." https://maps.nrel.gov/nsrdb-viewer/.

[51] A. Habte, M. Sengupta, and A. Lopez, "Evaluation of the National Solar Radiation Database (NSRDB): 1998-2015," 2017. https://www.nrel.gov/docs/fy17osti/67722.pdf.

[52] Pacific Northwest National Laboratory, "GridLAB-D." http://www.gridlabd.org/.

[53] National Renewable Energy Laboratory, "System Advisory Model." https://sam.nrel.gov/.

[54] PJM Interconnection, "PJM load forecast report," 2018. https://www.pjm.com/-%0A/media/library/reports-notices/load-forecast/2018-load-forecast-report.ashx.

[55] PJM Interconnection, "2018/2019 RPM base residual auction results," 2015. https://www.pjm.com/-/media/markets-ops/rpm/rpm-auction-info/2018-2019-baseresidual-auction-report.ashx?la=en

[56] PJM Interconnection, "2014 PJM Reserve Requirement Study," tech. rep., 2014.

[57] PJM Interconnection, "2019/2020 RPM base residual auction planning period parameters," 2016. https://www.pjm.com/-/media/markets-ops/rpm/rpm-auction-info/2019-2020-rpmbra-planning-parameters-report.ashx?la=en.

[58] C. Byers, T. Levin, and A. Botterud, "Capacity market design and renewable energy: Performance incentives, qualifying capacity, and demand curves," *The Electricity Journal*, vol. 31, no. 1, pp. 65–74, 2018. doi:https://doi.org/10.1016/j.tej.2018.01.006.

[59] D. B. Patton, "Comments of David B. Patton, Ph.D., regarding state policies affecting eastern RTOs," 2017. https://www.ferc.gov/CalendarFiles/20170426150115-Patton,%5C%0APotomac%5CEconomics.pdf.

[60] S. Newell, R. Carroll, A. Kaluzhny, K. Spees, K. Carden, N. Wintermantel, and A. Krasny, "Estimation of the market equilibrium and economically optimal reserve margins for the ERCOT region," 2018. http://www.ercot.com/content/wcm/lists/143980/10.12.2018_ERCOT_MERM_Report_%0AFinal_Draft.pdf.

[61] M. Watson, "PJM's winter 2019 experience reinforces need for reserve pricing reform," 2019. https://www.spglobal.com/platts/en/market-insights/latest-news/electricpower/031819-pjms-winter-2019-experience-reinforces-need-for-reserve-pricing-reform.

[62] PJM Energy Price Formation Senior Task Force, "Price formation," 2018. https://www.pjm.com/-/media/committees-groups/task-forces/epfstf/20181214/20181214-item-04-price-formation-paper.ashx.

[63] United States Department of Energy, "Staff report to the Secretary on electricity markets and reliability," 2017. https://energy.gov/sites/prod/files/2017/08/f36/Staff%5C%0AReport%5Con%5CElectricity%5CMarkets%5Cand%5CReliability_0.pdf.

[64] M. B. Roth and P. Jaramillo, "Going nuclear for climate mitigation: An analysis of the cost effectiveness of preserving existing US nuclear power plants as a carbon avoidance strategy," *Energy*, vol. 131, pp. 67–77, 2017. doi:https://doi.org/10.1016/j.energy.2017.05.011.

[65] J. D. Jenkins, "What's killing nuclear power in US electricity markets? Drivers of wholesale price declines at nuclear generators in the PJM Interconnection," 2018. http://ceepr.mit.edu/files/papers/2018-001.pdf.

[66] PJM Interconnection, "Fuel security analysis: A PJM resilience initiative," 2018. https://www.pjm.com/-/media/library/reports-notices/fuel-security/2018-fuel-securityanalysis.ashx.

[67] D. Burillo, M. V. Chester, B. Ruddell, and N. Johnson, "Electricity demand planning forecasts should consider climate non-stationarity to maintain reserve margins during heat waves," *Applied Energy*, vol. 206, no. June, pp. 267–277, 2017. doi:https://doi.org/10.1016/j.apenergy.2017.08.141.

[68] F. R. Fonseca, P. Jaramillo, M. Bergés, and E. Severnini, "Seasonal effects of climate change on intra-day electricity demand patterns," tech. rep., 2019. doi:https://doi.org/10.1007/s10584-019-02413-w.

[69] S. N. Chandramowli and F. Felder, "Impact of climate change on electricity systems and markets–A review of models and forecasts," *Sustainable Energy Technologies and Assessments*, vol. 5, pp. 62–74, 2014. doi:https://doi.org/10.1016/j.seta.2013.11.003.

[70] M. A. Cook, C. W. King, F. T. Davidson, and M. E. Webber, "Assessing the impacts of droughts and heat waves at thermoelectric power plants in the United States using integrated regression, thermodynamic, and climate models," *Energy Reports*, vol. 1, pp. 193–203, 2015. doi:https://doi.org/10.1016/j.egyr.2015.10.002.

[71] M. T. Craig, S. Cohen, J. Macknick, C. Draxl, O. J. Guerra, M. Sengupta, S. E. Haupt, B.-M. Hodge, and C. Brancucci, "A review of the potential impacts of climate change on bulk power system planning and operations in the United States," *Renewable and Sustainable Energy Reviews*, vol. 98, pp. 255–267, 2018. https://doi.org/10.1016/j.rser.2018.09.022.

[72] Intergovernmental Panel on Climate Change, "Climate change 2014 synthesis report summary for policymakers," 2018. https://www.ipcc.ch/site/assets/uploads/2018/02/AR5_SYR_FINAL_SPM.pdf.

[73] Intergovernmental Panel on Climate Change, "Global warming of 1.5 ºC: An IPCC special report on the impacts of global warming of 1.5 ºC above pre-industrial levels and related global greenhouse gas emission pathways," 2018. https://www.ipcc.ch/sr15/.

[74] PJM Interconnection, "Analysis of Operational Events and Market Impacts During the January 2014 Cold Weather Events," 2014. https://www.pjm.com/~/media/library/reports-notices/weather-related/20140509-analysis-of-operational-events-and-market-impacts-during-the-jan-2014-cold-weather-events.ashx.

[75] J. Mays, D. Morton, and R. P. O'Neill, "Asymmetric Risk and Fuel Neutrality in Capacity Markets," *SSRN Electronic Journal*, 2019. doi:https://doi.org/10.2139/ssrn.3330932.

Chapter 4 Dynamic Operating Reserve Procurement Improves Scarcity Pricing in PJM*

Abstract

Competitive electricity markets can procure reserve generation through a market in which the demand for reserves is administratively established. A downward sloping or stepped administrative demand curve is commonly termed an operating reserve demand curve (ORDC). We propose a dynamic formulation of an ORDC with generator forced outage probabilities conditional on ambient temperature to implement scarcity pricing in a wholesale electricity market. This formulation improves on common existing methods used by wholesale market operators to articulate ORDCs by explicitly accounting for a large source of observed variability in generator forced outages, whereby for a fixed load, more reserves are required during times of extreme heat and cold to maintain a constant risk of reserve shortage. Such a dynamic ORDC increases social welfare by $17.1 million compared to current practice in the PJM Interconnection during a high load week in a welfare-maximizing electricity market with co-optimized procurement of energy and reserves. A dynamic ORDC increases reserve prices under scarcity conditions, but has minimal effects on total market payments. The results are directly relevant to the modeled two-settlement electricity market in PJM, which is currently undergoing enhancements to its ORDC.

* This paper was published as L. Lavin, S. Murphy, B.Sergi, and J. Apt, "Dynamic operating reserve procurement improves scarcity pricing in PJM," *Energy Policy*, 2020. 147: 111857. doi:https://doi.org/10.1016/j.enpol.2020.111857 .

Abbreviations and acronyms

BGE-PEP	Baltimore Gas and Electric and Potomac Electric Power Company model zone
CC	Combined cycle natural gas generator type
COPT	Capacity outage probability table
CT	Combustion turbine natural gas generator type
DA	Day-ahead
DASR	Day-ahead scheduling reserves
DOM	Dominion model zone
DR	Demand response
DS	Diesel generator type
ERCOT	Electricity Reliability Council of Texas
FERC	Federal Energy Regulatory Commission
GADS	Generating Availability Data System
GW	Gigawatt
HD	Hydroelectric and pumped storage generator type
HE	Hour ending
LDA	Load delivery area
LMP	Locational marginal price
LOLP	Loss of load probability, the probability net load exceeds available generation
MAD	Mid-Atlantic Dominion reserve subzone
MW	Megawatt
MWh	Megawatt-hour
MRR	Minimum reserve requirement
NERC	North American Electric Reliability Corporation
NU	Nuclear generator type
ORDC	Operating reserve demand curve
PJM	The PJM Interconnection, the largest system operator by load in North America
PPL-METED	Pennsylvania Power and Light and Metropolitan Edison Company model zone
RT	Real-time
ST	Steam turbine generator type
UC/ED	Unit commitment economic dispatch
VOLL	Value of lost load, the valuation of a unit of unmet load

4.1 Introduction

Operating reserves are crucial for power system reliability. Traditionally, system operators use heuristics for determining the quantity and types of operating reserves needed to maintain system reliability. A common heuristic is to hold sufficient reserves to guard against the largest single contingency (termed the N-1 contingency), generally a large generator or transmission line. Following the expansion of competitive wholesale electricity markets after restructuring, increasing effort has

been put on translating reserve heuristics into competitively procured services; for example, by articulating a demand curve for a reserve product based on the marginal value it provides in increased reliability. This paper quantifies how better accounting for the probability of generator contingencies improves reserves procurement.

Competitive wholesale electricity markets commonly determine market clearing prices paid by loads to generators as locational marginal prices (LMPs) using unit commitment and economic dispatch (UC/ED) optimization models. These models minimize the as-bid cost of supplying generation to serve load, subject to relevant commitment and security constraints; prices are determined as the locational marginal cost of serving load and theoretically support efficient dispatch [1]. Academic research on policies and trends affecting electricity loads, generator characteristics, and generator costs therefore commonly employs similar mixed-integer linear programs to estimate the policy or trend's effect [2–4]. Due to both public data availability limitations and computational limitations in solving large mixed-integer programs at increasing levels of geographic and temporal granularity researchers necessarily make simplifications and approximations in these models; for example, by limiting geography to a zonal representation of a market [5] and by employing a vertical demand curve for operating reserves as a function of key sources of uncertainty to address reliability-related security constraints [6]. Our research improves common simplifications in procurement of the optimal quantity of operating reserves in both previous academic research and under current and proposed practice in the PJM Interconnection LLC (PJM), whose data we use for our model cases to increase the direct policy relevance of conclusions.

PJM operates the largest wholesale electricity market by load in North America, serving 65 million customers in 13 mid-Atlantic states. PJM administers a competitive two-settlement wholesale electricity market, with settlements at day-ahead and real-time. PJM selects competitive generation offers to meet expected load at least cost, subject to relevant physical constraints. To maintain reliability under uncertainty in load and generation availability, PJM holds a minimum quantity of primary operating reserves in real-time, historically 150% of its single largest contingency (usually about 2.1 gigawatts (GW) of capacity, approximately equivalent to the output of two large nuclear reactors), with synchronized primary reserves equivalent to the single largest contingency (usually

about 1.4 GW) [7].

System operators like PJM have long understood that probabilistic methods can be used to better quantify the optimal quantity and type of reserves to hold across different timeframes [8]. However, implementing these methods requires either extensive use of computationally intensive stochastic optimization methods [9] or improved heuristics for deterministically representing demand for reserves deriving from uncertainties that affect reliability. Pursuing the latter approach requires both identification and quantification of uncertainties as well as a method for implementing desired reserve procurement in competitive wholesale markets.

Traditionally, important uncertainties include large conventional generator failures and load forecast error, with the more recent addition of uncertainty in the availability of variable generation like wind power [10, 11]. Of these uncertainties the greatest risk to the operator's ability to reliably serve load on short time horizons is generally unscheduled conventional generator failures. Consequently, research has focused on developing heuristics for integrating the probability of different generator availability states into market clearing algorithms; for example, by enumerating the probability of each possible single and double generation failure [12]. These failure probabilities are traditionally calculated by assuming generators can be modeled as unconditionally independent random variables based on historical data [13].

Given a heuristic for reserve requirements, wholesale market operators in North America commonly integrate reserve procurement into competitive electricity markets via a demand curve. Reserve procurement in many markets has historically been administrative (equivalently conceptualized as a vertical demand curve), with a fixed quantity of reserves procured and a very high penalty pricing mechanism triggered during shortfalls. A more sophisticated and increasingly popular method for reserve procurement is a sloped or step-wise operating reserve demand curve (ORDC). An ORDC approach recognizes the optimal quantity of reserves to hold depends on a probability distribution of near real-time uncertainties, which may be represented in markets as demand curves reflecting declining marginal value for reserves based on the probability they will be needed multiplied by the cost of the reserve shortage action they avoid [14]. The cost of reserve shortage actions is sometimes referred to as the value of lost load (VOLL), which may reflect either the cost of unserved energy (i.e.,

shedding firm load) or other emergency operator actions taken under reserve shortage, and therefore is often referred to by system operators as a penalty factor associated with those actions. This paper uses the terms VOLL and penalty factor interchangeably, assuming both reflect emergency actions taken by the system operator under reserve shortage conditions.

Recent research demonstrates that PJM's conventional generators are much more likely to fail during extreme temperature events [15,16]. In other words, generator failures are not well represented by unconditionally independent random variables. Further, extreme temperatures are correlated with high load hours. In a separate analysis, accounting for this temperature dependence was estimated to increase PJM's total capacity requirement for resource adequacy planning purposes by 6% in order to cover peak summer and winter loads [17]. An increase in necessary capacity for planning purposes suggests modifications to operating reserve demand curves may also be needed to ensure this capacity is available during hours with high loss of load probability (LOLP). Previous research has estimated the effect of using conditional distributions for wind forecast error on the dynamic articulation of an ORDC to support improved scarcity pricing in a realistic wholesale electricity market [11]; we extend this research by applying similar methods for conventional generator forced outages.

Analysis using PJM's Generating Availability Data System (GADS) database shows that generation losses greater than PJM's primary synchronized reserve requirement have occurred in approximately 1% of hours since 1995. While these large generation losses typically occur in only a single hour per day, days with multiple such hours also occur. For example, during the Polar Vortex of 2014, 15 hours experienced generation losses greater than the operating reserve target on the most extreme day (January 7). A bar plot of these large-loss hours is presented against the prevailing ambient temperature in Figure 4.1. Such large losses on operational timescales suggest that holding constant levels of reserves or conditioning reserves only on historical (e.g. annual- or seasonal-average) generator failure probabilities is not an optimal strategy for PJM; instead, the optimal quantity of reserves is better modeled as conditional on ambient temperature.

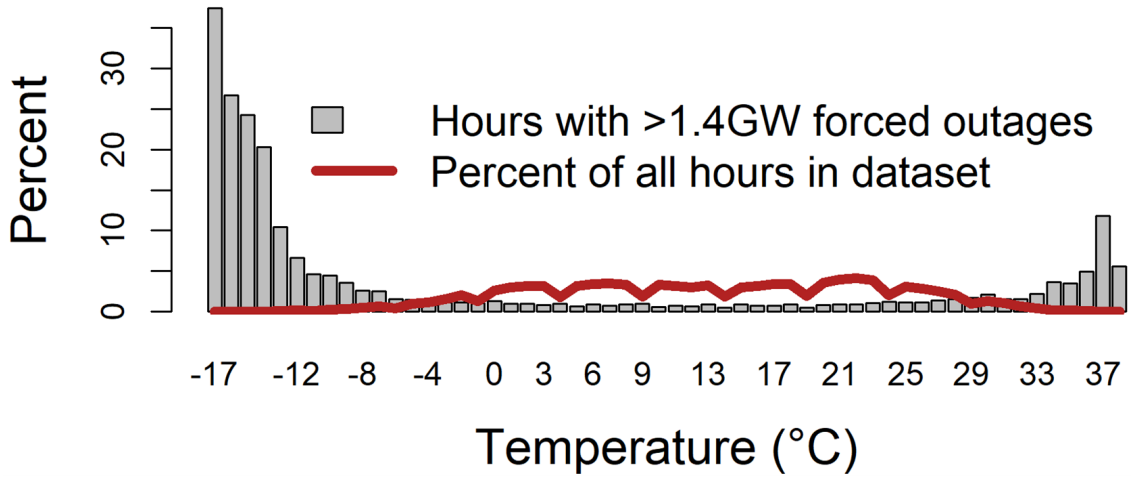

Figure 4.1: The proportion of hours at a given ambient temperature in which generator losses greater than PJM's primary synchronized minimum reserve requirement (set at 1.4 GW for figure) occurred. Calculations use PJM's Generating Availability Data System (GADS) database, processed as described in [16]. Data period: 1995-2018Q1.

In a competitive wholesale market, operational challenges during extreme events should be endogenously reflected in prices, providing incentives for additional reserves. However, historically PJM has procured additional reserves through selective operator intervention. Recognizing that operator actions taken in support of reliability can unduly mute energy and reserve market price signals, PJM submitted a proposal for reserve pricing reform to the Federal Energy Regulatory Commission (FERC) in March 2019 [18]. Among other enhancements, PJM proposes a stepped demand curve for operating reserves, rather than the current vertical demand curve it employs.[1] This ORDC would provide price incentives for the provision of operating reserves in excess of the current fixed procurement target, reducing the need for operator intervention to maintain reliability.

We contribute to the goal of reducing the need for administrative interventions by demonstrating the benefits that may be obtained by articulating ORDCs that account for day-ahead generator failure probabilities conditioned on forecast ambient temperature. We accomplish this by drawing on the temperature dependence of generator availability identified in [16] to set operating reserve requirements in PJM's day-ahead market. This differs from PJM's current proposal before FERC,

[1]PJM added a shorter fixed second step and optional second step extension to its primary synchronized reserve demand curve in 2017 and 2014, respectively, in response to a FERC order requiring implementation of transient reserve shortage pricing [19]. We implement and display the fixed part of this second step in Figure 4.4.

which develops 24 fixed ORDCs for six daily time blocks across four seasons based on empirical distributions of historical uncertainties in load, wind, solar, forced outages, and net interchange over applicable time intervals [18]. While calculating historical average generator forced outages by time block will capture some temperature dependence, it may not demand sufficient reserves during truly extreme conditions. We employ PJM as a test case due to its current policy relevance, but expect the results to be broadly relevant to market operators with similar two-settlement market designs and proportion of conventional generation resources.

4.2 Methods

We employ a zonal unit commitment and economic dispatch (UC/ED) optimization model to simulate generation, reserves, and locational marginal prices (LMPs) in PJM. We employ a zonal model due to lack of sufficient publicly available data for nodal representation of PJM's high voltage transmission system and our policy focus on zonally procured operating reserves. Zonal models are common in academic literature but may underestimate the effect of transmission constraints on prices, particularly under stressed system conditions. Temporal resolution is hourly, with prices interpreted as the dual variable of the model's load-balance constraint when relaxed to a linear program for the pricing run.[2]

The main distinction between our approach and UC/ED models used by system operators in the United States (including PJM) is our ability to dynamically estimate and implement an ORDC based on forecast hourly ambient temperatures, described in more detail below. Because energy and reserves are co-optimized in commitment and dispatch, our ORDC formulation affects the simulated zonal energy market clearing prices, in addition to reserve prices. Our approach is similar to that employed by Zhou and Botterud [11], but with a focus on generator forced outage rates instead of wind power variability for parameterization of the ORDC.

[2]This approach to price formation is consistent with current practice in PJM, though PJM's models have more detailed representation of the transmission system for determining LMPs and constraints around the ability of generators to set clearing price based on their type of bid submission and market power mitigation tests. Alternative approaches to price formation within an optimization market clearing price framework are possible and have been recently discussed by PJM [19].

4.2.1 Articulation of a Dynamic ORDC

Operating reserves are a subset of ancillary services that reflect the ability of generators to help maintain system stability and serve load given uncertainty in the future availability of generation and load. Power system operators procure or maintain multiple types of operating reserves associated with different uncertainties (e.g., over- and under-forecasting) and timescales. In competitive wholesale markets, like PJM, some portion of these reserves may be procured and paid through bid-based market mechanisms. Historically PJM's ancillary services have included bid-based processes for regulation (5 minute response time) and three operating reserve products: primary synchronized and non-synchronized reserves (10 minute response time) in its real-time (RT) market and day-ahead (DA) scheduling (secondary) reserves (DASR, 30 minute response time) in the day-ahead market (Table 5.1). PJM has one additional set of sub-zonal primary reserve procurement constraints for the Mid-Atlantic Dominion sub-zone, reflecting significant East-West transmission constraints across the Appalachian mountain range. PJM's proposal before FERC seeks to better align performance and prices for primary and secondary reserves between the day-ahead and real-time markets [18].

Table 4.1: Current operating reserve products in PJM. "N-1" is the single largest contingency (typically a generator or transmission line).

	Minutes	**Market**	**Minimum Reserve Requirement**
1) Primary synchronized	≤ 10	Real-time	$\geq$ N-1
1A) Sub-zonal primary synchronized	≤ 10	Real-time	1) in sub-zone + most limiting interface into sub-zone $\geq$ N-1
2) Primary non-synchronized	≤ 10	Real-time	Together with (1) $\geq$ 1.5*(N-1)
2A) Sub-zonal primary non-synchronized	≤ 10	Real-time	2) in sub-zone + most limiting interface into sub-zone $\geq$ 1.5*(N-1)
3) Day-ahead scheduling	≤ 30	Day-ahead	$\geq$ forecast error

As evidenced by current practice in the Electricity Reliability Council of Texas, Inc. (ERCOT) [20] and proposals in PJM [18], ORDCs are an increasingly popular approach to reserve procurement in a competitive market. This demand curve may take different shapes in different markets, but is generally based on underlying probabilities of lost load for a quantity of held reserves. The probability of lost load is converted to a price by multiplying by a penalty factor based on an assumed cost of emergency reserve shortage actions. Given that an ORDC quantifies the marginal value of held operating reserves, co-optimizing energy and reserves with an ORDC allows for trade-offs between the utility of holding additional reserves with their cost.

A key parameter for determining the shape of the ORDC is thus the LOLP, which is itself a function of the uncertainty in the availability of generators and load at some time t in the future. Current practice for system operators like ERCOT with LOLP-based ORDCs generally parameterizes held reserve forecast error as a normal distribution within season-hour blocks, calculating a mean and standard deviation for the distribution from historical data within the season-hour block [21]. Implicitly, this assumes that conventional generator failures are invariant to ambient conditions within seasonal-hour block, and invariant to changes in fleet composition relative to historical data within seasonal-hour blocks, implying a single capacity outage probability table (COPT) for the conventional generator fleet. However, Murphy et al. [15] demonstrates that conventional generator failures increase during extreme temperatures and depend on generator type. We therefore draw on Murphy et al. [16] to compute temperature-dependent forced outage rates, which we use to create more accurate conditional distributions of available capacity of conventional generators in each hour. This allows us to more accurately forecast hourly generation margins and corresponding LOLP. We model LOLP solely as a function of conventional generator forced outages, since they are the focus of the paper and largest reason for unexpected inability of generation to serve load. Inclusion of load, wind (the focus in [11]), solar, and net interchange forecast errors are also possible and is implicit in LOLPs derived from historical reserve forecast errors like those used by ERCOT [21], but is not addressed in this work. The expected generation margin for the upcoming day forms the basis of our day-ahead ORDC. The generation margin is determined as follows:

1. Determine expected online generators based on forecast net load (i.e., gross load net of wind and solar generation), scheduled conventional generator availability, and a merit-order of conventional generator marginal costs (Figure 4.2).

2. Create a COPT for expected online generators in each modeled hour conditional on forecast ambient temperature. Generator-level ambient temperatures are based on a geographic match against the closest weather station for which we obtain data (see Table B-2 in Appendix B.3 for assignments).

3. Convert the COPT into a probability distribution of generation margins for the associated time

period of reserve procurement using temperature data along with the temperature-dependent generator availability determined in [16] (Figure 4.3).[3]

4. Translate the cumulative probability distribution of generation margins into a ten-segment stepwise ORDC, where the probability of generation availability being less than the midpoint of each segment determines the expected LOLP on the segment. Segment-wise LOLPs are multiplied by a penalty factor applied to being unable to serve load to determine the utility (i.e., price) of reserves on the segment. We assume each of the ten segments are equivalent in length (in megawatts, MW) and the cutoff point for the final segment calculated as the point where the LOLP is ≤ 0.00001 (i.e., less than 0.001% chance that this amount of expected online generation capacity will be unavailable in the hour, see Figure 4.4).

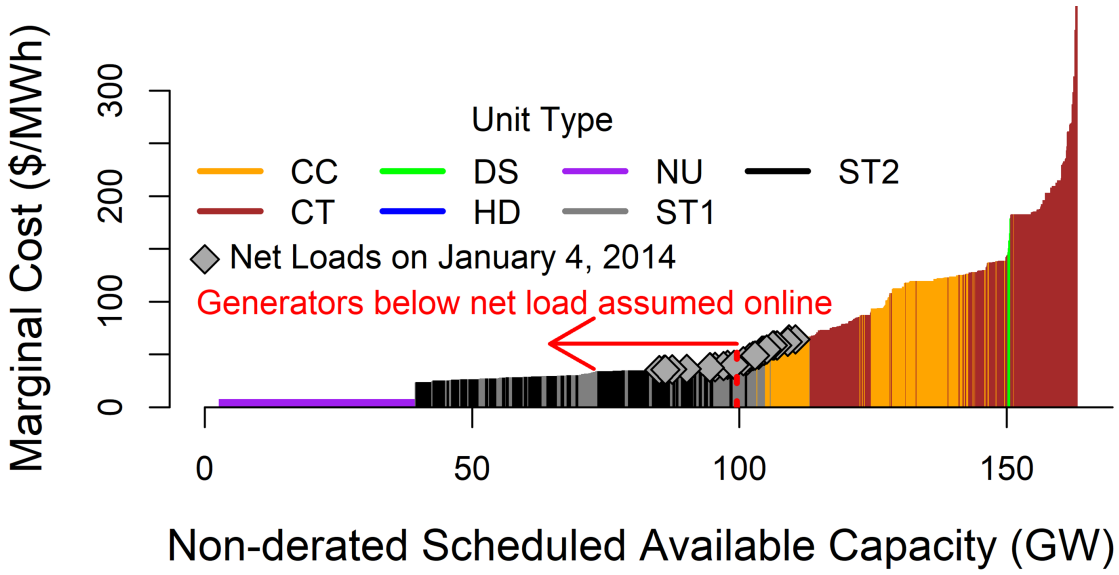

Figure 4.2: Example of step 1 of determining the generation margin: expected online generators based on marginal cost merit order. CC is combined cycle (natural gas-fired), CT is combustion turbine (natural gas-fired), ST1 and ST2 are steam turbine (coal) of two different unit sizes (ST1 $\leq$ 500MW), HD is hydroelectric, NU is nuclear, and DS is diesel.

[3]Figure 4.3 is based on hourly generator failure probabilities, timescale is adjusted to that of the relevant reserve product in PJM in the presented model cases.

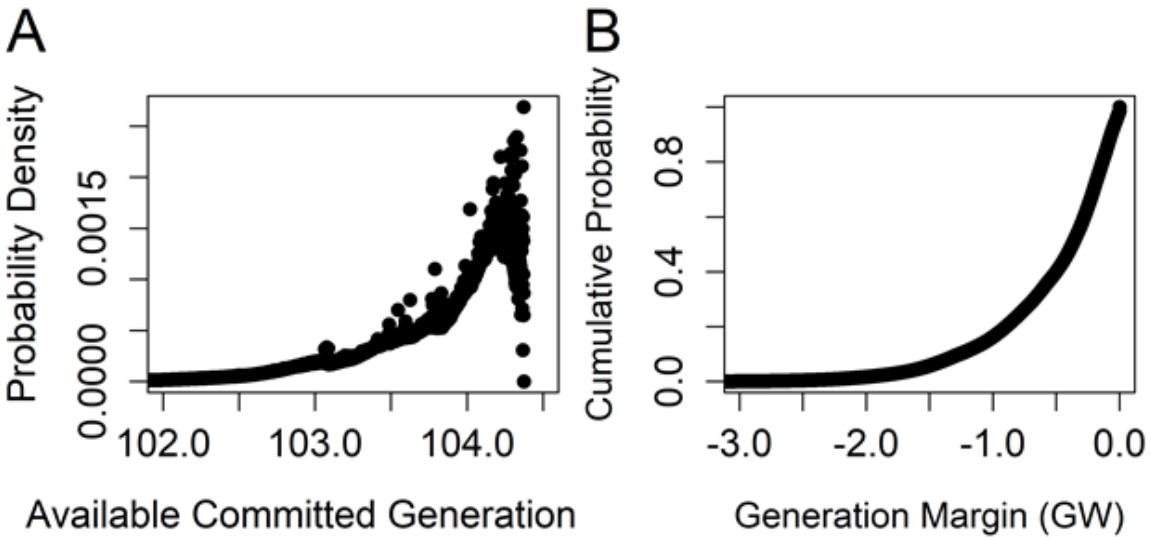

Figure 4.3: Example of steps 2-3 of determining the generation margin: Panel A shows the hourly available generation probability distribution in 1 MW increments for an example hour with 104.4 GW of net load; Panel B shows the cumulative probability that the generation margin is more than the x-axis value below net load after one hour.

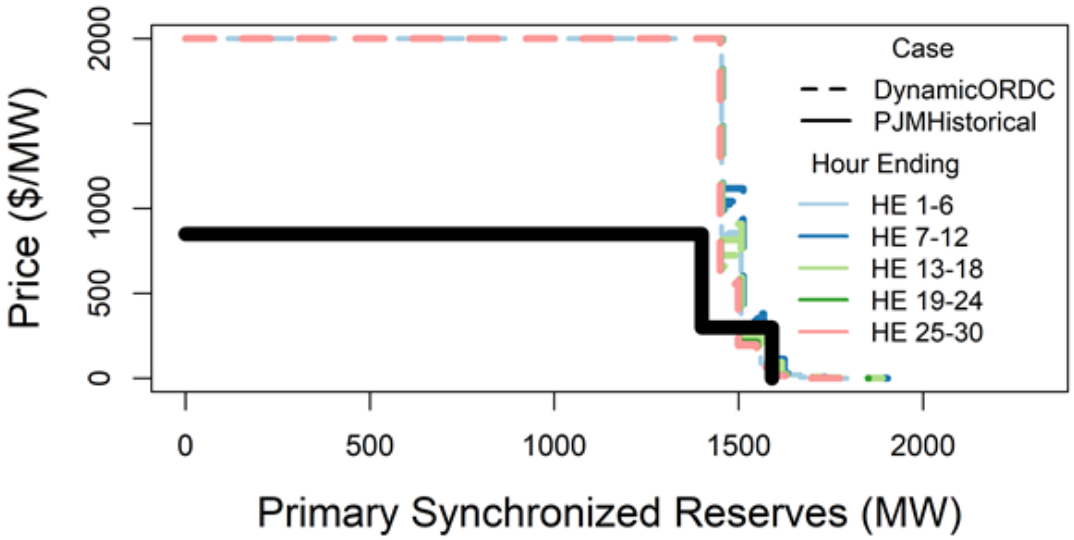

Figure 4.4: Example of step 4 of determining the generation margin: ten-segment hourly ORDCs for January 4, 2014 with 6-hour look-ahead period to avoid end-effects in generator commitments. Solid black line is current PJM practice. Description of DynamicORDC case is in Section 4.3.

The ORDC is parameterized in our electricity market model following these steps, and generation is then committed and dispatched to co-optimize energy and reserves.

4.2.2 Implementation differences from proposed PJM ORDCs

Section 4.2.1 describes our general approach to articulating a dynamic ORDC. Because of the paper's use of PJM data and PJM's recent proposal to modify its ORDC, this section highlights how our approach could be modified to work with PJM's. PJM proposed 24 season-hour block ORDCs, with each ORDC constructed from the empirical distribution of net load forecast errors in the previous three years. Net load forecast error is defined as:

$$NLE_t = (L_t - W_t - S_t - NSI_t) - (L_{t-x} - W_{t-x} - S_{t-x} - NSI_{t-x}) + \sum_{t-x}^{t} FOR_{tmp} + RR_t \quad (4.1)$$

Where $t \in$ 0,1,...23, hours beginning, $tmp \in$ timepoints, the temporal resolution at which PJM records thermal forced outages, and x=offset, the temporal offset for recording forecast error (assumed to be 1 hour in Figure 4.5). NLE is net load error, L is load, W is wind generation, S is solar generation, NSI is net scheduled interchange, FOR are forced outages of thermal units, and RR is the regulation requirement (equation modified from Tribulski et al., 2018 [18]).

Implementing our proposed dynamic ORDC modifies the $\sum_{t-x}^{t} FOR_{tmp}$ component of NLE_t. Instead of using the empirical average $\sum_{t-x}^{t} FOR_{tmp}$ in the relevant season-hour block for all hours t in the block, the forced outage component could either be drawn from the dynamic conditional distribution calculated specifically for hour t as described in Section 4.2.1, or the forced outage conditional distribution for hour t could be separately convolved with the net load error distribution excluding forced outages. An ORDC could then be articulated based on the new distribution of net load errors for hour t.

The difference between our method and PJM's for calculating the probability of generator forced outages is highlighted by Figure 4.5 for the 2014 Polar Vortex week. We compare the empirical probability distribution of hourly total unscheduled generation forced outages reported in PJM GADS for Hours Beginning 7-10 (i.e., 7AM-11AM ET) during winter months (December, January, and February) in the three preceding years (2011-2013) with the daily distributions for those same hours produced using our dynamic ORDC method. We select these hours for the figure because

they are the most important for winter reliability and are defined by PJM to be the basis for the "Winter Block 5" ORDC in their proposal to FERC. The figure shows that while more typical days (like January 10, 2014) are a good match for the empirical distribution, the conditional probability distribution of forced outages on very cold (and high load) days like January 7, 2014 is significantly rightward shifted compared to the full historical empirical distribution in the season-hour block, justifying articulation of a dynamic ORDC from the conditional distribution. As further confirmation of this shift's validity, observed hourly quantities of forced outages for 7-11AM ET during the Polar Vortex week in PJM GADS are displayed as individual ticks along Figure 4.5's horizontal axis, showing the observed rightward skew of forced outages during the peak of the Polar Vortex.

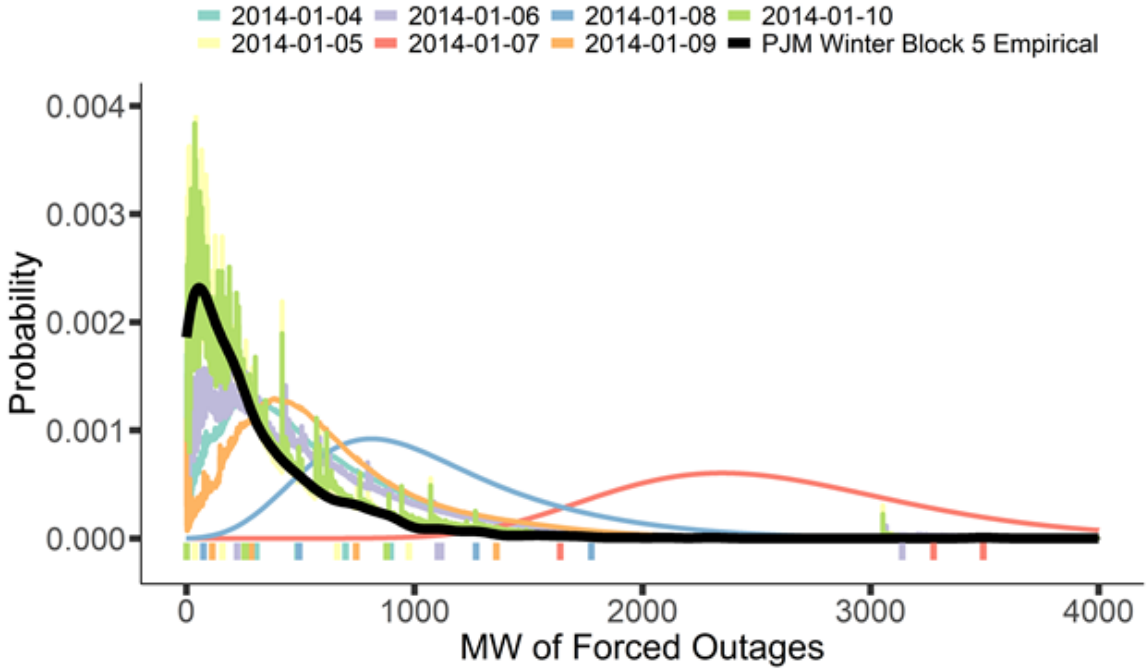

Figure 4.5: Distribution of temperature-conditional forced outage probabilities for hours beginning 7-10 during each day of the 2014 Polar Vortex week compared to empirical distribution for the three preceding winters. Ticks along bottom of the horizontal axis show the observed quantities of forced outages in these hours.

4.2.3 Electricity market model

To quantify the effects of different ORDC parameterizations we formulate a reduced-form zonal UC/ED model of PJM as a mixed-integer linear program in Pyomo, solved using IBM's CPLEX

version 12.9 solver through the Python application programming interface [22]. PJM is modeled as five zones linked by six transmission lines, as used by Lueken and Apt (2014) (Figure 4.5). To account for the effect of uncertainty in historical cases on dispatch and pricing the model is configured to run in three stages, all at hourly resolution. In the first stage all generators are committed using historical day-ahead forecast load. A second stage allows re-commitment of fast-starting generators (assumed to be CT and DS units only) using the day-of load forecast. Finally, a dispatch-only run against actual loads enables re-dispatch as a linear program after fixing commitment decisions for determination of prices. The model co-optimizes energy and reserves simultaneously at all stages, where the utility of reserves may be articulated as a stepwise dynamic ORDC as described in Section 2.1.

Three operating reserve products are included in the model: primary synchronized reserves, primary non-synchronized reserves, and day-ahead scheduling reserves (DASR), mimicking the major operating reserve products competitively procured in PJM's markets (Table 5.1).[4] We additionally model the non-Western zones as a single reserve sub-zone, mimicking PJM's Mid-Atlantic Dominion sub-zone. Market clearing prices are reported as the dual variable of the zonal load-balance constraint. Additional constraints on reserves and transmission flows yield dual variables interpretable as zonal reserve and transmission congestion prices.

The objective of the UC/ED model is to minimize the negative utility of serving firm electricity loads (i.e., maximize social welfare). The utility of held operating reserves is the penalty factor multiplied by the probability of a lost load event the held operating reserves would mitigate, as articulated by the ORDC. Production costs include starting and operating generators, equivalent to the negative utility of served load assuming a constant value associated with unserved energy. The penalty factor is assumed equal to the cost of unserved energy. The unit commitment objective

[4]PJM also has markets for regulation, which responds on shorter timescales. We do not include regulation procurement in our model.

function is:

$$MIN \sum_{t,g,z,gs} [SD_{t,g,z,gs} * GENMC_{g,gs}] + \sum_{t,g} [NLC_g * COMMIT_{t,g}] + \sum_{t,g} [SC_g * START_{t,g}] + \sum_{t,z} [UE_{t,z}] + \sum_{t,g} [TXMW_{t,l} * TXH_{t,l}] - \sum_{t,s,p} [RSP_{t,s,p} * SR_{t,s,p}] \quad (4.2)$$

where $t \in T$ is the set of hourly timepoints, $g \in G$ is the set of generators, $z \in Z$ is the set of zones used to represent the PJM balancing area, $l \in L$ is the set of transmission lines, p ∈ P are the types of operating reserves specified for PJM in Table 5.1, and gs ∈ GS are linearized segments of a generator's marginal heat rate curve. SD is the dispatch on each segment of the generator's offer curve, $GENMC$ are generator marginal costs of burning fuel (in piecewise linear segments), NLC are generator no-load costs (i.e., fuel burn associated with having the generator online regardless of its power output), $COMMIT$ is a binary variable indicating whether a generator is committed, SC are additional generator start-up costs, $START$ is a binary variable indicating a generator start, and UE is the cost of unserved energy.[5] TXH is the marginal cost associated with scheduling flows $TXMW$ between zones, and RSP is the utility of held operating reserves on each segment $s \in S$ of the ORDC, with SR indicating the reserves held on that segment. Data for the parameters of the unit commitment objective is provided in Appendix B.

Constraints in the model enforce serving forecast or actual zonal loads at each respective stage, subject to generator scheduled availability, thermal generator commitment parameters (minimum up/down time, minimum online power output, ramp rate, etc.), generator ability to provide operating reserves, and inter-zonal transmission flows (Appendix B.2). Wind and solar resources connected to the bulk power system are modeled as zero marginal cost variable resources without commitment parameters nor ability to provision reserves, though they can be curtailed. Additional constraints govern total generation, minimum and maximum hourly generation, and ramping capability of zonal hydro resources (Appendix B.3). The ability of generators to provide synchronized reserves is

[5]Equivalent to the price cap for model purposes: $2000/MWh in PJM. This is also modeled as the strike price for emergency demand response resources.

constrained based on the lesser of their ramping limit or maximum output within the timeframe of the reserve quantity, while non-synchronized reserves are limited to generators with fast-start capability (only CT and DS units are assumed to have this ability). Self-scheduled resources, modeled to include most coal and all nuclear units, are assigned zero ramp rate and thus cannot provision synchronized reserve nor set price (Appendix B.2). A full mathematical formulation of the UC/ED model is provided in Appendix A.

In the final dispatch run commitment decisions are fixed, so the model is relaxed to a linear program with no-load and startup costs excluded from the objective function. Zonal clearing prices can be determined as the dual variables of the zonal load balance constraint. To reflect PJM practice, the constraint on minimum generation level is relaxed for CTs [23], allowing these generators to set the clearing price even if dispatched at their minimum online power output.

4.2.4 Data development

We draw on numerous public data sources to realistically parameterize a reduced-form zonal representation of PJM's footprint. Zones are the same as those used in Lueken and Apt [5] and shown in Figure 4.6. Day-ahead, day-of, and actual loads are obtained from historical data and assigned to our model zones based on PJM LDA aggregation. Thermal generator offers are cost-based, with variation primarily resulting from matching gas-fired generators against the nearest gas price hub for that delivery day in the historical data. Hydro generation faces additional constraints based on historical maximum, minimum, and total generation on similar days. Nuclear and coal-fired units are assumed to self-schedule to better match the quantity of economic offers observed in anonymized PJM bid data; less flexible generation units commonly self-schedule in PJM [24]. Scheduled generator outages for historical days are obtained via PJM GADS data. Further details on matching for developing thermal generator offers, forced outage probabilities, renewable generation, loads, and other data are in Appendix B.

4.2.5 Validation

Selection of validation weeks

To validate the behavior of our model we run it with historical data for January 4-10, 2014 and October 19-25, 2017, as described below. We choose these timeframes to represent an extreme and normal week, respectively. Additional validation uses the entire month of October 2017 and is discussed in Appendix C.

The January 4-10, 2014 period included a polar vortex event that caused extreme stress on PJM's system due to increased generator outages and load.[6] Accordingly, this period represents a time of high importance for reserve procurement since these stresses may result in the operator taking high-cost actions to preserve the ability to serve firm load under contingencies.

The October 19-25, 2017 period is chosen as a recent shoulder season week to help validate the model's ability to reproduce PJM prices and dispatch under more typical system conditions. An extension of this validation week to the full month of October 2017 is included in Appendix C.

[6]During the January 2014 polar vortex event the fleetwide forced outage rate reached 22% compared to a winter average forced outage rate of 7%. In addition, PJM experienced its second-highest winter peak load of approximately 143 GW, approximately 35% higher than typical in January [25].

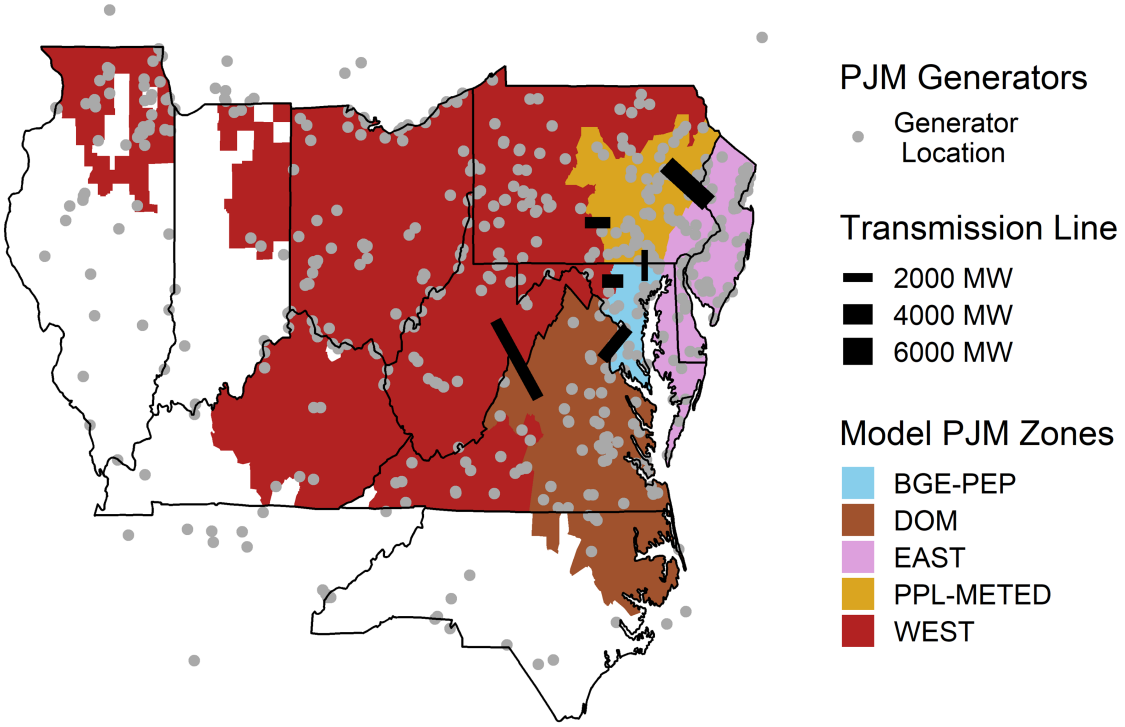

Figure 4.6: Model PJM zones, generator locations, and transmission links between zones. Generator locations include all generators for which we have data reported to PJM GADS (1995-2018Q1). Generators are assigned to the zone they are geographically located within or nearest. Transmission flow limits based on PJM data [26].

Model performance validation: generation and LMPs

Figure 4.7 illustrates total generation by fuel type dispatched by the model in order to meet electricity demand for the week of January 4-10, 2014 (A) and October 19-25, 2017 (B).

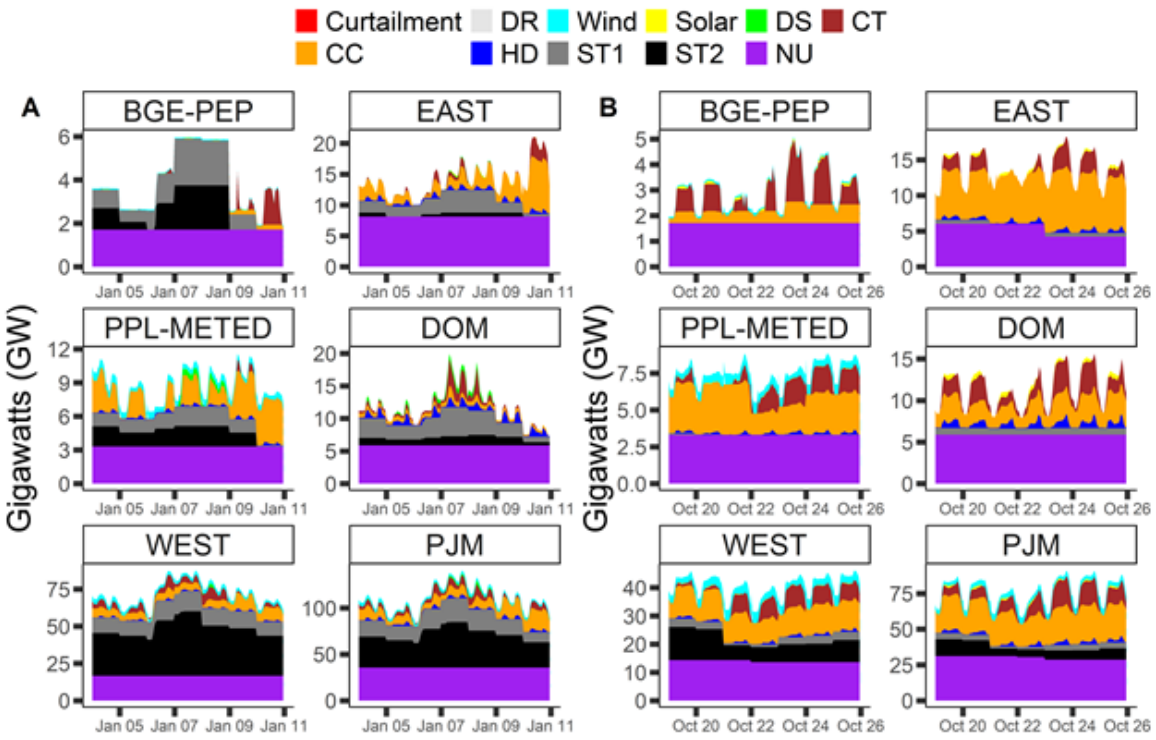

Figure 4.7: Generation by zone and fuel type for Jan. 4-10, 2014 (A) and Oct. 19-25, 2017 (B). Aggregate dispatch for the entire PJM system is given by 'PJM'. Only wind and solar resources may be curtailed, though no curtailment is observed in cases with reserves procurement. See Figure 4.2 for definitions of each fuel type.

During the high load January 2014 week the model fully dispatches nuclear and hydro resources as well as available solar and wind. The model relies heavily on coal due to high spot prices for natural gas during this week, making coal units inframarginal relative to gas. During the very highest load hours on January 7 and 8 (the peak of the 2014 polar vortex), the model requires the higher priced gas generation to serve load. In contrast, during the lower load and lower natural gas price October week, more gas resources are inframarginal.

To validate model results we compare modeled zonal LMPs with the historical day-ahead LMPs reported by PJM for these zones from PJM [27]. Modeled cases assume historical PJM practice for reserve procurement and pricing. Figure 4.8 shows the zonal LMPs associated with the above dispatch profile for January 4-10, 2014 (A) and October 19-25, 2017 (B). LMPs are the dual variables of the zonal load-balance constraints, reflecting the cost of needing 1 MWh of additional generation

to serve load in each zone. Modeled LMPs during the January week are lower than observed, particularly in the West zone during the peak of the polar vortex. Reasons for these differences may include differences in modeled vs. actual gas generator bids as well as the model's reduced-form representation of PJM's transmission topology and interchange. The model's match with reported prices is closer during the lower load Oct. 2017 week, with little difference in average hourly PJM-wide LMP ($26.3/MWh modeled vs. $27.6/MWh reported) and hourly root-mean square error under 11% (see PJMHistorical in Table 4.2). In both cases we consider the model well-validated enough to serve our primary purpose, which is estimating the magnitude of reserve quantity and price changes due to changes in formulation of ORDCs to reflect ambient conditions.

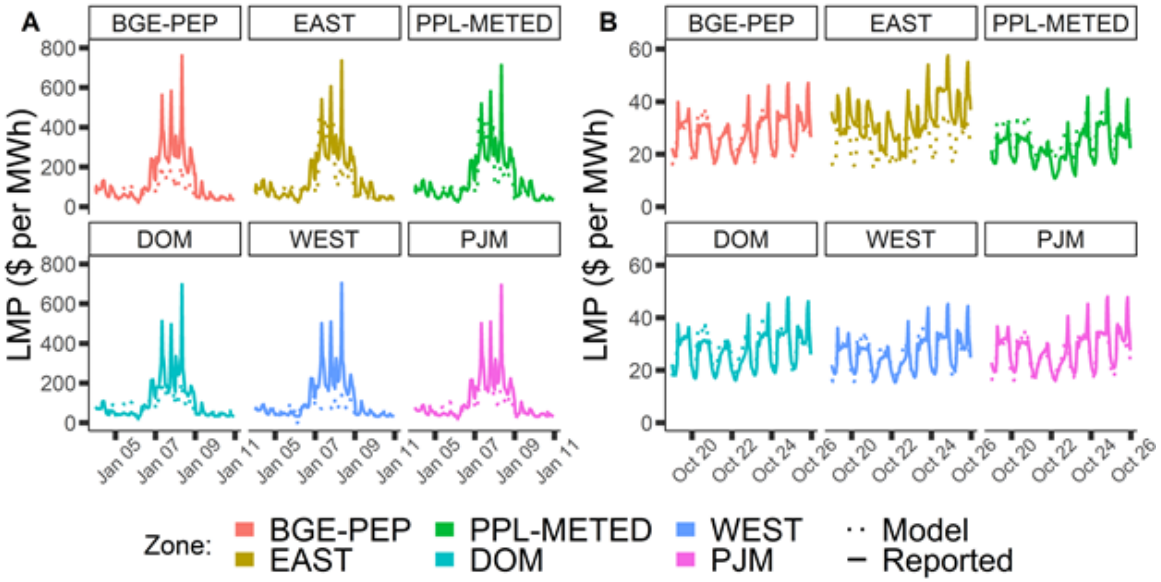

Figure 4.8: Hourly zonal LMPs for Jan. 4-10, 2014 (A) and Oct. 19-25, 2017 (B).

4.3 Results

4.3.1 Choice of model runs

To highlight the effect of a dynamic ORDC on energy and reserve procurement and pricing, we run four sets of cases for the two weeks.

1. *NoReserves*: No requirement for operating reserves. While unrealistic, given uncertainty in generation availability, this case sets a lower bound on cost for reserve procurement and

associated increases in electricity market clearing prices due to lack of reserves when actual loads deviate from forecast.

2. *PJMHistorical*: Mimics historical PJM practice prior to PJM's current proposal before FERC. This case sets the primary synchronized reserve minimum reserve requirement (MRR) to 1.4 GW and day ahead scheduling reserves (DASR, secondary reserves) MRR equivalent to 6% of forecast hourly load. The penalty factor for being below the MRR for primary reserves is \$850/MWh; an additional 190 MW step beyond the MRR at \$300/MWh penalty factor is included for primary reserves. Secondary reserves (DASR) do not have a penalty factor under current PJM practice and are cleared the highest offer price [28], in our model based on lost opportunity costs.

3. *SimpleORDC*: This case is most like PJM's proposed ORDC formulation before FERC, though with notable differences. It implements a ten-segment ORDC based on an unconditional fleet-average generator forced outage rate. The first segment of the ORDC for each reserve product sets the MRR to the current PJM MRR (e.g., 1.4 GW synchronized primary reserves); subsequent ORDC segments are based on look-ahead time period (10-30 minutes) generation margins calculated as described in Section 2 assuming unconditional forced outage rates. The maximum penalty factor for reserve shortage is assumed to be \$2000/MWh. Notable differences with actual PJM proposal are that this case only accounts for forced outages (PJM additionally considers wind, solar, load, and net interchange forecast uncertainty) and is based on historical unconditional forced outage rates without empirical distributions subset by the 24 proposed hour-season blocks used by PJM.

4. *DynamicORDC*: Implements a dynamic ten-segment ORDC that considers changes in generator failure probabilities driven by ambient conditions. The first segment of the ORDC for each reserve product sets the MRR to the current PJM MRR (e.g., 1.4 GW synchronized primary reserves); subsequent ORDC segments are based on look-ahead time period (10-30 minutes) generation margins calculated as described in Section 2 assuming conditional forced outage rates. The maximum penalty factor for reserve shortage is assumed to be \$2000/MWh. The

major difference between this case and SimpleORDC is in considering dynamic generator failure probabilities, which differ most from invariant failure probabilities during extreme temperature hours (see Figure 4.1).

Additionally, because procurement of operating reserves can increase prices, particularly during conditions of significant system-wide shortage, we compare the modeled payments for operating reserves to those for congestion, another payment that increases during shortage conditions as transmission lines are heavily loaded. While we model only six transmission interfaces, these interfaces drive a meaningful portion of PJM's annual congestion costs, with the AEP-DOM, 5004/5005, Bedington-Black Oak, and AP South interfaces all ranking among the top 10 on PJM's transmission system during January-September 2018 and accounting for 18% of all congestion costs [29].

4.3.2 Effects of using a dynamic ORDC

The hourly zonal prices for both modeled weeks for the four cases described above are shown in Table 4.2. Cases differ only in their reserve procurement strategies, with changes in prices resulting from both how reserve procurement affects ability to handle differences between forecast and actual load as well as the cost of reserve procurement itself. The cost of additional reserve procurement, when justified by higher ex-ante generator failure probabilities articulated via the reformulated ORDC, has attendant benefits that are not revealed in dispatch prices because we do not model observed unit-level failures between commitment and dispatch in historical data. To highlight the expected value of these benefits, Figure 4.11 shows expected social welfare benefits associated with changes to reserve procurement costs and loss-of-load probability between the cases with reserve procurement.

Price differences resulting from actual loads exceeding the day-ahead and day-of load forecasts used in commitment decisions are most pronounced in the January NoReserves case, where only emergency demand response resources are available to balance load forecast errors that cannot be served by online generators. When this occurs the market clears at the price cap of $2000/MWh, demonstrating the importance of operating reserves as a hedge against uncertainty. A similar effect

of lower magnitude occurs in the PJMHistorical case, where a reserve shortage condition rather than emergency demand response sets price below the cap. The two ORDC cases that hold variable quantities of operating reserves result in lower prices, as they are better able to cover deviations between forecast and actual loads with additional reserves.

While the SimpleORDC and DynamicORDC cases produce similar average prices during both modeled weeks, the DynamicORDC case holds additional quantities of reserves on high load, low temperature days like January 7-8, 2014. Holding additional reserves results in higher prices that reflect the need for higher marginal cost generators, both to serve load and to provision sufficient reserves to mitigate the increased probabilities of generator forced outages associated with extreme conditions (see also Figure 4.10). The expected benefits of this procurement are highlighted in Figure 4.11, which shows the additional reserve procurement during the Polar Vortex reduces the probability of generator forced outages resulting in a reserve shortage, more than offsetting the cost of the additional reserve procurement.

Table 4.2: Pricing summary statistics for modeled cases. Root mean square errors are a measure of how accurately each case reproduces day-ahead PJM prices.

	Reported	**No Reserves**	**PJMHistorical**	**SimpleORDC**	**DynamicORDC**
January 4-10, 2014 LMP ($/MWh)					
Mean	109.1	104	93.7	83.1	83.2
Standard Deviation	102.3	241	91.8	43.6	43.2
Median	61.5	69.9	71.7	72.1	72.8
Min	31.6	-1260	0.5	0.5	0.5
Max	694	2000	1030	194	193
January 4-10, 2014 Root Mean Square Error compared to Reported					
Mean ($/MWh)		65.5	41.7	42.8	42.5
% of Reported Mean		60.0%	38.3%	39.2%	39.0%
October 19-25, 2017 LMP ($/MWh)					
Mean	27.6	26.0	26.3	26.4	26.3
Standard Deviation	6.8	5.9	5.9	6.0	6.0
Median	28.3	27.0	28.2	28.3	28.0
Min	16	15.1	15.3	15.1	15.1
Max	48	35.3	34.6	34.6	35.2
October 19-25, 2017 Root Mean Square Error compared to Reported					
Mean ($/MWh)		2.9	2.8	2.8	2.8
% of Reported Mean		10.5%	10.1%	10.1%	10.1%

To give context for the magnitude of the additional reserve payments associated with imple-

mentation of an ORDC, we compare the total weekly primary synchronized reserve payments due to ORDC penalty factors in the four cases with zonal congestion. We show these results for the high load, high cost January week in Figure 4.9 (for the October week, see Appendix C), reporting only the primary synchronized reserve penalty factor costs, as primary synchronized reserves are the most binding and highest priced reserve product procured in the model. During an extreme week like January 4-10, 2014, the additional payments associated with primary synchronized reserve penalty factors are considerably less than congestion costs (though congestion costs are also well above average during an extreme week).

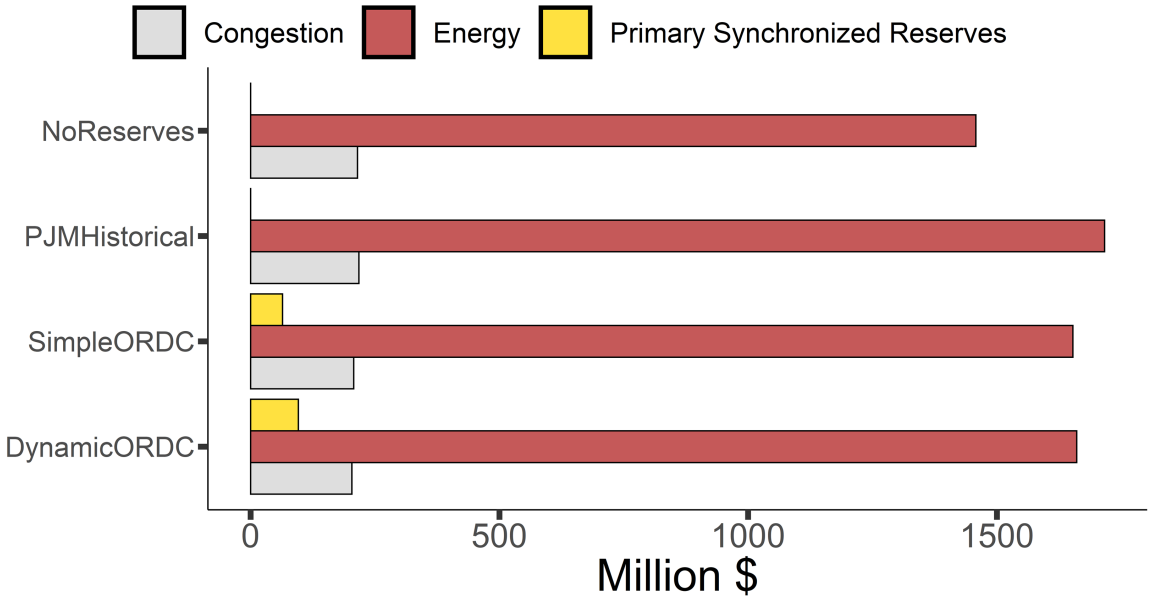

Figure 4.9: Total congestion payments, total energy payments, and ORDC-related primary synchronized reserve payments for the modeled January 2014 week. NoReserves and PJMHistorical cases do not have an ORDC and therefore have no ORDC-related reserve payments, though they may still have non-zero reserve prices due to lost opportunity costs.

To give a sense of how the ORDC affects price formation, Figure 4.10 shows the hourly marginal cost (penalty factor) on the ORDC of primary synchronized reserves, as well as the total quantity of reserves procured, for the January 2014 week across the four cases. The NoReserves case has no reserve requirements, and therefore procures no reserves with $0 penalty factor. In the PJMHistorical case, the 1.4 GW MRR and 190 MW extended step for primary synchronized reserves are fully procured during commitment runs. However, non-zero reserve penalty factors may occur in dispatch

when actual load exceeding forecast requires re-dispatch of synchronized reserves to serve the additional load resulting in a reserve shortage. This occurs with some frequency during the Polar Vortex week. In the two cases with ORDCs, reserves have value to the system that depends on the price of the ORDC segment, and procurement and the associated penalty factor can vary accordingly to maximize social welfare. The extreme cold temperatures during January 4-10, 2014 highlight the difference between a simple ORDC and the dynamic ORDC: in the dynamic case, increasing reserve procurement is required during cold hours to achieve the same level of reliability due to increased generator failure probabilities, while the simple ORDC reserve requirements change only with the changing magnitude of forecast load. This process plays out during the highest load day of the polar vortex (Jan. 7), when the DynamicORDC case holds higher levels of reserves. Alternatively, both cases may accept lower levels of reliability when the cost of bringing additional generators online is high, particularly if the peak load event necessitating committing additional generators and reserves is of short duration. Again, this effect is exacerbated in the DynamicORDC case as more reserves are required for equivalent reliability given cold temperatures, resulting in greater price spikes for shorter duration peak loads on Jan. 8 than in the SimpleORDC case. The high reserve penalty factor in all three reserve-procuring cases on Jan. 6 results from actual load exceeding day-of forecast.

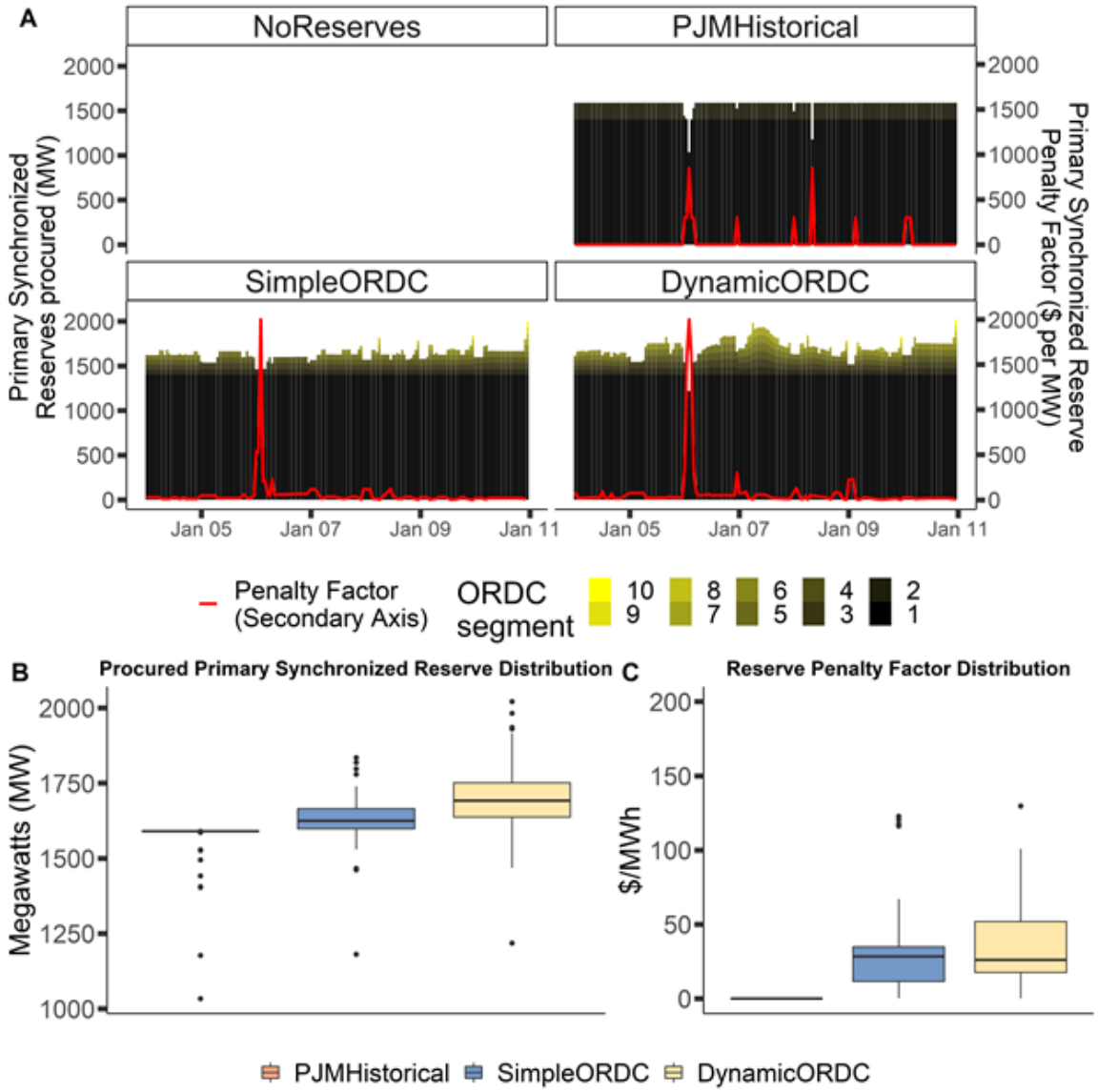

Figure 4.10: (A) Quantity of primary synchronized reserves procured by ORDC segment (shaded vertical line segments; left axis, in MW) by hour for Jan. 4-10, 2014, along with ORDC penalty factor for primary synchronized reserves (red line; right axis, in $/MW). (B) and (C) show the distributions of reserve procurement and penalty factors for the week for the three reserve-procuring cases.

To quantify the benefits of a dynamic ORDC, Figure 4.11 compares the hourly probability of reserve shortage and relative social welfare of the three cases with reserve procurement. Comparisons in Figure 4.11 assume the DynamicORDC method perfectly parameterizes generator failure probabilities and that its other assumptions (e.g., use of a MRR to define reserve shortage condition)

are justified, though the DynamicORDC case itself is not perfect due to deviations in actual from forecast load used in commitment decisions. An example of a large under-forecast of actual load occurs on Jan. 6 as the polar vortex begins, resulting in high probabilities of reserve shortage in all cases. We do not include the NoReserves case in this comparison as procuring no reserves results in a reserve shortage every time a generator fails or load is under forecast, making it both unrealistic and highly suboptimal.

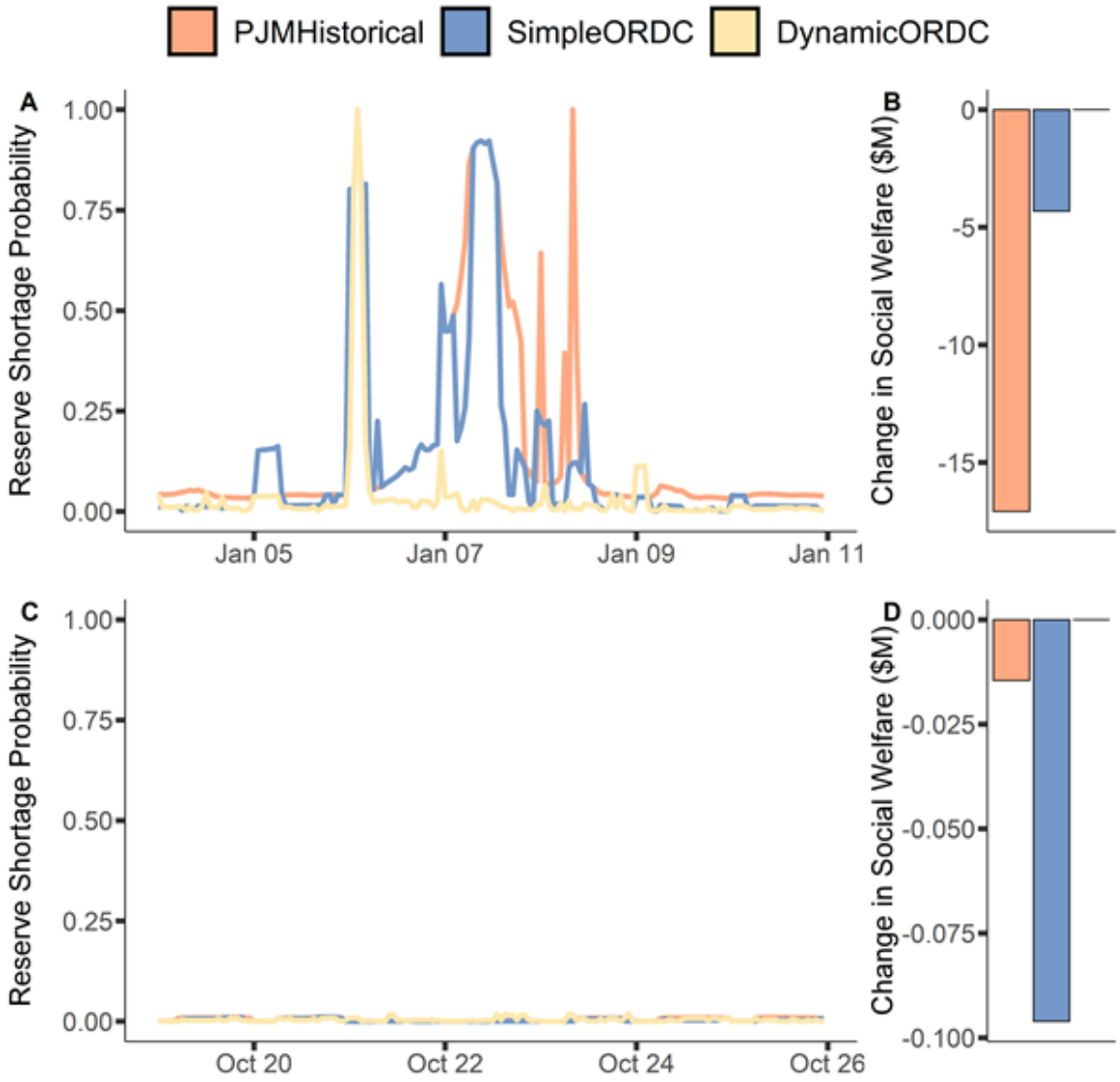

Figure 4.11: Hourly reserve shortage probability (A,C) and decrease in weekly social welfare of different reserve procurement strategies compared to the DynamicORDC case (B,D) during January 4-10, 2014 and Oct 19-25, 2017 weeks, respectively. Note the different y-axis scales on weekly change in social welfare plots (B,D).

Compared to assuming the DynamicORDC as best practice, the PJMHistorical and SimpleORDC reserve procurement approaches are suboptimal by $17.1 million and $4.3 million, respectively, during the January 2014 week (Figure 4.11B). The comparative suboptimality of these reserves procurement strategies is primarily due to their not accounting for increased failure probabilities of PJM's conventional generator fleet under abnormally low temperatures, resulting in higher probability of a

reserve shortage condition. As shown in Figure 4.11A, the risk of experiencing a reserve shortage during the peak of the 2014 polar vortex on Jan. 7 using both PJMHistorical and SimpleORDC reserve procurement exceeds that of the DynamicORDC approach by over 90%. In contrast, during the October 2017 week, both PJMHistorical and SimpleORDC adequately procure reserves (Figure 4.11C) because generator failure probabilities are generally below that of the unconditional annual average due to the mild ambient conditions; in fact, slight suboptimality (Figure 4.11D) results from paying for over-procurement of reserves relative to quantities in the DynamicORDC case.

Recalling generators also benefit from the increased reserve prices shown in Figure 4.9, the dynamic ORDC case shows benefits for both generators and consumers during a polar vortex week. Both increases in generator payments and social welfare are on the order of millions to low tens of millions of dollars, about 1% of cleared energy payments during the week. Realized benefits for consumers may be even larger than estimated if the assumed \$2000/MWh reserve shortage penalty factor underestimates the cost of reserve shortage actions taken by the system operator or additional sources of uncertainty beyond load forecast are incorporated.

4.4 Conclusions and Policy Implications

Building on recent research demonstrating that generator reliability depends on temperature, we propose a dynamic formulation of an ORDC to implement scarcity pricing in a wholesale electricity market. Our model co-optimizes the procurement of energy and reserves to meet electricity demand, subject to uncertainty in the next day's availability of dispatchable generation, variable renewables, and realized load. We validate our model's price formation during two historical weeks, comparing its performance to three alternative approaches for procuring operating reserves, two of which mimic historical and proposed PJM practice.

We find that a dynamic ORDC increases reserve procurement and prices during hours with heightened risk of generator forced outages, reducing the probability of a reserve shortage during the 2014 polar vortex. More accurate quantification of generator loss-of-load probabilities leads to a more accurate operating reserve valuation, increasing social welfare by \$17.1 million and \$4.3

million compared to historical practice and an ORDC with unconditional generator force outages, respectively, during a historical cold, high load week (Figure 4.11). Neither PJM's historical practice nor its current ORDC proposal before FERC capture hourly variation in failure probabilities of its conventional generator fleet. Our approach offers a method for incorporating the effects of ambient temperature on this relationship in PJM and could be applied in other wholesale electricity markets. In the specific case of PJM, the dynamic ORDC approach could be straightforwardly incorporated with proposed practice by convolving the conditional distribution of generator forced outages with other sources of net load error underlying ORDC articulation at each timepoint.

A practical consideration in these proposals not included in our model's objective is balancing accuracy with complexity; partially for this reason PJM proposes 24 ORDCs as six four-hour blocks across four seasons rather than articulating an ORDC from conditional distributions at each timepoint. However, it is not always hot during midday summer hours nor cold on winter mornings; it is the subset of extreme hours during those seasons that pose risk to the system and are associated with higher generator forced outage rates (see Figure 4.1 and Figure 4.6). A dynamic ORDC is well-approximated by the simpler approach under seasonal expected conditions when conditional and unconditional generator forced outage rates are similar, as evidenced by results in Figure 4.11C-D. However, a dynamic ORDC also captures demand for additional operating reserves under extreme conditions like those of the January 2014 polar vortex without the need for discretionary operator intervention, creating benefits for both consumers from reduced probability of reserve shortages and generators through increased co-optimized reserve and energy payments. A dynamic ORDC may therefore offer a market-based method for extending reserve procurement under extreme conditions, enhancing the benefits an ORDC would bring to PJM.

References

[1] W. Hogan, "Competitive Electricity Market Design: A Wholesale Primer," *Mimeo*, pp. 1–59, 1998. papers://c89c297f-464b-44b4-9d34-c7682ac4a388/Paper/p361.

[2] J. K. Szinai, C. J. Sheppard, N. Abhyankar, and A. R. Gopal, "Reduced grid operating costs and renewable energy curtailment with electric vehicle charge management," *Energy Policy*, vol. 136, p. 111051, 2020. doi:https://doi.org/10.1016/j.enpol.2019.111051.

[3] M. Craig, O. J. Guerra, C. Brancucci, K. A. Pambour, and B. M. Hodge, "Valuing intra-day coordination of electric power and natural gas system operations," *Energy Policy*, vol. 141, December 2019, p. 111470, 2020. doi:https://doi.org/10.1016/j.enpol.2020.111470.

[4] D. C. Steinberg, D. A. Bielen, and A. Townsend, "Evaluating the CO2 emissions reduction potential and cost of power sector re-dispatch," *Energy Policy*, vol. 112, September 2017, pp. 34–44, 2018. doi:https://doi.org/10.1016/j.enpol.2017.10.003.

[5] R. Lueken and J. Apt, "The effects of bulk electricity storage on the PJM market," *Energy Systems*, vol. 5, no. 4, pp. 677–704, 2014. doi:https://doi.org/10.1007/s12667-014-0123-7.

[6] M. T. Craig, P. Jaramillo, and B. M. Hodge, "Carbon dioxide emissions effects of grid-scale electricity storage in a decarbonizing power system," *Environmental Research Letters*, vol. 13, no. 1, 2018. doi:https://doi.org/10.1088/1748-9326/aa9a78.

[7] V. Stefanowicz, "Real-Time Reserves," 2018. https://www.pjm.com/-/media/committees-groups/committees/oc/20180501/20180501-item-33-real-time-reserves.ashx.

[8] L. T. Anstine, R. E. Burke, J. E. Casey, R. Holgate, R. S. John, and H. G. Stewart, "Application of Probability Methods to the Determination of Spinning Reserve Requirements for the Pennsylvania-New Jersey-Maryland Interconnection," *IEEE Transactions on Power Apparatus and Systems*, vol. 82, no. 68, pp. 726–735, 1963. doi:https://doi.org/10.1109/TPAS.1963.291390.

[9] A. Papavasiliou and S. S. Oren, "Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network," *Operations Research*, vol. 61, no. 3, pp. 578–592, 2013. doi:https://doi.org/10.1287/opre.2013.1174.

[10] M. A. Ortega-Vazquez and D. S. Kirschen, "Estimating the spinning reserve requirements in systems with significant wind power generation penetration," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 114–124, 2009. doi:https://doi.org/10.1109/TPWRS.2008.2004745.

[11] Z. Zhou and A. Botterud, "Dynamic scheduling of operating reserves in co-optimized electricity markets with wind power," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 160–171, 2014. doi:https://doi.org/10.1109/TPWRS.2013.2281504.

[12] F. Bouffard and F. D. Galiana, "An Electricity Market With a Probabilistic Spinning Reserve Criterion," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 300–307, 2004. doi:https://doi.org/10.1109/TPWRS.2003.818587.

[13] PJM, "2018 PJM Reserve Requirement Study," 2018. https://www.pjm.com/-/media/committees-groups/subcommittees/raas/20181004/20181004-pjm-reserve-requirement-study-draft-2018.ashx.

[14] W. W. Hogan, "Electricity scarcity pricing through operating reserves," *Economics of Energy and Environmental Policy*, vol. 2, no. 2, pp. 65–86, 2013. doi:https://doi.org/10.5547/2160-5890.2.2.4.

[15] S. Murphy, J. Apt, J. Moura, and F. Sowell, "Resource adequacy risks to the bulk power system in North America," *Applied Energy*, vol. 212, no. June 2017, pp. 1360–1376, 2018. doi:https://doi.org/10.1016/j.apenergy.2017.12.097.

[16] S. Murphy, F. Sowell, and J. Apt, "A time-dependent model of generator failures and recoveries captures correlated events and quantifies temperature dependence," *Applied Energy*, vol. 253, no. July, p. 113513, 2019. doi:https://doi.org/10.1016/j.apenergy.2019.113513.

[17] S. Murphy, L. Lavin, and J. Apt, "Resource adequacy implications of temperature-dependent electric generator availability," *Applied Energy*, vol. 262, p. 114424, 2020. doi:https://doi.org/10.1016/j.apenergy.2019.114424.

[18] J. Tribulski, P. M. Flynn, and R. J. Collins, "Enhanced Price Formation in Reserve Markets of PJM Interconnection, L.L.C.," 2018. urlhttps://pjm.com/directory/etariff/FercDockets/4036/20190329-el19-58-000.pdf.

[19] PJM Energy Price Formation Senior Task Force, "Price formation," 2018. https://www.pjm.com/-/media/committees-groups/task-forces/epfstf/20181214/20181214-item-04-price-formation-paper.ashx.

[20] R. Surendran, "Scarcity Pricing in ERCOT," 2014. https://www.ferc.gov/CalendarFiles/20160629114652-3-FERC2016_ScarcityPricing_ERCOT_ResmiSurendran.pdf.

[21] W. W. Hogan, "Back Cast of Interim Solution B + to Improve Real-Time Scarcity Pricing Whitepaper," 2013. http://www.ercot.com/content/gridinfo/resource/2013/mktanalysis/White_Paper_BackCastofInterimSolutionB+toImproveRe.pdf.

[22] IBM, "No Title," 2019. www.cplex.com.

[23] A. Giacomoni and A. Long, "Exploring the Impacts of Price Formation Enhancements in PJM's Wholesale Energy Markets PJM," 2019. https://www.ferc.gov/CalendarFiles/20190625100414-2-PJMFERCTechnicalConferenceJune2019FINAL.pdf.

[24] S. Shparber, N. Mullins, and S. Llp, "Customer Focused and Clean Power Markets for the Future," Tech. Rep. November, 2018. https://windsolaralliance.org/wp-content/uploads/2018/11/WSA_Market_Reform_report_online.pdf.

[25] PJM Interconnection, "Analysis of Operational Events and Market Impacts," January, 2014. https://www.pjm.com/~/media/library/reports-notices/weather-related/20140509-analysis-of-operational-events-and-market-impacts-during-the-jan-2014-cold-weather-events.ashx.

[26] PJM, "Data Miner 2 Day Ahead Interface Flows and Limits," 2019. https://dataminer2.pjm.com/feed/da_interface_flows_and_limits/definition.

[27] PJM, "Data Miner 2," 2019. https://www.pjm.com/markets-andoperations/etools/data-miner-2.aspx.

[28] L. Morelli, "Day-Ahead Scheduling Reserve Operating Reserve Demand Curve," 2018. https://www.pjm.com/-/media/committees-groups/task-forces/epfstf/20180608/20180608-item-03c-day-ahead-scheduling-reserve-operating-reserve-demand-curve.ashx.

[29] Monitoring Analytics, "State of the Market Report for PJM," 2017. https://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2017.shtml.

[30] EPA, "EPA NEEDS Database v6," 2019. https://www.epa.gov/airmarkets/national-electric-energy-data-system-needs-v6.

[31] EPA, "EMC: Continuous Emission Monitoring Systems," 2019. https://www.epa.gov/emc/emc-continuous-emission-monitoring-systems.

[32] PJM, "PJM Manual 11: Scheduling Operations," 2010. http://www.pjm.com/~/media/training/nerc-certifications/m11v45-scheduling-operations.ashx.

[33] P. Deaver and R. Macdonald, "Updating Thermal Power Plant Efficiency Measures and Operational Characteristics for Production Cost Modeling California Energy Commission California Energy Commission," January, 2019. https://ww2.energy.ca.gov/2019publications/CEC-200-2019-001/CEC-200-2019-001.pdf.

[34] U. Energy Information Administration, "Form 860, final 2018 data," 2019. https://www.eia.gov/electricity/data/eia860/.

[35] Dominion, "Dominion Energy Power Purchase Agreements," 2018. https://www.dominionenergy.com/library/domcom/media/about-us/making-energy/renewable-generation/va-nc-renewables-map.pdf?la=en&modified=20191003201334.

[36] U.S. Energy Information Administration, "Form 923, final 2017 data," 2018. https://www.eia.gov/electricity/data/eia923/.

[37] NREL, "Annual Technology Baseline," 2018. https://atb.nrel.gov/.

A Model Formulation Appendix

A.1 Sets

$t \in T$, timepoints (hourly in current model)

$g \in G$, conventional generators

$h \in H \subseteq G$, subset of hydroelectric generators

$gsz \in GSZ \subseteq G$, subset of generators in constrained sub-zones

$s \in S$, segments of operating reserve demand curve (10 in current model)

$z \in Z$, zones in power system (5 zones used to represent PJM based on public data [26])

$sz \in SZ \subseteq Z$, sub-zones with additional reserve constraints in model (all non-West zones in PJM based on PJM's Mid-Atlantic Dominion sub-zone definition)

$l \in L$, transmission lines connecting zones in power system (6 lines interconnecting 5 zones in current model)

$gs \in GS$, linearized segments of generator heat rate curves (4 segments for thermal generators corresponding to 25% increments of power output in current model)

A.2 Parameters

$LOAD_{t,z}$, hourly zonal gross load; day-ahead forecast in first stage, day-of forecast in second stage, and actual in final (dispatch) stage

$WCAP_z$, installed capacity of utility-scale wind in zone

$SCAP_z$, installed capacity of utility-scale solar in zone

$WCF_{t,z}$, hourly maximum capacity factor of utility-scale wind in zone

$SCF_{t,z}$, hourly maximum capacity factor of utility-scale solar in zone

$MINHYD_{t,z}$, hourly minimum quantity of zonal hydroelectric generation

$MAXHYD_{t,z}$, hourly maximum quantity of zonal hydroelectric generation

$RHYD_{t,z}$, hourly ramping capability (up or down) of zonal hydroelectric generation

$TOTHYD_z$, total zonal hydroelectric generation in modeled time period

$PMIN_g$, minimum power output of generator if online and providing power

SC_g, startup cost of generator

$CSPIN_g$, boolean (0/1) of whether generator can provide synchronized reserves

$CNSPIN_g$, boolean (0/1) of whether generator can provide non-synchronized reserves (i.e., has fast-start capability)

$MINUP_g$, minimum online time of generator once started (integer)

$MINDOWN_g$, minimum offline time of generator if shut down (integer)

NLC_g, no-load cost of online generator

$SA_{t,g}$, fraction of generator scheduled to be available in hour to provide power (accounts for previously scheduled generator outages)

$CINIT_{t=1,g}$, initialization of commitment boolean for generators

$UPINIT_{t=1,g}$, initialization of how long generator has been online (if committed)

$DOWNINIT_{t=1,g}$, initialization of how long generator has been offline (if not committed)

$C_{g,z}$, generator capacity, with assignment to its zone

$R_{g,z}$, hourly ramp rate of generator (assumed to be equal in up and down directions)

$RSUL_{g,z}$, maximum power output a generator can achieve in an hour where it starts up

$RSDL_{g,z}$, minimum power output from which a generator can shut down in an hour

$PSRSMW_{t,s}$, length (in MW) of operating reserve segment for primary synchronized reserves

$PNSRSMW_{t,s}$, length (in MW) of operating reserve segment for primary non-synchronized reserves

$SECRSMW_{t,s}$, length (in MW) of operating reserve segment for secondary reserves

$RSP_{t,s}$, price of operating reserves on segment

$PSCALAR$, ratio of time horizon for holding primary reserves to model temporal resolution in commitment and dispatch

$SSCALAR$, ratio of time horizon for holding secondary reserves to model temporal resolution in commitment and dispatch

$RSCALAR$, ratio of time horizon between secondary and primary reserves

$TXFROM_{t,l}$, zone from which transmission line originates

$TXTO_{t,l}$, zone to which transmission line goes

$TXTOCAPACITY_{t,l}$, hourly maximum capacity of transmission line flowing to zone

$TXFROMCAPACITY_{t,l}$, hourly maximum capacity of transmission line flowing from zone

$TXH_{t,l}$, incremental cost of scheduling flows on transmission line

GSL_{gs}, fraction of generator capacity on marginal cost segment

$GENMC_{g,gs}$, marginal cost of generation on each generator segment for each generator

$UE_{t,z}$, cost of unserved energy

A.3 Decision Variables

$wgen_{t,z}$, hourly wind output used to serve zonal load

$sgen_{t,z}$, hourly solar output used to serve zonal load

$curtail_{t,z}$, hourly curtailment of wind and solar in zone

$d_{t,g,z}$, hourly generator dispatch with its assigned zone

$sd_{t,g,z,gs}$, hourly generator dispatch on each segment of its marginal cost curve

$psr_{t,g}$, hourly primary synchronized reserves provided by generator

$pnsr_{t,g}$, hourly primary non-synchronized reserves provided by generator

$secr_{t,g}$, hourly secondary reserves provided by generator

$pssr_{t,s}$, hourly reserves provided on each segment of primary synchronized operating reserve demand curve

$pszsr_{t,s}$, hourly reserves provided on each segment of sub-zonal primary synchronized operating reserve demand curve

$pnssr_{t,s}$, hourly reserves provided on each segment of primary non-synchronized operating reserve demand curve

$secsr_{t,s}$, hourly reserves provided on each segment of secondary operating reserve demand curve

$txmw_{t,l}$, hourly power flow on transmission line

A.4 Integer Decision Variables (commitment runs only)

$commit_{t,g}$, unit commitment of generator (binary)

$start_{t,g}$, startup decision of generator (binary)

$shut_{t,g}$, shutdown decision of generator (binary)

A.5 Constraints

$$WCAP_z * WCF_{t,z} \geq wgen_{t,z} \tag{A.1}$$

$$SCAP_z * SCF_{t,z} \geq sgen_{t,z} \tag{A.2}$$

$$curtail_{t,z} == SCAP_z * SCF_{t,z} - sgen_{t,z} + WCAP_z * WCF_{t,z} - wgen_{t,z} \tag{A.3}$$

$$\sum_{h}^{H}\sum_{t}^{T} d_{t,h,z} + \sum_{h}^{H}\sum_{t}^{T} psr_{t,h} == TOTHYD_z \tag{A.4}$$

$$MAXHYD_{t,z} \geq \sum_{h}^{H} d_{t,h,z} \tag{A.5}$$

$$\sum_{h}^{H} d_{t,h,z} \geq MINHYD_{t,z} \tag{A.6}$$

$$\sum_{h}^{H} d_{t-1,h,z} + RHYD_{t,z} \geq \sum_{h}^{H} d_{t,h,z}, \forall t \geq 2 \tag{A.7}$$

$$\sum_{h}^{H} d_{t,h,z} \geq \sum_{h}^{H} d_{t-1,h,z} - RHYD_{t,z}, \forall t \geq 2 \tag{A.8}$$

$$TXTOCAPACITY_{t,l} \geq txmw_{t,l} \tag{A.9}$$

$$TXFROMCAPACITY_{t,l} \leq txmw_{t,l} \tag{A.10}$$

$$C_{g,z} * commit_{t,g} * SA_{t,g} \geq d_{t,g,z} \tag{A.11}$$

$$d_{t,g,z} \geq C_{g,z} * commit_{t,g} * SA_{t,g} * PMIN_g \tag{A.12}$$

$$d_{t,g,z} \geq d_{t-1,g,z} - R_{g,z} * commit_{t-1,g} + shut_{t,g} * RSDL_{g,z}, \forall t \geq 2 \tag{A.13}$$

$$d_{t-1,g,z} + R_{g,z} * commit_{t-1,g} + start_{t,g} * RSUL_{g,z} \geq d_{t,g,z}, \forall t \geq 2 \tag{A.14}$$

$$commit_{t,g} - CINIT_{0,g} == start_{t,g} - shut_{t,g}, t == 1 \tag{A.15}$$

$$1 - CINIT_{0,g} \geq start_{t,g}, t == 1 \tag{A.16}$$

$$1 - commit_{t-1,g} \geq start_{t,g}, \forall t \geq 2 \tag{A.17}$$

$$1 - CINIT_{0,g} \geq shut_{t,g}, t == 1 \tag{A.18}$$

$$1 - commit_{t-1,g} \geq shut_{t,g}, \forall t \geq 2 \tag{A.19}$$

$$commit_{t,g} - commit_{t-1,g} == start_{t,g} - shut_{t,g} \forall t \geq 2 \tag{A.20}$$

$$SA_{t,g} == commit_{t,g}, \forall SA_{t,g} == 0 \tag{A.21}$$

$$\begin{gathered} commit_{t,g} \geq CINIT_g \\ \forall MINUP_g \geq t + UPINIT_{0,g}, \forall MINUP_g > t \geq 1, \forall SA_{t,g} \neq 0 \end{gathered} \tag{A.22}$$

$$commit_{t,g} \geq \sum_{t-MINDOWN_g}^{t-1} start_{t,g} \quad \forall t \geq MINUP_g \tag{A.23}$$

$$CINIT_g \geq commit_{t,g} \quad \forall MINDOWN_g \geq t + DOWNINIT_{0,g}, \forall MINDOWN_g > t \geq 1 \tag{A.24}$$

$$1 - \sum_{t-MINDOWN_g}^{t-1} shut_{t,g} \geq commit_{t,g}, \forall t \geq MINDOWN_g \tag{A.25}$$

$$GSL_{g,s} * C_{g,z} * commit_{t,g} * SA_{t,g} \geq sd_{t,z,g,gs} \tag{A.26}$$

$$d_{t,g,z} == \sum_{gs}^{GS} sd_{t,z,g,gs} \tag{A.27}$$

$$\sum_{z}^{Z} [C_{g,z} * commit_{t,g} * SA_{t,g} - d_{t,g,z}] * CSPIN_g \geq psr_{t,g} \tag{A.28}$$

$$R_{g,z} * PSCALAR \geq psr_{t,g} \tag{A.29}$$

$$\sum_{z}^{Z} [C_{g,z} * PMIN_g * (1 - commit_{t,g})] * CNSPIN_g \geq pnsr_{t,g} \tag{A.30}$$

$$RSCALAR * [psr_{t,g} + pnsr_{t,g}] \geq secr_{t,g} \tag{A.31}$$

$$\sum_{z}^{Z} C_{g,z} - d_{t,g,z} \geq secr_{t,g} \tag{A.32}$$

$$\sum_{g}^{G} psr_{t,g} \geq \sum_{s}^{S} PSSR_{t,s} \tag{A.33}$$

$$\sum_{gsz}^{GSZ} psr_{t,gsz} + [txmw_{t,l} - TXTOCAPACITY_{t,l} | TXTO_{t,l} == sz] + [TXFROMCAPACITY_{t,l} - txmw_{t,l} | TXFROM_{t,l} == sz] \geq \sum_{s}^{S} PSZSR_{t,s} \tag{A.34}$$

$$\sum_{g}^{G} (pnsr_{t,g} + psr_{t,g}) \geq \sum_{s}^{S} PNSSR_{t,s} \tag{A.35}$$

$$\sum_{g}^{G} secr_{t,g} \geq \sum_{s}^{S} SECSR_{t,s} \tag{A.36}$$

$$PSRSMW_{t,s} \geq pssr_{t,s} \tag{A.37}$$

$$PSRSMW_{t,s} \geq pszsr_{t,s} \tag{A.38}$$

$$PNSRSMW_{t,s} \geq pnssr_{t,s} \tag{A.39}$$

$$SECRSMW_{t,s} \geq secsr_{t,s} \tag{A.40}$$

$$\sum_{g}^{G} [d_{t,g,z}] + wgen_{t,z} + sgen_{t,z} + \sum_{l}^{L} [txmw_{t,l} | TXTO_{t,l} = z] - \sum_{l}^{L} [txmw_{t,l} | TXFROM_{t,l} = z] == LOAD_{t,z} \tag{A.41}$$

A.6 Objectives

$$MIN \sum_{t,g,z,gs}^{T,G,Z,GS} [sd_{t,g,z,gs} * GENMC_{g,gs}] + \sum_{t,g}^{T,G} [SC_g * start_{t,g}] + \sum_{t,g}^{T,G} [NLC_g * commit_{t,g}] + \sum_{t,z}^{T,Z} UE_{t,z} + \sum_{t,l}^{T,L} txmw_{t,l} * TXH_{t,l} - \sum_{t,s}^{T,S} RSP_{t,s} * pssr_{t,s} - \sum_{t,s}^{T,S} RSP_{t,s} * pnssr_{t,s} - \sum_{t,s}^{T,S} RSP_{t,s} * secsr_{t,s} \quad (A.42)$$

$$MIN \sum_{t,g,z,gs}^{T,G,Z,GS} [sd_{t,g,z,gs} * GENMC_{g,gs}] + \sum_{t,z}^{T,Z} UE_{t,z} + \sum_{t,l}^{T,L} txmw_{t,l} * TXH_{t,l} - \sum_{t,s}^{T,S} RSP_{t,s} * pssr_{t,s} - \sum_{t,s}^{T,S} RSP_{t,s} * pnssr_{t,s} - \sum_{t,s}^{T,S} RSP_{t,s} * secsr_{t,s} \quad (A.43)$$

B Data Description Appendix

B.1 Assignment of PJM Load Delivery Areas (LDAs) to zones

PJM data are often defined at the level of a Load Delivery Area (LDA), generally corresponding to historical electric distribution company service territories within PJM's footprint. For purposes of our study, we assign these LDAs to modeled transmission zones, and assign other parameters (e.g., installed wind and solar capacity, load, etc.) at the LDA level before aggregating up to the zonal level. Generators located geographically outside the PJM footprint recorded in the PJM Generating Availability Data System (GADS) database are assigned to the geographically closest PJM zone.

Our assignment of PJM LDAs to the zones used in this study is in Table B-1. The results of this zonal assignment are shown in more detail in Figure B-1.

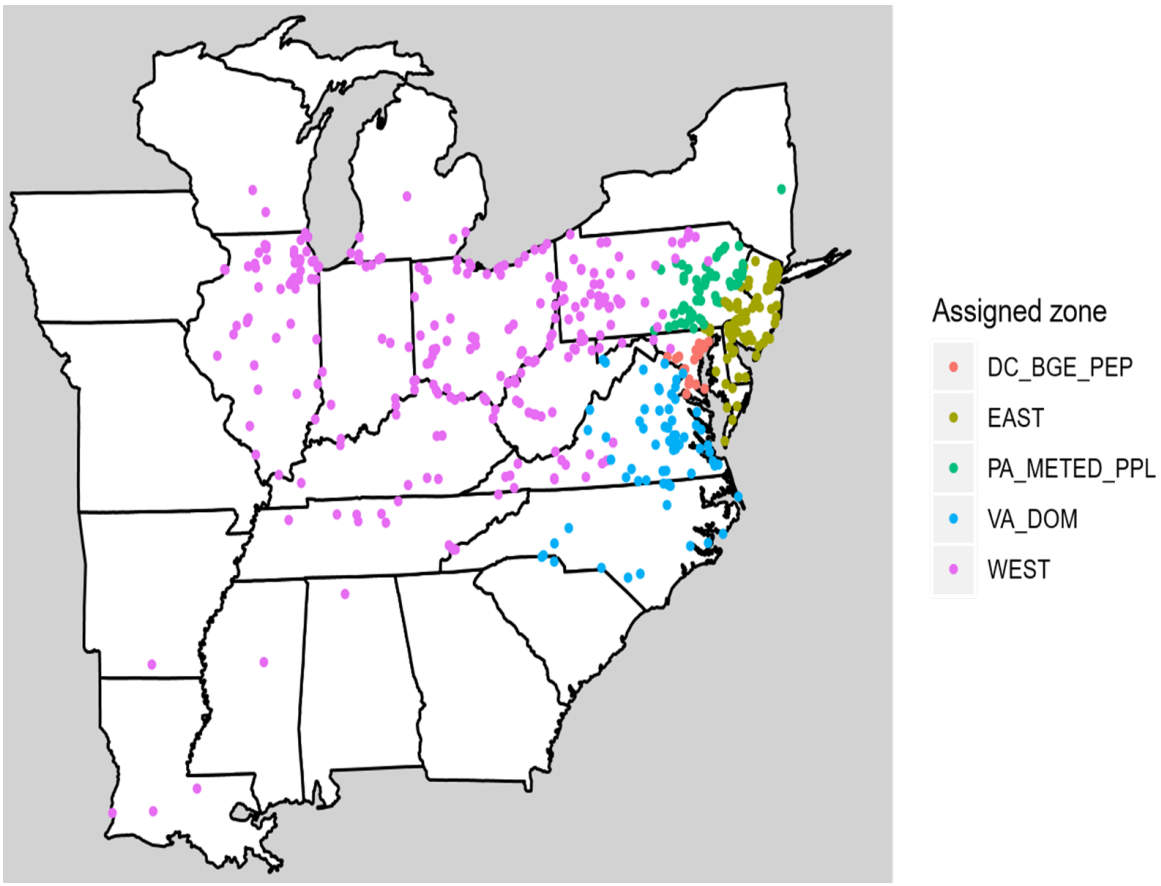

Figure B-1: Model zone assignment of generators in PJM. Generators geographically outside the PJM footprint but contracted to serve load and provide capacity in PJM are assigned to their closest zone.

Table B-1: Assignment of PJM LDAs to model zones.

PJM LDA	Model Zone
Delmarva Power and Light (DPL)	EAST
Atlantic City Electric (AE)	EAST
Dominion Energy (DOM)	DOM
Pennsylvania Power and Light (PPL)	PPL-METED
Metropolitan Edison (METED)	PPL-METED
Pennsylvania Electric Company (PENELEC)	WEST
Allegheny Power (APS)	WEST
Public Service Enterprise Group (PSEG)	EAST
Baltimore Gas and Electric (BG&E)	BGE-PEP
Jersey Central Power and Light (JCPL)	EAST
American Electric Power (AEP)	WEST
Eastern Kentucky Power Cooperative (EKPC)	WEST
Potomac Electric Power Company (PEPCO)	BGE-PEP
Rockland Electric (RE)	EAST
Duke Energy Ohio and Kentucky (DEOK)	WEST
Dayton Power and Light (DAY)	WEST
American Transmission Service Inc. (ATSI)	WEST
Duquesne Light Company (DLCO)	WEST
Commonwealth Edison (COMED)	WEST
Philadelphia Electric Company (PECO)	EAST

B.2 PJM Conventional generation fleet

We use the PJM GADS database to define the conventional generator fleet serving PJM and to develop operating histories of all generators contained therein.

PJM GADS provides generator geographic location information, installed capacity, generator online and retirement (if applicable) dates, scheduled outages, and fuel/unit type. We obtain and check additional generator information as follows. As a check on installed generator capacity, we match GADS data against the Environmental Protection Agency's (EPA) National Electricity Energy Database System (NEEDS) at the plant level (i.e., by ORIS plant code) [30]. Where a match between ORIS plant codes is obtained and installed capacity values differ, we use the NEEDS value. We take ramp rates and preliminary minimum up/down time for hydro, gas-fired and diesel-fired generators by unit type from Lueken and Apt [5]. As an additional check on minimum up and downtime we calculate the minimum number of hours a generator was on or offline based on 2014 EPA Continuous

Emissions Monitoring Systems (CEMS) data matched by ORIS plant code [31]. If the minimum offline or online time in 2014 is less than our assumed value, the generator(s) associated with that ORIS plant code are assigned the lesser offline or online time based on CEMS.

Market participants in PJM may self-schedule their generator [32], and commonly choose to do so with less flexible units, affecting clearing prices [24]. To improve representation of this behavior in the absence of revealed self-schedule data at the plant level we assume coal-fired generators are self-scheduled daily at their most economic feasible set-point and nuclear generators are self-scheduled at their maximum available capacity when economic; as a result, these resources are not capable of providing primary reserves nor load-following. This assumption is not a technical limitation of these resources in providing those services; rather, it is meant to approximate the quantity and type of self-scheduled resources observed in PJM.

We derive thermal generator fully loaded heat rates by unit type based on EPA [30]. Given fully loaded heat rates, we develop thermal generator heat rate curves based on Deaver and Macdonald [33] as four-segment piecewise linear functions. We assume nuclear units have constant heat rates. Hydroelectric units and demand response do not have heat rates.

B.3 Additional considerations for hydroelectric generators

Hydroelectric generators face additional constraints not applicable to thermal generators, e.g., seasonal operating constraints for environmental, recreational, agricultural, or other reasons related to maintaining river flow exogenous to power generation. Additionally, total hydroelectric generation may vary based on water inflows (e.g., precipitation). For these reasons, we implement four additional constraints that bind zonal hourly hydroelectric ramp, minimum, and maximum generation, as well as total generation for the time period of the model run. These parameters are based on four years (2015-2018) of data on total hydroelectric dispatch in PJM available from PJM [27]. Each hour is constrained to the highest max, highest min, and highest ramp observed in the four years of historical data for that hour, and the total daily generation is equivalent to the four-year average of total generation for the modeled day. This likely under-constrains hydroelectric dispatch; for example, by not capturing annual variation in precipitation, but better represents the reality that

hydro dispatches more during high-load than low-load hours in historical data than assuming a fleet-average capacity factor.

We do not differentiate between different types of hydroelectric generators, i.e. pumped storage vs. reservoir (the existence of considerable pumped storage generation as a fraction of total hydroelectric generation in PJM is part of the reason more dispatch is observed during higher load hours).

Table B-2: Assignment of WBANs to PJM LDAs.

PJM LDA	WBAN Code	WBAN Name
Delmarva Power and Light (DPL)	93721	BALTIMORE-WASHINGTON INTL AIRPORT
Atlantic City Electric (AE)	13735	MILLVILLE MUNICIPAL AIRPORT
Dominion Energy (DOM)	13740	RICHMOND INTERNATIONAL AIRPORT
Pennsylvania Power and Light (PPL)	14778	WILLIAMSPORT REGIONAL AIRPORT
Metropolitan Edison (METED)	14737	LEHIGH VALLEY INTERNATIONAL AIRPORT
Pennsylvania Electric Company (PENELEC)	14736	ALTOONA-BLAIR COUNTY AIRPORT
Allegheny Power (APS)	13736	MGTN RGNL-WL B HART FD AIRPORT
Southern Electric Reliability Council (SERC)	13737	NORFOLK INTERNATIONAL AIRPORT
Public Service Enterprise Group (PSEG)	14734	NEWARK LIBERTY INTL AIRPORT
Baltimore Gas and Electric (BG&E)	93721	BALTIMORE-WASHINGTON INTL AIRPORT
Jersey Central Power and Light (JCPL)	14734	NEWARK LIBERTY INTL AIRPORT
American Electric Power (AEP)	3860	TRI-STATE/M.J.FERGUSON FIELD AIRPORT
Midwest Independent System Operator (MISO)	94846	CHICAGO O'HARE INTERNATIONAL AIRPORT
Eastern Kentucky Power Cooperative (EKPC)	93820	BLUE GRASS AIRPORT
Potomac Electric Power Company (PEPCO)	13734	RONALD REAGAN WASHINGTON NATL
Rockland Electric (RE)	14734	NEWARK LIBERTY INTL AIRPORT
Duke Energy Ohio and Kentucky (DEOK)	93814	CINCINNATI/NORTHERN KENTUCKY
Dayton Power and Light (DAY)	93815	J.M.COX DAYTON INTERNATIONAL AIRPORT
American Transmission Service Inc. (ATSI)	14820	CLEVELAND-HOPKINS INTL AIRPORT
Duquesne Light Company (DLCO)	94823	PITTSBURGH INTERNATIONAL AIRPORT
Commonwealth Edison (COMED)	94846	CHICAGO O'HARE INTERNATIONAL AIRPORT
Philadelphia Electric Company (PECO)	13739	PHILADELPHIA INTERNATIONAL AIRPORT
New York Independent System Operator (NYISO)	4725	GREATER BINGHAMTON/EA LINK FI

B.4 Assignment of ambient temperature to PJM generators for determining outage probability

In our DynamicORDC case, forced outage rates of conventional generators are conditional on temperature. We must therefore match a temperature to each generator. Matched temperatures done at the LDA level, with generators within an LDA all being assigned the same ambient temperature based on the matched Weather Bureau Army Navy (WBAN) station for their LDA. LDA-WBAN matches used in the model are listed in Table B-2.

B.5 Initialization of conventional generators

When initializing a model run we must determine whether a generator has a scheduled outage, and separately whether it begins the model period on forced outage. In both cases, we use PJM GADS to obtain this information at the unit level. Generators that do not report forced or scheduled outages are assumed to be available, and have their probability of future forced outage modeled as in Murphy et al. [16] in the DynamicORDC case and at the fleetwide average forced outage rate in the SimpleORDC case.

B.6 PJM solar and wind generation

Wind and solar hourly generation shapes are developed as in Murphy et al. [17]. For years outside 2007-2012 for which Murphy et al. [17] models wind generation, the hourly average capacity factor for the date in the 2007-2012 modeled years is applied. We use a single wind and solar generation shape for all of PJM. We include only utility-scale wind and solar generation as behind-the-meter generation is assumed to be endogenous to metered load.

Wind and solar zonal installed capacity by date is based on EIA-860 [34]. EIA identifies installation at the state and county levels; we include all installed capacity in the states of Illinois, Indiana, Ohio, and Kentucky in our WEST zone, Pennsylvania in the PPL-METED zone, Delaware and New Jersey in the EAST zone, Maryland and Washington, D.C. in the BGE-PEPCO zone, and Virginia in the DOM zone. Due to considerable levels of solar installation in North Carolina, we additionally include installed solar capacity in the North Carolina counties of Anson, Sampson, Lenoir, Wayne, Northampton, Pasquotank, Currituck, Martin, Washington, Hertford, Beaufort, Bertie, Gates, Halifax, Pitt, Perquimans, Edgecombe, Greene, Alexander, Mecklenburg, Wake, Moore, and Chatham as part of the DOM zone, overall approximately matching the level of installed utility scale solar capacity reported by Dominion in 2018 [35].

B.7 Generator commitment and dispatch costs

Fuel costs for all nuclear generators are assumed to be \$0.5/million British Thermal Units (mmbtu), and for all diesel generators \$13/mmbtu. Hydroelectric units do not have fuel costs.

Demand response is dispatched as an emergency resource prices reach the market cap (assumed $2000/MWh in our cases for consistency with existing PJM price cap).

For coal-fired generators, we obtain monthly fuel costs for regulated utilities that report a fuel cost from EIA-923 [36] at the plant level and assign them to matched generators on ORIS plant code. For generators that do not report fuel costs or for which we do not obtain an ORIS code match in states with other generators reporting fuel costs, we assign the state quantity-weighted average fuel price to the generator. For states with fully deregulated generation sectors where no generators report coal fuel prices (in PJM this includes MD, DC, PA, NJ, and DE), generators are assigned the Virginia quantity-weighted average fuel price. Virginia is chosen to better maintain consistency in the eastern vs. western parts of PJM.

For gas-fired units (CC and CT) we repeat the EIA-923 matching process for regulated generators outlined above. However, for unregulated or unmatched gas generators we instead base the fuel cost on matching against the daily spot price of the nearest hub, plus a delivery adder of $0.40/mmbtu to convert to burnertip price. Additionally, we allow regulated generators for which we obtain EIA-923 contract data to increase their bid to the daily spot price of the nearest hub (plus delivery adder) if it exceeds their reported monthly average fuel cost. Included gas price hubs for matching gas commodity price are Chicago Citygates (commodity price for COMED generators), Dominion South (PENELEC, APS), MichCon (DEOK, DAY, ATSI), TCO (AEP), TETCO M3 (PPL, METED, PECO, DLCO), TGP 500L (EKPC), Transco Z6 (non-NY) (AE, BG&E, DPL, JCPL), Transco Z6 (NY) (RE, PSEG), and Transco Z5 (non-WGL) (DOM, PEPCO). We do not differentiate between gas generators that hold different types of contracts for either commodity or delivery (i.e., firm vs. interruptible contracts). We also do not model dual fuel capability; gas generators can only run on natural gas at the applicable burnertip price.

We include variable operations and maintenance (V O&M) in generator bids by unit type based on NREL [37]. We take start-up cost capacity multipliers for generators by unit type from Lueken and Apt [5].

B.8 Transmission topology

We assign PJM LDAs to five interconnected transmission zones as defined in Table B-1. Transmission zones are interconnected by six transmission lines, with hourly maximum line flows defined as per PJM interface limit data taken from PJM [26]. This transmission topology is used because of the public availability of data on maximum and average interface limits at some of the most limiting transmission interfaces in PJM and is the same topology used in Lueken and Apt [5]. Transfers with external balancing areas (e.g., New York ISO, Duke Energy) are ignored. Transmission lines between zones are assumed to have a hurdle rate of \$0.25/MW, meant to be broadly consistent with the incremental losses associated with flowing power over longer distances.

B.9 Loads

We obtain hourly historical metered ("actual") load data used in dispatch and pricing runs for PJM by LDA from 2006-2017 from PJM's DataMiner2 [27]. We then aggregate LDA loads to zones as per the LDA-zone assignment in Table B-1. If a LDA entered PJM after 2006 (EKPC entered in 2013, DEOK in 2012, and ATSI in 2011) its load prior to entry is modeled and included as equivalent in shape to the most similar fully observed zone at a magnitude equal to the ratio of loads between the unobserved LDA and the matched LDA during the observed time period after entry. For commitment runs we use PJM-wide day-ahead and day-of forecast loads, assigned pro-rata to model zones with the actual load for the same hour obtained as described in the prior paragraph.

C Additional Model Validation Runs Appendix

As further validation of the model's ability to reproduce PJM prices, we provide a figure and summary statistics for a run of all days in October 2017.

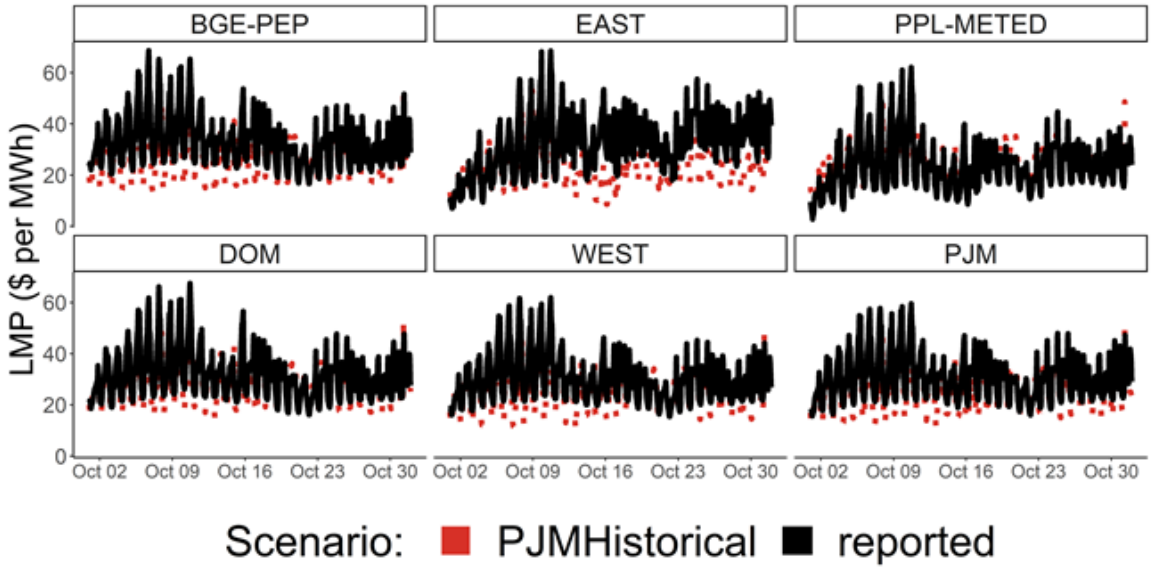

Figure C-1: Full October 2017 run.

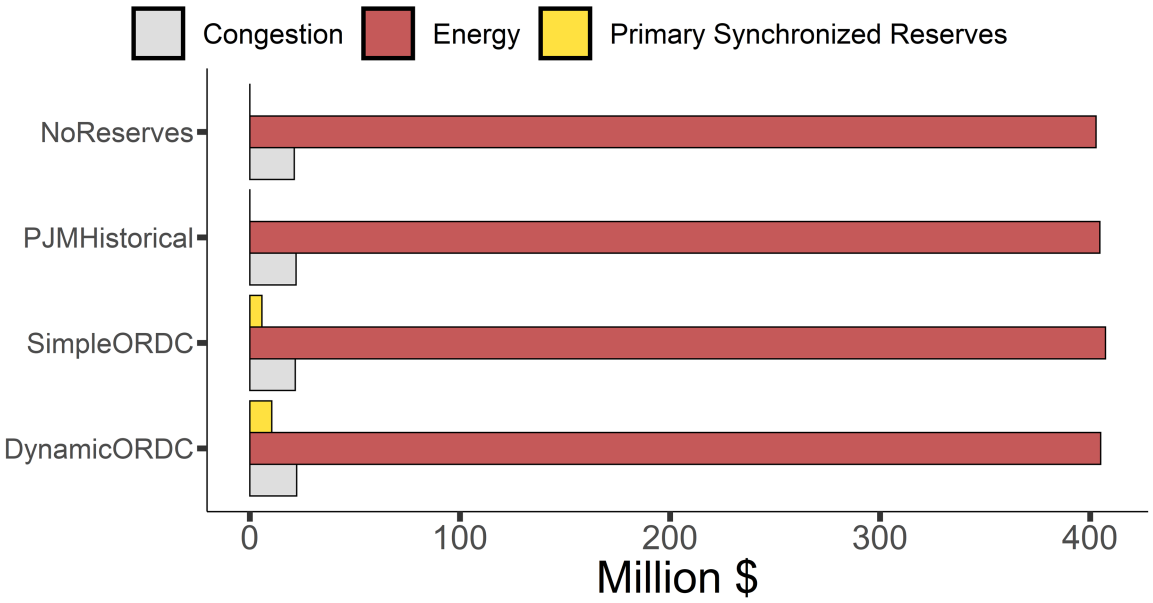

Figure C-2: Total congestion payments, total energy payments, and ORDC-related primary synchronized reserve payments for the modeled October 2017 week.

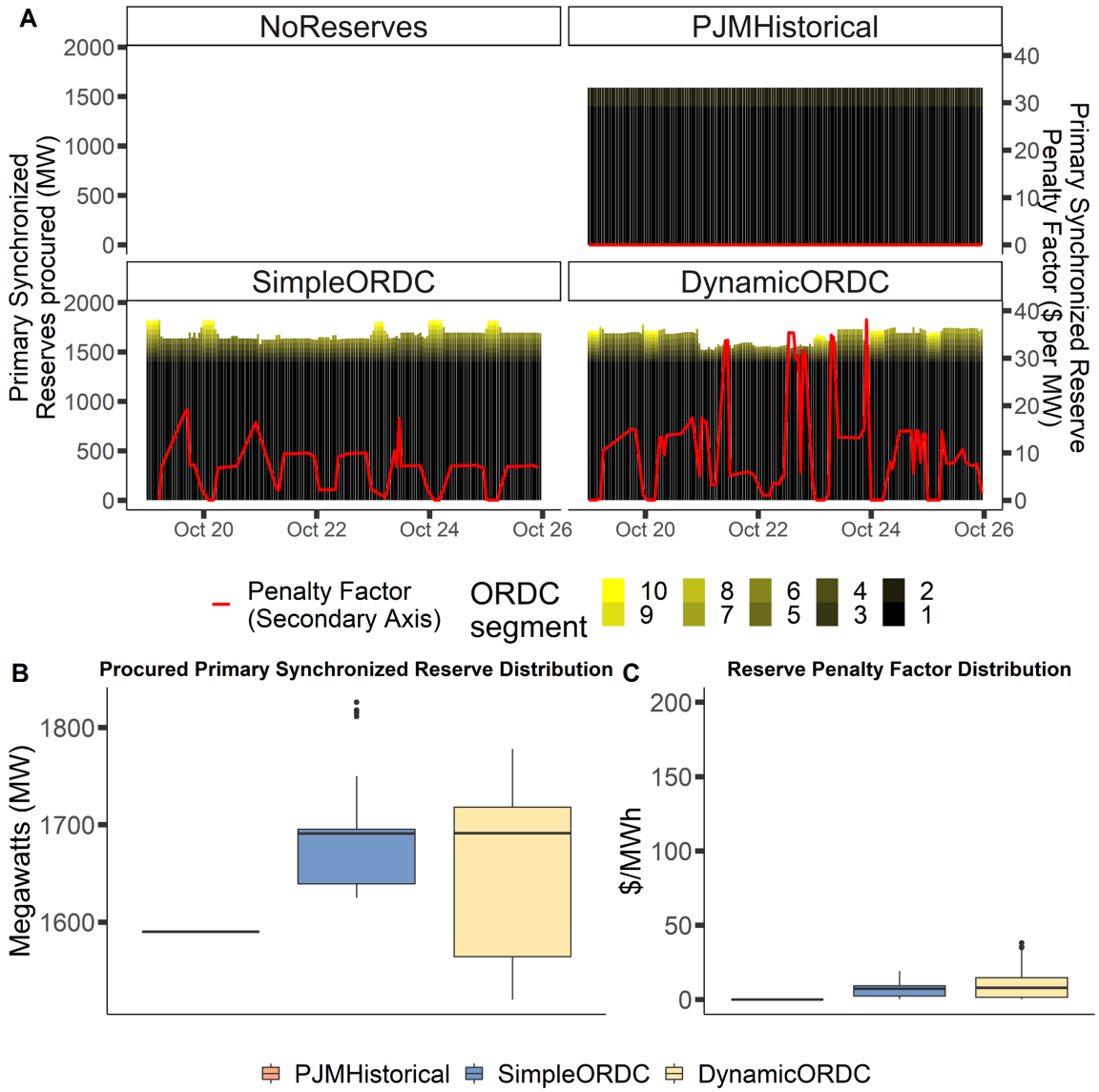

Figure C-3: (A) Quantity of primary synchronized reserves procured by ORDC segment (shaded vertical line segments; left axis, in MW) by hour for Oct. 19-25, 2017, along with ORDC penalty factor for primary synchronized reserves (red line; right axis, in $/MW). (B) and (C) show the distributions of reserve quantities and penalty factors.

Table C-1: Pricing summary statistics October 2017.

	Reported	PJMHistorical
October 2017 LMP ($/MWh)		
Mean	30.2	24.8
Standard Deviation	8.6	8.4
Median	29.6	25.2
Min	15.6	11.0
Max	59.6	49.5
October 2017 Root Mean Square Error compared to Reported		
Mean ($/MWh)		6.1
% of Reported Mean		20.1%

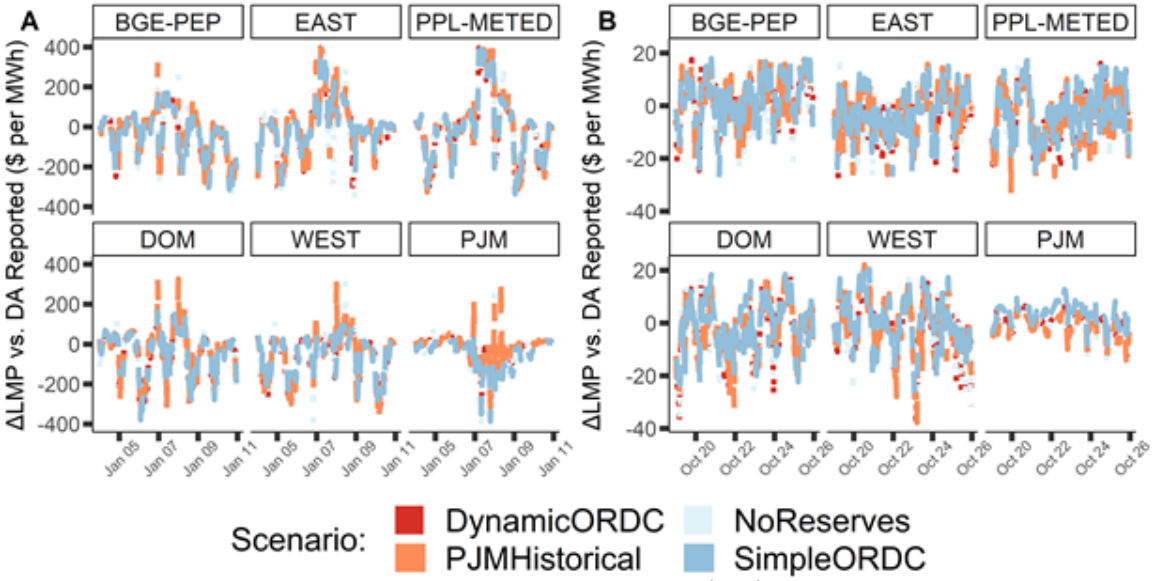

Figure C-4: Deviations in model LMP vs. Day-Ahead (DA) reported. Negative deviations indicate higher reported price, positive higher model price.

Chapter 5 Market Power Challenges and Solutions for Electric Power Storage Resources*

Abstract

Energy storage can enable low-carbon power and resilient power systems. However, market design is critical if a transition to renewables and storage is to result in low costs for customers. Pivotal suppliers with energy storage resources (ESRs) can achieve supernormal profits when allowed to fully participate and set clearing prices in wholesale electricity markets. Additional strategic profit from offers inconsistent with marginal costs can hurt competition and increase customer payments, hindering ongoing transitions to high shares of low marginal cost renewable generation and ESRs in electricity markets. We classify three strategies identified by our bi-level model for achieving additional strategic profits: (1) increased ESR discharge bids, (2) decreased ESR charge bids, and (3) cross-product manipulation to benefit other resources owned by the pivotal ESR supplier. We examine cases on a 25-bus test system with 67% average renewable energy generation where the ESR is commonly pivotal due to congestion. We observe under some circumstances the ESR owner can increase its energy market profits from \$10-20/MWh discharged when competitive to \$40-250/MWh discharged when strategic. Most increased profit comes from cross-product manipulation aimed at increasing prices to benefit a large co-located or hybridized zero marginal cost wind generator owned by the same entity. Marginal cost-based offer caps commonly applied to other resources could be extended to include ESRs' intertemporal opportunity costs limit, but these caps do not fully mitigate manipulative cross-product strategies. Relative inframarginal ESR offers over co-optimized time intervals with energy limits can be used to manipulate clearing quantities and prices and should be closely monitored when ESRs are pivotal suppliers. Requiring inframarginal offer uniformity over co-optimized time intervals shows promise as a policy remedy.

* This paper is available online as Carnegie Mellon Electricity Industry Center working paper 21-02.[1] It is being prepared for journal submission. It benefited from research assistance by Nik Zheng.

[1] https://www.cmu.edu/ceic/research-publications/ceic_21_02-esr-policy.pdf

Abbreviations and acronyms

CAISO	California Independent System Operator
CC	Combined Cycle
CT	Combustion Turbine
DA	Day-ahead
DCOPF	Direct Current Optimal Power Flow
EPEC	Equilibrium Program with Equilibrium Constraints
ERCOT	Electricity Reliability Council of Texas
ESR	Energy Storage Resource
FERC	Federal Energy Regulatory Commission
IMM	Independent Market Monitor
KKT	Karush-Kuhn-Tucker
MILP	Mixed Integer Linear Program
MPEC	Mathematical Program with Equilibrium Constraints
NREL	National Renewable Energy Laboratory
NYISO	New York Independent System Operator
RT	Real-time
RTPV	Rooftop Photovoltaic
RTS-GMLC	Reliability Test System - Grid Modernization Lab Consortium
PHS	Pumped Hydroelectric Storage
SPP	Southwest Power Pool
VRE	Variable Renewable Energy

5.1 Introduction

Mid-century decarbonization pathways commonly increase the quantity and share of final energy demand supplied by electricity [1, 2]. A highly decarbonized and expanded electricity sector requires rapid transition from current generation mixes, with most pathways relying heavily on declining costs of variable renewable energy (VRE) technologies [3] and energy storage technologies [4] that enable better instantaneous matching of supply and demand [5]. While standard electricity market designs are theoretically consistent with this transition [6], a key question is what reforms to existing electricity market designs are complementary with high penetrations of variable, low emission, and low marginal cost resources [7–10].

Consistent with integrating higher quantities of variable renewables in the electricity sector, standalone, stationary energy storage resources (ESRs) and ESRs sharing an interconnection with another generator ("hybrids") [11] make up an increasingly large portion of interconnection queues in competitive North American wholesale electricity markets and are expected to increase in coming

years [12]. ESRs and hybrids are energy-limited and shift load and generation in time rather than generating electricity. These differences, combined with low variable operating costs for lithium-ion technologies likely to dominate near-term ESR deployment [13], mean marginal cost based competitive ESR and hybrid bids generally reflect the intertemporal opportunity costs of ESR usage [14]. Existing North American competitive electricity markets have experience with a limited number of pumped hydroelectric storage (PHS) units sharing these general operational characteristics [15]. However, ESRs and hybrids are forecast to be more numerous, modular in sizing and deployment, and more readily dispatchable than PHS in coming years, enabling different use cases than PHS [16] and requiring modifications to existing rules to enable their full participation in competitive electricity markets [17].

We contribute to this discussion by identifying profit-maximizing bidding strategies for ESR- or hybrid-owning entities in a realistic multi-node, two-settlement electricity market with high penetration of variable, renewable, low marginal cost resources. Methods for identifying bidding strategies are essential to maintaining competitive electricity markets and delivering customers low-cost, reliable electricity service with high shares of VRE and ESRs.

We develop a bi-level model with a profit-maximizing supplier in the upper level and the market operator minimizing the as-bid cost of serving load in the lower level. A bi-level model allows full participation of resources with the ability to endogenously set locational marginal prices (LMPs) in nodal wholesale electricity markets [18], and is commonly referred to as a type of "price-maker" model when applied to ESRs [19]. This approach is consistent with previous approaches to modeling ESR market participation with ability to set clearing prices [20, 21]. The model can be used to identify bidding strategies. The major contribution of our research is to extend the policy relevance of previous research focused on solving stylized cases of ESRs exercising market power as a pivotal supplier [20–22]. We do this by classifying three market manipulation strategies on a high VRE nodal test system and suggesting directions for monitoring and mitigating these strategies.

5.2 Methods

We develop a bi-level optimization model reformulated as a mixed-integer linear program (MILP). This approach can be conceived as finding an equilibrium in a leader-follower Stackelberg game applied to electricity markets. In the upper level the leading generation- and ESR-owning entity submits bids to maximize the expected joint profits of its portfolio of resources. In the lower level the follower market operator minimizes the as-bid cost of serving electricity loads, subject to physical constraints on power flow and generator operational parameters. To focus on the properties of resource offers in one market we exclude security constraints, markets for ancillary services, and demand-side offers other than ESR charging loads (model formulation in Appendix A).

We increase the policy relevance of our cases compared to bi-level models on single-node test systems by modeling a congested high VRE nodal network, allowing us to observe congestion-related cross-product strategies of particular concern in electricity markets [23, 24]. We further extend previous work using bi-level models in electricity markets [21, 25, 26] by allowing hybridization of ESR and other generators located at the same bus as a single resource in bidding.

5.2.1 Multiple settlement functionality

North American wholesale markets commonly have two settlement intervals: a day-ahead (DA) forward market co-optimized for the subsequent day at hourly resolution and a higher temporal resolution (often five minutes) real-time (RT) market for settling deviations from the DA market with more limited look ahead temporal co-optimization. Our cases are commonly run DA with perfect foresight, but we include multiple settlement functionality in the model to allow sensitivity analysis under uncertainty with DA bids cleared against RT actual load and generation with limited bidding recourse (Appendix B).

5.2.2 Offer constraints and mitigation

In all cases unless otherwise noted generators are assumed to offer all available generation at marginal cost (Appendix B). For variable renewable generators with zero fuel cost, we additionally

assume zero variable O&M and no effect of subsidies (e.g., wind Production Tax Credit [27]) on marginal cost, so these resources offer at $0/MWh. Because ESR offers are largely opportunity cost-based, no constraints are placed on ESR offers in cases without ESR-specific offer mitigation constraints.

To investigate the efficacy of ESR offer mitigation we develop two approaches. First, we mitigate day-ahead offers based on an ex-ante price forecast, disallowing offers from exceeding the expected competitive clearing price. We show this approach does little to mitigate the most profitable strategic bidding by entities owning both ESRs and generation, which can use relative rather than absolute price offers for energy-limited resources. Second, we propose a mitigation framework based on uniform bidding for co-optimized temporal intervals. This framework is more effective in mitigating cross-product manipulation but requires careful consideration to avoid over-mitigation and allow ESRs to capture option value.

5.3 Data

To achieve germane results we implement our model on the National Renewable Energy Laboratory (NREL)-modified version of the IEEE RTS-96 test system: the Reliability Test System Grid Modernization Lab Consortium (RTS-GMLC) [28]. The RTS-GMLC test system updates an older IEEE test system primarily by modernizing the generation fleet to include more gas-fired and renewable resources. Renewable and load profiles are based on three zonal locations in the United States southwest, though we retain only a single 25-bus zone for model cases in this paper (Figure 5.1)

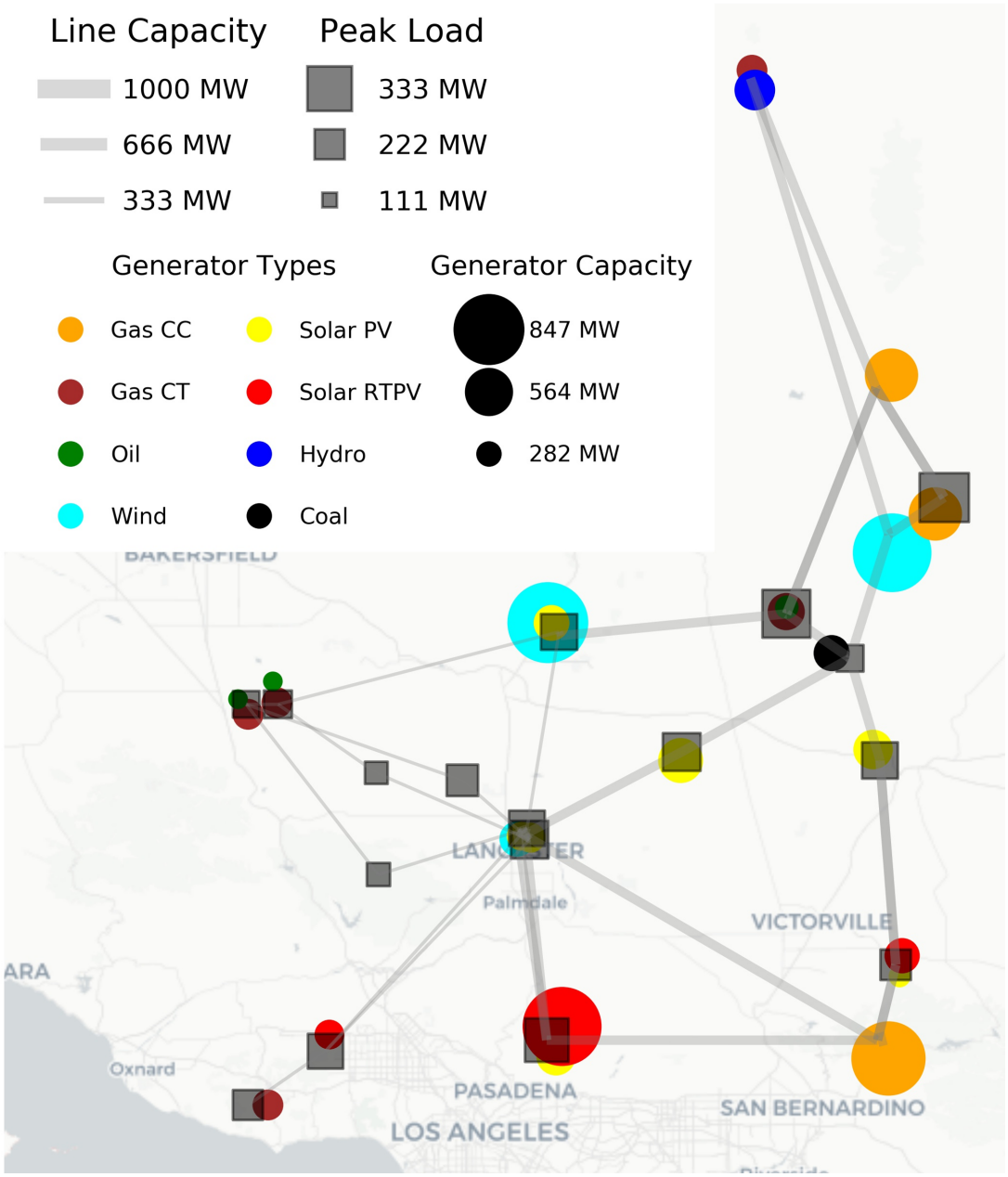

Figure 5.1: Modified NREL RTS-GMLC case nodes, transmission lines, generation capacity, and peak loads. The three RTS-GMLC zones use geographic data based on Arizona Public Service Company (AZPS), Nevada Energy, and the Los Angeles Department of Water and Power (LADWP), though they are not intended to represent existing infrastructure. We retain only the displayed 25 buses in LADWP for cases to decrease runtime and because LADWP has the highest average (67%) and instantaneous renewables penetrations and the most congestion in the model year. RTPV is rooftop photovoltaic, CT combustion turbine, CC combined cycle.

We modify the RTS-GMLC source data to exclude native ESR and add ESRs with user-specified

capacity, duration, and round-trip efficiency (assumed 85%) at select buses. These ESRs may also be hybridized with generators owned by the same entity at the same node. Day-ahead data are hourly resolution. Real-time data are 5-minute resolution, but are reformatted to equivalent hourly average load and renewable generation for the two-settlement model. The model is run as a sequence of co-optimized 24-hour resolution days to reduce solution time compared to co-optimization of a longer timeframe, and because this mimics DA markets. Constraints enforce a single daily cycle for ESR dispatch as a heuristic for degradation in the absence of more sophisticated degradation incorporation [14]. Initial and final ESR SOC are constrained to be zero to enforce continuity between days in sequential runs.

5.3.1 Summary of 25-bus data

Figure 5.2 shows January DA hourly average load, net load, and generation by resource for the reduced 25-bus version of the RTS-GMLC data without ESRs. A comparison to RT data used for multiple settlement functionality is included in Appendix B. The test system has high renewable generation compared to current United States averages [29], but these renewable penetrations are commonly met or exceeded in forward looking decarbonization pathways [2].

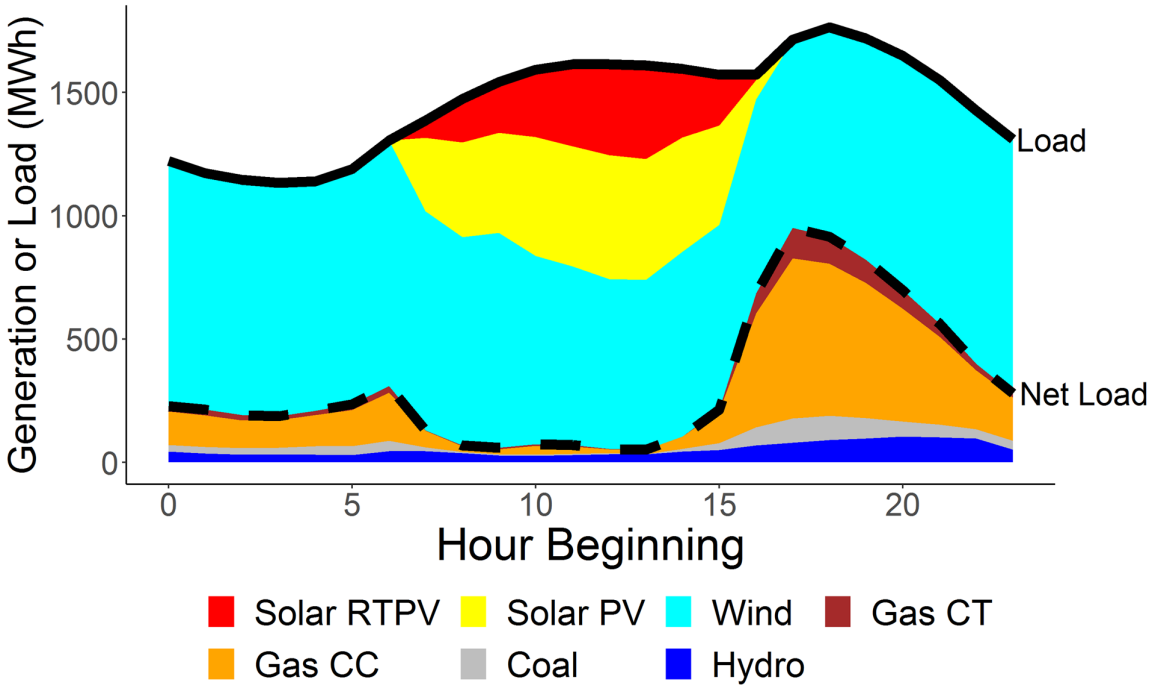

Figure 5.2: Hourly day-ahead average load, net load, and generation by type for RTS-GMLC data without energy storage in Zone 3 for the modeled month, January. RTPV is rooftop photovoltaic, CT is combustion turbine, CC is combined cycle. Oil-fired generators are also included but not dispatched during the modeled month.

The combination of high quantities of available zero marginal cost renewable generation concentrated at a few buses with large wind generators and higher load hours often results in transmission congestion. The 175MW line connecting buses 03 and 09 is most often congested when large quantities of wind generation are available from the 847MW of installed wind capacity at bus 03. Bus 03 is the lowest LMP bus when there is congestion on the line connecting buses 03 and 09 (Figure 5.3).

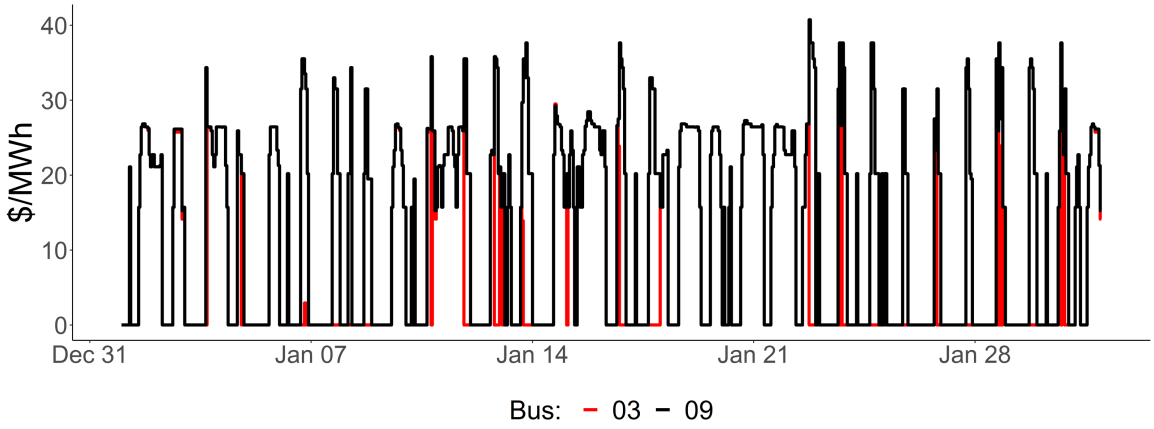

Figure 5.3: Hourly LMPs without ESRs at the two buses in the test system linked by an often-congested line for the entire modeled month. LMPs for all buses are included in Appendix B.

5.4 Results

Three ESR bidding strategies increase the profit of a portfolio of resources owned by a strategic entity.

1. Increase discharge offers when pivotal to increase LMP at ESR bus when discharging;
2. Decrease charge offers when pivotal to decrease LMP at ESR bus when charging;
3. Increase ESR charge or discharge offers to increase LMP at buses where the strategic entity owns other generation.

Of these strategies the third, a cross-product strategy, is the most profitable when strategic ESRs are co-located or hybridized with a large renewable generator at a commonly congested bus. It is also the most difficult to mitigate against (Section 5.4.4).

To assess the additional profit associated with strategic ESR bids in a portfolio of resources we use two metrics: additional total portfolio profit $\Delta\pi^p$ over an applicable time interval and ESR incremental per-MWh discharged portfolio profit $\frac{\Delta\pi^p}{MWh}$. These metrics are defined:

$$\Delta\pi^p = \pi^{SS} - \pi^{NSS} \tag{5.1}$$

$$\frac{\Delta\pi^p}{MWh} = \frac{\pi^{SS} - \pi^G}{\sum sd^S} \tag{5.2}$$

Where π^{SS} is the profit of the strategic entity's portfolio p of generators with ability to bid its ESRs strategically (strategic storage is SS), π^{NSS} is the strategic entity's portfolio profit when it does not bid ESRs strategically (non-strategic storage is NSS), and π^G is the non-ESR generator profit in the non-strategic case. sd is total discharge over the applicable time interval of the ESR, S.

Assumptions include generators being mitigated to offer at marginal cost, but no offer mitigation applied to opportunity costs for ESRs, and the strategic entity has perfect foresight of load, renewable generation, and offers by other suppliers. Results in Section 5.4.1-5.4.2 using these assumptions set an upper bound on strategic profit of ESR bidding decisions for an assumed system and strategic ownership parameterization. Section 5.4.3 investigates relaxing some perfect foresight assumptions. Section 5.4.4 explores monitoring and mitigation frameworks for the most profitable perfect foresight strategies.

5.4.1 Demonstrating the three strategies

We parameterize two cases to demonstrate the applicability of the three strategies assuming perfect foresight of the market clearing problem (Table 5.1).

Table 5.1: Demonstration case parameterization. Differences between cases are in **bold**.

Case Label	Generators and Loads	Storage Sizing	Storage Location	Other Owned Generators	Time Period and Resolution
Case A: "ESR Only"	All NREL-RTS LADWP	300MW/900MWh	Bus 03	None	Hourly Day-ahead, January 2020
Case B: "ESR+Wind"	All NREL-RTS LADWP	300MW/900MWh	Bus 03	**Wind (847MW), bus 03**	Hourly Day-ahead, January 2020

In both cases the model is configured and run in two ways: "competitive" and "strategic." Competitive is equivalent to a cost-minimizing linear program under the assumption all resources offer at marginal cost. Competitive ESRs are dispatched to minimize production costs. Strategic assumes ESRs and other generators owned by the strategic entity[2] submit offers to maximize the

[2]As applicable in the case per Table 5.1, though recall generators are constrained to offer at marginal cost.

strategic generation-owning entity's profit knowing the market operator will minimize as-bid costs of serving load. Strategic uses the full functionality of our model to find a profit-maximizing solution to within a pre-set MILP optimality gap (1% in these cases).

Differences between the strategic and competitive solution in the profits accrued by the strategic entity are quantified as $\Delta\pi^p$. Figures show LMPs only at the bus where the ESR is installed unless otherwise noted.

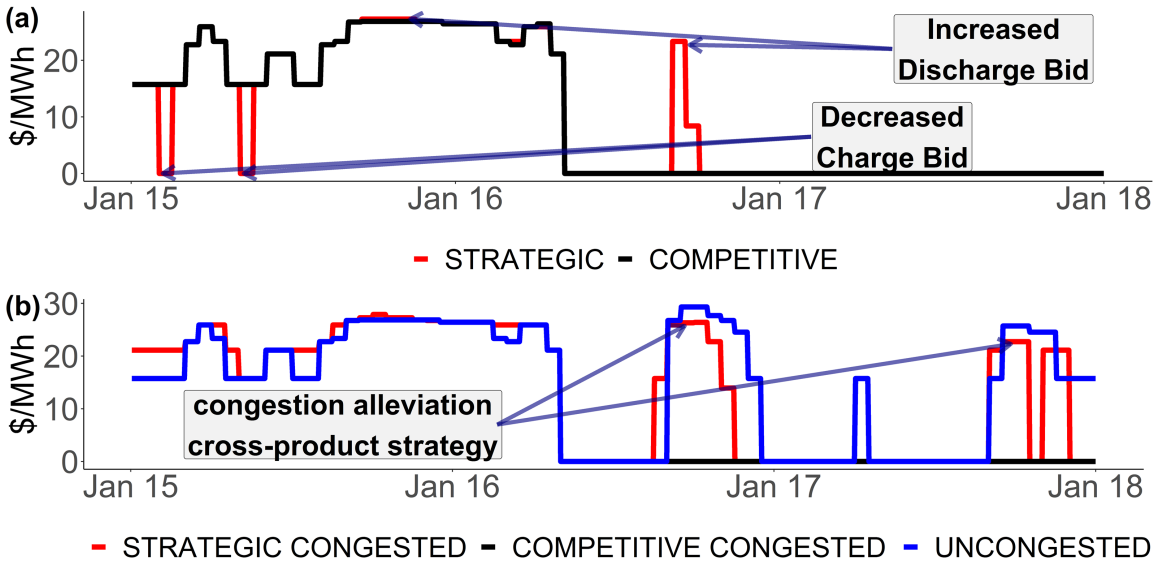

Figure 5.4: Comparison of strategic and competitive LMPs at the bus (03) with ESR for an illustrative subset of modeled January days. In case A the ESR is able to set price and increase profits when pivotal through two strategies: increasing its discharge bid and decreasing its charge bid. In case B the strategic entity also owns a 847MW wind generator. Bus 03 is congested in many hours due to the wind generation and transmission limits, so prices are often below the system lambda (blue). Strategies used in case A still appear (though decreasing price for charging only makes sense when little wind is available), but the most profitable strategy is a cross-product strategy to alleviate congestion that results in the greatest increases in clearing price.

The total profits accrued in case A and B for the full month by the ESR and wind generator are compared in Figure 5.5.

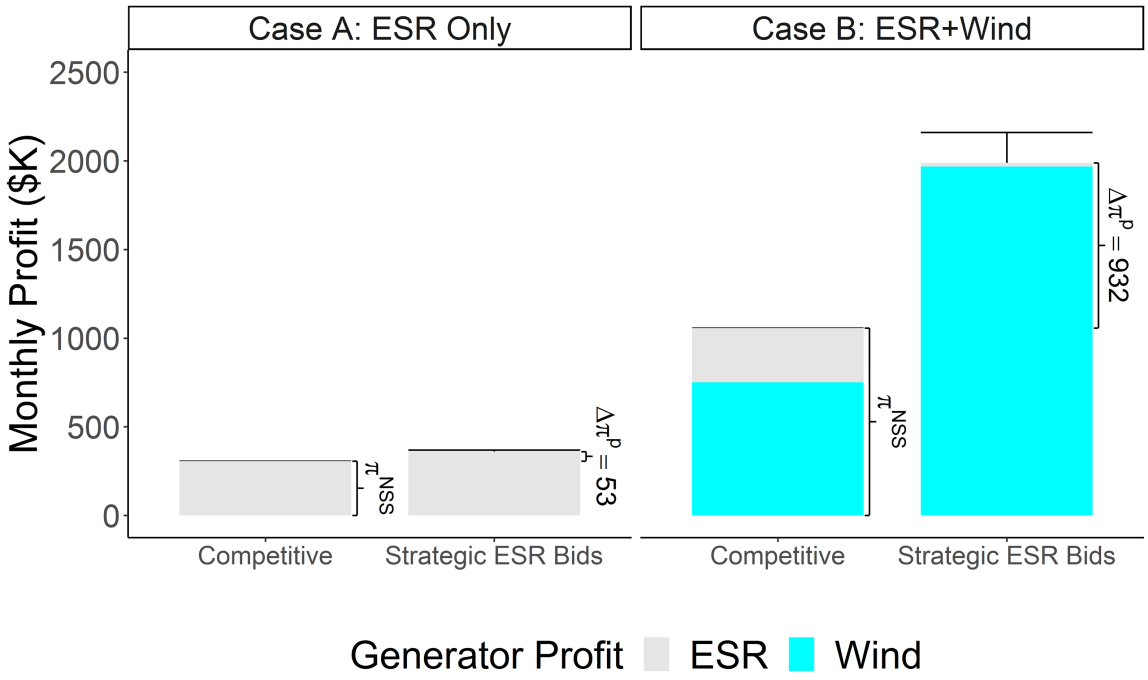

Figure 5.5: Comparison of case profits for month of January when dispatched competitively vs. strategically. In case A the strategic entity only owns the ESR; in case B it also owns the wind generator and maximizes the joint profits by modifying ESR bids to increase revenues received by the wind generator. Uncertainty bars show the aggregate upper bound optimality gap for strategic MILP. Competitive cases are cost-minimizing LPs and have no optimality gap.

5.4.2 Sensitivity to ESR parameterization and hybridization

This section adds sensitivity analysis on how ESR sizing and hybridization affects strategic profits. In these sensitivity cases we maintain the perfect foresight assumption, so results set an upper bound on portfolio profit from ESR bidding.

Figure 5.6 shows ESR capacity and duration sensitivity analysis. The only changes to the case B parameterization are ESR capacity and duration. Because the cross-product strategy is profitable only when wind generation exceeds storage charging load, sensitivities consider ESR capacity installations up to 500 MW.[3] Per-MWh profits associated with storage ownership $\frac{\pi^{NSS}}{\sum sd^{S}}$ are \$10-20/MWh ESR discharge in all competitive cases, while strategic incremental profits $\frac{\Delta\pi^{p}}{MWh}$ are \$40-250/MWh ESR discharge. Increased profits largely result from cross-product manipulation that decreases ESR revenues, but increases clearing price and thus wind revenues by more.

[3]the wind generator's capacity factor is 53% in the month; 448 MW average wind generation.

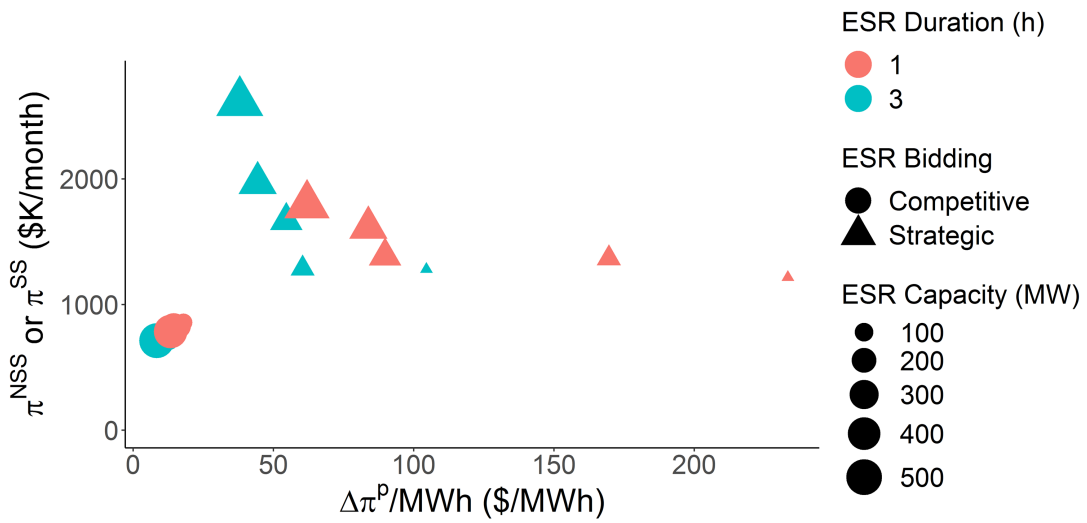

Figure 5.6: Profit sensitivities show in all cases strategic bidding increases joint profits compared to the competitive outcome where the ESR is dispatched to minimize production costs. Increasing ESR capacity and duration generally exhibit decreasing marginal value (lower $/MWh profits) under the perfect foresight assumption.

Motivated by increasing prominence of wind-ESR and solar-ESR hybrid generators in North American interconnection queues [12], Figure 5.7 compares case B with a hybrid made up of the same wind and ESR. The hybrid differs from co-located in two ways: (1) it makes a single, unconstrained offer, and (2) the hybrid cannot dispatch more than the rated capacity (847MW) of its wind generator as an assumed interconnection limit, whereas the co-located resources can both dispatch at their full rated capacity.

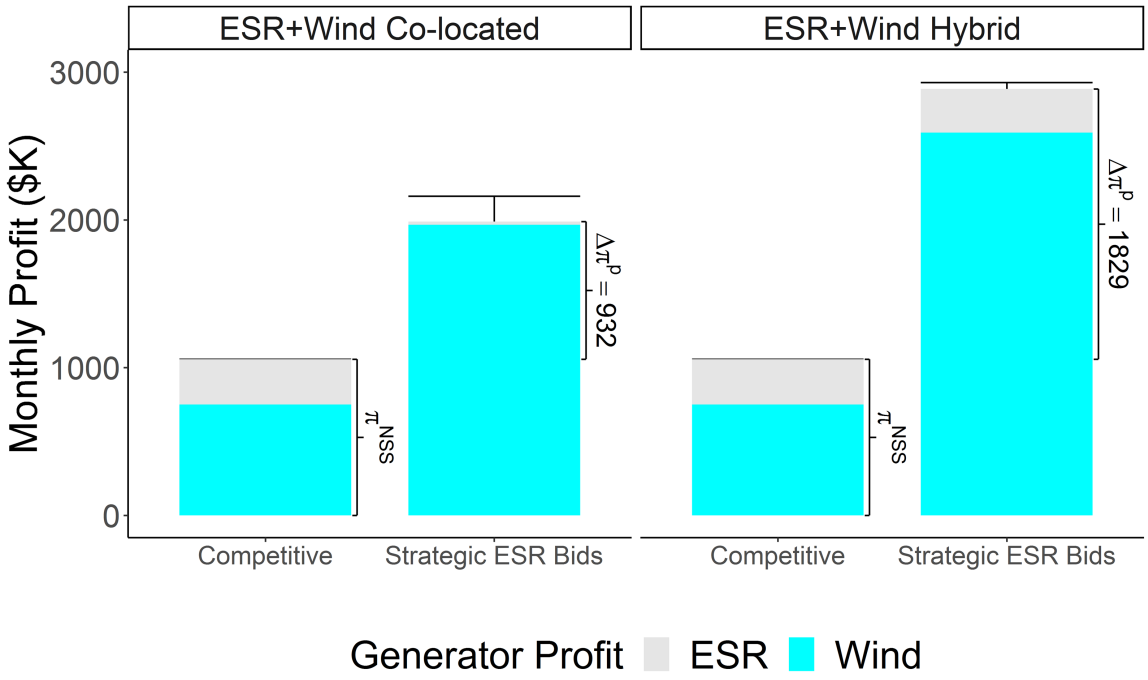

Figure 5.7: Comparison of co-located and hybrid strategic profits with the competitive solution. Co-located case is the same as Case B in Section 5.4.1.

The hybrid achieves more profit than the co-location because of the removal of offer constraints on the wind generator, previously offered at its marginal cost of \$0/MWh even when owned by the strategic entity. The new, larger hybrid generator can exercise additional ability to alleviate congestion affecting the LMP at bus 03 and set a higher price. The results illustrate how hybridization could be used to enable additional bidding latitude not afforded to a generator or ESR if hybridization allows the resource a new classification with fewer bidding restrictions.

5.4.3 Incorporating uncertainty

Previous cases were day-ahead, hourly resolution under the assumption the strategic entity has perfect foresight of bus-level system loads, price-quantity bids by other generators, and knowledge how the market operator will minimize production costs and set prices. This section considers whether the strategic entity could achieve some of the perfect information profit when uncertainties in available generation and loads are incorporated. When fixing its profitable pivotal supply bid quantities in DA prior to realization of actual load and generation in RT, the strategic entity can

maintain some profit (Figure 5.8, Appendix B for detail on strategy under uncertainty).

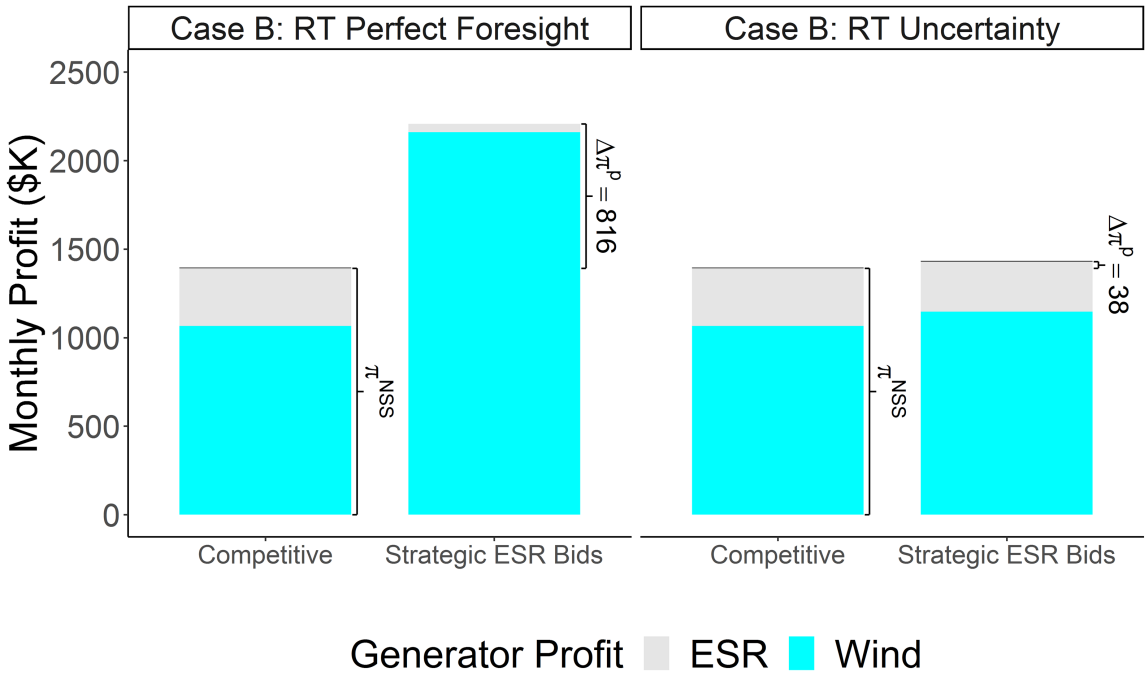

Figure 5.8: Strategic profits are reduced under uncertainty. Implemented bidding strategy under uncertainty is explained in Appendix B.

Figure 5.8 is only an existence proof for potential market manipulation, and real traders have in markets have developed profitable strategies under uncertainty. Market monitoring can still benefit from considering the profit maximizing strategy assuming perfect foresight, since tools for mitigating the three strategies identified in Section 5.4.1 will still be applicable under uncertainty.

5.4.4 Analytical consideration for ESR market monitoring and offer mitigation

In this section we extend results to a discussion of detection and mitigation of ESR offers intended to manipulate market prices.

A common tool for mitigating offers in existing wholesale electricity markets are offer caps. North American electricity markets generally have both a market-wide offer cap and resource level offer mitigation using estimation and verification of marginal costs. Extending offer mitigation to ESRs and hybrids often proceeds from this framework, assuming adding intertemporal opportunity costs[4]

[4]Clearing prices endogenously account for lost opportunity costs (LOC) for co-optimized products in wholesale

to the marginal cost framework will account for the energy-limits of ESRs. Results in this section suggest an absolute offer cap for ESRs, even assuming an agreeable framework for estimation of ex-ante opportunity costs, misses important aspects of the ability of ESRs to exercise market power when energy-limit constraints are binding.

Over co-optimized temporal intervals with market operator incorporation of ESR energy limit constraints time-varying relative ESR offers can be used to manipulate market clearing prices and quantities, even with a stringent absolute offer cap. The mathematical insight is the effect of energy-limited ESR offers on clearing prices in all co-optimized time periods depends on the difference between inframarginal ESR offers and the generation production costs the ESR displaces, meaning the relative ESR offers across time periods are what matters in this setting. When pivotal the ESR can make use of this fact to change its dispatch, and thus pricing, based on its relative offers, even when its absolute offers are constrained to be inframarginal in all time periods. Appendix C provides a proof and example in a co-optimized, two-time period setting where a fully charged ESR's energy-limit constraint is binding (i.e., it can only dispatch in one of the two time periods, but not both), showing that when the energy-limit constraint binds the difference between inframarginal ESR offers and the generation production costs the ESR displaces will determine when the limited energy in the ESR is dispatched, and thus how it will affect clearing prices.

To demonstrate the lack of efficacy of an absolute offer cap with binding ESR energy-limits, the case B parameterization is run for a single day (1 January) with DA data. The ESR is constrained to a single cycle of charge and discharge during the day, so the energy-limit binds. An ex-ante bus 03 LMP-based offer cap is developed as the market clearing price from the fully competitive solution. The model is then re-run with an additional set of constraints requiring ESR discharge offers to be less than or equal to the ex-ante LMP offer cap. This approach is similar to marginal cost-based mitigation for generator offers, and more stringent than mitigation using a uniform estimate of daily opportunity costs equivalent to the Nth (N=ESR duration) highest price hour, as would be

electricity markets. Without co-optimization LOCs would have to be reflected in offers. As an example of the effect of product co-optimization, a generator required to provide upward reserves for security reasons cannot simultaneously clear that capacity in an energy market and, if it would be more profitable to provide energy, incurs a LOC, which will then be reflected in the reserve clearing price. This makes the generator at minimum indifferent between providing upward reserve and energy. This is distinct from intertemporal opportunity costs considered for energy-limited resources like ESRs.

suggested by optimal price-taker dispatch of a single daily ESR cycle. Perfect information profits remain above the competitive solution (Figure 5.9).

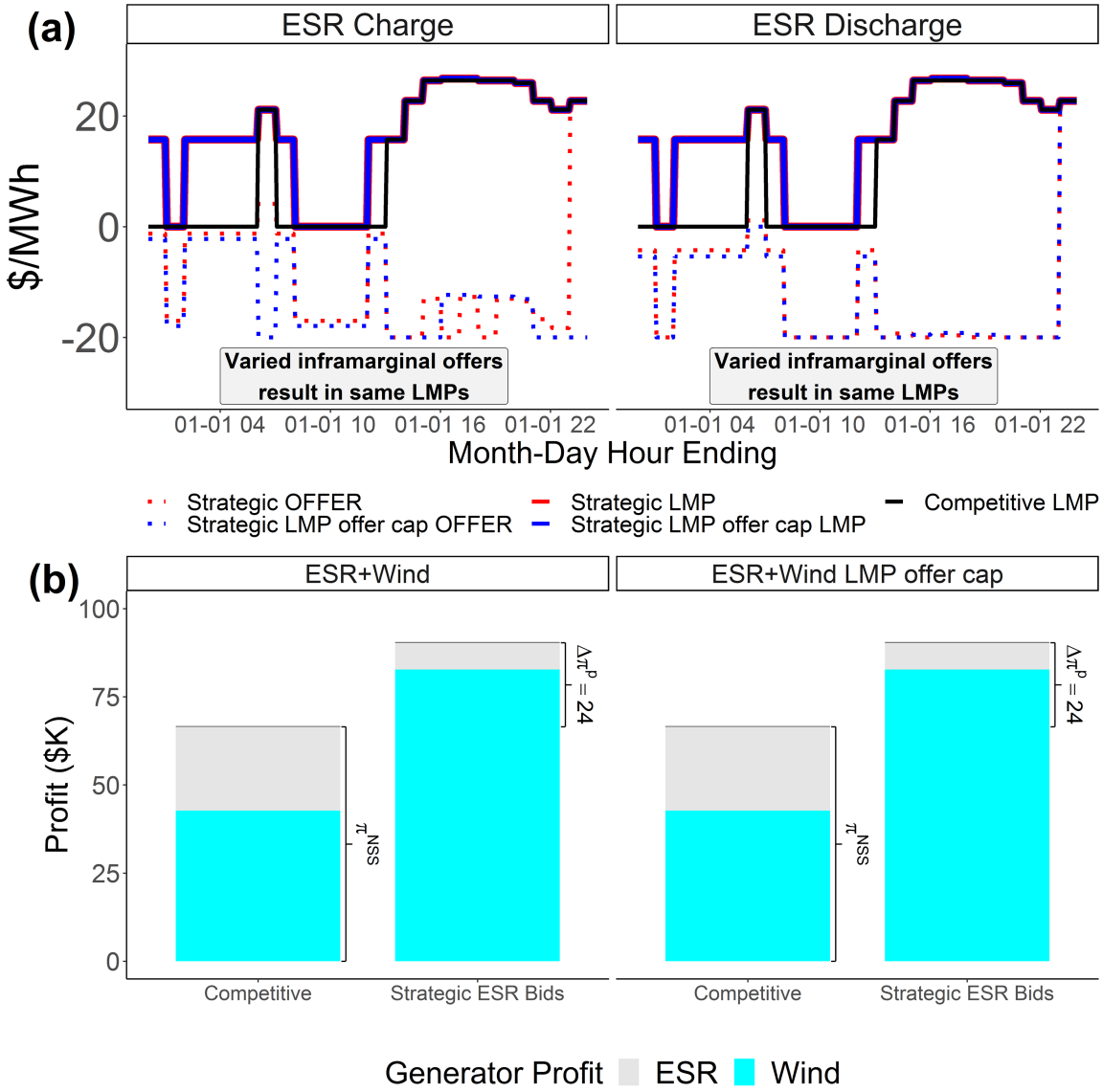

Figure 5.9: ESR and wind owning strategic entity offers for a single day (1 January) at hourly resolution (a). In the "offer cap LMP" case offers are constrained to be no greater than the ex-ante competitive LMP, but are allowed to be negative. Reduced ability of the ESR to set price does not eliminate strategic profit (b).

The ESR's ability to maintain strategic profits in cases like the one displayed in Figure 5.9 with inframarginal offers is different than for an inframarginal generator offer. Energy-limited resources face an additional energy-limit constraint in temporally co-optimized dispatch that generators do

not. Energy-limited ESRs can make use of this constraint when binding to affect clearing prices and quantities based on relative, inframarginal offers because, unlike generators, the energy-limit means they cannot be dispatched in all time periods.

That ESRs' ability to manipulate dispatch and pricing depends on relative offers over temporally co-optimized intervals suggests a direction for monitoring and mitigation: focus on relative offers. A simple, restrictive mitigation technique could involve requiring a temporally fixed, or "uniform" ESR offer for all co-optimized time periods. Running an additional case where the ESR is constrained to a single offer for charge and discharge over co-optimized intervals on the same example day as Figure 5.9 produces Figure 5.10.

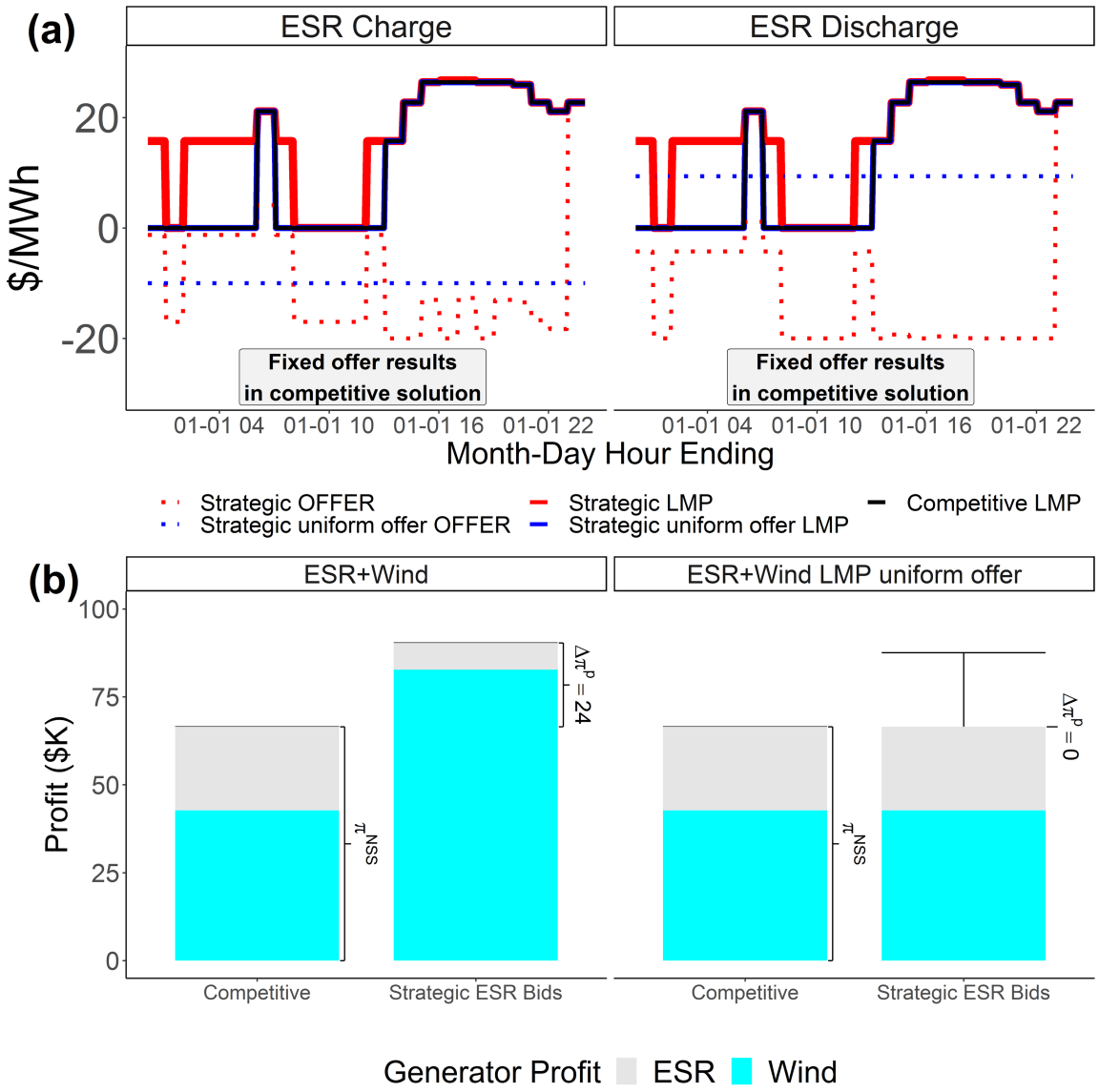

Figure 5.10: ESR and wind owning strategic entity offers for a single day (1 January) at hourly resolution (a). In the "uniform offer" case offers are constrained to a single value for charge and discharge, respectively. The uniform offer LMP is the same as the competitive LMP and is overplotted. The uniform offer reduces strategic profits to be equivalent to the competitive solution for this day (b).

In 5.10, the uniform offer requirement reproduces the competitive solution because the same two conditions as in Appendix C are met: the offer is inframarginal in all time periods and the energy limits on ESR dispatch are binding. Mathematically, the ESR offer SO_t must be lower than the generation costs G_t it displaces in all time periods, $SO_t < G_t, \forall t$, and the ESR dispatch $\sum sd_t$ must

be limited; $\sum sd_t < sd_t^{max} * T$, where T is the number of time periods. Under these two conditions the uniform offer disallows the ESR modifying the perceived social welfare gain of its energy-limited dispatch with different bids in different hours. With a uniform offer, the perceived social welfare gain is the difference between the offer and the production costs it replaces, resulting in discharge replacing generation the highest production cost time periods and charge increasing generation the lowest production cost time periods, as in the competitive case.

Requiring a uniform offer has potential drawbacks of over-mitigation, particularly if the offer is also required to be inframarginal, which the above cases and Appendix C require. A pivotal ESR not required to offer inframarginally can offer $SO_t \geq G_t$ for at least one t and set price in a time interval or intervals with a uniform bid above marginal cost. Offering $SO_t > G_t, \forall t$ will withhold capacity. A bid above marginal generation costs in all time periods could be justified by opportunity costs realized beyond the co-optimized time horizon. However, limiting the ability of ESRs to update a uniform DA offer in shorter time horizon balancing or RT markets risks over mitigation by reducing ESR option value if prices are higher (or lower) than expected. Even with these concerns, that inframarginal uniform offers mitigate the modeled optimal strategy of a profit-maximizing strategic entity with perfect foresight and binding ESR energy-limits to increase its portfolio's profits suggests a direction for ESR market monitoring and design considerations.

5.5 Discussion

Energy storage resources can facilitate integration of high levels of variable renewable energy, and market design must recognize their unique characteristics to prevent storage from manipulating high VRE, low marginal cost markets. This work highlights how price-making ESRs' ability to increase load and state of charge limitations enables additional latitude in bidding to manipulate dispatch and pricing and suggests ways that market monitors can avoid manipulation.

Modeling on a high VRE test system under perfect foresight suggests the most profitable strategy involves cross-product manipulation by bidding an ESR to raise prices and thus revenues received by other generators in the strategic entity's portfolio. ESRs can more commonly be pivotal at or

near buses with high VRE generation and congestion than system-wide – exactly where ESRs would be installed to reduce transmission related renewable energy curtailments [30] – suggesting those generation pockets deserve additional attention. While we model only energy markets, identified strategies could be extended to ancillary service markets providing nearer term profit opportunities for ESR participation [16, 31, 32]. Hybridization of ESRs with generators otherwise constrained to offer at low or zero marginal cost is another potential avenue for manipulating prices upward if bidding rules allow. Hybrid bidding rules are highly policy relevant given large quantities of renewable-ESR hybrid generators entering North American interconnection queues [12].

Strategic profits are limited with a deterministic bidding strategy under uncertainty about future loads and generation, but may be increased with more targeted or sophisticated strategies. Incorporating methods for predicting future market prices under uncertainty [33] and stochastic optimization [34] with a bi-level or other price-making model for ESRs suggest directions for future research. We assume fixed capacity and perfect information about the market operator's algorithm; both assumptions could be relaxed or extended to incorporate aspects of investment decision-making [35, 36] and algorithmic differences between forward and real-time markets, such as treatment of nonconvexities [37] and effects of financial products [24, 38].

We show ESRs can use their intertemporal energy limit constraint to change pricing and dispatch even with exclusively inframarginal offers. If the market operator respects ESRs' energy limits in optimizing its schedule and setting prices over multiple time intervals monitoring should look closely at relative inframarginal offers when ESRs are pivotal suppliers. We suggest uniform offer requirements as one approach. However, the assumption that market operators will respect energy limits through a SOC parameter or penalty is itself not a policy guarantee. The Electric Power Research Institute (EPRI) outlines four approaches for ESR participation in US wholesale markets under Federal Energy Regulatory Commission Order 841, including a self-SOC management option where the market operator considers only whether ESR offers are part of least-cost security constrained economic dispatch and ESRs are expected to self-update offers to maintain sufficient SOC to meet their schedules [39]. EPRI's modeling shows self-SOC management can have severe reliability implications if ESRs cannot meet their dispatch schedule due to SOC infeasibility in real-time [40].

Weighing reliability and market participation objectives highlights the broader point: market design reforms to accommodate the technical characteristics of ESRs must carefully consider objectives including competition, reliability, and rapid decarbonization. If updated monitoring and mitigation for ESRs and hybrids is not included alongside these objectives, ossified market design can undermine the benefits of competition and hinder rapid decarbonization using high shares of low marginal cost VRE and ESRs.

References

[1] J. H. Williams, A. DeBenedictis, R. Ghanadan, A. Mahone, J. Moore, W. R. Morrow, S. Price, and M. S. Torn, "The technology path to deep greenhouse gas emissions cuts by 2050: the pivotal role of electricity.," *Science (New York, N.Y.)*, vol. 335, pp. 53–9, Jan 2012. doi:https://doi.org/10.1126/science.1208365.

[2] J. H. Williams, R. A. Jones, B. Haley, G. Kwok, J. Hargreaves, J. Farbes, and M. S. Torn, "Carbon Neutral Pathways for the United States," *AGU Advances*, vol. 2, no. 1, 2021. doi:https://doi.org/10.1029/2020av000284.

[3] J. E. Bistline and D. T. Young, "Economic drivers of wind and solar penetration in the US," *Environmental Research Letters*, vol. 14, no. 12, 2019. doi:https://doi.org/10.1088/1748-9326/ab4e2d.

[4] N. Kittner, F. Lill, and D. M. Kammen, "Energy storage deployment and innovation for the clean energy transition," *Nature Energy*, vol. 2, no. 9, 2017. doi:https://doi.org/10.1038/nenergy.2017.125.

[5] A. Mileva, J. Johnston, J. H. Nelson, and D. M. Kammen, "Power system balancing for deep decarbonization of the electricity sector," *Applied Energy*, vol. 162, pp. 1001–1009, 2016. doi:https://doi.org/10.1016/j.apenergy.2015.10.180.

[6] W. W. Hogan, "Electricity wholesale market design in a low-carbon future," *Harnessing Renewable Energy in Electric Power*, pp. 113–136, 2010. https://scholar.harvard.edu/whogan/files/hogan_market_design_012310.pdf.

[7] E. Ela, M. Milligan, A. Bloom, A. Botterud, A. Townsend, and T. Levin, "Evolution of Wholesale Electricity Market Design with Increasing Levels of Renewable Generation," pp. 1–139, September, 2014. https://www.nrel.gov/docs/fy14osti/61765.pdf.

[8] P. L. Joskow, "Challenges for wholesale electricity markets with intermittent renewable generation at scale: The US experience," *Oxford Review of Economic Policy*, vol. 35, no. 2, pp. 291–331, 2019. doi:https://doi.org/10.1093/oxrep/grz001.

[9] K. Spees, S. A. Newell, W. Graf, and E. Shorin, "How States , Cities , and Customers Can Harness Competitive Markets to Meet Ambitious Carbon Goals," September, 2019. https://www.brattle.com/11975.

[10] C. Batlle, P. Rodilla, and P. Mastropietro, "Markets for Efficient Decarbonization: Revisiting Market Regulation and Design," *IEEE Power and Energy Magazine*, vol. 19, no. 1, pp. 20–28, 2021. doi:https://doi.org/10.1109/MPE.2020.3033393.

[11] CAISO, "Hybrid Resources Issue Paper," 2019. http://www.caiso.com/Documents/IssuePaper-HybridResources.pdf.

[12] W. Gorman, A. Mills, M. Bolinger, R. Wiser, N. G. Singhal, E. Ela, and E. O'Shaughnessy, "Motivations and options for deploying hybrid generator-plus-battery projects within the bulk power system," *Electricity Journal*, vol. 33, no. 5, p. 106739, 2020. doi:https://doi.org/10.1016/j.tej.2020.106739.

[13] M. Beuse, B. Steffen, and T. S. Schmidt, "Projecting the Competition between Energy-Storage Technologies in the Electricity Sector," *Joule*, vol. 4, no. 10, pp. 2162–2184, 2020. doi:https://doi.org/10.1016/j.joule.2020.07.017.

[14] G. He, Q. Chen, P. Moutis, S. Kar, and J. F. Whitacre, "An intertemporal decision framework for electrochemical energy storage management," *Nature Energy*, vol. 3, no. 5, pp. 404–412, 2018. doi:https://doi.org/10.1038/s41560-018-0129-9.

[15] E. Ela, B. Kirby, A. Botterud, C. Milostan, I. Krad, and V. Koritarov, "The Role of Pumped Storage Hydro Resources in Electricity Markets and System Operation," *Proceedings of the Hydro Vision International*, pp. 1–10, 2013. https://www.nrel.gov/docs/fy13osti/58655.pdf.

[16] P. Denholm, W. Cole, A. W. Frazier, K. Podkaminer, N. Blair, P. Denholm, W. Cole, A. W. Frazier, K. Podkaminer, N. Blair, S. C. Denholm, W. Cole, A. W. Frazier, K. Podkaminer, N. Blair, and T. Four, "Storage Futures Study The Four Phases of Storage Deployment : A Framework for the Expanding Role of Storage in the U.S. Power System," 2021. https://www.nrel.gov/docs/fy21osti/77480.pdf.

[17] FERC, "Electric Storage Participation in Markets Operated by Regional Transmission Organizations nad Independent System Operators," *Federal Energy Regulatory Commission Docket No. RM16-23-000*, 2018. https://ferc.gov/sites/default/files/2020-06/Order-841.pdf.

[18] C. Ruiz and A. J. Conejo, "Pool strategy of a producer with endogenous formation of locational marginal prices," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1855–1866, 2009. doi:https://doi.org/10.1109/TPWRS.2009.2030378.

[19] M. Miletić, H. Pandžić, and D. Yang, "Operating and investment models for energy storage systems," *Energies*, vol. 13, no. 18, pp. 1–33, 2020. doi:https://doi.org/10.3390/en13184600.

[20] H. Mohsenian-Rad, "Coordinated Price-Maker Operation of Large Energy Storage Units in Nodal Energy Markets," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 786–797, 2016. doi:https://doi.org/10.1109/TPWRS.2015.2411556.

[21] Y. Ye, D. Papadaskalopoulos, R. Moreira, and G. Strbac, "Investigating the impacts of price-taking and price-making energy storage in electricity markets through an equilibrium programming model," *IET Generation, Transmission and Distribution*, vol. 13, no. 2, pp. 305–315, 2019. doi:https://doi.org/10.1049/iet-gtd.2018.6223.

[22] E. Tómasson, M. R. Hesamzadeh, and F. A. Wolak, "Optimal offer-bid strategy of an energy storage portfolio: A linear quasi-relaxation approach," *Applied Energy*, vol. 260, no. May 2019, p. 114251, 2020. doi:https://doi.org/10.1016/j.apenergy.2019.114251.

[23] J. B. Cardell, C. C. Hitt, and W. W. Hogan, "Market power and strategic interaction in electricity networks," *Resource and Energy Economics*, vol. 19, no. 1-2, pp. 109–137, 1997. doi:https://doi.org/10.1016/s0928-7655(97)00006-7.

[24] C. Lo Prete, N. Guo, and U. V. Shanbhag, "Virtual Bidding and Financial Transmission Rights: An Equilibrium Model for Cross-Product Manipulation in Electricity Markets," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 953–967, 2019. doi:https://doi.org/10.1109/TPWRS.2018.2875624.

[25] E. Nasrolahpour, H. Zareipour, W. D. Rosehart, and S. J. Kazempour, "Bidding strategy for an energy storage facility," *19th Power Systems Computation Conference, PSCC 2016*, 2016. doi:https://doi.org/10.1109/PSCC.2016.7541016.

[26] X. Fang, V. Krishnan, and B. M. Hodge, "Strategic offering for wind power producers considering energy and flexible ramping products," *Energies*, vol. 11, no. 5, 2018. doi:https://doi.org/10.3390/en11051239.

[27] P. Brown, "US Renewable Electricity: How Does Wind Generation Impact Competitive Power Markets," *Congressional Research Service, Nov*, 2012. https://www.files.ethz.ch/isn/155870/R42818_20121107.pdf.

[28] C. Barrows, E. Preston, A. Staid, G. Stephen, J. P. Watson, A. Bloom, A. Ehlen, J. Ikaheimo, J. Jorgenson, D. Krishnamurthy, J. Lau, B. McBennett, and M. O'Connell, "The IEEE Reliability Test System: A Proposed 2019 Update," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 119–127, 2020. doi:https://doi.org/10.1109/TPWRS.2019.2925557.

[29] EIA, *United States Energy Information Administration January 2021 Monthly Energy Review*, vol. 0035. 2021. https://www.eia.gov/totalenergy/data/monthly/pdf/mer.pdf.

[30] P. Denholm and T. Mai, "Timescales of energy storage needed for reducing renewable energy curtailment," *Renewable Energy*, vol. 130, pp. 388–399, September 2019. doi:https://doi.org/10.1016/j.renene.2018.06.079.

[31] T. Lee, "Exploring Frequency Regulation Market Transformation," November, 2017. https://kleinmanenergy.upenn.edu/research/publications/energy-storage-in-pjm-exploring-frequency-regulation-market-transformation/.

[32] B. Xu, Y. Shi, D. S. Kirschen, and B. Zhang, "Optimal Battery Participation in Frequency Regulation Markets," *IEEE Transactions on Power Systems*, vol. 33, pp. 6715–6725, Nov 2018. doi:https://doi.org/10.1109/TPWRS.2018.2846774.

[33] B. Xu, A. Botterud, and M. Korpås, "Operational Valuation for Energy Storage under Multi-stage Price Uncertainties," *arXiv*, 2019. doi:https://doi.org/10.1109/cdc42340.2020.9304081.

[34] A. J. C. Morales, M. Carrión, and J. M., *Decision Making Under Uncertainty in Electricity Markets*. Springer, 2010. doi:https://doi.org/10.1007/978-1-4419-7421-1

[35] R. Fernández-Blanco, Y. Dvorkin, B. Xu, Y. Wang, and D. S. Kirschen, "Optimal Energy Storage Siting and Sizing: A WECC Case Study," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 733–743, 2017. doi:https://doi.org/10.1109/TSTE.2016.2616444.

[36] E. Nasrolahpour, H. Zareipour, and W. D. Rosehart, "Battery investment by a strategic wind producer: A scenario-based decomposition approach," *Electric Power Systems Research*, vol. 182, no. January, p. 106255, 2020. doi:https://doi.org/10.1016/j.epsr.2020.106255.

[37] Y. Ye, D. Papadaskalopoulos, J. Kazempour, and G. Strbac, "Incorporating Non-Convex Operating Characteristics into Bi-Level Optimization Electricity Market Models," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 163–176, 2020. doi:https://doi.org/10.1109/TPWRS.2019.2925317.

[38] A. Jha and F. A. Wolak, "Can Financial Participants Improve Price Discovery and Efficiency in Multi-Settlement Markets with Trading Costs?," *NBER Working Paper*, 2019. http://www.nber.org/papers/w25851.pdf.

[39] N. G. Singhal and E. G. Ela, "Pricing Impacts of State of Charge Management Options for Electric Storage Resources," in *2020 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–6, Aug 2020. doi:https://doi.org/10.1109/PESGM41954.2020.9282160.

[40] N. G. Singhal and E. Ela, "Incorporating Electric Storage Resources into Wholesale Electricity Markets While Considering State of Charge Management Options," tech. rep., CIGRE US National Committe 2019 Grid of the Future Symposium, 2019.

[41] SPP, "Dynamic Opportunity Cost Mitigated Energy Offer Framework For Electric Storage," *Southwest Power Pool Market Monitoring Unit*, 2018. https://www.spp.org/documents/58525/dynamic%20opp%20cost%20esr%20mitigated%20offer%20framework_20180824.pdf.

[42] S. Kasina, "Essays on Unit Commitment and Interregional Cooperation in Transmission Planning," 2017. http://jhir.library.jhu.edu/handle/1774.2/44712.

[43] J. Fortuny-Amat and B. McCarl, "A Representation and Economic Interpretation of a Two-Level Programming Problem," *The Journal of the Operational Research Society*, vol. 32, no. 9, pp. 783–792, 1981. doi:https://doi.org/10.2307/2581394.

[44] C. Root, H. Presume, D. Proudfoot, L. Willis, and R. Masiello, "Using battery energy storage to reduce renewable resource curtailment," *2017 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2017*, pp. 1–5, 2017. doi:https://doi.org/10.1109/ISGT.2017.8085955.

[45] A. Alanazi and A. Khodaei, "Optimal battery energy storage sizing for reducing wind generation curtailment," *IEEE Power and Energy Society General Meeting*, vol. 2018-Janua, no. July, pp. 1–5, 2018. doi:https://doi.org/10.1109/PESGM.2017.8274599.

[46] ERCOT, "ERCOT Nodal Protocols," 2020. http://www.ercot.com/content/wcm/libraries/218985/December_10__2020_Nodal_Protocols.pdf.

A Model Explanation and Formulation Appendix

A.1 Additional model explanation

The full formulation of the model includes ramping and an implementation of linearized unit commitment but excludes security constraints. Generally additional security constraints and markets make pivotal supplier conditions more common and increase opportunities for strategic bidding, while demand side offers reduce supply side market power. Minimizing as-bid cost of serving load ignoring ancillary service utility is equivalent to maximizing social welfare assuming a uniform value of marginal electricity demand at or above the market clearing price cap. By anticipating the market operator's decisions, the strategic leading entity can submit bids to increase profits from its portfolio of assets.

Configurable options in running cases allow implementation of offer mitigation constraints based on ex-ante price forecasts, similar to an approach proposed by Southwest Power Pool's (SPP) Market Monitoring Unit [41]. Other configurable options allow changes to temporal indexing of ESR offers in the upper level and running the model as a single level linear program with fixed cost-based bids to compare results to the fully competitive solution for a given parameterized case.

Configurable options in the model explored in the paper include:

1. Reformulation as a single level linear program for market clearing assuming fixed bids. Produces competitive solution when all resources are assumed to offer at marginal cost. (all results sections)

2. Hybridization of co-located ESR and generation resources owned by strategic bidder. (Section 5.4.2)

3. Multiple market settlements with reduced bidding and dispatch recourse in real-time. (Section 5.4.3)

4. Offer mitigation constraints including requiring ESR discharge offers not exceed an offer cap based on ex-ante expected market clearing prices, or requiring a single, uniform discharge or charge ESR offer over co-optimized temporal intervals. (Section 5.4.4)

A full formulation of the model follows.

A.2 Notation

The following model notation is also used for equations in Ch. 6.

Sets

$g \in G$, generators, indexed by g

$gc \in GC \subset G$, subset of generators owned by strategic entity, indexed by gc

$gnc \in GNC \subset G$, subset of generators not owned by strategic entity, indexed by gnc

$gs \in GS$, linearized segments of generator heat rate curves, indexed by gs

$guc \in GUC \subset G$, subset of unit commitment generators, indexed by guc

$gnuc \in GNUC \subset G$, subset of non-unit commitment generators, indexed by $gnuc$

$l \in L$, transmission lines connecting nodes in power system, indexed by l

$s \in S$ energy storage resources, indexed by s

$ss \in SS \subset S$, subset of storage owned by strategic entity, indexed by ss

$nss \in NSS \subset S$, subset of storage not owned by strategic entity, indexed by nss

$t \in T$, timepoints, indexed by t

$z \in Z$, buses in power system, indexed by z

Parameters

$CAP_{t,g}$, generator capacity, [MW]

$CAPA^{RT}_{t,g}$, Real-time available generation capacity of strategic entity [MW]

CE_s, energy storage charge efficiency, unitless

$CMAX_s$, maximum charge rate of energy storage resource [MW]

$CMAXO_{t,s}$, Maximum energy storage resource charge offer with offer mitigation [$/MWh]

$CDA_{t,s}$, Day-ahead energy storage resource charge for use in real-time [$/MWh]

DE_s, energy storage resource discharge efficiency, unitless

$DMAX_s$, maximum discharge rate of energy storage resource [MW]

$DMAXO_{t,s}$, Maximum energy storage resource discharge offer with offer mitigation [$/MWh]

$DDA_{t,s}$, Day-ahead energy storage resource discharge for use in real-time [$/MWh]

$GMC_{g,gs}$, marginal cost of generation on segment [$/MWh]

$GSL_{g,gs}$, fraction of generator capacity on marginal segment

$\Delta\lambda^{DA}_{t,z==SZL_s}$, Observed change in ESR bus clearing price in DA case when strategic vs. competitive (used only in RT cases)

$L_{t,z}$, gross load at bus [MWh]

NLC_g, No-load costs of committed generator ($/timepoint)

$PMIN_g$, Minimum online generation of committed unit commitment generator

$RBUS_t$, label of reference bus

RR_g, Generator ramp rate (MW/timepoint)

S_l, susceptance of transmission line [Siemens]

$SA_{t,g}$, fraction of generator capacity scheduled to be available

SC_g, Generator start-up costs [$/start]

$SMAX_s$, maximum state of charge of energy storage resource [MWh]

$TXFL_l$, zone or node from which transmission line originates

$TXTL_l$, zone or node to which transmission line goes

$TFCAP_l$, capacity of transmission line from zone or bus [MW]

$TTCAP_l$, capacity of transmission line to zone or bus [MW]

UE, Cost of unserved energy or offer cap [$/MWh]

$VMAX_z$, maximum voltage angle of bus (positive)

$VMIN_z$, minimum voltage angle of bus (negative)

ZL_g, Generator zone or bus label (assignment)

ZLS_g Energy storage resource zone or us label (assignment)

Decision Variables

$gd_{t,g}$, generator dispatch [MWh]

$gdn_{t,g}$, shutdown status of generator with unit commitment [0,1]

$gsd_{t,g,gs}$, dispatch on generator segment [MWh]

$gso_{t,guc,gs}$, offer on generator segment for generators with unit commitment [$/MWh]

$go_{t,gnuc}$, offer for generators without unit commitment [$/MWh]

$gopstat_{t,g}$, Online operating status of generator with unit commitment [0,1]

$gup_{t,g}$, Startup status of generator with unit commitment [0,1]

$nucgd_{t,g}$, non unit commitment generator dispatch [MWh]

$sc_{t,s}$, charging of energy storage resource [MWh]

$sd_{t,s}$, discharging of energy storage resource [MWh]

$soc_{t,s}$, state-of charge (SOC) of energy storage resource [MWh]. SOC is by definition determined by discharge and charge, so SOC-related terms are written as a summation of charge and discharge in subsequent equations.

$sofc_{t,s}$, Energy storage resource charge offer [$/MWh]

$sofd_{t,s}$, Energy storage resource discharge offer [$/MWh]

$txmwh_{t,l}$, real power flow on transmission line, [MWh]

$va_{t,z}$, bus voltage angle

$\alpha_{t,s}$, Energy storage resource charge lower bound dual variable

$\beta_{t,s}$, Energy storage resource discharge lower bound dual variable

$\gamma_{t,s}$, Tight energy storage resource operation constraint dual variable

$\nu_{t,s}^{max}/\nu_{t,s}^{min}$, State of Charge (SOC) constraint dual variable

χ_s, Final SOC dual variable

ξ_s, energy storage resource cycle constraint dual variable

$\mu_{t,l}^{max}/\mu_{t,l}^{min}$, transmission flow dual variables

$\psi_{t,z}^{max}/\psi_{t,z}^{min}$, voltage angle dual variables

$\lambda_{t,z}$, power balance equation dual variable; interpreted as locational marginal price (LMP)

$\eta_{t,s}$, SOC balance equation dual variable

$\phi_{t,g,}^{max}/\phi_{t,g}^{min}$, generator dispatch dual variables

$\varphi_{t,g,gs}^{max}/\varphi_{t,g,gs}^{min}$, generator segment dispatch dual variables

$\sigma_{t,g}^{up}/\sigma_{t,g}^{down}$, generator ramp up/down dual variables

$\pi_{t,g}$, unit commitment generator operating status change dual variable

$\omega_{t,g}^{max}/\omega_{l,g}^{min}$, non-unit commitment generator dispatch dual variables

A.3 Model formulation

Upper-level objective function

The upper-level objective function maximizes the profit of a single entity's competitively owned generators and ESRs.

$$\begin{aligned} MAX\ \lambda_{t,z} * (& \sum_{t,g \in GC \cap GUC, z == ZL_g}^{T,G} gd_{t,g} \\ + & \sum_{t,g \in GC \cap GNUC, z == ZL_g}^{T,G} nucgd_{t,g} + \sum_{t,s \in SS, z == ZLS_s}^{T,SS} sd_{t,s}) \\ & - \sum_{t,g \in GC \cap GUC, gs}^{T,G,GS} GMC_{g,gs} * gsd_{t,g,gs} \\ & - \sum_{t,s \in SS, z == ZLS_s}^{T,S} \lambda_{t,z} * sc_{t,s} \end{aligned} \tag{A.1A}$$

The upper-level objective function can optionally be configured to include no-load and startup

costs when the model is configured to include tight relaxed unit commitment (TRUC) [42].

$$
\begin{aligned}
MAX\ \lambda_{t,z} * (& \sum_{t,g \in GC \cap GUC, z == ZL_g}^{T,G} gd_{t,g} \\
+ & \sum_{t,g \in GC \cap GNUC, z == ZL_g}^{T,G} nucgd_{t,g} + \sum_{t,s \in SS, z == ZLS_s}^{T,SS} sd_{t,s}) \\
& - \sum_{t,g \in GC \cap GUC, gs}^{T,G,GS} GMC_{g,gs} * gsd_{t,g,gs} \\
& - \sum_{t,g \in GC \cap GUC}^{T,G} NLC_g * gopstat_{t,g} \\
& - \sum_{t,g \in GC \cap GUC}^{T,G} SC_g * gup_{t,g} \\
& - \sum_{t,s \in SS, z == ZLS_s}^{T,S} \lambda_{t,z} * sc_{t,s}
\end{aligned}
\tag{A.1B}
$$

Upper-level offer constraints

Generator offer constraints are either optional constraints implemented to reduce solution time or based on common market rules in North American wholesale electricity markets. Implemented offer constraints also prevent physical withholding of unit commitment generators, a common market rule.

Equation	Notes	Eq. No.
$2 * GMC_{g,gs} \geq gso_{t,g,gs} \geq GMC_{g,gs}$, $\forall g \in GC \cap GUC$	Generators may offer only up to two times their marginal cost and must offer at least their marginal costs. Helps reduce solution time.	(A.2)
$gso_{t,g,gs} \geq gso_{t,g,gs-1}$, $\forall g \in GC \cap GUC$	Each generator segment must offer at least the offer on the previous segment. Common wholesale market rule.	(A.3)
$UE \geq gso_{t,g,gs}$, $\forall g \in GC \cap GUC$	Generators may offer only up to the market cap, assumed \$2000/MWh in paper cases. Common wholesale market rule.	(A.4)
$0 == go_{t,g}$, $\forall g \in GC \cap GNUC$	Variable renewable generators are assumed to have zero marginal cost and offer. Helps reduce solution time and reflects low marginal cost of VRE.	(A.5)

Lower-level objective function

The lower-level objective function minimizes the as-bid cost of serving firm load

$$\forall \lambda \in arg[MIN \sum_{t,g,gs}^{T,GUC,GS} gso_{t,g,gs} * gsd_{t,g,gs} + \sum_{t,g}^{T,GNUC} go_{t,g} * nucgd_{t,g} + \sum_{t,s}^{T,SS} (sofd_{t,s} * sd_{t,s} - sofc_{t,s} * sc_{t,s})] \tag{A.6A}$$

As with the upper level objective, the lower level objective may be optionally configured to include TRUC terms.

$$\forall \lambda \in arg[MIN \sum_{t,g,gs}^{T,GUC,GS} gso_{t,g,gs} * gsd_{t,g,gs} + \sum_{t,g}^{T,GNUC} go_{t,g} * nucgd_{t,g} + \sum_{t,s}^{T,SS} sofd_{t,s} * sd_{t,s} - sofc_{t,s} * sc_{t,s}] + \sum_{t,g \in GUC}^{T,GC} SC_g * gup_{t,g} + \sum_{t,g \in GUC}^{T,GC} NLC_g * gopstat_{t,g} \tag{A.6B}$$

Lower-level constraints

Equation	Notes	Eq. No.
$L_{t,z} == \sum_{t,g,gs,z==ZL_g}^{T,GUC,GS} gsd_{t,g,gs}$ $+\sum_{t,g,z==ZL_g}^{T,GNUC} nucgd_{t,g} + \sum_{t,s,z==ZLS_s}^{T,S} sd_{t,s} - sc_{t,s} +$ $\sum_{t,l}^{T,L} (txmw_{t,l,TXTL_l==z} - txmw_{t,l,TXFL_l==z}) : \lambda_{t,z}$	Load-balance constraint	(A.7)
$va_{t,z==RBUS_t} = 0$	Voltage angle at reference bus	(A.8)
$DMAX_s * CMAX_s \geq DMAX_s * sc_{t,s}$ $+CMAX_s * sd_{t,s} : \gamma_{t,s}$	Tight storage dispatch	(A.9)
$sc_{t,s} \geq 0 : \alpha_{t,s}, sd_{t,s} \geq 0 : \beta_{t,s}$	Non-negative storage charge and discharge	(A.10)
$SMAX_s \geq \sum_{t}^{1,\ldots,t} (CE_s * sc_{t,s} - DE_s * sd_{t,s})$ $\geq 0 : \nu_{t,s}^{max}, \nu_{t,s}^{min}$	Sum of storage charge and discharge in previous timepoints stays below max SOC	(A.11)
$SMAX_s \geq \sum_{t}^{T} sd_{t,s} \geq 0 : \xi_s$	Limits storage discharge during a day to a single cycle	(A.12)
$\sum_{t}^{T} (CE_s * sc_{t,s} - DE_s * sd_{t,s}) == 0 : \chi_s$	Final storage SOC balance	(A.13)
$CAP_{t,g} * SA_{t,g} * gopstat_{t,g} \geq gd_{t,g}$ $\geq gmin_{t,g} : \phi_{t,g}^{max}, \phi_{t,g}^{min}, \forall g \in GUC$	Generator dispatch limited by max available capacity	(A.14)
$CAP_{t,g} * SA_{t,g} * GSL_{g,gs} \geq gsd_{t,g,gs}$ $\geq 0 : \varphi_{t,g,gs}^{max}, \varphi_{t,g,gs}^{min}, \forall g \in GUC$	Generator piecewise segment dispatch limited by max available capacity	(A.15)
$gd_{t,g} == \sum_{gs}^{GS} gsd_{t,g,gs} \forall g \in GUC$	Sum of generator segment dispatch equivalent to total generator dispatch	(A.16)
$CAP_{t,g} * SA_{t,g} \geq nucgd_{t,g} \geq 0$ $: \omega_{t,g}^{max}, \omega_{t,g}^{min}, \forall g \in GNUC$	Generator dispatch limited by max available capacity	(A.17)
$TTCAP_l \geq txmwh_{t,l} \geq TFCAP_l$ $: \mu_{t,l}^{max}, \mu_{t,l}^{min}$	Transmission flows bounded by positive and negative maxima	(A.18)
$VMAX_z \geq va_{t,z} \geq VMIN_z : \psi_{t,z}^{max}, \psi_{t,z}^{min}$	Voltage angle limits	(A.19)
$txmwh_{t,l} == S_l * (va_{t,z==TXTL_z} - va_{t,z==TXFL_z})$	DCOPF constraints	(A.20)

A set of additional optional constraints implement TRUC, associated minimum generation levels for online generators, and ramping limits.

Equation	Notes	Eq. No.
$gmin_{t,g} == PMIN_g * SA_{t,g} * CAP_{t,g} * gopstat_{t,g},$ $\forall g \in GUC$	Minimum generation scales with online capacity in TRUC	(A.21)
$gopstat_{t,g} - gopstat_{t-1,g} == gup_{t,g} - gdn_{t,g}$ $: \pi_{t,g}, \forall g \in GUC$	Unit commitment status changes only with startups and shutdowns	(A.22)
$gd_{t-1,g} - gmin_{t-1,g} + RR_g * gopstat_{t,g}$ $\geq gd_{t,g} - gmin_{t,g} : \sigma_{t,g}^{up}, \forall g \in GUC$	In TRUC upward operating status changes are bound by ramp rate	(A.23)
$gd_{t,g} - gmin_{t,g} + RR_g * gopstat_{t,g}$ $\geq gd_{t-1,g} - gmin_{t-1,g} : \sigma_{t,g}^{down}, \forall g \in GUC$	In TRUC downward operating status changes are bound by ramp rate	(A.24)

Optional constraints for implementing offer mitigation and binding day-ahead offers

The model may be configured to optionally include additional constraints for two settlement functionality or offer mitigation, and may also be configured to run as a cost-minimizing linear program.

Our approach to implementing offer mitigation requires developing an ex-ante price forecast to parameterize and mitigate against maximum offers, similar to an approach suggested by SPP's market monitor [41]. To achieve this we run a preliminary cost-minimizing dispatch of all resources as a linear program and extract the resulting bus LMPs. Offers are then mitigated against this price stream in each timepoint.

As one way of investigating the effect of uncertainty in loads and other offers when submitting strategic offers, we also allow the model to be configured to first optimize a DA run, then run a RT case where storage offers or quantities can be fixed to some of their DA values.

Equation	Notes	Eq. No.
$DMAXO_{t,s} \geq sofd_{t,s}$	Storage discharge offer mitigated to ex-ante max (set equal to LMP in Section 5.4.4)	(A.25)
$CMAXO_{t,s} \geq sofc_{t,s}$	Storage charge offer mitigated to ex-ante max	(A.26)
$sd_{t,s} == DDA_{t,s} (\forall t \vert DDA_{t,s} > 0, CAPA^{RT}_{t,g} \geq CMAX_s, \Delta\lambda^{DA}_{t,z==SZL_s} > 0)$	Real-time storage discharge is equal to DA discharge in time periods with pivotal DA dispatch and sufficient RT strategic wind generation	(A.27)
$sc_{t,s} == CDA_{t,s} (\forall t \vert CDA_{t,s} > 0, CAPA^{RT}_{t,g} \geq DMAX_s, \Delta\lambda^{DA}_{t,z==SZL_s} > 0)$	Real-time storage charge is equal to DA charge in time periods with pivotal DA dispatch and sufficient RT strategic wind generation	(A.28)

Derivation of KKT conditions and MPEC reformulation

Since the lower level problem is a linear program, it can be reformulated using its Karush-Kuhn-Tucker (KKT) conditions as an equivalent Mathematical Program with Equilibrium Constraints (MPEC). The four KKT conditions are stationarity, complementary slackness, primal feasibility, and dual feasibility. The lower-level LP is primal and dual feasible, so the two tasks for MPEC reformulation are to write and the stationarity and complementary conditions of the lower-level problem and add them as constraints. Pyomo allows for representation of complementarity constraints in the formulation of optimization problems. The lower level objective and pre-existing constraints

are unchanged.

Stationarity conditions

Stationarity conditions are derived by taking partial derivatives of the problem's Lagrangian function with respect to the decision variables, then constraining the resulting equations to be zero to ensure stationarity. These equations are written below.

$$gso_{t,g,gs} - \lambda_{t,z==ZL_g} + \phi^{max}_{t,g} - \phi^{min}_{t,g} + \varphi^{max}_{t,g,gs} - \varphi^{min}_{t,g,gs} == 0, \forall g \in GUC \tag{A.29}$$

$$SC_g - \pi_{t,g} == 0, \forall g \in GUC \tag{A.30}$$

$$NLC_g - SA_{t,g} * CAP_{t,g} * \phi^{max}_{t,g} + PMIN_g * SA_{t,g} * CAP_{t,g} * \phi^{min}_{t,g} + \pi_{t,g} == 0^5 \tag{A.31}$$

$$sofd_{t,s} + CMAX_s * \gamma_{t,s} - \beta_{t,s} - DE_s * \eta_{t,s} + \xi_s - \lambda_{t,z==SZL_s} == 0, \forall s \in SS \tag{A.32}$$

$$-sofc_{t,s} + DMAX_s * \gamma_{t,s} - \alpha_{t,s} + CE_s * \eta_{t,s} + \lambda_{t,z==SZL_s} == 0, \forall s \in SS \tag{A.33}$$

$$CMAX_s * \gamma_{t,s} - \beta_{t,s} + DE_s * (\chi_s - \sum_{t}^{t,...N_T} \nu^{max}_{t,s} - \nu^{min}_{t,s}) + \xi_s - \lambda_{t,z==SZL_s} == 0, \forall s \in NSS \tag{A.34}$$

[5]Eq. A.31 is extended to include ramp rate-related terms when TRUC is implemented

$$DMAX_s * \gamma_{t,s} - \alpha_{t,s} - CE_s * (\chi_s - \sum_{t}^{t,...N_T} \nu_{t,s}^{max} - \nu_{t,s}^{min}) + \lambda_{t,z==SZL_s} == 0, \forall s \in NSS \quad \text{(A.35)}$$

$$\eta_{t,s} - \eta_{t+1,s} + \nu_{t,s}^{max} - \nu_{t,s}^{min}, \forall t \leq N_t \quad \text{(A.36)}$$

$$\eta_{N_t,s} + \nu_{N_t,s}^{max} - \nu_{N_t,s}^{min} == 0 \quad \text{(A.37)}$$

$$\mu_{t,l}^{max} - \mu_{t,l}^{min} + \psi_{t,z}^{max} - \psi_{t,z}^{min} == 0 \quad \text{(A.38)}$$

$$-\lambda_{t,z==ZL_g} + \omega_{t,g}^{max} - \omega_{t,g}^{min} == 0 \quad \text{(A.39)}$$

Complementarity constraints

Complementary slackness conditions take the form $u_i * h_i(x) = 0$, $\forall i$. Such an equation is nonlinear as both terms contain decision variables. However, it may be rewritten as a complementarity constraint $0 \leq u_i \perp h_i(x) \geq 0$, which says either $u_i = 0$, or $h_i(x) = 0$, or both. These coupled constraints are then each themselves linear. The complementarity constraints used in the model are shown below.

$$0 \leq \gamma_{t,s} \perp DMAX_s * CMAX_s - DMAX_s * sc_{t,z} - CMAX_s * sd_{t,z} \geq 0 \quad \text{(A.40)}$$

$$0 \leq \alpha_{t,s} \perp sc_{t,z} \geq 0 \quad \text{(A.41)}$$

$$0 \leq \beta_{t,s} \perp sd_{t,z} \geq 0 \quad \text{(A.42)}$$

$$0 \leq \nu_{t,s}^{max} \perp SMAX_s - \sum_{t}^{1,\ldots,t} (CE_s * sc_{t,s} - DE_s * sd_{t,s}) \geq 0 \tag{A.43}$$

$$0 \leq \nu_{t,s}^{min} \perp \sum_{t}^{1,\ldots,t} (CE_s * sc_{t,s} - DE_s * sd_{t,s}) \geq 0 \tag{A.44}$$

$$0 \leq \xi_s \perp SMAX_s - \sum_{t}^{T} sd_{t,s} \geq 0 \tag{A.45}$$

$$0 \leq \mu_{t,l}^{max} \perp TTCAP_l - txmw_{t,l} \geq 0 \tag{A.46}$$

$$0 \leq \mu_{t,l}^{min} \perp txmw_{t,l} - TFCAP_l \geq 0 \tag{A.47}$$

$$0 \leq \psi_{t,z}^{max} \perp VMAX_z - va_{t,z} \geq 0 \tag{A.48}$$

$$0 \leq \psi_{t,z}^{min} \perp va_{t,z} - VMIN_z \geq 0 \tag{A.49}$$

$$0 \leq \phi_{t,g}^{max} \perp SA_{t,g} * CAP_{t,g} * gopstat_{t,g} - gd_{t,g} \geq 0 \tag{A.50}$$

$$0 \leq \phi_{t,g}^{min} \perp gd_{t,g} - gmin_{t,g} \geq 0 \tag{A.51}$$

$$0 \leq \varphi_{t,g,gs}^{max} \perp SA_{t,g} * CAP_{t,g} * GSL_{t,g,gs} - gsd_{t,g,gs} \geq 0 \tag{A.52}$$

$$0 \leq \varphi_{t,g,gs}^{min} \perp gsd_{t,g,gs} \geq 0 \tag{A.53}$$

$$0 \leq \sigma_{t,g}^{up} \perp gd_{t-1,g} - gmin_{t-1,g} + RR_g * gopstat_{t,g} - (gd_{t,g} - gmin_{t,g}) \geq 0 \tag{A.54}$$

$$0 \leq \sigma_{t,g}^{down} \perp gd_{t,g} - gmin_{t,g} + RR_g * gopstat_{t,g} - (gd_{t-1,g} - gmin_{t-1,g}) \geq 0 \tag{A.55}$$

$$0 \leq \omega_{t,g}^{max} \perp SA_{t,g} * CAP_{t,g} - nucgd_{t,g} \geq 0 \tag{A.56}$$

$$0 \leq \omega_{t,g}^{min} \perp nucgd_{t,g} \geq 0 \tag{A.57}$$

MILP reformulation using Big-M and strong duality

The objective function of the MPEC is still (A.1B) and remains non-linear because terms such as $\lambda_{t,z} * gd_{t,g}$ contain multiple decision variables (in the example, the LMP and generator dispatch are both decision variables). The following steps linearize the objective with an equivalent formulation by making use of the lower-level objective and complementarity conditions.

First, strong duality theory holds the objective of the primal problem is equivalent to the objective of the corresponding dual problem. The equivalence between the primal and dual objectives for the

lower level problem (recall the primal objective is eq. (A.6B)) at the optimum is

$$\begin{aligned}
&\sum_{t,g,gs}^{T,GUC,GS} gso_{t,g,gs} * gsd_{t,g,gs} + \sum_{t,g}^{T,GNUC} go_{t,g} * nucgd_{t,g} \\
&\sum_{t,s}^{T,S} (sofd_{t,s} * sd_{t,s} - sofc_{t,s} * sc_{t,s}) + \sum_{t,g\in GUC}^{T,GC} SC_g * gup_{t,g} + \sum_{t,g\in GUC}^{T,GC} NLC_g * gopstat_{t,g} \\
&= \sum_{t,s}^{T,S} DMAX_s * CMAX_s * (-\gamma_{t,s}) + \sum_{t,s}^{T,S} SMAX_s * (-\nu_{t,s}^{max}) + \sum_{s}^{S} SMAX_s * (-\xi_s) \\
&\quad + \sum_{t,l}^{T,L} TTCAP_{t,l} * (-\mu_{t,z}^{max}) + \sum_{t,l}^{T,L} TFCAP_{t,l} * (\mu_{t,z}^{min}) \\
&\quad + \sum_{t,z}^{T,Z} VMAX_z * (-\psi_{t,z}^{max}) + \sum_{t,z}^{T,Z} VMIN_z * (\psi_{t,z}^{min}) \\
&\quad + \sum_{t,z}^{T,Z} L_{t,z} * \lambda_{t,z} + \sum_{t,g,gs}^{T,G,GS} CAP_{t,g} * SA_{t,g} * GSL_{g,gs} * (-\varphi_{t,g,gs}^{max}) \\
&\quad + \sum_{t,g}^{T,G} CAP_{t,g} * SA_{t,g} * (-\omega_{t,g}^{max})
\end{aligned} \tag{A.58}$$

Using eq. (A.29), (A.32), and (A.33) to substitute for generator and storage offer-related decision variables and reformulate the primal objective in (A.58), as well as (A.5) to set $go_{t,g} = 0$,

$\forall g \in GC \cap GNUC$, the dual and primal objectives are equivalent to

$$\sum_{t,g,gs}^{T,GUC,GS} (\lambda_{t,z==ZL_g} - \phi^{max}_{t,g} + \phi^{min}_{t,g} - \varphi^{max}_{t,g,gs} + \varphi^{min}_{t,g,gs} - \sigma^{up}_{t,g} + \sigma^{down}_{t,g}) * gsd_{t,g,gs}$$
$$+ \sum_{t,g}^{T,GNUC} (\lambda_{t,z==ZL_g} - \omega^{max}_{t,g} + \omega^{min}_{t,g}) * nucgd_{t,g}$$
$$+ \sum_{t,g\in GUC}^{T,GC} SC_g, gup_{t,g} + \sum_{t,g\in GUC}^{T,GC} NLC_g * gopstat_{t,g} \tag{A.59}$$
$$+ \sum_{t,s\in SS}^{T,S} (\lambda_{t,z==SZL_s} - CMAX_s * \gamma_{t,s} + \beta_{t,s} - DE_s * (\chi_s - \sum_{t}^{t,...N_T} \nu^{max}_{t,s} - \nu^{min}_{t,s}) - \xi_s) * sd_{t,s}$$
$$+ \sum_{t,s\in SS}^{T,S} (-\lambda_{t,z==SZL_s} - DMAX_s * \gamma_{t,s} + \alpha_{t,s} + CE_s * (\chi_s - \sum_{t}^{t,...N_T} \nu^{max}_{t,s} - \nu^{min}_{t,s})) * sc_{t,s}$$

This allows us to set the dual objective from (A.58) equal to (A.59). Then, rearrange by moving

all linear terms to the left hand (dual objective) side of the equation.

$$
\begin{aligned}
&\sum_{t,s}^{T,S} DMAX_s * CMAX_s * (-\gamma_{t,s}) + \sum_{t,s}^{T,S} SMAX_s * (-\nu_{t,s}) + \sum_{s}^{S} SMAX_s * (-\xi_{t,s}) \\
&+ \sum_{t,l}^{T,L} TTCAP_{t,l} * (-\mu_{t,l}^{max}) + \sum_{t,l}^{T,L} TFCAP_{t,l} * (\mu_{t,l}^{min}) + \sum_{t,z}^{T,Z} VMAX_z * (-\psi_{t,z}^{max}) \\
&+ \sum_{t,z}^{T,Z} VMIN_z * (\psi_{t,z}^{min}) + \sum_{t,z}^{T,Z} L_{t,z} * \lambda_{t,z} + \sum_{t,g,gs}^{T,G,GS} CAP_{t,g} * SA_{t,g} * GSL_{g,gs} * (-\varphi_{t,g,gs}^{max}) \\
&+ \sum_{t,g}^{T,G} CAP_{t,g} * SA_{t,g} * (-\omega_{t,g,gs}^{max}) - \sum_{t,g \in GUC}^{T,GC} SC_g * gup_{t,g} - \sum_{t,g \in GUC}^{T,GC} NLC_g * gopstat_{t,g} \\
&= \sum_{t,g,gs}^{T,GUC,GS} (\lambda_{t,z==ZL_g} - \phi_{t,g}^{max} + \phi_{t,g}^{min} - \varphi_{t,g,gs}^{max} + \varphi_{t,g,gs}^{min} - \sigma_{t,g}^{up} + \sigma_{t,g}^{down}) * gsd_{t,g,gs} \\
&\qquad + \sum_{t,g}^{T,GNUC} (\lambda_{t,z==ZL_g} - \omega_{t,g}^{max} + \omega_{t,g}^{min}) * nucgd_{t,g} \\
&+ \sum_{t,s \in SS}^{T,S} (\lambda_{t,z==SZL_s} - CMAX_s * \gamma_{t,s} + \beta_{t,s} - DE_s * (\chi_s - \sum_{t}^{t,...N_T} \nu_{t,s}^{max} - \nu_{t,s}^{min}) - \xi_s) * sd_{t,s} \\
&+ \sum_{t,s \in SS}^{T,S} (-\lambda_{t,z==SZL_s} - DMAX_s * \gamma_{t,s} + \alpha_{t,s} + CE_s * (\chi_s - \sum_{t}^{t,...N_T} \nu_{t,s}^{max} - \nu_{t,s}^{min})) * sc_{t,s}
\end{aligned}
\tag{A.60}
$$

The relationship between the right hand side of (A.60) and the upper level objective (A.1B) can then be more straightforwardly seen after grouping the right hand side LMP-related terms, though showing this step is omitted. The remaining task is to substitute for the nonlinear, non-LMP related terms on the right hand side (e.g., $\phi_{t,g}^{min} * gsd_{t,g,gs}$ using complementarity relationships for the decision variables in (A.40)-(A.57), removing terms equal to zero, and cancelling resulting non-zero terms with left hand side terms where equivalent. The indexing of the right hand side across only strategic generators means, generally, strategic terms cancel and non-strategic generator and storage terms remain in the resulting objective. Finally, the generator marginal cost-related term is subtracted

from both sides of the equation so the right hand side reproduces eq. (A.1B).

$$
\begin{aligned}
&\sum_{t,z}^{T,Z} L_{t,z} * \lambda_{t,z} - \sum_{t,g\in GUC,gs}^{T,GNC,GS} CAP_{t,g} * SA_{t,g} * GSL_{g,gs} * \varphi_{t,g,gs}^{max} \\
&- \sum_{t,g\in GUC}^{T,GNC} CAP_{t,g} * SA_{t,g} * \phi_{t,g}^{max} - \sum_{t,g\in GNUC}^{T,GNC} CAP_{t,g} * SA_{t,g} * \omega_{t,g}^{max} \\
&- \sum_{t,g\in GUC,gs}^{T,G,GS} GMC_{g,gs} * gsd_{t,g,gs} - \sum_{t,g\in GUC}^{T,G} SC_g * gup_{t,g} - \sum_{t,g\in GUC}^{T,G} NLC_g * gopstat_{t,g} \\
&+ \sum_{t,s\in NSS}^{T,S} (-\gamma_{t,s} * DMAX_s * CMAX_s - \nu_{t,s}^{max} * SMAX_s) - \sum_{s\in NSS}^{S} (\xi_t * SMAX_s) \\
&= [\sum_{t,g\in guc,z==ZL_g}^{T,GC} gd_{t,g} + \sum_{t,g\in GNC,z==ZL_g}^{T,GC} nucgd_{t,g} + \sum_{t,s,z==ZLS_s}^{T,SS,Z} (sd_{t,s} - sc_{t,s})] * \lambda_{t,z} \\
&- \sum_{t,g\in GUC,gs}^{T,GC,GS} GMC_{g,gs} * gsd_{t,g,gs} - \sum_{t,g\in GUC}^{T,GC} SC_g * gup_{t,g} - \sum_{t,g\in GUC}^{T,GC} NLC_g * gopstat_{t,g}
\end{aligned}
\tag{A.61}
$$

The result is a linear left hand side of the equation equivalent to the original nonlinear objective. The left hand side of eq. (A.61) may now be used as a linear objective for solving the MILP.

Last, the complementarity constraints must be linearized. This can be done using the so-called "Big-M" method first described by Fortuny-Amat and McCarl [43]. This approach rewrites a complementarity constraint of the form $0 \leq u_i \perp h_i(x) \geq 0$ as a set of two constraints, $0 \leq u_i \leq M(1-\nu_i)$ and $0 \leq h_i(x) \leq M * \nu_i$, where M is a large enough constant to balance not limiting the feasible space of the problem[6] and ν_i is an auxiliary binary variable. Note the use of the binary variable linking the constraints means at least one of the two constraints must be equivalent to zero, thus satisfying the original complementary slackness condition. The rewritten transformations of each complementarity constraint are not included but can be provided on request. These transformations are automatically undertaken in the model code by use of the big-M implementation in Pyomo's generalized disjunctive programming library.

[6] If M is too large this can unnecessarily extend the feasible space and increase solution time, so choosing a value for M is something of an art; most values are 5000 in our implementation, reflecting upward limits on generator offers well in excess of the assumed \$2000/MWh bid cap.

The creation of a set of auxiliary binary variables for each constraint reformulation, along with the linearization of the objective, means the problem is now a MILP and can be readily solved by commercial solvers like CPLEX.

B Additional Data Description Appendix

B.1 Two-settlement functionality

We simplify two common aspects of two-settlement markets due to computational and data limitations and to better compare DA and RT results in cases incorporating uncertainty. First, DA and RT dispatch are commonly optimized using different algorithms; in particular, DA includes binary variables for unit commitment while RT is a linear program with fixed unit commitment. In the bi-level model the market operator's problem must be convex, so both settlements are cleared with the same lower level problem formulation and exclude or linearize (Appendix A) unit commitment. Second, we maintain the same temporal co-optimization across the entire day in both DA and RT, while market operators more commonly limit co-optimized look ahead to a single five-minute RT interval. Maintaining the same temporal co-optimization allows us to better compare the optimality of fixed DA ESR dispatch quantities when settled without recourse against RT deviations in load and generation. Because of the additional computational burden of co-optimizing 288 five-minute intervals instead of 24 hourly ones we reformat five-minute RT VRE generation and load data to hourly average equivalents and clear the RT market at hourly resolution. The primary purpose of these simplifications is to make DA and RT settlements more comparable for incorporating uncertainty in load, wind, and solar generation in sensitivity analysis.

B.2 RTS-GMLC generator offer data

Table B-1: Generator location, capacity, and offer data used in all cases.

BusID	Group	Capacity (MW)	Type	Units	Bid Segments	Bid Capacities (MW)	Marginal Cost ($/MWh)
301	U20	20	Oil CT	2	4	8,4,4,4	$87.26, $87.26, $99.7, $105.37
301	U55	55	Gas CT	2	4	21,11,11,11	$28.47, $28.47, $29.29, $43.74
302	U20	20	Oil CT	2	4	8,4,4,4	$87.26, $87.26, $99.7, $105.37
302	U55	55	Gas CT	2	4	21,11,11,11	$33.79, $33.79, $38.38, $38.64
307	U55	55	Gas CT	2	4	21,11,11,11	$27.89, $27.89, $29.22, $35.47
313	U355	355	Gas CC	1	4	170,61.67,61.67,61.67	$15.73, $15.73, $26.76, $33.75
315	U12	12	Oil ST	5	4	5,2.33,2.33,2.33	$75.44, $75.44, $100.4, $124.1
315	U55	55	Gas CT	3	4	21,11,11,11	$26.4, $26.4, $26.9, $31.86
316	U155	155	Coal	1	4	62,31,31,31	$21.12, $21.12, $21.29, $27.27
318	U355	355	Gas CC	1	4	170,61.67,61.67,61.67	$26.85, $26.85, $27.28, $31.53
321	U355	355	Gas CC	1	4	170,61.67,61.67,61.67	$22.73, $22.73, $25.91, $33.95
322	U55	55	Gas CT	2	4	21,11,11,11	$23.33, $23.33, $26.5, $27.89
323	U355	355	Gas CC	2	4	170,61.67,61.67,61.67	$26.43, $26.43, $30.28, $31.73
303	WIND	847	Wind	1	1	As available	$0
308	RTPV	100.9	Solar RTPV	1	1	As available	$0
309	WIND	148.3	Wind	1	1	As available	$0
310	PV	103.3	Solar PV	1	1	As available	$0
312	PV	189.2	Solar PV	1	1	As available	$0
313	PV	93.3	Solar PV	1	1	As available	$0
313	RTPV	806	Solar RTPV	1	1	As available	$0
314	PV	144.3	Solar PV	1	1	As available	$0
317	WIND	799.1	Wind	1	1	As available	$0
319	PV	188.2	Solar PV	1	1	As available	$0
320	PV	51.6	Solar PV	1	1	As available	$0
322	U50	50	Hydro	1	1	As available	$0
324	PV	152.3	Solar PV	1	1	As available	$0

While the supply curve will change based on generation availability even when all generators offer at marginal cost, and congestion may result in different online generators and marginal cost than suggested by intersecting a single supply and demand curve, Figure B-1 constructs an average supply curve for the modeled month and compares to average and peak demand.

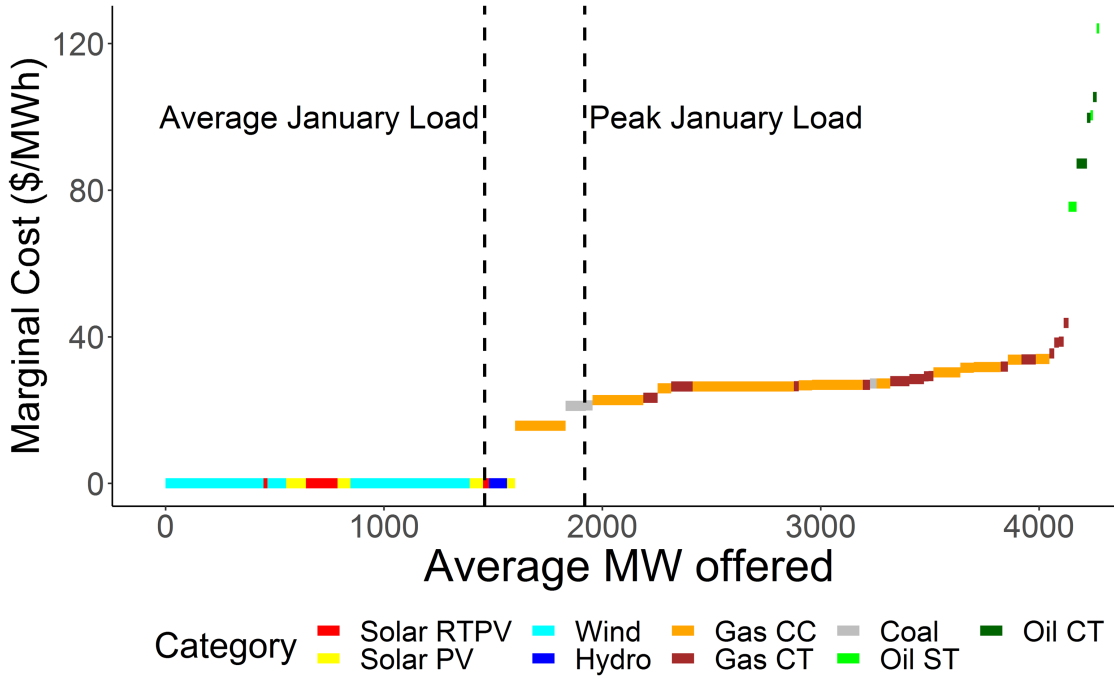

Figure B-1: January supply curve. Variable renewable energy resources displayed at average hourly generation for the month. Supply curve ignores congestion and curtailment.

B.3 LMPs for all 25 buses in cases without ESRs

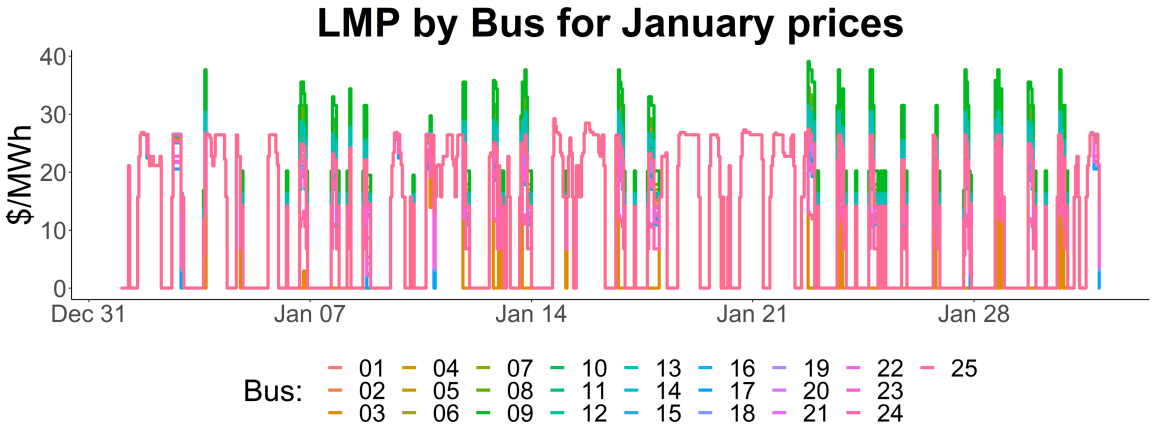

Figure B-2: LMP at all 25 buses.

B.4 Loads and renewable generation day-ahead forecast error distributions

The model uses DA RTS-GMLC data to settle the forward market and RT RTS-GMLC data to settle the operational balancing market. Changes in dispatch and pricing result from deviations in

RT data from DA expectations, termed forecast error, for load, solar PV, solar rooftop PV (RTPV), and wind resources. As a result, load net of PV and wind, defined as net load, shows the summed deviation of all forecast errors and is also included in the empirical distribution of forecast errors for the modeled month shown in Figure B-3.

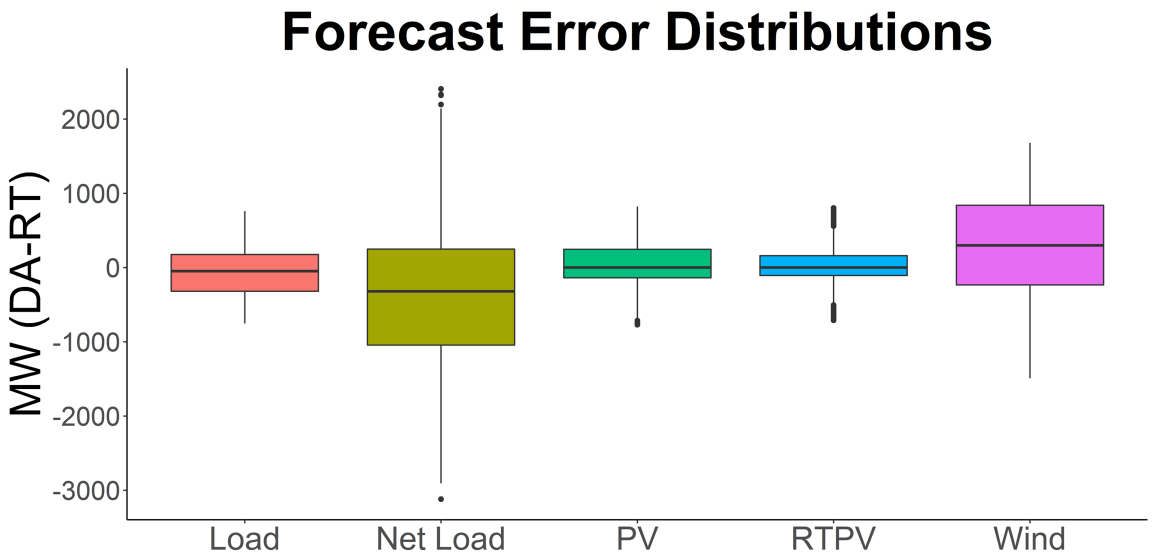

Figure B-3: Distribution of differences between hourly DA forecast and RT actuals of renewable generation by type (solar PV, solar rooftop PV, Wind) and load in MW for the Zone 3 RTS-GMLC test system. Net load is load net of VRE (i.e., wind and solar PV, including rooftop PV).

B.5 Additional information on Results Section 5.4.1 Demonstrating the Three Strategies

In case A (ESR only, Figure 5.5) the strategic entity owns only the 300MW/900MWh ESR. It implements two strategies to increase its profit: increasing the applicable LMP at its bus when discharging, and decreasing the LMP at its bus when charging. Full results are shown in Table B-2 to break out additional profits earned by the strategic entity and its effect on payments for serving firm load compared to the competitive case.

Table B-2: Case A (ESR Only) results for month of January. Changes are shown in red to indicate increased costs to consumers and in green to indicate increased profit for the strategic entity.

A: Category	B: Competitive	C: Strategic	D: Change
1: Load Payments[a]($M)	$5.223	$5.259	+0.7%
2: Storage Charging Cost ($M)	-$.048	-$0.39	-18.8%
3: Storage Discharging Revenue ($M)	$.355	$0.399	+12.4%
4: Storage Profit ($M)	$\pi^{p,NSS} = \$.306$	$\pi^{p,SS} = \$.360$	$\Delta\pi^{p} = \$.056$, +17.6%
5: Storage Profit ($/MWh discharged)	$\frac{\pi^{p,NSS}}{\sum sd^{S}} = \$9.9$	$\frac{\pi^{p,SS}}{\sum sd^{S}} = \$11.6$	$\frac{\Delta\pi^{p}}{MWh} = \$1.7$, + 17.6%

[a] Load payments are calculated as the inner product of bus clearing price and bus load

The strategies employed in case A (ESR Only) achieve additional profit when the ESR is pivotal: the inclusion of its charging load or discharging generation changes which generator(s) is/are[7] marginal and set price at bus 03. By adjusting its ESR bid the strategic entity can increase or decrease the cost of marginal supply at the cleared quantity of generation.

Case B extends case A by including the large (847 MW) wind generator at bus 03 in the strategic entity's portfolio ("ESR+Wind"). The strategic objective is to maximize the joint profits of the wind and ESR by modifying ESR bids; wind is constrained to offer at no more than its $0/MWh marginal cost. There is often congestion on a transmission line interconnecting bus 03 to the rest of the system (Figure 5.3) due to the large amount of zero marginal cost wind generation in many hours relative to available transmission capacity. Because of the prevalent congestion, the ESR can be used by a strategic entity when pivotal to alleviate congestion and increase price at bus 03. When wind generation exceeds additional ESR charging load the ability to increase price is a profitable cross-product strategy. Notably, since co-locating of ESRs with VRE is often suggested as a welfare-improving strategy due to reduction in curtailment and increased deliverability of low-cost and low-emission VRE [30, 44, 45] or to take advantage of incentives for ESR charging from VRE like the Investment Tax Credit [12], Table B-3 shows that at least some of this welfare for a given co-located ESR installation[8] could be absorbed by strategic bidding.

[7] When the system is congested there may be more than one marginal generator; equivalently, how clearing prices change depends on where load is added to the system.

[8] The installation of the ESR may still be welfare improving compared to a no-ESR system even with strategic bidding, but we do not investigate investment decision-making.

Table B-3: Case B (ESR+Wind) profits for month of January.

A: Category	B: Competitive	C: Strategic	D: Change
1: Load Payments[a]($M)	$5.223	$5.369	+2.8%
2: Storage Profit ($M)	$.306	$0.040	-86.9%
3: Storage Profit ($/MWh discharged)	$9.9	$1.4	-86.9%
4: Wind Profit ($M)	$0.751	$2.002	+166.6%
5: Incremental Profit Associated with Storage Ownership ($M)	$\pi^{p,NSS} = \$.306$	$\Delta\pi^{p} = \$1.292$	+368.7%
6: Incremental Profit Associated with Storage Ownership ($/MWh discharged)	$\frac{\pi^{p,NSS}}{\sum sd^{S}} = \$9.9$	$\frac{\pi^{p,SS}}{\sum sd^{S}} = \$46.3$	+368.7%

[a] Load payments are calculated as the inner product of bus clearing price and bus load

B.6 Additional information on Results Section 5.4.2

The sensitivity analysis in Figure 5.6 shows that under the perfect foresight assumption additional ESR capacity and duration have declining marginal value, as seen in the decreasing monthly incremental profit. Increased capacity has more effect than increased duration on total profits, suggesting the ability to offer more capacity in a single time period is a larger contributor to perfect foresight ESR profits than how often an offered quantity can be accepted and dispatched. Profits do not scale linearly with capacity as the quantity offer need to be pivotal and change clearing price in a given time period will be system dependent, particularly in the presence of congestion.

Figure 5.7 compares hybrid and co-located profits. Hybrid profits are higher because of additional assumed ability to submit a higher bid incorporating the wind with the ESR offer as a hybrid. Importantly, in practice different bidding rules can be applied to different types of resources. For example, section 4.4.9.3 of the Electricity Reliability Council of Texas' (ERCOT) December 2020 nodal protocols [46] only allows generators to update energy offer curves prior to an operating hour, but specifically provides a carve out for ESRs to update offers until any time prior to intra-hourly RT Security Constrained Economic Dispatch (SCED), giving ESRs additional offering latitude.[9] Market operators and monitors should think carefully about whether hybridization allows additional bidding latitude not usually afforded to one of the hybridized resources individually.

[9]Nodal Protocol Revision Request 1058 would update language to allow all generators to update energy offer curves at any time prior to RT SCED, and is available online at http://www.ercot.com/mktrules/issues/NPRR1058#keydocs

B.7 Additional information on Results Section 5.4.3

To investigate incorporation of uncertainty the model is configured to run as two temporally sequential settlements. The settlements can be conceived as a financial forward market where entities submit bids, and real-time operational market where bids are fixed and any deviations from the forward market are settled. The setup is similar to two-settlement day-ahead and real-time wholesale market settlement periods in North American markets, with the additional assumptions that the market operator clears the market using the same algorithm for the same temporal resolution at each settlement interval[10] and ignoring financial products (e.g., virtual bidding).

Maintaining strategic profits under uncertainty is a delicate balance. If an offer expected to change the clearing price does not, the strategic entity makes no additional profit and may lose profit compared to market operator dispatch if the ESR is charged or discharged with non-zero opportunity cost compared to dispatch in alternative time periods. Because of this reality and fact that the congestion alleviation strategy depends on supplying an appropriate quantity of ESR charge or discharge to the market to change clearing price and increase wind revenue, we implement the following assumptions and strategy:

1. The strategic entity is assumed to have perfect foresight of the forward DA market (equivalently, it could update its forward offers prior to real-time based on full knowledge of how changing offers would change the forward market clearing).

2. The strategic entity has perfect foresight of its own wind generation in RT, but otherwise does not update its DA expectation of other loads and generation.

3. The strategic entity knows which of its DA offers changed prices. It self-dispatches exactly that DA optimal quantity in the RT market if it has sufficient wind generation (>300 MW ESR installed capacity in modeled case case) to ensure a price-setting offer will still increase joint profits.

[10] Day-ahead markets are commonly cleared for the full day at hourly resolution and incorporate unit commitment, while real-time markets are cleared at 5-minute resolution with more limited forward temporal co-optimization and assume fixed unit commitment, eliminating binary decision variables.

4. The market operator dispatches any ESR capacity not self-dispatched to minimize production costs, subject to cycling constraints.

Taken together, this heuristic enables increased focus on pivotal hours under uncertainty, but is not an upper bound on strategic profits under uncertainty. A strategic entity with a more accurate updated forecast for load and generation in RT than simply maintaining its DA forecast would do better in our model. This strategy is reflected mathematically in the implemented constraints in Appendix A reproduced below.

$$sd_{t,s} == DDA_{t,s}(\forall t | DDA_{t,s} > 0, CAPA^{RT}_{t,g} \geq CMAX_s, \Delta\lambda^{DA}_{t,z==SZL_s} > 0) \tag{B.1}$$

$$sc_{t,s} == CDA_{t,s}(\forall t | CDA_{t,s} > 0, CAPA^{RT}_{t,g} \geq DMAX_s, \Delta\lambda^{DA}_{t,z==SZL_s} > 0) \tag{B.2}$$

Results compare the profit earned by the DA optimized bids fixed in the real-time balancing market to the perfect information real-time strategy as well as a fully competitive approach where the strategic entity offers at marginal cost. Results comparing these three strategies for Section 5.4.1's Case B parameterization for the same month of data are shown in Figure 5.8.

C Mathematical Exposition on ESR Offers Appendix

C.1 Example derivation

To show ESRs face different constraints relevant to offer mitigation, consider the two time period (indexed by $t \in$ 1,2) dispatch cost minimization problem in eq. (C.1-C.8). The market operator's objective in this problem is to minimize the costs of serving residual demands D_t using a generator G and an ESR S. For simplicity assume the ESR enters the two time periods fully charged at state of charge SOC and has sufficient capacity to fully discharge in either time period. To reduce constraints in the problem assume non-negative output from G is unbounded but comes at constant costs C_1 and C_2 per unit of output in each of the two time periods. Generation from G is denoted g_t, from S by sd_t, offers from S are denoted sd_t, and applicable dual variables for each constraint follow the

colon in eq. (C.3-C.8). The two time period problem is formulated:

$$min_{g_1,g_2,sd_1,sd_2} C_1 g_1 + C_2 g_2 + SO_1 sd_1 + SO_2 sd_2 \tag{C.1}$$

$$\text{S.T. } D_1 = g_1 + sd_1 : \lambda_1 \tag{C.2}$$

$$D_2 = g_2 + sd_2 : \lambda_2 \tag{C.3}$$

$$g_1 \geq 0 : \phi_1 \tag{C.4}$$

$$g_2 \geq 0 : \phi_2 \tag{C.5}$$

$$sd_1 \geq 0 : \alpha_1 \tag{C.6}$$

$$sd_2 \geq 0 : \alpha_2 \tag{C.7}$$

$$SOC \geq sd_1 + sd_2 : \eta \tag{C.8}$$

Note that eq. (C.8) assumes the market operator will monitor SOC in formulating its problem. The monitored SOC assumption may not hold in frameworks allowing self-monitoring SOC for ESR and hybrid market participation [12]. The Lagrangian for the two time period problem is:

$$\begin{aligned}\mathcal{L}(g, sd, \lambda, \alpha, \phi, \eta) = C_1 g_1 + C_2 g_2 + SO_1 sd_1 + SO_2 sd_2 - \lambda_1(g_1 + sd_1 - D_1) \\ -\lambda_2(g_2 + sd_2 - D_2) - \phi_1 g_1 - \phi_2 g_2 - \alpha_1 sd_1 - \alpha_2 sd_2 - \eta(SOC - sd_1 - sd_2)\end{aligned} \tag{C.9}$$

At the optimum the Karush-Kuhn-Tucker (KKT) conditions will hold. The KKT conditions include stationarity for primal variables in the objective, which are the generator (g_t) and storage dispatch (sd_t) in each time period. These stationarity conditions are written

$$\frac{\Delta\mathcal{L}(\overrightarrow{g}, \overrightarrow{sd}, \overrightarrow{\lambda}, \overrightarrow{\alpha}, \overrightarrow{\phi}, \eta)}{\Delta g_1} = C_1 - \lambda_1 - \phi_1 = 0 \tag{C.10}$$

$$\frac{\Delta\mathcal{L}(\overrightarrow{g}, \overrightarrow{sd}, \overrightarrow{\lambda}, \overrightarrow{\alpha}, \overrightarrow{\phi}, \eta)}{\Delta g_2} = C_2 - \lambda_2 - \phi_2 = 0 \tag{C.11}$$

$$\frac{\Delta\mathcal{L}(\overrightarrow{g}, \overrightarrow{sd}, \overrightarrow{\lambda}, \overrightarrow{\alpha}, \overrightarrow{\phi}, \eta)}{\Delta sd_1} = SO_1 - \lambda_1 - \alpha_1 + \eta = 0 \tag{C.12}$$

$$\frac{\Delta\mathcal{L}(\overrightarrow{g}, \overrightarrow{sd}, \overrightarrow{\lambda}, \overrightarrow{\alpha}, \overrightarrow{\phi}, \eta)}{\Delta sd_2} = SO_2 - \lambda_2 - \alpha_2 + \eta = 0 \tag{C.13}$$

The two generator-related stationarity conditions have no shared decision variables, and, when combined with the complementary slackness conditions for the inequalities (C.4) and (C.5), can be used to show G will set a non-zero price $C_t = \lambda_t$ in either time period if dispatched. However, the same does not hold for the ESR stationarity conditions, which both contain the SOC dual variable η and can be substituted and rewritten

$$SO_2 - SO_1 - \lambda_1 - \alpha_2 + \alpha_1 = 0 \tag{C.14}$$

Eq. (C.14) itself provides the critical insight: the effect of ESR offers on clearing prices λ_t in all time periods is a function of the relative storage offers SO_t in each time period. When pivotal the ESR can make use of this fact to change its dispatch, and thus pricing, based on its relative offers, even when its absolute offers are constrained to be inframarginal $C_t > SO_t$ in all time periods.

Under the additional assumptions the ESR is a pivotal supplier in each hour individually ($D_t < SOC$) but not both hours jointly ($\sum D_t > SOC$), we can show the minimum revenues π^{ESR} accrued assuming an inframarginal non-negative ESR offer ($C_t > SO_t \geq 0$) in both hours are:

$$\min \pi^{ESR} = \begin{cases} (SOC - D_1)C_2, SO_2 - SO_1 > C_2 - C_1 \\ (SOC - D_2)C_1, SO_2 - SO_1 < C_2 - C_1 \end{cases} \tag{C.15}$$

The proof of this result using the above assumptions and problem defined in Eq. (C.1)-(C.8) as well as an explanation of the solution(s) when $SO_2 - SO_1 = C_2 - C_1$ is in Appendix C.2. An example suffices to show there exists practical relevance. Assume $SOC = 50$, $D_1 = 10$, $D_2 = 45$, $C_1 = 20$, $C_2 = 25$. Under these conditions if the ESR offers its full quantity $SOC = 50$ to the market without a price offer (equivalently, $SO_2 = SO_1 = 0$), the market operator will minimize

dispatch costs by using as much of the ESR as feasible in the higher cost time period 2, and use the remainder in time period 1, so $sd_1^*, sd_2^*, g_1^*, g_2^*$=5,45,5,0 and the objective value using Eq. (3) is 100. Price is set by G at 20 in time period 1, so ESR revenues are at least $\lambda_1^* sd_1^* = 20 * 5 = 100$. However, by filling in eq. (C.15) for the assumed parameterization:

$$\min \pi^{ESR} = \begin{cases} 1000, SO_2 - SO_1 > 5 \\ 100, SO_2 - SO_1 < 5 \end{cases} \tag{C.16}$$

1000 is greater than 100, so the ESR can guarantee a greater minimum profit by offering $SO_2 - SO_1 > 5$. Because achieving this profit depends on the relative ESR offers, whether ESR offers are capped based on an ex-ante maximum is irrelevant to the profits in eq. (C.16): so long as the range of allowable offers exceeds 5 the ESR can guarantee the higher minimum revenue by offering $SO_2 - SO_1 > 5$. The dispatch assuming this offer is $sd_1^*, sd_2^*, g_1^*, g_2^*$=10,40,0,5, with the minimum payments for dispatching G now being 5*25=125, greater than the optimal value of 100 when the ESR did not make a price offer.[11] The difference $SO_2 - SO_1 > 5$ in the offers makes the market operator perceive ESR dispatch is more valuable in time period 1 than time period 2, so using as much ESR as feasible under SOC and demand constraints in time period 1 minimizes the perceived total dispatch cost objective. Assuming perfect information a pivotal ESR supplier can exploit this fact by submitting appropriate temporally differentiated offers to change optimal dispatch and pricing.

C.2 Derivation with additional assumptions

Assume as in Appendix B that a ESR enters a two time period model with full state of charge SOC and can be discharge fully in either time period. Assume this state of charge SOC is sufficient to serve residual demand D_t in either time period individually, but not both; $SOC > D_t, SOC < \sum D_t$. Additionally, assume the ESR must offer its available discharge capability into the market at a price less than the offer of the generator G in either time period; $SO_t < C_t$. This last assumption

[11] The value of the objective when $SO_2 - SO_1 > 5$ will also change based on the change in perceived costs of discharging the ESR, but for simplicity assume the actual costs of the ESR do not change, just the offer. Then the only change in actual production costs comes from the changing dispatch of the generator.

guarantees the market operator will discharge the ESR within the two time period window, so Eq. (C.8) can be rewritten as the equality Eq. (C.1.9) and we have the below modifications of the model presented in Section 5.4.4:

$$\text{Assume } SO_t < C_t, SOC > D_t, SOC < \sum D_t \tag{C.1.1}$$

$$\min C_1 g_1 + C_2 g_2 + SO_1 sd_1 + SO_2 sd_2 \tag{C.1.2}$$

$$\text{S.T. } D_1 = g_1 + sd_1 : \lambda_1 \tag{C.1.3}$$

$$D_2 = g_2 + sd_2 : \lambda_2 \tag{C.1.4}$$

$$g_1 \geq 0 : \phi_1 \tag{C.1.5}$$

$$g_2 \geq 0 : \phi_2 \tag{C.1.6}$$

$$sd_1 \geq 0 : \alpha_1 \tag{C.1.7}$$

$$sd_2 \geq 0 : \alpha_2 \tag{C.1.8}$$

$$SOC = sd_1 + sd_2 : \eta \tag{C.1.9}$$

Deriving the KKT conditions for this problem and substituting yields

$$C_1 - \lambda_1 - \phi_1 = 0 \tag{C.1.10}$$

$$C_2 - \lambda_2 - \phi_2 = 0 \tag{C.1.11}$$

$$SO_2 - SO_1 - \lambda_2 + \lambda_1 - \alpha_2 + \alpha_1 = 0 \tag{C.1.12}$$

$$g_1 \phi_1 = 0 \tag{C.1.13}$$

$$g_2 \phi_2 = 0 \tag{C.1.14}$$

$$sd_1 \alpha_1 = 0 \tag{C.1.15}$$

$$sd_2 \alpha_2 = 0 \tag{C.1.16}$$

$$g_t, sd_t, \phi_t, \alpha_t \geq 0 \tag{C.1.17}$$

Using the assumptions and Eq. (C.1.3-C.1.4) offers only three feasible solutions: either $sd_1 = D_1$, $sd_2 = D_2$, or $sd_t < D_t$. The ESR cannot fully serve residual demand in both time periods because $SOC < \sum D_t$ and must fully discharge per eq. (C.1.9), so no other option remains. We can then proceed to solve these three cases (1)-(3) in parallel, below:

(1) $sd_1 = D_1$	(2) $sd_2 = D_2$	(3) $sd_t < D_t$	
$g_1 = 0; g_2 = SOC - D_1 > 0$	$g_1 = SOC - D_2 > 0; g_2 = 0$	$g_1 = SOC - D_2 > 0; g_2 = SOC - D_1 > 0$	(C.1.18)
$\phi_2 = 0; \lambda_2 = C_2$	$\phi_1 = 0; \lambda_1 = C_1$	$\phi_1 = 0; \lambda_1 = 0; \phi_2 = C_1; \lambda_2 = C_2$	(C.1.19)
$sd_t > 0; \alpha_t = 0$	$sd_t > 0; \alpha_t = 0$	$sd_t > 0; \alpha_t = 0$	(C.1.20)
$SO_2 - SO_1 + \lambda_1 = C_2$	$SO_2 - SO_1 - \lambda_2 = -C_1$	$SO_2 - SO_1 = C_2 - C_1$	(C.1.21)
$SO_2 - SO_1 = C_2 - C_1 + \phi_1 \geq C_2 - C_1$	$SO_2 - SO_1 = C_2 - C_1 - \phi_2 \leq C_2 - C_1$		(C.1.22)
$SO_2 - SO_1 = C_2 - C_1 + \phi_1$	$SO_2 - SO_1 = C_2 - C_1 - \phi_2$		(C.1.23)
(1) $SO_2 - SO_1 \geq C_2 - C_1$	(2) $SO_2 - SO_1 \leq C_2 - C_1$	(3) $SO_2 - SO_1 = C_2 - C_1$	(C.1.24)

As suggested by case (3) ($sd_t < D_t$), the cost-minimizing system operator will be indifferent between the dispatch solutions when $SO_2 - SO_1 = C_2 - C_1$, as they will produce equivalent as-bid

total costs. We are thus left with optimal solutions for $\{sd_1^*, sd_2^*, g_1^*, g_2^*, \lambda_1^*, \lambda_2^*\}$ of:

(1) $SO_2 - SO_1 > C_2 - C_1$	(2) $SO_2 - SO_1 < C_2 - C_1$	(3) $SO_2 - SO_1 = C_2 - C_1$	
$\{sd_1^*, sd_2^*, g_1^*, g_2^*, \lambda_1^*, \lambda_2^*\} =$ $\{D_1, SOC - D_1, 0, D_2 - SOC + D_1, SO_1, C_2\}$	$\{sd_1^*, sd_2^*, g_1^*, g_2^*, \lambda_1^*, \lambda_2^*\} =$ $\{SOC - D_2, D_2, D_1 - SOC + D_2, 0, C_1, SO_2\}$	$\{sd_1^*, sd_2^*, g_1^*, g_2^*, \lambda_1^*, \lambda_2^*\} =$ $\{D_1 - g_1^*, D_2 - g_2^*, D_1 - sd_1^*, D_2 - sd_2^*, C_1, C_2\}$	(C.1.25)

The profits accrued by the ESR are the dot product of the clearing price and ESR discharge in each time period; $\pi^{ESR} = \sum \lambda_t sd_t$. Substituting values for each of the optimal solutions from eq. (C.1.25) we have

(1) $SO_2 - SO_1 > C_2 - C_1$	(2) $SO_2 - SO_1 < C_2 - C_1$	(3) $SO_2 - SO_1 = C_2 - C_1$	
$\pi^{ESR} = \lambda_1^* sd_1^* + \lambda_2^* sd_2^*$	$\pi^{ESR} = \lambda_1^* sd_1^* + \lambda_2^* sd_2^*$	$\pi^{ESR} = \lambda_1^* sd_1^* + \lambda_2^* sd_2^*$	(C.1.26)
$\pi^{ESR} = SO_1 D_1 + (SOC - D_1)C_2$	$\pi^{ESR} = (SOC - D_2)C_1 + SO_2 D_2$	$\pi^{ESR} = C_1 sd_1^* + C_2 sd_2^*$	(C.1.27)

If we further assume for simplicity the ESR offers SO_t will not be lower than 0 and exclude the third case, we get Eq. (C.17):

$$\min \pi^{ESR} = \begin{cases} (SOC - D_1)C_2, SO_2 - SO_1 > C_2 - C_1 \\ (SOC - D_2)C_1, SO_2 - SO_1 < C_2 - C_1 \end{cases} \tag{C.1.28}$$

Chapter 6 Hybridizing Resources to Profitable Decongest Generation Pockets*

Abstract

We investigate strategic energy storage resource (ESR) and hybrid resource offering in a competitive market formulated as a bi-level optimization model. We incorporate novel parameterization of hybrid wind-ESR resources making a single offer to the market. The bi-level formulation is converted to a mathematical program with equilibrium constraints (MPEC) using its Karush–Kuhn–Tucker optimality conditions, then converted to a mixed-integer linear programming model using strong duality theory and the Big-M method. We find that hybrid offer strategies increase strategic generation company profits and system costs primarily by decongesting the system to increase the clearing price at the hybrid resource's bus. Hybridization helps the strategic company avoid existing marginal cost-based offer mitigation, yielding supernormal profits that can be maintained under uncertainty. Results are obtained on a numerical example based on the IEEE NREL-RTS modified test system with 25-bus network and high renewable penetration, demonstrating relevance on a realistic test system.

* This paper, written with research assistance from Nik Zheng, is in preparation for journal submission.

6.1 Introduction

6.1.1 Background

The participation of energy storage resources (ESRs) and hybrid resources (pairing renewable or conventional power plants with ESRs) in North American wholesale electricity markets are expected to increase significantly in coming years. According to a U.S. Energy Information Administration's (EIA) report, more than 1000 Megawatts (MW) of large-scale battery storage was operational in United States before 2020, with another 3616 MW planned by project developers for beginning operation between 2020 to 2023 [1]. Independent System Operator (ISO) and Regional Transmission Organization (RTO) wholesale market interconnection queues have tens of GW of additional ESRs, with as much as 50% of wind and solar capacity entering the queue hybridized with ESRs in some markets [2]. While existing North American competitive electricity markets have experience with larger capacities of pumped storage hydroelectric (PHS) units sharing these general operational characteristics [3], Federal Energy Regulatory Commission (FERC) Order 841 finds updates are needed for full ESR participation. Order 841 requires FERC-jurisdictional ISOs and RTOs to develop and file compliance plans allowing full participation of ESRs in providing wholesale market products of which they are technically capable, including the ability to submit competitive offers and set clearing price in these markets [4]. The combination of rising ESR and hybrid penetration with evolving models for participation in markets make it essential to investigate how these resources may behave in competitive markets to maximize their individual or a resource portfolio's profits. Identification of strategies deleterious to social welfare can then inform ongoing policy discussions at market operators and independent market monitors regarding market rules for ESRs and hybrids. We highlight a particular strategy, which we term ϵ-decongestion, that can be used by hybridized resources to mask their marginal costs and manipulate congested bus prices upward.

6.1.2 Approach

This paper proposes a bi-level optimization formulation for maximizing the profit of a strategic generation company's (genco) portfolio of resources with inclusions of ESRs hybridized with generation

resources. In the lower level the ISO, acting as the market operator, minimizes the as-bid cost of serving firm load. This leader-follower setup is generally termed Stackelberg competition. ESRs not in the strategic genco's portfolio do not submit offers and are dispatched by the market operator subject to applicable operational constraints to minimize production costs. Additional offer rules applied to the genco's resources in the upper level reflect typical market rules, including monotonic increasing offer curves and cost-based offer mitigation for thermal and variable generators. ESR cycling is limited to one daily cycle to reflect common operational practice for reducing degradation and to allow cases to be structured as a series of co-optimized hourly resolution sequential days, similar to day-ahead (DA) markets.

The bi-level problem is reformulated as a mathematical program with equilibrium constraints (MPEC) by use of the Karush-Kuhn-Tucker (KKT) conditions of the lower level problem. Resulting stationarity and complementary slackness conditions may be included with the original problem formulation to obtain a nonlinear, nonconvex MPEC. The MPEC is reformulated as a MILP by use of strong duality to linearize the objective function and the Fortuny-Amat and McCarl [5] ("Big-M") approach to recasting complementarity constraints with the addition of scalars and auxiliary binary variables. This approach is common for similar problems in the existing literature [6, 7] and admits globally optimal solutions with use of existing commercial solvers under careful domain-relevant selection of Big-M bounds and computational limitations on the number of auxiliary binary variables included.

6.1.3 Literature review

Numerous studies investigate the ability of conventional generation [8], variable generation [6, 9], aggregators [10], prosumers [11, 12], and other arrangements of resources and loads to profitably manipulate prices as strategic players. To formulate and solve strategic games, rules are needed for the strategic player's ability to make valid quantity, price, or price-quantity offers and beliefs about how other participants in the electricity market will bid and operate [13]. Incorporating physical laws governing power flows in nodal electricity market clearing with congested transmission lines adds known additional opportunities for profitable strategies [14].

Energy limited resources like ESRs require additional intertemporal constraints to properly model and have made ESRs a common focus of optimization methods in recent years. Many studies develop methods for a single, social welfare maximizing entity to optimally build and dispatch ESRs at the bulk system and distribution level [15–17]. Optimal ESR capacity expansion and dispatch may account for existing policy [18], ancillary services necessary for grid stability [19], and incorporate uncertainty in load and variable generation forecasts through techniques including stochastic optimization [20, 21]. However, maximizing social welfare may not maximize profit for an ESR-operating entity in a competitive market where it can bid and set market clearing prices. ESRs' value as a hedge against uncertainty with increasing variable renewable penetration [22] means, by the same logic, ESRs may have substantial latitude in setting price as a pivotal supplier. Using ESRs to maximize social welfare allows modeling ESRs as price-takers, ignoring the ability of a single ESR or portfolio of resources with ESRs to coordinate offers to maximize profit in a market context. Our approach extends literature considering ESRs as a price-maker [23], particularly within a strategically coordinated portfolio [24].

Multiple empirical approaches exist for investigating Stackelberg competition between a leading ESR-owning genco and following market operator in both investment and operational decision-making [23, 25, 26]. Resulting bilevel problems are often reformulated as mathematical programs with equilibrium constraints (MPEC) [27]. The MPEC remains nonconvex but may conditionally admit analytical solutions through iteration of the strategic player's optimal strategy [8]. To investigate ESR operations in imperfectly competitive markets many use some version of this approach, often with additional reformulation of the MPEC as a MILP using strong duality and Big-M [15, 24, 26, 28–31]. Alternatives to the common Big-M reformulation applied to investigating strategic ESR behavior have recently been developed to improve commitment-related simplifications associated with requiring convexity in the lower-level problem [31] and reduce solution time [32]. But these methods require reformulations generally less easily solved with the heuristics implemented in existing commercial solvers, and in [32] constrain offer price-quantity pairs to pre-set discretized values to enable reformulation with disjunctions.

Among existing similar literature employing similar MILP reformulations for bi-level problems

with ESRs, [30] shows ESRs can increase their profit in an imperfectly competitive market, but does not consider ownership of a portfolio of resources. [28] and [31] look more closely at the effects of congestion on strategic ESR profitability in a multi-node system with often-binding transmission constraints, demonstrating the importance of congestion to electricity market competition [14] in the context of ESR strategic profitability. [29] develops optimal bids for a single hybrid resource made up of pumped hydro storage and wind generation, but does not extend its case to a multi-bus system with inclusion of battery ESRs. [24] is among recent research to extend the MPEC approach as an Equilibrium Program with Equilibrium Constraints (EPEC), allowing solving for multiple profit-maximizing ESR owners and demonstrating the importance of their collaboration in maximizing joint profits. However, solving EPECs at hourly resolution for a full day with multiple players and many buses is computationally intensive, requiring alternative methods to the Big-M approach to admit an analytically tractable reformulation as a sequence of iterated MPECs until a Nash Equilibrium is found.

We follow the above research by implementing a MPEC approach to investigate ESRs in imperfectly competitive markets. However, we hybridize ESRs with generators relevant to common configurations entering wholesale market interconnection queues. In conjunction with offer constraints and a realistic high renewable penetration congested nodal test system, this extension enables the MPEC approach to have additional relevance in ongoing discussions on hybrid participation in electricity markets.

6.1.4 Contributions

The main contributions of this paper are:

1. ESRs may be configured as hybrid assets with other generators at the same bus. Hybrids submit a single offer and, like ESRs but unlike generators, are assumed not subject to marginal cost-based offer mitigation. By default hybrid generators are constrained to have only sufficient interconnection capacity to dispatch to the maximum nameplate output of the hybridized generator, excluding ESR capacity. We show optimal hybrid offers are different under these assumptions than co-located generators with separate offers, differ when congestion

in incorporated, and can be maintained with a simple deterministic offering strategy under uncertainty.

2. Wholesale electricity markets currently place restrictions on competitive offers, often including disallowing physical withholding of available capacity. We extend previous research showing the potential for market manipulation by physically withholding low marginal cost generation in [10] by disallowing physical withholding.

3. Inclusion of a daily cycling constraint offers an alternative to inclusion of more detailed degradation functions for marginal ESR discharge [33, 34] and reflects a common heuristic recommended by original equipment manufacturers and in use by ESR owners in today's markets.

4. Use of a 25-bus zone from NREL-RTS test system data enables realistic parameterization of the model on a nodal test system with high renewable penetration and congestion, better reflecting conditions under which high levels of ESR and hybrid deployment are expected in the coming decade. Incorporation of congestion extends the practical relevance of theory on market manipulation to profitably decongest a nodal electricity market in [14, 35, 36].

Together these cases and contributions highlight a profitable strategy for hybridization in generation pockets: ϵ-decongestion. ϵ-decongestion involves hybridizing a resource to mask variable generation marginal costs in a generation pocket, then bidding consistent with the expected system lambda to incrementally ("ϵ") decongest the line and set the LMP at the hybrid's bus equivalent to the system lambda.

6.2 Model formulation and assumptions

Sets, parameters, and decision variables in this model are a subset of those described in Chapter 5, Appendix A. This section includes only the mathematical formulation of the constraints and model reformulation used in this paper.

6.2.1 Assumptions

The main assumptions of our model are:

1. Thermal and renewable generator offers are mitigated to marginal cost, including when owned by the strategic genco. For simplicity renewable generator marginal costs are assumed to be zero. ESR and hybrid offers are not mitigated to marginal cost due to the lack of existing opportunity cost-based offer mitigation for the intertemporal opportunity costs that comprise most of these resources' optimal offers in arbitrage applications in current markets [37].

2. The genco has perfect foresight of loads and offers of other generators in the day-ahead (DA) market. In sensitivity analysis we clear optimal DA bids against real-time (RT) actual load and variable generation data without recourse to show even deterministic hybrid offering strategies remain profitable under uncertainty. Deterministic offering assumptions are relaxed in related literature through use of stochastic optimization techniques [26, 38]; incorporating these approaches could further increase hybrid profits toward the theoretical maximum.

3. Generators submit monotonically increasing stepwise offers, consistent with current market practice.

4. Generators and ESRs must submit an offer for available generation, consistent with must-offer obligation rules generally applied to resources receiving resource adequacy payments in markets. This assumption limits the ability of low marginal cost resources (including ESRs) to engage in strategic physical withholding highlighted in previous research [10]; economic withholding strategies are still allowable within bid constraints.

5. The transmission network incorporates direct current optimal power flow (DCOPF) on transmission links between buses. Optimal clearing prices are determined at each node as locational marginal prices (LMPs) for this configuration and may include congestion, but not losses. DCOPF is commonly used as it allows a linear (and thus, convex) representation of transmission constraints in the lower-level problem. Congestion is demonstrated in other studies [26] and our work to be important to strategic ESR value.

6.2.2 Upper-level formulation

Objective Function

$$MAX\ \lambda_{t,z} * (\sum_{t,g \in GC \cap GUC, z == ZL_g}^{T,G} gd_{t,g} + \sum_{t,g \in GC \cap GNUC, z == ZL_g}^{T,G} nucgd_{t,g} + \sum_{t,s \in SS, z == ZLS_s}^{T,S} sd_{t,s}) - \sum_{t,g \in GC \cap GUC, gs}^{T,G,GS} GMC_{g,gs} * gsd_{t,g,gs} - \sum_{t,s \in SS, z == ZLS_s}^{T,S} \lambda_{t,z} * sc_{t,s} \tag{6.1}$$

Generator Offer Constraints

$$2 * GMC_{g,gs} \geq gso_{t,g,gs} \geq GMC_{g,gs}, \forall g \in GC \cap GUC \tag{6.2}$$

$$gso_{t,g,gs} \geq gso_{t,g,gs-1}, \forall g \in GC \cap GUC \tag{6.3}$$

$$UE \geq gso_{t,g,gs}, \forall g \in GC \cap GUC \tag{6.4}$$

$$0 == go_{t,g}, \forall g \in GC \cap GNUC \tag{6.5}$$

The upper-level objective maximizes the total profit of a single genco's assets. Offer constraint equation 6.2 helps reduce solution time by limiting feasible generator offers. Offer constraints equations 6.3-6.5 require increasing offers for generators submitting multiple price-quantity pairs for different segments of the generator equation 6.3, cap offers to a pre-determined market maximum equation 6.4, and enforce renewable generators are assumed to have \$0/MWh marginal cost equation 6.5. These constraints reflect commonly implemented market rules in current wholesale electricity markets.

Hybridization

Offers for the variable generation component of a hybridized resource have their offer mitigation constraint in equation 6.5 removed, and the generation offer is set equal to the storage discharge offer:

$$sofd_{t,s} == go_{t,g}, \forall g, s \in G, S \cap H \tag{6.6}$$

6.2.3 Lower-level formulation

Objective Function

$$\begin{aligned} \forall \lambda \in arg[MIN \sum_{t,g,gs}^{T,GUC,GS} gso_{t,g,gs} * gsd_{t,g,gs} \\ + \sum_{t,g}^{T,GNUC} go_{t,g} * nucgd_{t,g} + \sum_{t,s}^{T,SS} (sofd_{t,s} * sd_{t,s} - sofc_{t,s} * sc_{t,s})] \end{aligned} \tag{6.7}$$

System Level Constraints

$$\begin{aligned} L_{t,z} == \sum_{t,g,gs,z==ZL_g}^{T,GUC,GS} gsd_{t,g,gs} + \sum_{t,g,z==ZL_g}^{T,GNUC} nucgd_{t,g} \\ + \sum_{t,s,z==ZLS_s}^{T,S} (sd_{t,s} - sc_{t,s}) + \sum_{t,l}^{T,L} (txmw_{t,l,TXTL_l==z} - txmw_{t,l,TXFL_l==z}) \end{aligned} \tag{6.8}$$

$$va_{t,z==RBUS_t} = 0 \tag{6.9}$$

Storage Constraints

$$DMAX_s * CMAX_s \geq DMAX_s * sc_{t,s} + CMAX_s * sd_{t,s} \tag{6.10}$$

$$sc_{t,s} \geq 0;\ sd_{t,s} \geq 0 \tag{6.11}$$

$$SMAX_s \geq \sum_{t}^{1,\ldots,t} (CE_s * sc_{t,s} - DE_s * sd_{t,s}) \geq 0 \tag{6.12}$$

$$SMAX_s \geq \sum_{t}^{T} sd_{t,s} \geq 0 \tag{6.13}$$

$$\sum_{t}^{T} (CE_s * sc_{t,s} - DE_s * sd_{t,s}) == 0 \tag{6.14}$$

$$soc_{1,s} - (CE_s * sc_{1,s} - DE_s * sd_{1,s}) == 0 \tag{6.15}$$

Generator Constraints

$$CAP_{t,g} * SA_{t,g} * gopstat_{t,g} \geq gd_{t,g} \geq gmin_{t,g}, \forall g \in GUC \tag{6.16}$$

$$CAP_{t,g} * SA_{t,g} * GSL_{g,gs} \geq gsd_{t,g,gs} \geq 0, \forall g \in GUC \tag{6.17}$$

$$gd_{t,g} == \sum_{gs}^{GS} gsd_{t,g,gs}, \forall g \in GUC \tag{6.18}$$

$$CAP_{t,g} * SA_{t,g} \geq nucgd_{t,g} \geq 0, \forall g \in GNUC \tag{6.19}$$

Hybrid Constraints

$$CAP_{t,g} \geq sd_{t,s} + gd_{t,g}, \forall g, s \in G, S \cap H \tag{6.20}$$

$$ZL_g == ZLS_s, \forall g, s \in G, S \cap H \tag{6.21}$$

Transmission Constraints

$$TTCAP_l \geq txmwh_{t,l} \geq TFCAP_l \tag{6.22}$$

$$VMAX_z \geq va_{t,z} \geq VMIN_z \tag{6.23}$$

$$txmwh_{t,l} == S_l * (va_{t,z==TXTL_z} - va_{t,z==TXFL_z}) \tag{6.24}$$

The lower level objective minimizes the as-bid cost of serving load, the market operator's problem. In addition to load-balance, transmission, generator, and ESR constraints, equation 6.10 links storage charge and discharge to disallow simultaneous charge and discharge without use of a binary variable. This allows continuous operation of storage across its entire dispatch range but avoids integer commitment decisions. Equations 6.12-6.15 constrain ESR SOC, cycling, and set end and beginning period SOC to 0, respectively. Equation 6.20 limits dispatch of a hybrid resource to the nameplate capacity of the generation resource; equation 6.21 enforces that only resources located at the same bus may be hybridized. DCOPF constraints linearize power flow while respecting Kirchhoff's first and second laws.

6.2.4 Derivation of stationarity conditions

$$gso_{t,g,gs} - \lambda_{t,z==ZL_g} + \phi^{max}_{t,g} - \phi^{min}_{t,g} + \varphi^{max}_{t,g,gs} - \varphi^{min}_{t,g,gs} == 0, \forall g \in GUC \tag{6.25}$$

$$go_{t,g} - \lambda_{t,z==ZL_g} + \omega^{max}_{t,g} - \omega^{min}_{t,g} == 0, \forall g \in GNUC \tag{6.26}$$

$$\begin{aligned} sofd_{t,s} + CMAX_s * \gamma_{t,s} - \beta_{t,s} - DE_s * (\sum_{t}^{t,...N_T} \nu^{max}_{t,s} - \nu^{min}_{t,s}) \\ +\xi_s + DE_s * \chi_s - \lambda_{t,z==SZL_s} == 0, \forall s \in SS \end{aligned} \tag{6.27}$$

$$-sofc_{t,s} + DMAX_s * \gamma_{t,s} - \alpha_{t,s} + CE_s * (\sum_{t}^{t,...N_T} \nu_{t,s}^{max} - \nu_{t,s}^{min})$$
$$-CE_s * \chi_s + \lambda_{t,z==SZL_s} == 0, \forall s \in SS \tag{6.28}$$

$$\mu_{t,l}^{max} - \mu_{t,l}^{min} + \psi_{t,z}^{max} - \psi_{t,z}^{min} == 0 \tag{6.29}$$

$$-\lambda_{t,z==ZL_g} + \omega_{t,g}^{max} - \omega_{t,g}^{min} == 0 \tag{6.30}$$

Stationarity is one of the KKT conditions for MPEC reformulation, and proceeds by taking partial derivatives of the problem's Lagrangian with respect to all decision variables and setting them to zero. The state-of-charge decision variable is a function of the charge and discharge decision variables in prior times as written in equation 6.12, and is removed and rewritten in terms of those primal variables and their associated dual variables in equations 6.27 and 6.28.

6.2.5 Complementary constraints

$$0 \leq \gamma_{t,s} \perp DMAX_s * CMAX_s - DMAX_s * sc_{t,z} - CMAX_s * sd_{t,z} \geq 0 \tag{6.31}$$

$$0 \leq \alpha_{t,s} \perp sc_{t,z} \geq 0 \tag{6.32}$$

$$0 \leq \beta_{t,s} \perp sd_{t,z} \geq 0 \tag{6.33}$$

$$0 \leq \nu_{t,s}^{max} \perp SMAX_s - \sum_{t}^{1,...,t} (CE_s * sc_{t,s} - DE_s * sd_{t,s}) \geq 0 \tag{6.34}$$

$$0 \leq \nu_{t,s}^{min} \perp \sum_{t}^{1,...,t} (CE_s * sc_{t,s} - DE_s * sd_{t,s}) \geq 0 \tag{6.35}$$

$$0 \leq \xi_s \perp SMAX_s - \sum_{t}^{T} sd_{t,s} \geq 0 \tag{6.36}$$

$$0 \leq \phi_{t,g}^{max} \perp SA_{t,g} * CAP_{t,g} - gd_{t,g} \geq 0 \tag{6.37}$$

$$0 \leq \phi_{t,g}^{min} \perp gd_{t,g} \geq 0 \tag{6.38}$$

$$0 \leq \varphi_{t,g,gs}^{max} \perp SA_{t,g} * CAP_{t,g} * GSL_{t,g,gs} - gsd_{t,g,gs} \geq 0 \tag{6.39}$$

$$0 \leq \varphi_{t,g,gs}^{min} \perp gsd_{t,g,gs} \geq 0 \tag{6.40}$$

$$0 \leq \omega_{t,g}^{max} \perp SA_{t,g} * CAP_{t,g} - nucgd_{t,g} \geq 0 \tag{6.41}$$

$$0 \leq \omega_{t,g}^{min} \perp nucgd_{t,g} \geq 0 \tag{6.42}$$

$$0 \leq \mu_{t,l}^{max} \perp TTCAP_l - txmwh_{t,l} \geq 0 \tag{6.43}$$

$$0 \leq \mu_{t,l}^{min} \perp txmwh_{t,l} - TFCAP_l \geq 0 \tag{6.44}$$

$$0 \leq \psi_{t,z}^{max} \perp VMAX_z - va_{t,z} \geq 0 \tag{6.45}$$

$$0 \leq \psi_{t,z}^{min} \perp va_{t,z} - VMIN_z \geq 0 \tag{6.46}$$

The complementary slackness conditions are the other KKT conditions we must derive and add

to the reformulation of the lower-level problem as a MPEC.

6.2.6 Linearized lower-level objective function

Using strong duality theory, the objective function of primal problem is equal to the objective function of the corresponding dual problem at the optimum. For the lower level objective function, the dual problem objective function can thus be rewritten:

$$
\begin{aligned}
MIN \sum_{t,g,gs}^{T,GUC,GS} gso_{t,g,gs} * gsd_{t,g,gs} + \sum_{t,g}^{T,GNUC} go_{t,g} * nucgd_{t,g} \\
\sum_{t,s}^{T,S} (sofd_{t,s} * sd_{t,s} - sofc_{t,s} * sc_{t,s}) \\
= \sum_{t,s}^{T,S} DMAX_s * CMAX_s * (-\gamma_{t,s}) + \sum_{t,s}^{T,S} SMAX_s * (-\nu_{t,s}^{max}) + \sum_{s}^{S} SMAX_s * (-\xi_s) \\
+ \sum_{t,l}^{T,L} TTCAP_{t,l} * (-\mu_{t,z}^{max}) + \sum_{t,l}^{T,L} TFCAP_{t,l} * (\mu_{t,z}^{min}) \\
+ \sum_{t,z}^{T,Z} VMAX_z * (-\psi_{t,z}^{max}) + \sum_{t,z}^{T,Z} VMIN_z * (\psi_{t,z}^{min}) \\
+ \sum_{t,z}^{T,Z} L_{t,z} * \lambda_{t,z} + \sum_{t,g,gs}^{T,G,GS} CAP_{t,g} * SA_{t,g} * GSL_{g,gs} * (-\varphi_{t,g,gs}^{max}) \\
+ \sum_{t,g}^{T,G} CAP_{t,g} * SA_{t,g} * (-\omega_{t,g}^{max}) = B
\end{aligned}
\tag{6.47}
$$

where we define the now linearized right hand side to be the linear polynomial B. Using the

complementary constraints equations 6.31-6.46 to rewrite:

$$\begin{aligned}
&\sum_{t,g,gs}^{T,GUC,GS} gso_{t,g,gs} * gsd_{t,g,gs} + \sum_{t,g}^{T,GNUC} go_{t,g} * nucgd_{t,g} \\
&\quad + \sum_{t,s}^{T,SS} (sofd_{t,s} * sd_{t,s} - sofc_{t,s} * sc_{t,s}) \\
= &\sum_{t,g,gs}^{T,GUC,GS} (\lambda_{t,z==ZL_g} - \phi^{max}_{t,g} + \phi^{min}_{t,g} - \varphi^{max}_{t,g,gs} + \varphi^{min}_{t,g,gs}) * gsd_{t,g,gs} \\
&\quad + \sum_{t,g}^{T,GNUC} (\lambda_{t,z==ZL_g} - \omega^{max}_{t,g} + \omega^{min}_{t,g}) * nucgd_{t,g} \\
&\quad + \sum_{t,s\in SS}^{T,S} [\lambda_{t,z==SZL_s} - CMAX_s * \gamma_{t,s} \\
&+\beta_{t,s} + DE_s * (\chi_s - (\sum_{t}^{t,...N_T} \nu^{max}_{t,s} - \nu^{min}_{t,s}) - \xi_s))] * sd_{t,s} \\
&\quad + \sum_{t,s\in SS}^{T,S} [-\lambda_{t,z==SZL_s} - DMAX_s * \gamma_{t,s} \\
&+\alpha_{t,s} - CE_s * (\chi_s - (\sum_{t}^{t,...N_T} \nu^{max}_{t,s} - \nu^{min}_{t,s}))] * sc_{t,s} \\
= &\sum_{t,g,gs}^{T,GUC,GS} (\lambda_{t,z==ZL_g} * gsd_{t,g,gs} - CAP_{t,g} * SA_{t,g} * GSL_{g,gs} * \phi^{max}_{t,g}) \\
&\quad + \sum_{t,g}^{T,GNUC} (\lambda_{t,z==ZL_g} * nucgd_{t,g} - CAP_{t,g} * SA_{t,g} * \omega^{max}_{t,g}) \\
&+ \sum_{t,ss}^{T,SS} (\lambda_{t,z==SZL_s} * (sd_{t,s} - sc_{t,s}) - DMAX_s * CMAX_s * \gamma_{t,s} - SMAX_s * \nu^{max}_{t,s})
\end{aligned} \tag{6.48}$$

We can linearize upper level objective function using linear polynomial B defined in equations

6.47 and 6.48:

$$\begin{aligned}
&\sum_{t,g\in GUC,z==ZL_g}^{T,GC}[gd_{t,g}+\sum_{t,g\in GNC,z==ZL_g}^{T,GC}nucgd_{t,g}\\
&+\sum_{t,s,z==ZLS_s}^{T,SS,Z}(sd_{t,s}-sc_{t,s})]*\lambda_{t,z}-\sum_{t,g\in GUC,gs}^{T,G,GS}GMC_{g,gs}*gsd_{t,g,gs}=B\\
&+\sum_{t,g,gs}^{T,GUC,GS}CAP_{t,g}*SA_{t,g}*GSL_{g,gs}*\phi_{t,g}^{max}+\sum_{t,g}^{T,GNUC}CAP_{t,g}*SA_{t,g}*\omega_{t,g}^{max}\\
&+\sum_{t,s}^{T,SS}(DMAX_s*CMAX_s*\gamma_{t,s}+SMAX_s*\nu_{t,s}^{max})-\sum_{t,g\in GUC,gs}^{T,GC,GS}GMC_{g,gs}*gsd_{t,g,gs}
\end{aligned}\tag{6.49}$$

The objective on the right-hand side of 6.49 is now linearized.

6.3 Experiment design

All cases are implemented on the Reliability Test System Grid Modernization Lab Consortium (RTS-GMLC) [39], an updated version of the RTS-96 73-bus test system [40] developed by the National Renewable Energy Laboratory and released publicly on Github. Updates to the RTS-96 system include increasing the proportion of gas-fired and renewable generation, reducing transmission capacities to have more congestion, developing heat rate curves for modern gas-fired generation, developing renewable profiles for wind and solar generators based off Southwestern USA temporally coincident profiles, and modifying loads to reflect a temporally coincident profile for the same geography. A diagram of buses, loads, and generation from the test system used in our cases is included below.

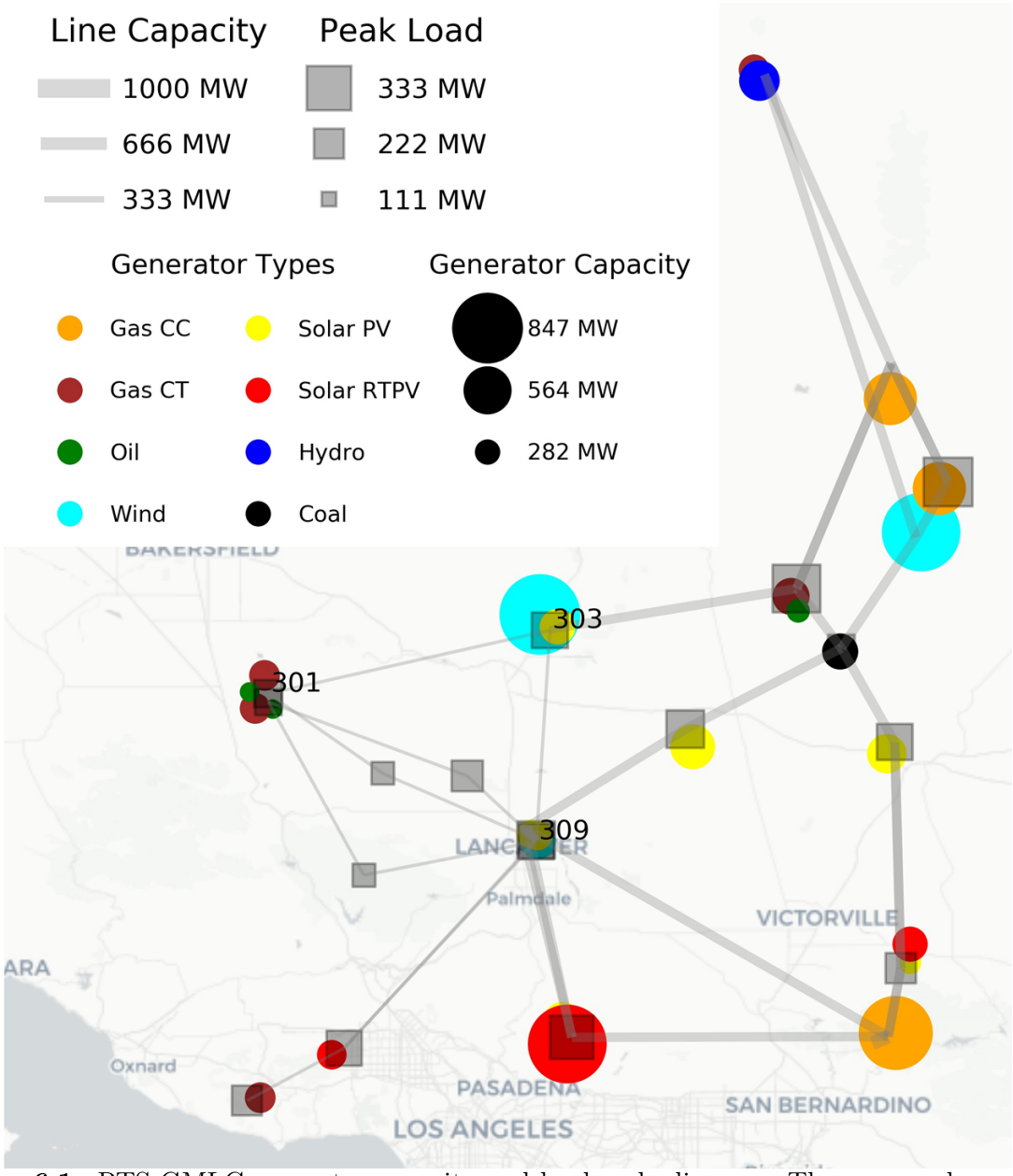

Figure 6.1: RTS-GMLC generator capacity and load node diagram. Three areas draw on data from Arizona Public Service (Zone 1), Nevada Power (Zone 2), and Los Angeles Department of Water and Power (Zone 3), though they do not represent existing infrastructure. We retain only the displayed Zone 3 in cases. Buses 301, 303, and 309 are indicated because they are discussed in more detail in cases.

In our cases we remove the RTS-GMLC ESR at bus 313 and add case-specific ESR and hybrid resources. To decrease solution time without loss of generality we reduce the RTS-GMLC test system

to one of its three areas, choosing to focus on the highest renewable penetration and congestion area, Zone 3. This zone, loosely based on the Los Angeles Department of Water and Power (LADWP) geography, has 25 buses and includes a large wind generator (847 MW) at the often-congested bus 303. We typically choose to co-locate and hybridize an ESR installation with the bus 303 wind generator to investigate congestion effects, further enumerated in Table 7.2. Parameters for other generators and loads shared across all cases are included in Table 7.1 and Figure 6.2.

Table 6.1: Zone 3 RTS-GMLC Generating Unit Data.

BusID	Group	Capacity (MW)	Type	Units	Bid Segments	Bid Capacities (MW)	Marginal Cost ($/MWh)
301	U20	20	Oil CT	2	4	8,4,4,4	$87.26, $87.26, $99.7, $105.37
301	U55	55	Gas CT	2	4	21,11,11,11	$28.47, $28.47, $29.29, $43.74
302	U20	20	Oil CT	2	4	8,4,4,4	$87.26, $87.26, $99.7, $105.37
302	U55	55	Gas CT	2	4	21,11,11,11	$33.79, $33.79, $38.38, $38.64
307	U55	55	Gas CT	2	4	21,11,11,11	$27.89, $27.89, $29.22, $35.47
313	U355	355	Gas CC	1	4	170,61.67,61.67,61.67	$15.73, $15.73, $26.76, $33.75
315	U12	12	Oil ST	5	4	5,2.33,2.33,2.33	$75.44, $75.44, $100.4, $124.1
315	U55	55	Gas CT	3	4	21,11,11,11	$26.4, $26.4, $26.9, $31.86
316	U155	155	Coal	1	4	62,31,31,31	$21.12, $21.12, $21.29, $27.27
318	U355	355	Gas CC	1	4	170,61.67,61.67,61.67	$26.85, $26.85, $27.28, $31.53
321	U355	355	Gas CC	1	4	170,61.67,61.67,61.67	$22.73, $22.73, $25.91, $33.95
322	U55	55	Gas CT	2	4	21,11,11,11	$23.33, $23.33, $26.5, $27.89
323	U355	355	Gas CC	2	4	170,61.67,61.67,61.67	$26.43, $26.43, $30.28, $31.73
303	WIND	847	Wind	1	1	As available	$0
308	RTPV	100.9	Solar RTPV	1	1	As available	$0
309	WIND	148.3	Wind	1	1	As available	$0
310	PV	103.3	Solar PV	1	1	As available	$0
312	PV	189.2	Solar PV	1	1	As available	$0
313	PV	93.3	Solar PV	1	1	As available	$0
313	RTPV	806	Solar RTPV	1	1	As available	$0
314	PV	144.3	Solar PV	1	1	As available	$0
317	WIND	799.1	Wind	1	1	As available	$0
319	PV	188.2	Solar PV	1	1	As available	$0
320	PV	51.6	Solar PV	1	1	As available	$0
322	U50	50	Hydro	1	1	As available	$0
324	PV	152.3	Solar PV	1	1	As available	$0

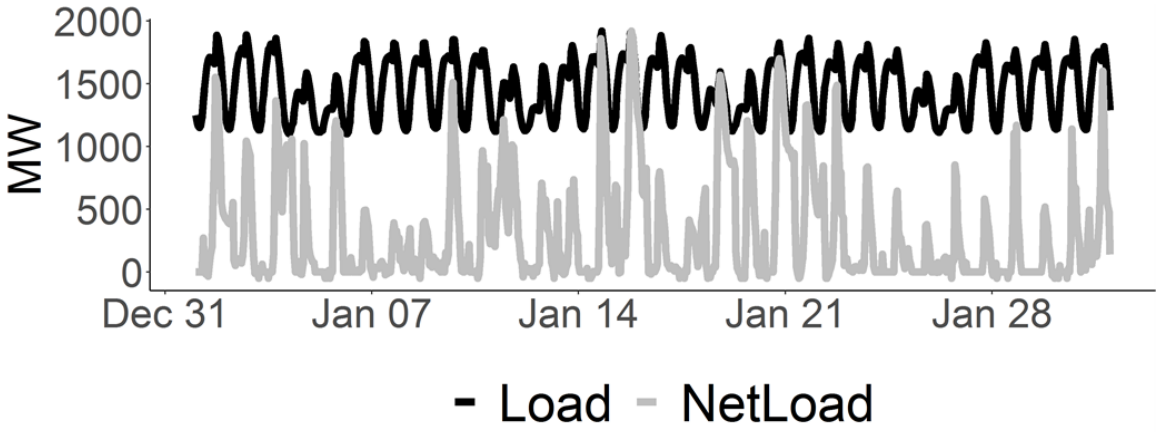

Figure 6.2: One month of hourly DA load and net load for zone 3. DA data is used exclusively in all cases other than those additionally clearing DA bids without recourse using RT data. Net load excludes non-curtailed generation from wind, solar PV, and solar rooftop PV (RTPV) resources.

As seen in Figure 6.2, net load is commonly zero or near-zero for many consecutive hours. As a result clearing prices are commonly \$0/MWh (ESR and hydro resources are not excluded from net load, but are assumed \$0/MWh marginal cost).

6.4 Case study results

We consider eight cases; four competitive, four strategic. The competitive cases are co-located (COL-) and hybrid (H-) reference (-REF) cases assuming all offers at marginal cost. Strategic cases demonstrate the effects of including (1) resource co-location (COL-PF) vs. hybridization with perfect foresight (H-PF), (2) increased (10x) transmission capacity so the system is uncongested with perfect foresight (H-UNC-PF), and (3) fixed bids from deterministic DA expectations cleared without recourse against RT actual load and renewable generation (H-UN-PF). Strategic cases are typically compared to their corresponding competitive reference case unless otherwise noted. In all cases the strategic genco owns the 847MW wind generator at bus 303; in hybrid cases the wind generator is hybridized with the ESR.

Table 6.2: Cases in main text. PF is Perfect Foresight.

CASE	ESR BUS	*CMAX*	*SMAX*	STRATEGIC?	HYBRID?	PF?	TRANSMISSION?
COL-REF	303	300	900	F	F	T	1x
H-REF	303	300	900	F	T	T	1x
H-UNC-REF	303	300	900	F	T	T	10x
H-UN-REF	303	300	900	F	T	F	1x
COL-PF	303	300	900	T	F	T	1x
H-PF	303	300	900	T	T	T	1x
H-UNC-PF	303	300	900	T	T	T	10x
H-UN-PF	303	300	900	T	T	F	1x

All simulations are performed in Pyomo 5.6.9 and CPLEX 12.9.0.0 on an Intel Xeon CPU E5-2680 v3 2.50GHz desktop with 12 cores and 64 GB memory. The MILP gap for strategic cases is set at 1.00% for solving each daily case for the modeled month. The cumulative monthly optimality gap in strategic cases is shown as an uncertainty between the optimal value and the upper bound in Figure 6.3, 6.5, and 6.7. Competitive cases are run as LPs and do not have an optimality gap.

Where applicable, total monthly profits for the genco are reported as π^{SS} for cases where ESRs are bid strategically and π^{NSS} for the compared competitive reference case where they are not, so

$$\Delta\pi^{p} = \pi^{SS} - \pi^{NSS} \tag{6.50}$$

Is then the incremental portfolio profit associated with strategic bidding compared to a reference competitive case.

Figure 6.3 compares co-located and hybrid profits when bid strategically to their corresponding reference cases. Because the resource capacities and availability are identical, the competitive reference cases result in the same solution and profit. What changes is how the resources are offered when strategic.

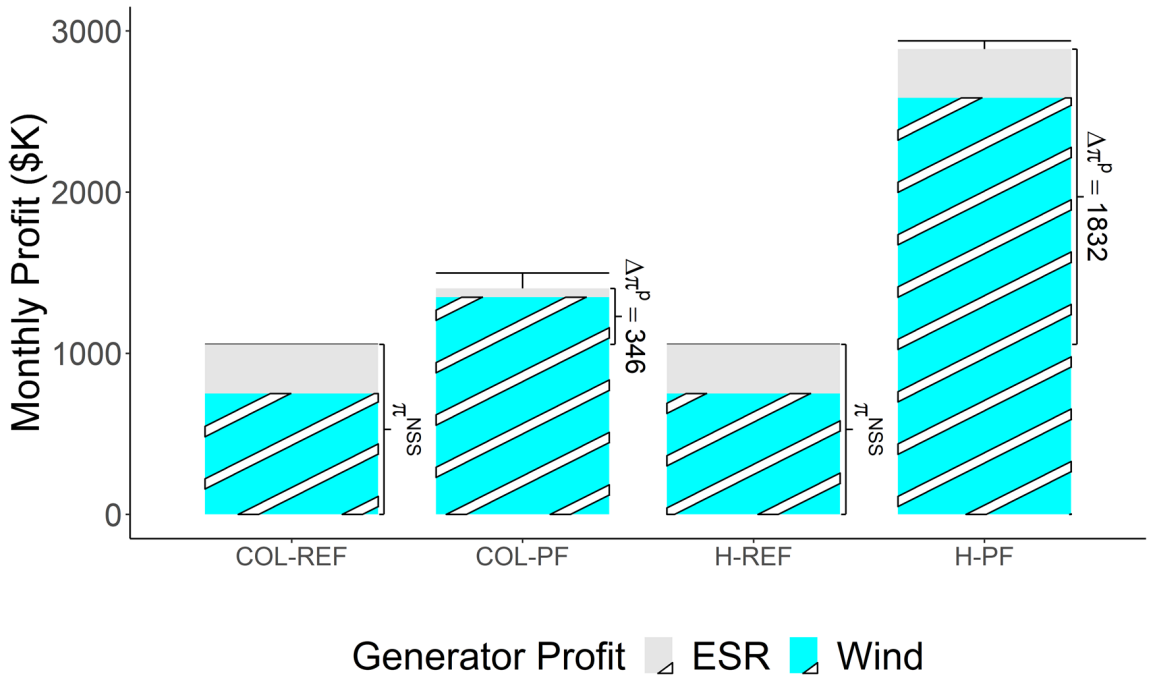

Figure 6.3: Comparison of monthly profits for co-located and hybrid cases. Additionally bidding latitude to offer wind generation above marginal cost when hybridized increases portfolio profits. Uncertainty bars show cumulative optimality gap for cases run as MILPs.

Figure 6.4 highlights why additional supernormal profits are available to hybrids compared to co-location (Figure 6.3). When co-located the wind resource is bid separately and constrained to offer at marginal cost. When hybridized the hybrid resource makes a single, unconstrained offer. The strategic hybrid-owning genco uses this bidding latitude to raise the clearing price at bus 303, where the hybrid is located.

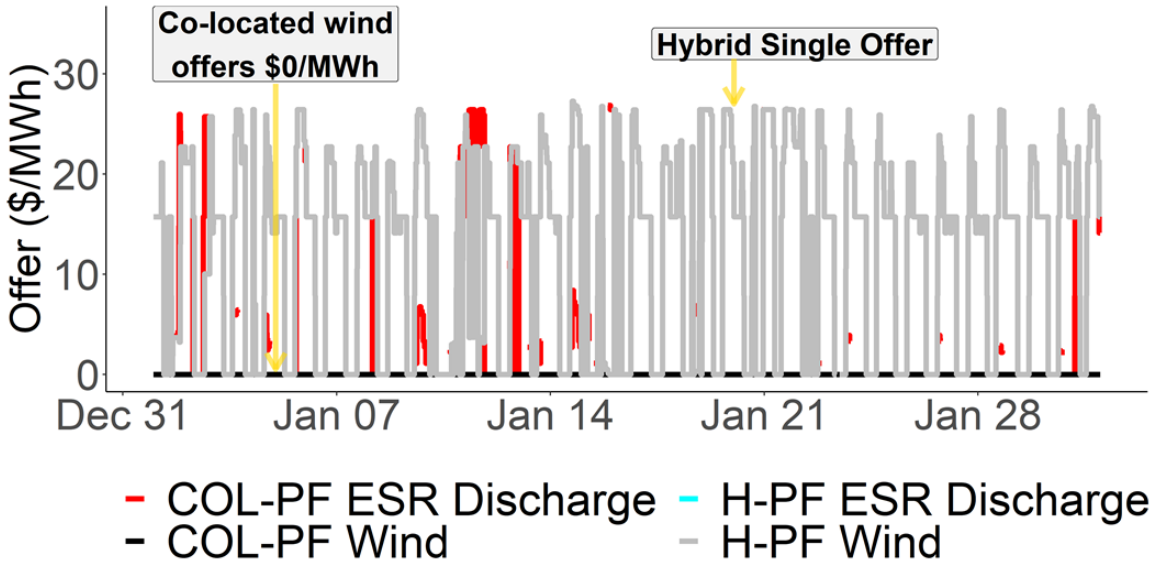

Figure 6.4: One month of ESR discharge and wind dispatch offers in co-located and hybrid perfect foresight cases. ESR charge offers are not shown.

An important part of the parameterization of the previous cases is that the transmission line between bus 303 and 309 is often constrained, resulting in a bus 303 LMP lower than the system lambda. As suggested in [35], this creates a situation where a strategic genco in congested generation pocket can bid up to the system lambda and increase clearing price at its bus; a strategy we refer to in this case as "ϵ-decongestion." Figure 6.5 demonstrates the importance of this strategy by comparing monthly profit results for the base H-PF case with those of the H-UNC-PF case, which has 10x the transmission capacity and therefore no congestion. In the uncongested case the competitive profits π^{NSS} are greater, since bus 303's competitive clearing price is higher without congestion. However, in the congested case the additional profit from strategic bidding $\Delta\pi^{p}$ increases.

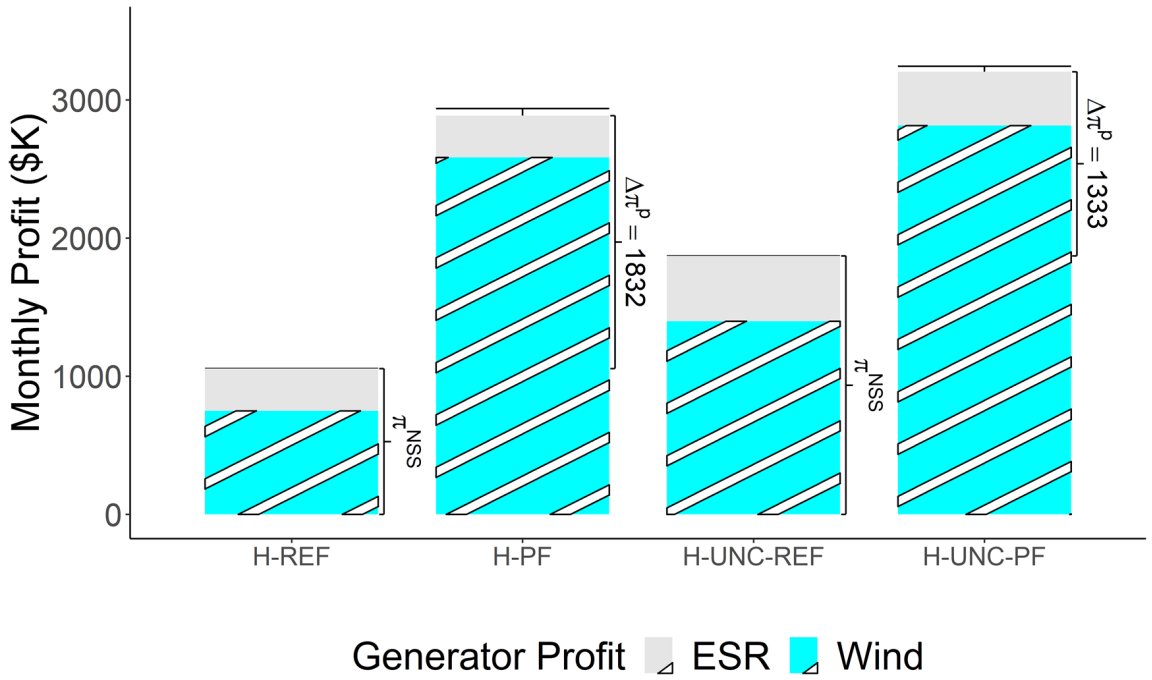

Figure 6.5: Additional profits are higher for a congested than uncongested case, demonstrating the relevance of ϵ-decongestion in strategic hybrid bidding.

Figure 6.6 compares the bus 301 and 303 clearing price distributions for the reference and strategic perfect foresight congested and uncongested cases to provide additional insight on results reported in Figure 6.5. In all cases prices are commonly \$0/MWh, but average prices at bus 303 are considerably higher in strategic (-PF) cases than reference (-REF) cases. Uncongested prices are higher than congested. In H-REF the bus 301 clearing prices are higher than the bus 303 clearing prices due to congestion; in H-UNC-REF without congestion they are the same. In H-PF the prices at bus 301 and 303 are again equal, but this time due to ϵ-decongestion from strategic bidding.

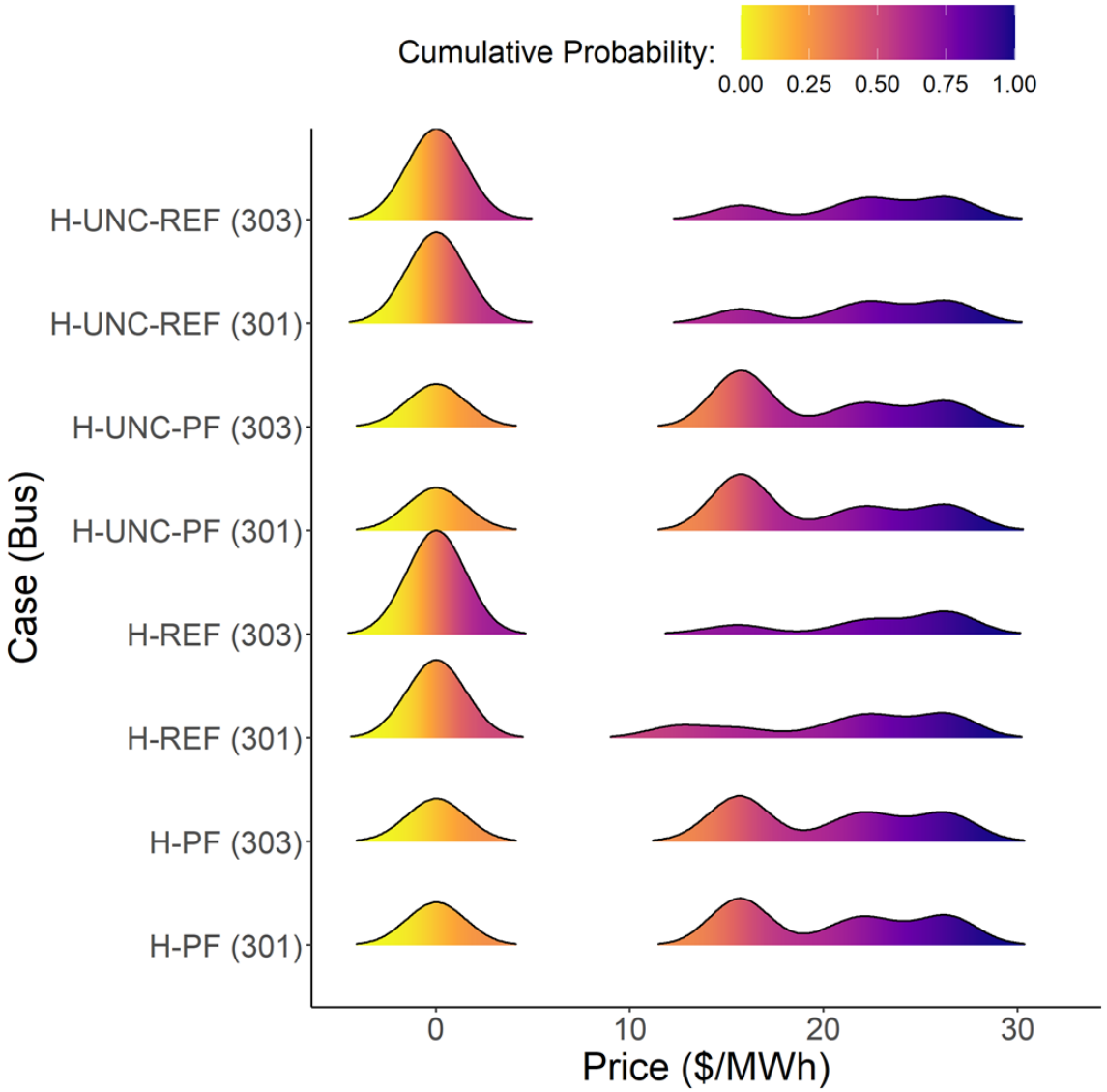

Figure 6.6: Bus 303 clearing price comparison. H-REF prices are lower at bus 303 than 301, but with strategic bidding in the H-PF case ϵ-decongestion produces the same, higher clearing prices at both buses.

In the final cases we investigate the ability to maintain strategic profits under uncertainty by fixing DA hybrid bids without recourse when actual load and variable generation availability are realized in RT. That is, we fix DA offers without recourse by setting

$$sodf_{t,s} == DADO_{t,s} \tag{6.51}$$

$$sodc_{t,s} == DACO_{t,s} \tag{6.52}$$

Where $DADO_{t,s}$ is the optimal DA discharge offer, $DACO_{t,s}$ is the optimal DA charge offer, and $sofd_{t,s}$ and $sofc_{t,s}$ are the DA discharge and charge offer decision variables, as previously defined. While a sophisticated genco could undoubtedly do better in forecasting when its optimal offer than using a single deterministic DA forecast, even this strategy maintains profits for the hybrid-owning genco under uncertainty. The results demonstrate the potential importance of hybridization as profitable way to decongest generation pockets and raise clearing prices, for example, by incrementally decongesting the pocket with an appropriate bid, termed ϵ-decongestion.

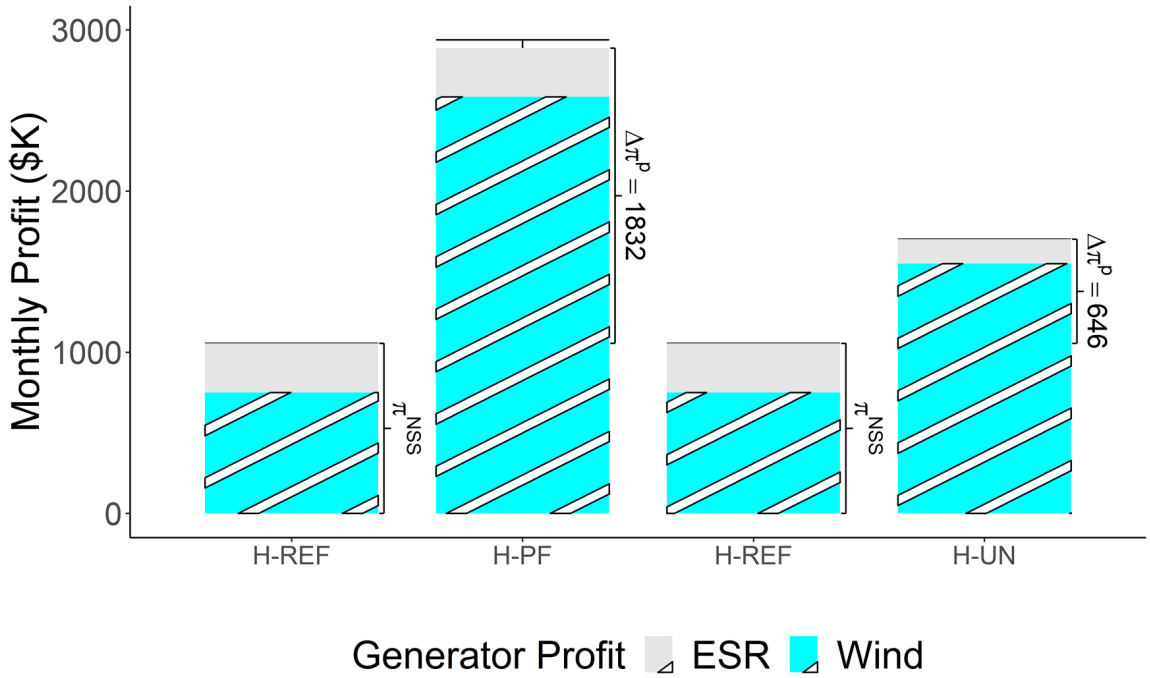

Figure 6.7: Deterministic hybrid bids reduce profit compared to perfect foresight but retain considerable positive profits under uncertainty in load and variable generation availability.

6.5 Conclusion

A bi-level optimization with reformulation to a MPEC and MILP for analytical tractability is commonly used to investigate potential strategic behavior in competitive wholesale electricity

markets, but has not previously been used to consider important aspects of ESR and hybrid resource bidding in a realistic congested test system with high renewable penetration. We demonstrate the importance of considering hybridization with ESRs to manipulate prices with strategic bids, particularly by ϵ-decongestion. Analytical results are computationally tractable for a 25-bus test system with high renewables penetration, showing that ESRs located at a congested node with wind generation may be considerably more profitable for the strategic player if hybridized to allow offering a single bid, averting marginal cost-based offer mitigation.

Future work will focus on extensions allowing implementation of offer mitigation to inform current policy processes. Other extensions may include more sophisticated incorporation of uncertainty using stochastic programming methods. The MPEC formulation may be extended to an EPEC with multiple strategic players or conjectural variation alternatives to the assumption all other players bid at known marginal costs. A final direction of research may incorporate optimal siting of ESRs and hybrids in the test system to maximize strategic player profits, potentially with options for the strategic player to own additional existing assets in the system as part of its portfolio, rather than pre-assigning the asset ownership structure.

References

[1] EIA, "Battery Storage in the United States: An Update on Market Trends," *United States Energy Information Administration*, p. 23, July 2020. https://www.eia.gov/analysis/studies/electricity/batterystorage/pdf/battery_storage.pdf.

[2] W. Gorman, A. Mills, M. Bolinger, R. Wiser, N. G. Singhal, E. Ela, and E. O'Shaughnessy, "Motivations and options for deploying hybrid generator-plus-battery projects within the bulk power system," *Electricity Journal*, vol. 33, no. 5, p. 106739, 2020. doi:https://doi.org/10.1016/j.tej.2020.106739.

[3] E. Ela, B. Kirby, A. Botterud, C. Milostan, I. Krad, and V. Koritarov, "The Role of Pumped Storage Hydro Resources in Electricity Markets and System Operation," *Proceedings of the Hydro Vision International*, pp. 1–10, 2013. https://www.nrel.gov/docs/fy13osti/58655.pdf.

[4] FERC, "Electric Storage Participation in Markets Operated by Regional Transmission Organizations nad Independent System Operators," *Federal Energy Regulatory Commission Docket No. RM16-23-000*, 2018. https://www.ferc.gov/whats-new/comm-meet/2018/021518/E-1.pdf.

[5] J. Fortuny-Amat and B. McCarl, "A Representation and Economic Interpretation of a Two-Level Programming Problem," *The Journal of the Operational Research Society*, vol. 32, no. 9, pp. 783–792, 1981. doi:https://doi.org/10.2307/2581394.

[6] X. Fang, V. Krishnan, and B. M. Hodge, "Strategic offering for wind power producers considering energy and flexible ramping products," *Energies*, vol. 11, no. 5, 2018. doi:https://doi.org/10.3390/en11051239.

[7] S. Pineda and J. M. Morales, "Solving linear bilevel problems using Big-Ms: Not all that glitters is gold," *arXiv*, vol. 34, no. 3, pp. 2469–2471, 2018. doi:https://doi.org/10.1109/TPWRS.2019.2892607.

[8] B. F. Hobbs, C. B. Metzler, and J. S. Pang, "Strategic gaming analysis for electric power systems: An MPEC approach," *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 638–645, 2000. doi:https://doi.org/10.1109/59.867153.

[9] P. Twomey and K. Neuhoff, "Wind power and market power in competitive markets," *Energy Policy*, vol. 38, no. 7, pp. 3198–3210, 2010. doi:https://doi.org/10.1016/j.enpol.2009.07.031.

[10] N. Azizan Ruhi, K. Dvijotham, N. Chen, and A. Wierman, "Opportunities for price manipulation by aggregators in electricity markets," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5687–5698, 2018. doi:https://doi.org/10.1109/TSG.2017.2694043.

[11] X. Xiao, J. Wang, R. Lin, D. J. Hill, and C. Kang, "Large-scale aggregation of prosumers toward strategic bidding in joint energy and regulation markets," *Applied Energy*, vol. 271, no. May, p. 115159, 2020. doi:https://doi.org/10.1016/j.apenergy.2020.115159.

[12] S. Ramyar, A. L. Liu, and Y. Chen, "Power Market Model in Presence of Strategic Prosumers," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 898–908, 2020. doi:https://doi.org/10.1109/TPWRS.2019.2937887.

[13] C. Ruiz, S. J. Kazempour, and A. J. Conejo, "Equilibria in futures and spot electricity markets," *Electric Power Systems Research*, vol. 84, no. 1, pp. 1–9, 2012. doi:https://doi.org/10.1016/j.epsr.2011.09.019.

[14] J. B. Cardell, C. C. Hitt, and W. W. Hogan, "Market power and strategic interaction in electricity networks," *Resource and Energy Economics*, vol. 19, no. 1-2, pp. 109–137, 1997. doi:https://doi.org/10.1016/s0928-7655(97)00006-7.

[15] H. Chen, Y. Yu, Z. Hu, H. Luo, C. W. Tan, and R. Rajagopal, "Energy storage sharing strategy in distribution networks using bi-level optimization approach," *arXiv*, no. Llm, 2017. doi:https://doi.org/10.1109/PESGM.2017.8274595.

[16] H. Pandzic, Y. Wang, T. Qiu, Y. Dvorkin, and D. S. Kirschen, "Near-Optimal Method for Siting and Sizing of Distributed Storage in a Transmission Network," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2288–2300, 2015. doi:https://doi.org/10.1109/TPWRS.2014.2364257.

[17] R. Fernández-Blanco, Y. Dvorkin, B. Xu, Y. Wang, and D. S. Kirschen, "Optimal Energy Storage Siting and Sizing: A WECC Case Study," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 733–743, 2017. doi:https://doi.org/10.1109/TSTE.2016.2616444.

[18] Y. M. Atwa and E. F. El-Saadany, "Optimal allocation of ESS in distribution systems with a high penetration of wind energy," *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1815–1822, 2010. doi:https://doi.org/10.1109/TPWRS.2010.2045663.

[19] M. Nick, R. Cherkaoui, and M. Paolone, "Optimal allocation of dispersed energy storage systems in active distribution networks for energy balance and grid support," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2300–2310, 2014. doi:https://doi.org/10.1109/TPWRS.2014.2302020.

[20] C. Abbey and G. Joós, "A stochastic optimization approach to rating of energy storage systems in wind-diesel isolated grids," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 418–426, 2009. doi:https://doi.org/10.1109/TPWRS.2008.2004840.

[21] M. Dicorato, G. Forte, M. Pisani, and M. Trovato, "Planning and operating combined wind-storage system in electricity market," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 209–217, 2012. doi:https://doi.org/10.1109/TSTE.2011.2179953.

[22] H. Akhavan-Hejazi and H. Mohsenian-Rad, "Optimal operation of independent storage systems in energy and reserve markets with high wind penetration," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 1088–1097, 2014. doi:https://doi.org/10.1109/TSG.2013.2273800.

[23] M. Miletić, H. Pandžić, and D. Yang, "Operating and investment models for energy storage systems," *Energies*, vol. 13, no. 18, pp. 1–33, 2020. doi:https://doi.org/10.3390/en13184600.

[24] M. Dolányi, K. Bruninx, G. Deconinck, and E. Delarue, "Strategic operation of storage in energy markets: an EPEC approach," *Energy Systems Integration & Modeling Group Working Paper Series No. EN2019-10*, September, 2019. https://www.mech.kuleuven.be/en/tme/research/energy_environment/Pdf/wp-en2019-10.

[25] Q. Huang, Y. Xu, and C. Courcoubetis, "Stackelberg competition between merchant and regulated storage investment in wholesale electricity markets," *Applied Energy*, vol. 264, no. October 2019, p. 114669, 2020. doi:https://doi.org/10.1016/j.apenergy.2020.114669.

[26] H. Mohsenian-Rad, "Coordinated Price-Maker Operation of Large Energy Storage Units in Nodal Energy Markets," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 786–797, 2016. doi:https://doi.org/10.1109/TPWRS.2015.2411556.

[27] S. A. Gabriel, A. J. Conejo, J. D. Fuller, B. F. Hobbs, and C. Ruiz, *Complementarity Modeling in Energy Markets*, vol. 1. 2013.

[28] Y. Ye, D. Papadaskalopoulos, R. Moreira, and G. Strbac, "Investigating the impacts of price-taking and price-making energy storage in electricity markets through an equilibrium programming model," *IET Generation, Transmission and Distribution*, vol. 13, no. 2, pp. 305–315, 2019. doi:https://doi.org/10.1049/iet-gtd.2018.6223.

[29] A. V. Ntomaris and A. G. Bakirtzis, "Optimal bidding for risk-averse hybrid power station producers in insular power systems: An MPEC approach," *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe, ISGT-Europe 2017 - Proceedings*, vol. 2018-Janua, pp. 1–6, 2017. doi:https://doi.org/10.1109/ISGTEurope.2017.8260169.

[30] E. Nasrolahpour, H. Zareipour, W. D. Rosehart, and S. J. Kazempour, "Bidding strategy for an energy storage facility," *19th Power Systems Computation Conference, PSCC 2016*, 2016. doi:https://doi.org/10.1109/PSCC.2016.7541016.

[31] Y. Wang, Y. Dvorkin, R. Fernández-Blanco, B. Xu, and D. S. Kirschen, "Impact of local transmission congestion on energy storage arbitrage opportunities," *arXiv*, 2017. doi:https://doi.org/10.1109/PESGM.2017.8274471.

[32] Y. Ye, D. Papadaskalopoulos, J. Kazempour, and G. Strbac, "Incorporating Non-Convex Operating Characteristics into Bi-Level Optimization Electricity Market Models," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 163–176, 2020. doi:https://doi.org/10.1109/TPWRS.2019.2925317.

[33] G. He, Q. Chen, P. Moutis, S. Kar, and J. F. Whitacre, "An intertemporal decision framework for electrochemical energy storage management," *Nature Energy*, vol. 3, no. 5, pp. 404–412, 2018. doi:https://doi.org/10.1038/s41560-018-0129-9.

[34] D. Tran and A. M. Khambadkone, "Energy management for lifetime extension of energy storage system in micro-grid applications," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1289–1296, 2013. doi:https://doi.org/10.1109/TSG.2013.2272835.

[35] S. S. Oren, "Economic inefficiency of passive transmission rights in congested electricity systems with competitive generation," *Energy Journal*, vol. 18, no. 1, pp. 63–83, 1997. doi:https://doi.org/10.5547/ISSN0195-6574-EJ-Vol18-No1-3.

[36] S. Stoft, "Financial Transmission Rights Meet Cournot: How TCCs Curb Market Power," *The Energy Journal*, vol. 20, no. 1, pp. 1–23, 1999. http://www.jstor.org/stable/41322816.

[37] H. Cui, S. Member, and F. Li, "Bi - Level Arbitrage Potential Evaluation for Grid - Scale Energy Storage Considering Wind Power and LMP," pp. 1–11, 2016. doi:https://doi.org/10.1109/TSTE.2017.2758378

[38] E. Tómasson, M. R. Hesamzadeh, and F. A. Wolak, "Optimal offer-bid strategy of an energy storage portfolio: A linear quasi-relaxation approach," *Applied Energy*, vol. 260, no. May 2019, p. 114251, 2020. doi:https://doi.org/10.1016/j.apenergy.2019.114251.

[39] C. Barrows, E. Preston, A. Staid, G. Stephen, J. P. Watson, A. Bloom, A. Ehlen, J. Ikaheimo, J. Jorgenson, D. Krishnamurthy, J. Lau, B. McBennett, and M. O'Connell, "The IEEE Reliability Test System: A Proposed 2019 Update," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 119–127, 2020. doi:https://doi.org/10.1109/TPWRS.2019.2925557.

[40] C. Grigg, et al. "The IEEE Reliability Test System-1996. A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee," vol. 14, no. 3, 1999. doi:https://doi.org/10.1109/59.780914.

Chapter 7 Quantifying the Complementarity Benefit of Transmission and Electricity Storage for Incorporation in Resource Adequacy*

Abstract

Resource contributions to power systems reliability at the margin depend on the other resources already on the system. Commonly-used Effective Load Carrying Capability (ELCC) calculations in resource adequacy modeling account for this dependency, but rarely include transmission interfaces. We show the simplification of excluding transmission interfaces when computing energy storage resource ELCCs is often not appropriate. Sequential Monte Carlo analysis of ELCCs shows energy storage interacts with transmission in a way generators do not because energy storage can use transmission capacity to charge from external generation for later discharge. We develop a two-part screening criteria to identify when storage charging from external generation using transmission interfaces will affect its contribution to resource adequacy. The relevance of these criteria is first demonstrated on a simple test system, and then for a realistic parameterization of a 20-zone representation of the Midcontinent System Operator's (MISO) control area. ELCCs for one- to six-hour duration energy storage resources are decreased 25%-71% by halving available transmission capacity in the MISO test system and increased 0-17% by doubling available transmission capacity, while generator ELCCs are unaffected. Results are directly policy relevant to centralized capacity markets with zonal clearing prices attributing quantity contributions to generation and storage resources using ELCC methods.

* This paper, written with research assistance from Nik Zheng, is in preparation for journal submission.

7.1 Introduction

Power systems must provide sufficient generation supply on a forward basis to reliably serve load; this is generally termed resource adequacy (RA). Given new generation resources cannot be added instantaneously and uncertainty about future load forecasts and generator availability, probabilistic methods are used to quantify whether a power system achieves a desired level of RA. A common RA target for North American system operators is "one-day-in-ten-years" [1]. This target assumes available generation resources insufficient to serve load in expectation one day every ten years is the optimal trade-off between additional generator cost and value. With additional assumptions RA target metrics can be converted to an equivalent planning reserve margin (PRM) quantifying the percent excess of generation capacity above expected annual system peak load. The historically common equivalent PRM heuristic is still used today in the North American Electric Reliability Corporation (NERC) target and assessed reserve margins for its constituent areas [2].

Planners must choose among resource options for incremental additions even though resource adequacy is a system property. Quantifying resource-level contributions is also financially important because competitive capacity remuneration mechanisms often compensate individual resources. To better compare resource contributions Garver introduced effective load carrying capability (ELCC) in 1966 [3]. ELCC quantifies the additional firm load an added resource can serve while maintaining the same level of reliability as before the resource and load were added. For conventional generators with failures assumed time invariant and uncorrelated, the ELCC can be calculated by knowing the Equivalent Forced Outage Rate Demand (EFORd) of the generator; i.e., how often in expectation the generator's rated capacity is unavailable when needed to serve load. This heuristic avoids computationally intensive resource level ELCC calculations incorporating time-dependent failures and operating constraints.

In recent years the rising penetration of variable renewable energy (VRE) with output correlated with both other VRE and load has renewed interest in using ELCC-based methods to properly credit VREs in RA. For example, the Midcontinent Independent System Operator (MISO) calculates a resource class-average ELCC for its wind and solar capacity [4], and the PJM Interconnection (PJM)

recently filed for Federal Energy Regulatory Commission (FERC) approval to use ELCC methods to quantify VRE and energy storage resource (ESR) contributions in its capacity market [5].

The ELCC of an additional variable or energy-limited resources depends on the correlation of its availability with load and the existing resource portfolio [6]. The existing resource portfolio effect on ELCC is seen in literature quantifying the declining marginal capacity value of solar PV [7], wind [8], and energy storage resources (ESRs) [9] as a function of increasing capacity of the same resource across multiple system configurations [10]. ESRs can partially counteract declining VRE capacity credit [11], particularly in solar PV-heavy systems given solar PV's diurnal generation profile [12, 13]. Another solution for integrating VRE is transmission capacity expansion to increase VRE deliverability [14,16]. If ESRs increase VRE capacity credit and vice-versa there is a quantifiable portfolio complementarity benefit between VRE and ESRs [17].

The motivating logic behind this research is that RA complementarity applies to transmission capacity in addition to different types of generators and energy storage resources. We show transmission capacity and ESRs have a unique complementarity in multi-zone resource adequacy when the transmission capacity can be used to charge ESRs at an earlier time to increase reliability at a later time, increasing overall use of the transmission capacity. Generators, which do not charge and discharge, are not complemented by transmission capacity in this way. Multi-area RA assessments incorporating limiting transmission interfaces show reliability benefits of transmission [18, 19], but are uncommon in existing planning processes in large, multi-area power systems due to computational intensity. We therefore propose new heuristics power system planners can use to identify when ESRs will charge from generators in an external zone using available transmission capacity for later discharge to increase reliability at a time when interzonal transmission capacity is constrained. In this scenario, if transmission capacity is reduced and the ESR could not charge from external generation, its resource adequacy contribution is reduced. This fact should be reflected in the ESR's capacity credit or ELCC by incorporating a limiting transmission interface when it affects ESR charge and discharge. Because ELCC methods are increasingly commonly used to credit resource-level capacity contributions in competitive markets, our results showing material effects of transmission capacity when it constrains charge for later reliability-relevant discharge of ESRs on their ELCCs are directly

policy relevant to considering increasing penetrations of ESRs in multi-area reliability planning.

The simplest criteria for identifying when transmission capacity should be incorporated in a RA assessment of ESRs is to know the ESRs are charged by generation resources outside the zone where the ESR is located for reliability purposes. However, because power flows according to physical laws and future resource availability and loads are uncertain, it is difficult to identify ex-ante which generation resources will charge ESRs. We therefore propose two quantifiable metrics for determining whether increased use of a transmission interface will change ESR capacity credit enough to include in a RA assessment. We call these the "spatial" and "temporal" criteria: there must be reliability value (e.g., decreased loss-of-load-probability) in moving energy in both space (i.e., across available existing transmission capacity) and time (i.e., from a time when transmission capacity is constrained to a time when it is not) for transmission capacity to complement ESRs.

1. Zones connected by a transmission interface have sufficiently different loss-of-load-probability (LOLP) or other target RA metric time series ("spatial" criterion); and

2. The transmission interface is constrained during the hours with different zonal LOLP, but not constrained during hours in the preceding storage duration window, allowing ESR charge in the earlier hour for discharge in the later hour ("temporal" criterion). This criterion can be quantitatively screened by comparing loss-of-load event durations to ESR discharge duration at maximum rated capacity.

Subsequent sections refer to these two criteria as the (1) spatial and (2) temporal criterion or criteria.

We use a state-of-the-art RA model to run sequential Monte Carlo simulations parameterizing uncertainty in time-sequential resource failure and recovery probabilities on a two-zone, two-hour test system to demonstrate the relevance of these criteria. We then run simulations on an annual 20-zone representation of MISO's footprint also used in the National Renewable Energy Laboratory's (NREL) Seams study [14] to demonstrate practical and policy relevance on a realistic coordinated power system with intra-balancing authority transmission interfaces. Results are quantified by changing ESR ELCCs at several quantities of transmission capacity. Increasing ESR ELCCs with added

transmission capacity show when ESRs can better use incremental transmission capacity to charge for later discharge, thereby avoiding the need for either more generation capacity or peak-period transmission capacity connecting to a zone with reliability concerns.

The 20-zone MISO numerical simulations show 17 of 51 (33%) of inter-zonal transmission interfaces considered link zones with sufficiently different Expected Unserved Energy (EUE). Different zonal EUEs indicate transmission constraints, because otherwise the higher reliability (lower EUE) zone would export to the lower reliability (higher EUE) zone to improve system reliability. Sufficiently different is defined to be 5% of the highest zonal EUE. This suggests these 17 lines are most relevant to include in RA studies if computationally feasible, while other transmission interfaces could be excluded.

In base case MISO simulations using data developed for the National Renewable Energy Laboratory's Seams study [14] one- and four-hour duration ESRs in transmission-constrained zones pass the temporal criteria. Base case one-hour ESR average ELCCs are 17% lower than those in a transmission capacity-doubling case, four-hour ESR average ELCCs are 3% lower. ESRs of 6+ hour duration can serve all loss-of-load events from full state-of-charge, so the temporal criterion does not apply and 6+ hour duration ESR ELCC does not increase with additional transmission capacity above base case ratings. However, sensitivity analysis shows longer duration ESR ELCCs are much lower with reduced transmission capacity ratings compared to the base case between zones: 4-hour and 6-hour ESR average ELCCs are 19% and 21% when available transmission capacity is halved, compared to 86% and 92% base case values. For 1-hour ESRs average ELCC is 9% when transmission capacity is halved compared to a 34% base case value. Overall, this means halving transmission capacity reduces one- to six-hour duration ESR ELCCs by 25-71%, highlighting the complementarity benefit of increased transmission capacity with ESR capacity credit. A similar complementarity is not realized for the solar PV resource used as a generation comparison. Because ESRs shift generation in time to better match supply and demand they can have a complementarity with existing transmission capacity through higher usage not realized by generators.

Results demonstrate the increasing importance of considering limiting transmission interfaces in resource adequacy with ESRs. ESRs using transmission capacity to charge from external generation

for later reliability-relevant discharge require moving energy in space and time to maximize their resource adequacy contribution, as measured by ELCC. Our proposed spatial and temporal criteria help system planners screen when charging ESRs across a major, reliability-limiting transmission interface is likely to occur and therefore should be incorporated in ESR ELCC calculations.

7.2 Methods

Power systems models make trade-offs in granularly representing temporal horizon, network topology, and generator operations. Planners, operators, and researchers therefore commonly link models to iteratively evaluate power systems at multiple resolutions [14]. This research uses the National Renewable Energy Laboratory's Probabilistic Resource Adequacy Suite (PRAS). A high-level depiction of model resolutions contextualizing PRAS is shown in Figure 7.1.

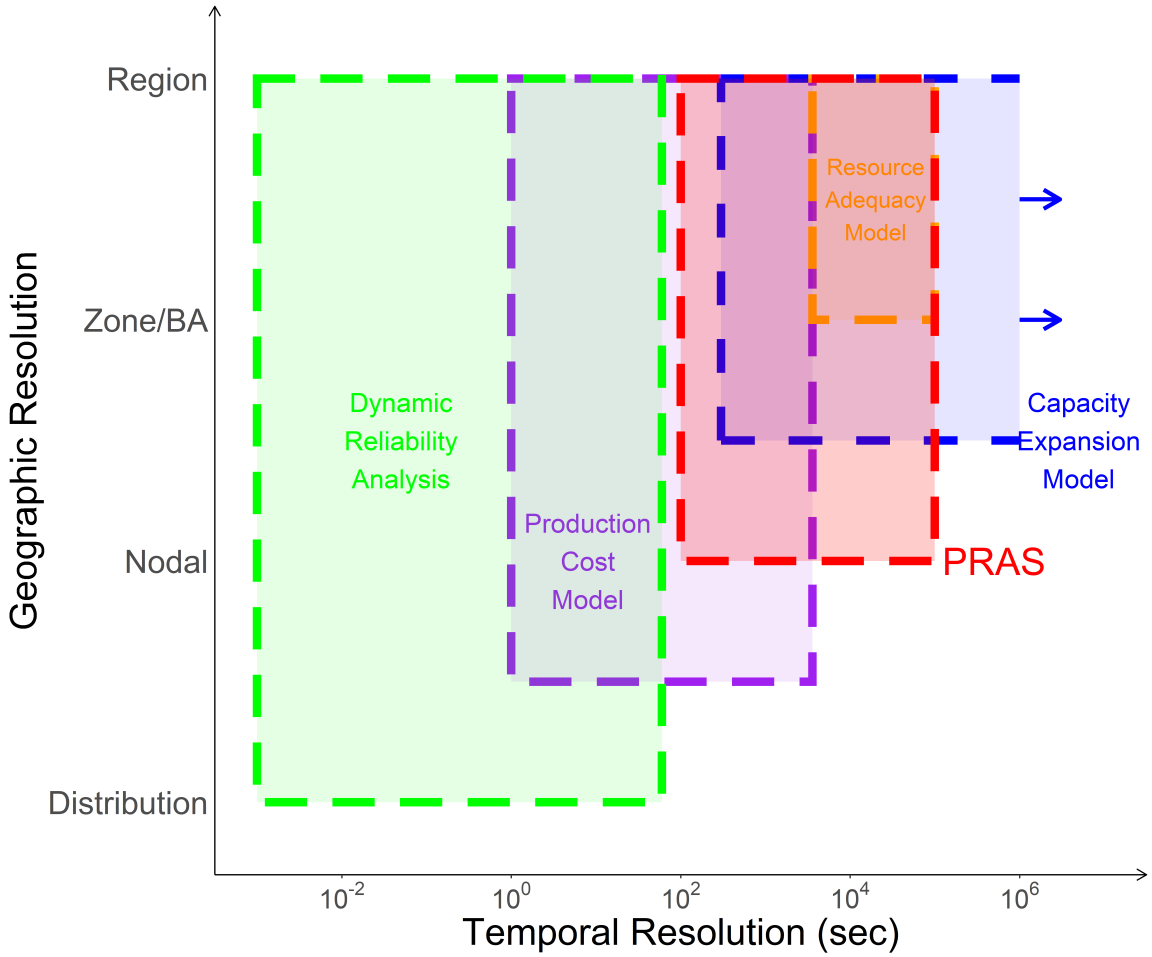

Figure 7.1: Typical ranges of geographic and temporal operational resolution of bulk power systems models. Operational detail of resources is also important and not depicted. Temporal resolution is commonly non-chronological in resource adequacy models. Capacity expansion models also include resource investment and retirement decisions at coarser resolution, the figure refers only to their endogenous operational resolution. BA is Balancing Authority.

In a typical model taxonomy RAMs serve as a link between more temporally granular production cost models (PCMs) and less granular Capacity Expansion Models (CEMs). PCMs are primarily concerned with dispatching a fixed resource portfolio to maximize social welfare. CEMs select an optimal future portfolio of resources to maximize social welfare, but generally with less geographic, resource, and temporal detail for evaluating resource investment and retirement decisions than included in PCMs. RAMs typically take portfolios from a CEM and screen them with additional detail in high (net) load time periods to ensure they meet target resource adequacy metrics before

the portfolio's operation is modeled in more detail in a PCM. If a portfolio is deficient in a target RA metric, the CEM solution can be iterated or further constrained to produce an appropriate portfolio. Likewise, a portfolio deemed sufficient by the RAM can be modeled in more detail by the PCM, and iterated if additional PCM detail reveals deficiencies in operations (e.g., insufficient generator ramping capabilities).

RAMs use probabilistic methods to increase sampling of the most reliability-relevant tail events while reducing computational burdens compared to PCMs. Common simplifying assumptions in RAMs include time-invariant generation failure and recovery probabilities, enabling homogeneous Markov models [20], and ignoring chronologically linked generator operational constraints like unit commitment and ramp rates. RAMs often model resource availability and load non-chronologically to better focus on a larger sample of high loss-of-load-probability events most relevant to evaluating a resource portfolio's long-run sufficiency. For example, the PJM Interconnection's approach fits weekly distributions of loads based on historical data normalized to future peak load forecasts, then evaluates the sufficiency of a resource availability probability distribution for serving load [21]. Net load duration curve [22] and convolutional methods [21] can enforce additional correlation between different types of resource availability and load, improving ability to model variable generation contributions [23] and correlated generator failures without the chronological computational burden of a PCM. An assumption that makes these non-chronological abstractions more accurate is that intertemporal constraints on resource availability rarely bind. This assumption can be inaccurate for energy-limited resources, whose availability to dispatch is a function of how the resource has dispatched in preceding hours (e.g., storage or demand response with limited duration). If duration and charge/discharge behavior are expected to limit resource adequacy contributions, more computationally intensive sequential Monte Carlo methods may be warranted.

Because of our focus on ESRs, we perform sequential Monte Carlo evaluation by running all cases using the National Renewable Energy Laboratory's Probabilistic Resource Adequacy Suite (PRAS).[1] PRAS extends traditional single zone RA tools by allowing limited representation of a transmission system and replaces convolution methods with sequential Monte Carlo draws for calculating the

[1] https://nrel.github.io/PRAS/

contribution of resources with temporally linked availability like energy storage. Each Monte Carlo draw parameterizes a time series of generator, storage, and transmission availability based on failure and recovery probabilities. Resulting resource adequacy is then evaluated using all Monte Carlo draws.

PRAS dispatches available ESRs based on a greedy algorithm demonstrated in Evans et al. [24] to be Expected Unserved Energy (EUE)-minimizing. ESRs are charged when generation is available and immediately begin discharging at their fully rated capacity when there is a loss of load event until either their state of charge (SOC) is 0 or the event ends. This does not minimize Loss of Load Expectation (LOLE), Loss of Load Hours (LOLH), or Loss of Load Probability (LOLP)-based metrics (see Table 7.1 for definitions), all commonly used as target RA metrics for interpreting the "one-day-in-ten-years" heuristic [25, 26]. EUE considers the magnitude of unserved energy events and treats MWh of shortfall equally, so it differs as an minimization objective from strictly event-based metrics. EUE is generally more computationally intensive because it requires calculating the probability of different magnitudes of shortfall, not just the probability a shortfall event occurs. A greedy algorithm that minimizes EUE by discharging ESRs as soon as shortfall occurs will increasingly diverge from LOLE, LOLH, or LOLP-based metrics as ESR capacity increases.

Table 7.1: Definition of Common RA Target Metrics.

Metric	**Acronym**	**Equation**[a]
Expected Unserved Energy or Expected Energy Not Served	EUE or EENS	$\sum_{h=1}^{H} \sum_{s}^{S} P(s)$
Loss of Load Expectation	LOLE	$\sum_{d=1}^{D} P(G_d < L_d)$
Loss of Load Hours	LOLH	$\sum_{h}^{H} P(G_h < L_h)$
Loss of Load Probability	LOLP	$P(G_t < L_t)$

[a] L is Load, G is available generation capacity. h indexes hours, s indexes generator availability states, d indexes days, t is a discrete time interval. P is probability.

Alternative objectives for optimizing ESR dispatch, such as profit maximization in applicable wholesale energy and ancillary service markets [27] or retail tariffs [28], integrating and maximizing charging from variable or low-carbon resources [29–31], and minimizing ESR degradation or operational costs [32] are also not considered in ESR dispatch decisions. While outside our scope, this research highlights the potential importance of multiple objectives in optimizing ESR dispatch to

assessing ESRs' capacity contribution.

PRAS is first configured in each case to run a network-wide resource adequacy assessment and record results. PRAS is then configured to calculate effective load carrying capability (ELCC) for an additional resource.[2] To calculate resource ELCC we add either a 100MW or 500MW nameplate capacity resource in each case, in each zone with a seeded number of sequential Monte Carlo draws. The additional resource capacity is chosen to be small enough (<1% of installed capacity) to approximate a marginal addition to the system, but large enough for PRAS evaluation at 1MW ELCC precision. Results should be interpreted as an average ELCC for a small, added resource of a specific type and location at a maximum precision of 1MW. ELCCs are reported as a percent of the added nameplate capacity to better compare across resource additions and cases. The number of Monte Carlo draws is chosen to balance runtime against precision. PRAS employs a bisection method for calculating resource ELCC. The bisection method sequentially tightens upper and lower bounds on additional load that can be served by the evaluated resource until the lower and upper bounds of the target RA metric (e.g., EUE) are no longer statistically significantly different under a hypothesis test at a user-input p-value (default, 0.05). PRAS returns the ELCC bounds for the added resource at the hypothesis test's significance level for the seeded Monte Carlo draws.

A more detailed explanation of how PRAS incorporates energy-limited resources and calculates ELCC is available with its online documentation.[3]

7.3 Data

We examine two cases: (1) a two-zone, two-time period model to demonstrate the model, and (2) a 20-zone MISO model to test results on a realistic power system.

The two-zone model has one zone with a much higher reserve margin than the other. The two zones are linked by case-varying transmission capacity. The second time period is higher load than the first, with the ratio adjusted in cases. The purpose of the two-zone model is to demonstrate the relevance of the spatial and temporal criteria.

[2]The type of added resource, e.g. 1-hour or 6-hour ESR, or solar PV, depends on the case. The rest of this paragraph refers to a "resource" as a term to encompass ESRs of varying duration and different types of generation resources.
[3]https://github.com/NREL/PRAS/blob/master/docs/getting-started.md

The 20-zone MISO model uses information on zonal balancing authorities assigned to MISO's footprint in NREL's Seams Study data [14]. Gross zonal hourly loads are available for a weather-normalized 2012[4] delivery year; the hourly averages by month are shown in Figure 7.2.

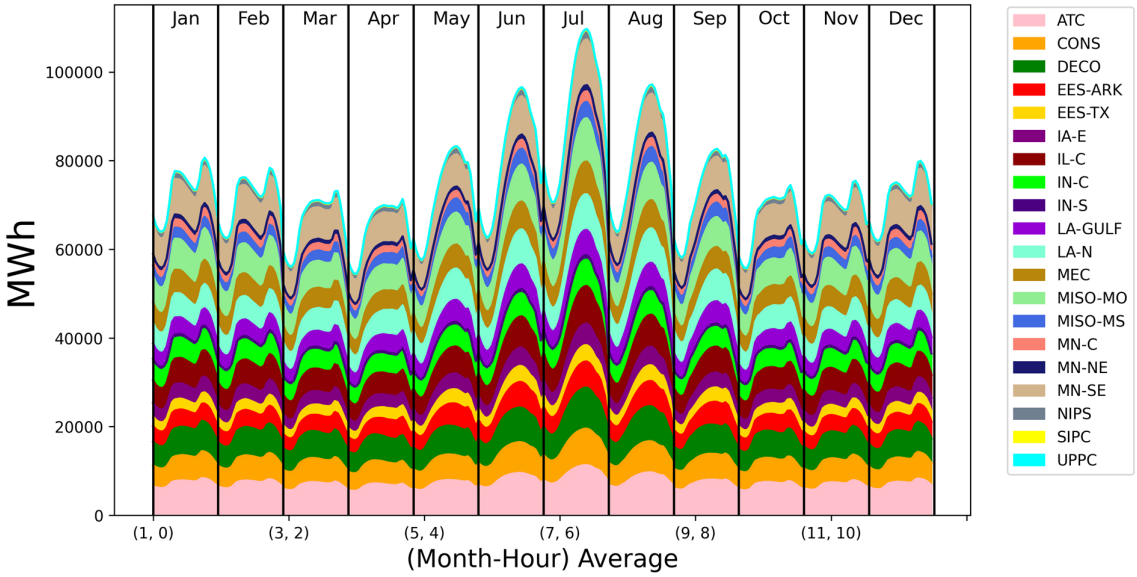

Figure 7.2: Gross hourly average loads for 2012 weather-normalized delivery year and included MISO zones. Names for zonal abbreviations are in Appendix A. Geographic locations of zones are in Figure 7.3.

Zonal shapefiles are not publicly available, so boundaries of MISO zones are mapped to a public shapefile of the entire MISO footprint[5] based on coordinates of renewable energy profiles assigned to the zones in Seams data provided by contacts at MISO.[6] Zones without either a MISO load profile or renewable energy profiles are excluded in results. The resulting zonal geography is Figure 7.3.

[4]2012 was a leap year with 8784 hours. We exclude February 29, 2012 data and evaluate all annual cases for 8760 hours in 2012.

[5]https://hifld-geoplatform.opendata.arcgis.com/datasets/9d1099b016e5482c900d657f06f3ac80_0

[6]Figueroa-Acevedo et al. [15] discusses data development for the Seams study, including renewable profiles and transmission representation of the Eastern Interconnect. Online documentation is available at https://www.nrel.gov/analysis/seams.html. However, at the time of writing the final Seams study and associated data are not yet fully publicly available.

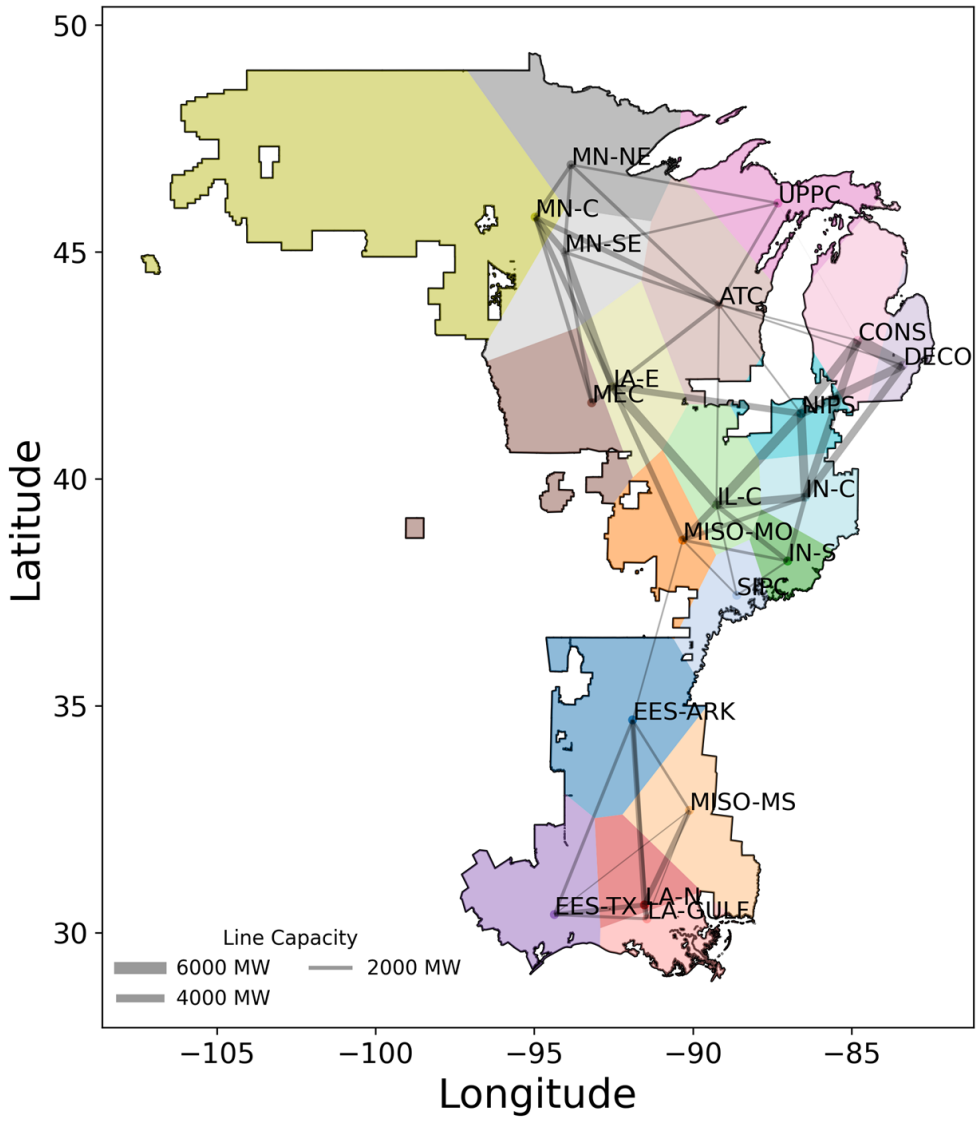

Figure 7.3: Zonal polygons and inter-zonal transmission capacity in retained 20-zone MISO model.

We also received Seams study data for the MISO footprint on interzonal transmission capacities, the individual generation resources assigned to each zone, as well as regional assignment of installed renewable energy capacity. Only resources assigned in Seams data to the 20 zones in Figure 7.3 are included in model results. Two external zones (AECIZ and CBPC-NIPCO) have transmission, generation, and load Seams data and are included in resource adequacy calculations, but they are not assigned renewable capacity and their RA results are not reported. Conventional generator parameters for MISO cases are shown in Table 7.2. Assigned equivalent forced outage rates (eFOR)

incorporate both planned and unplanned outages, and for simplicity assume planned outages are random. The same generator eFORs are used in the two-zone model, but total installed capacities differ.

Table 7.2: Conventional Generator Capacity and Failure Parameters.

Generator Type	**Total Capacity (MW)**	**eFOR (%)**
CC	23,140	10.88
Coal_ST	61,282	12.16
Gas_GT	24,260	6.37
Hydro	2,612	2.40
Nuclear	13,288	10.80
Biomass	615	13.97
Gas_IC	16	4.58
Gas_ST	4,323	15.00
Oil_ST	671	12.93
Oil_IC	41	4.58
PS	2,000	13.05
Waste HT_ST	18	2.40
Oil_GT	78	7.58

Zonal renewable penetration uses the wind-heavy, 40% renewable penetration by energy case developed in MISO's Renewable Integration Impact Assessment (RIIA) [33]. Additional information on RIIA scenarios is in Appendix A. Zonal installations of each renewable resource type are shown in Table 7.2. Renewable capacity is allocated from the MISO Local Resource Zone at which it is defined in RIIA scenarios to Seams zones pro rata with the Seams zone's fraction of LRZ load. Translating between the 20 retained Seams zones and 10 MISO LRZs is required because MISO data and scenarios for planning are for MISO-defined LRZs, while Seams uses a different definition of zones constituting the MISO footprint. Zonal renewable profiles use the highest capacity factor site. The far-right column in Table 7.3 also reports the top-100 gross load hour capacity factor as indicative of the resource's ELCC on a system where gross and net peak load hours are similar. Zonal variable generation capacity factors differ due to differences in resource quality; in particular, Seams zones in the southern portion of MISO's footprint typically have lower wind resource quality than zones in central and northern MISO. Differences in resource quality are generally reflected in both existing installed capacity and the RIIA wind-heavy, 40% renewable penetration scenario for

which installed capacities by LRZ are shown in Table 7.3.

Table 7.3: Renewable Generation Installation in MISO Base Case.

Renewable Generator Type	MISO LRZ	Constituent Seams Zones	Renewable Installed Capacity (MW)	Annual Capacity Factor (best site, same order as Constituent Seams Zones, %)	Top-100 gross load hours Capacity Factor (same order as Constituent Seams Zones, %)
Utility Wind	1	MN-C, MN-NE, MN-SE	20,100	49%, 50%, 49%	22%, 25%, 29%
Utility Solar	1	MN-C, MN-NE, MN-SE	2,700	17%, 17%, 16%	54%, 51%, 47%
Distributed Solar	1	MN-C, MN-NE, MN-SE	1,700	17%, 16%, 15%	55%, 53%, 52%
Utility Wind	2	ATC, UPPC	3,400	50%, 55%	26%, 47%
Utility Solar	2	ATC, UPPC	600	16%, 15%	46%, 42%
Distributed Solar	2	ATC, UPPC	1,200	16%, 15%	48%, 44%
Utility Wind	3	MEC, IA-E	21,900	47%, 51%	22%, 26%
Utility Solar	3	MEC, IA-E	1,400	18%, 17%	52%, 50%
Distributed Solar	3	MEC, IA-E	900	17%, 16%	54%, 52%
Utility Wind	4	IL-C, SIPC	4,800	47%, 47%	23%, 23%
Utility Solar	4	IL-C, SIPC	1,400	17%, 17%	49%, 46%
Distributed Solar	4	IL-C, SIPC	800	17%, 16%	50%, 48%
Utility Wind	5	MISO-MO	1,000	45%	10%
Utility Solar	5	MISO-MO	800	17%	49%
Distributed Solar	5	MISO-MO	700	17%	49%
Utility Wind	6	IN-S, IN-C, NIPS	3,100	45%, 43%, 46%	18%, 20%, 24%
Utility Solar	6	IN-S, IN-C, NIPS	5,000	17%, 16%, 16%	46%, 43%, 45%
Distributed Solar	6	IN-S, IN-C, NIPS	1,400	16%, 16%, 16%	48%, 44%, 46%
Utility Wind	7	CONS, DECO	6,200	51%, 44%	36%, 27%
Utility Solar	7	CONS, DECO	800	16%, 16%	42%, 43%
Distributed Solar	7	CONS, DECO	1,700	16%, 15%	44%, 46%
Utility Wind	8	EES-ARK	100	47%	23%
Utility Solar	8	EES-ARK	4,000	17%	46%
Distributed Solar	8	EES-ARK	700	16%	48%
Utility Wind	9	LA-GULF, LA-N, EES-TX	0	42%, 46%, 40%	13%, 14%, 14%
Utility Solar	9	LA-GULF, LA-N, EES-TX	4,800	17%, 17%, 17%	44%, 47%, 51%
Distributed Solar	9	LA-GULF, LA-N, EES-TX	2,100	17%, 16%, 17%	44%, 48%, 52%
Utility Wind	10	MISO-MS	100	35%	8%
Utility Solar	10	MISO-MS	1,800	17%	43%
Distributed Solar	10	MISO-MS	500	16%	43%

The additional renewable capacity creates capacity excess compared to the assumed 2012 load profiles. We therefore adjust load profiles to a MISO-wide planning reserve margin (PRM) of 18% for general consistency with MISO's current capacity-based planning reserve margin [34]. Renewables are counted toward PRM at their capacity factor in the top-100 gross load hours as a computationally fast approximation of their ELCC, conventional generators are counted at their nameplate capacity. We do not assume any installed ESR capacity in the base case; ESR capacity is added and its ELCC evaluated as specified in cases. We include a set of sensitivity cases with 30 GW of installed 6-hour duration ESR capacity in the main results. In these cases and other sensitivities with ESRs the ESR capacity is assigned to zones pro-rata with zonal renewable capacity and the MISO-wide PRM is re-adjusted, for simplicity to ignore resource type-specific retirement decisions, to 18% by adding load. Monte Carlo draws simulate 1,000 runs of all 8760 annual hours in MISO cases.

7.4 Results

Results are first presented for a two-zone test system for two hours to show the importance of the temporal and spatial criteria to an ESR-transmission complementarity benefit in resource adequacy. The complementarity benefit is quantified as a higher ELCC for ESRs when transmission capacity is increased, showing the ESRs can use the additional transmission capacity to access generation in another zone and charge for later discharge. Comparing with a generator demonstrates when the complementarity is unique to the charging and discharging behavior of ESRs. A subsequent resource adequacy assessment of the annual (8760) 20-zone MISO representation demonstrates the practical relevance of results.

7.4.1 Two-zone model

The two-zone, two-hour model has one zone ("Zone 1") with a lower reserve margin than the other ("Zone 2"). Load is higher in hour 2 than hour 1 so a 1-hour duration ESR can charge in the earlier hour and discharge in the next hour. The two zonal reserve margins are adjusted across the three sets of cases shown in Figures 7.4-7.6 to change relative zonal LOLP and demonstrate the spatial

criterion. The two zones are connected by a transmission line. The capacity of the transmission line is changed in runs within each set of cases to allow excess transmission capacity in neither hour, only hour 1, or both hours, demonstrating the relevance of the temporal criterion.

All two-zone, two-hour model runs use 1,000,000 Monte Carlo draws of generator, storage, and transmission availability in the two zones. Generation resources fail and recover randomly with the eFORs defined in Table 7.2. For simplicity ESRs and transmission are assumed perfectly available, so Monte Carlo draws only reflect different time series of generator availability. ELCCs are calculated using the same 1,000,000 draws for a 100MW addition of an evaluated resource (e.g., 1-hour duration ESR) in each of the two zones and reported as a percent ELCC of the resource's max ELCC of 100MW:

$$\%ELCC = \frac{ELCC(MW)}{MaxELCC(MW)} \tag{7.1}$$

7.4.2 Demonstrating conditions with two-zone model

Three sets of cases demonstrate the spatial criterion in Figure 7.4-Figure 7.6 by adjusting zonal reserve margin and hourly load to show how differences in zonal reliability metrics interact with interzonal transmission capacity to affect resource ELCC. Transmission capacity connecting the two zones is then varied within each set of cases (see panel B of Figure 7.4-Figure 7.6) to test the applicability of the temporal criteria. The cases in Figure 7.4-Figure 7.6 have homogeneous, temporally heterogeneous, and temporally and zonally heterogeneous reserve margins, respectively.

In the Figure 7.4 case both zones have a lot of internal excess generation capacity in both hours (homogenous zonal reserve margins). Resulting LOLP in both zones is 0% irrespective of interzonal transmission capacity. The spatial criterion does not apply because zonal LOLPs are equal.

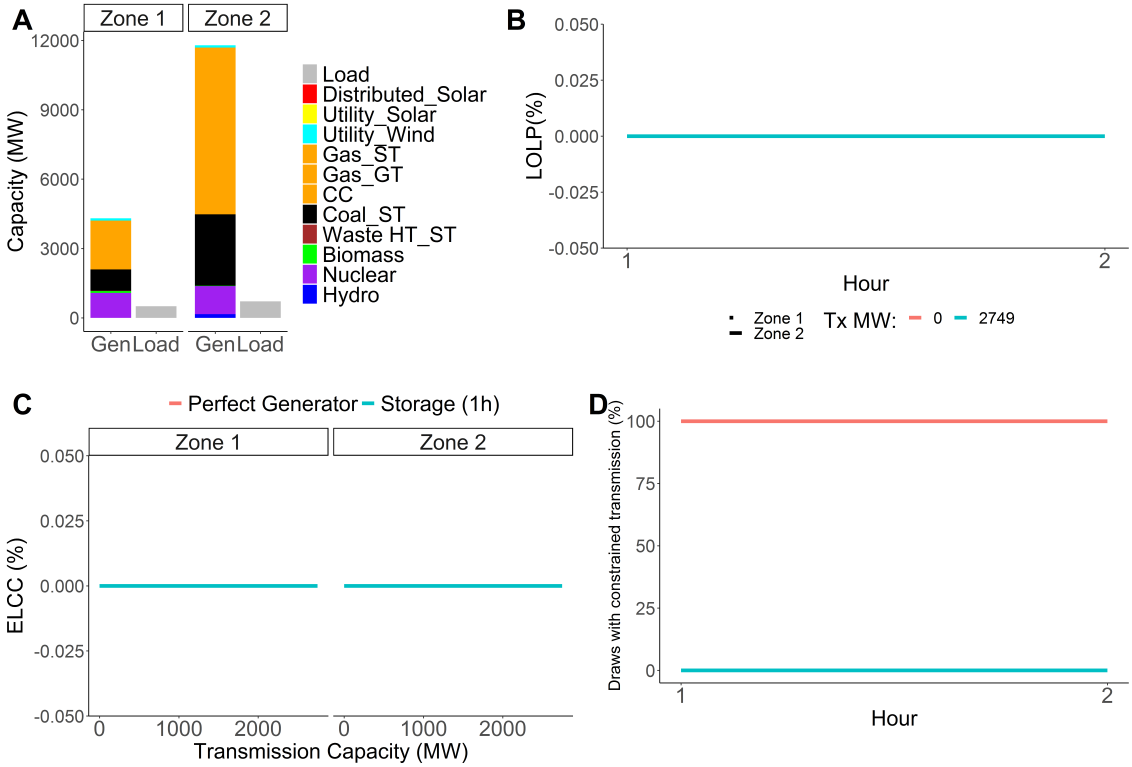

Figure 7.4: Both zones have large quantities of excess internal generation capacity compared to load (A). This results in 0.00% LOLP in both modeled hours (B), even without a transmission link between zones (red line in (B) and (D)). Both 1-hour duration storage and a perfectly available generation resource have 0% ELCC (C) because the system is already perfectly reliable within rounding error precision.

In the Figure 7.5 case both zones have a lot of excess internal generation capacity in hour 1, but insufficient internal generation capacity in the very high load hour 2 (temporally heterogeneous zonal reserve margins). The spatial criterion does not apply because LOLP is different across time (100% hour 2, 0% hour 1) but not space (same in both zones for the same hour).

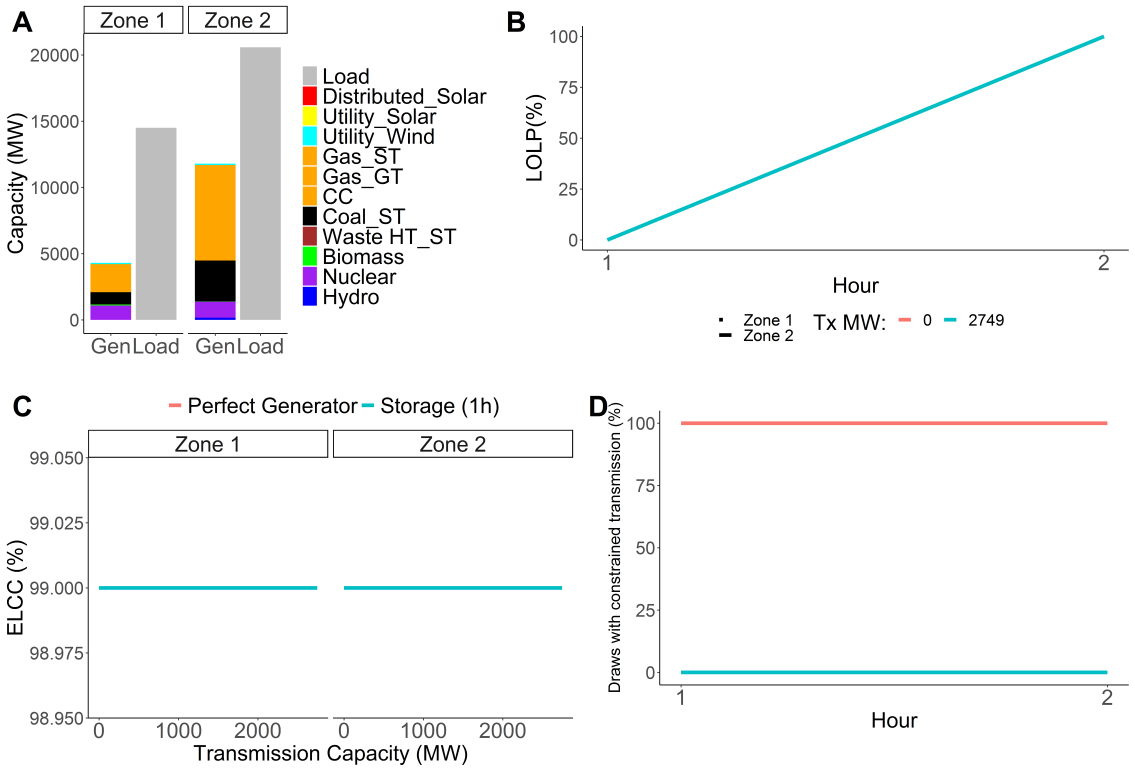

Figure 7.5: Both zones are deficient in installed generation capacity compared to peak load (A) in hour 2, but not hour 1 (B). This results in 0.00% LOLP in hour 1 but 100% LOLP in hour 2 in both zones (B). Transmission does not constrain generators' ability to serve load (D). Both 1-hour storage and a perfectly available generation resource have 99%±1% ELCC irrespective of which zone they are in (C) because they have the same ability to serve load in hour 2.

In the set of cases in Figure 7.6 zone 2 has excess internal generation capacity in both hours, but zone 1 is deficient in internal generation capacity in both hours. Without a transmission link the LOLP in zone 1 is non-zero in hour 1 and 100% in the higher load hour 2. LOLP is 0% in both hours in zone 2. That LOLP is non-zero for only a single zone in a single hour means LOLP is both zonally and temporally heterogeneous. More transmission capacity allows the zone 2 generation capacity to serve load in zone 1 and affects resources' ELCCs.

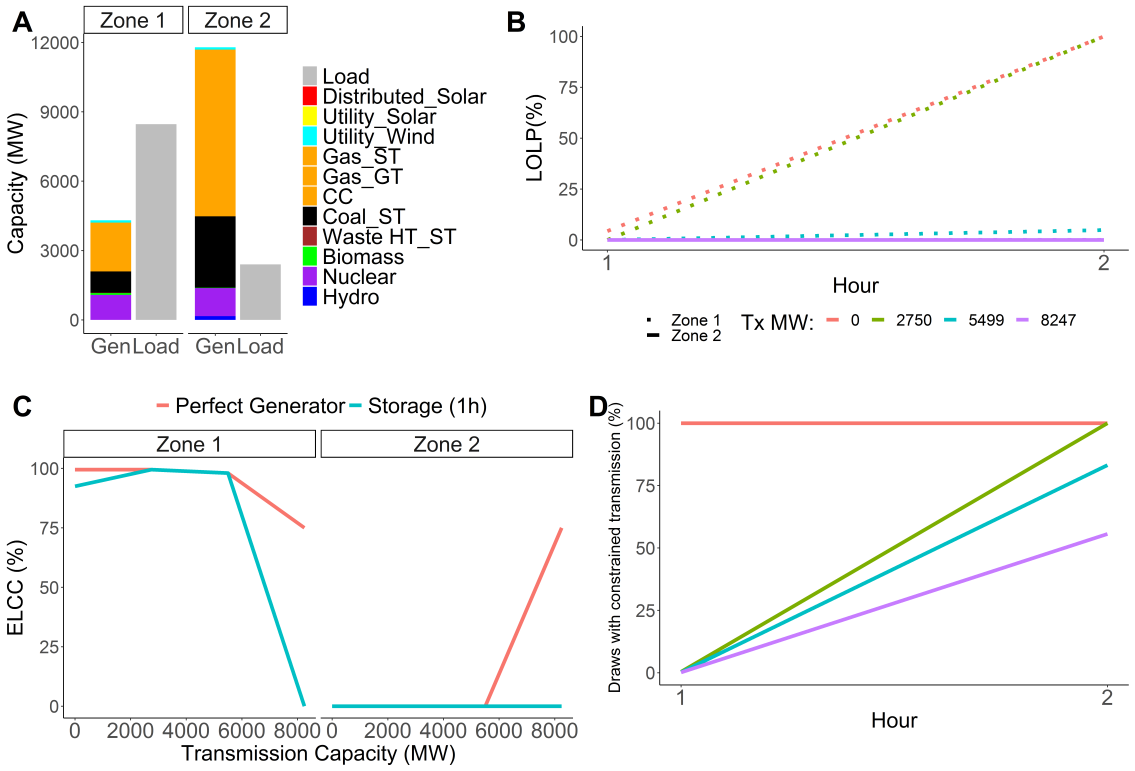

Figure 7.6: Zone 2 has significant excess internal generation capacity in both hours, while Zone 1 is deficient in internal generation capacity both compared to peak load (A) in hour 2 and, depending on generator failures and transmission availability, hour 1 load as well (B). A complementarity benefit is realized between storage and transmission that is not true of generators (C), see Figure 7.7 for additional detail.

The key insight in Figure 7.6 is how transmission capacity affects resource ELCC can be different for ESRs and generators (panel C). Unlike generators, ESRs must first charge to later discharge. The ESR ELCC depends on its ability to use available transmission to charge in an earlier hour without unserved energy[7] so it is available to later reduce unserved energy. Generators are not intertemporally constrained in RA models.[8] The intertemporal linking of ESR availability can result in an increasing ELCC "complementarity benefit" for ESRs with transmission availability for ESR charging that does not exist for generators, which do not need to charge.

The complementarity benefit is realized only under circumstances where shifting energy in both

[7]If the earlier hour has unserved energy, charging the ESR will increase total unserved energy assuming <100% roundtrip efficiency

[8]In operational models conventional generators are often modelled with commitment decisions reflecting minimum online, offline, and startup/shutdown constraints, which make their availability temporally linked. This detail is not included in RA models.

space (i.e., across zones to charge ESRs) and time (i.e., from an earlier to a later hour) is important. These are the circumstances where the two criteria apply. Figure 7.7 provides additional detail on when and why this occurs in the same two-zone, two-time period case introduced in Figure 7.6.

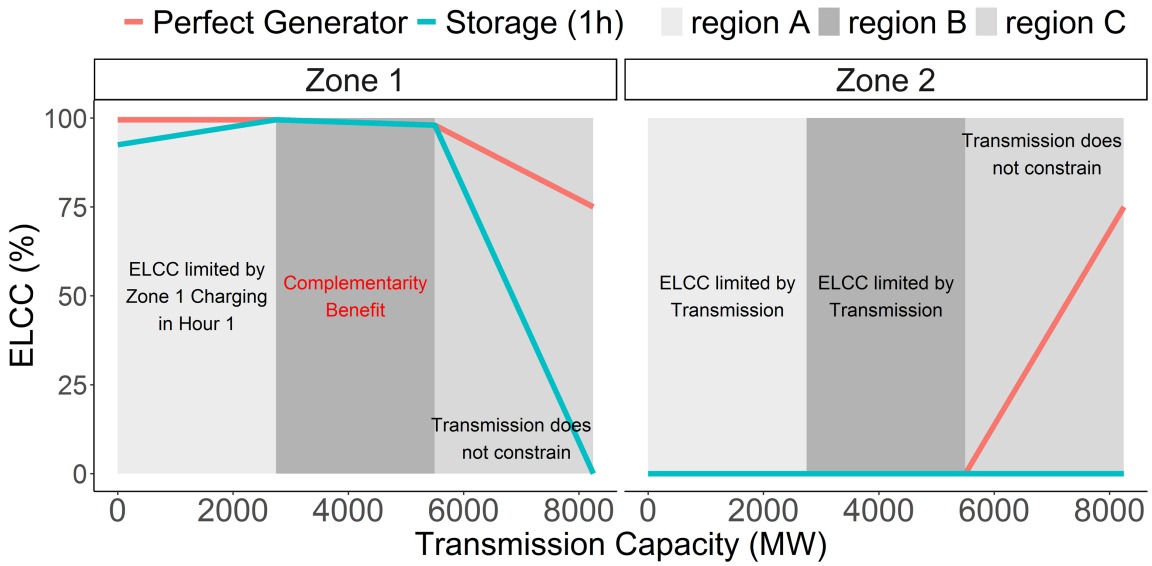

Figure 7.7: Increasing transmission capacity in region A increases ESR ELCC in the higher LOLP zone 1 because ESRs can increasingly be charged in hour 1 by generators in zone 2 for discharge in hour 2 in zone 1. In region B transmission capacity is always sufficient to charge zone 1 ESRs in hour 1 for discharge in hour 2 but always insufficient to reduce LOLP in hour 2 to 0%. Both criteria apply and the complementarity benefit is maximized. In region C transmission capacity is sufficient to use resources in Zone 2 to serve all load in Zone 1 in hour 2 directly without needing to shift energy from hour 1.

Figure 7.7 is proof of concept for the two criteria. First, unequal zonal LOLPs, the spatial criterion, are necessary for valuing shifting energy in space to the higher LOLP zone with transmission (region B vs. region C). Alone, additional ability to shift energy in space increases both generator and ESR ELCC. For there to be a unique benefit for ESRs, transmission availability must be different in the earlier and later hour. This temporal criterion means ESRs complement available transmission capacity by using available transmission capacity to shift energy in time (region A vs. region B), increasing overall use of the transmission capacity. Generators do not shift energy in time; therefore, they cannot increase overall use of transmission capacity by charging in an earlier time period for discharge in a later time period, and do not realize the same complementarity with transmission capacity.

Extending the spatial criterion to more than two zones is straightforward: identify zones with different reliability metrics, indicating a binding transmission constraint. Extending the temporal criterion to more than two timepoints is less straightforward because event lengths are heterogeneous and ESRs can have different durations. Figure 7.8 demonstrates a method for generalizing the temporal criteria by comparing the probability of different duration loss-of-load events to the duration of the ESR at different levels of installed transmission capacity. ESR ELCC is approximated by the percent of event durations shorter in time than the ESR duration. In the case of a 1-hour ESR, it can serve only 1-hour events (note 0-hour "events" mean no loss of load).

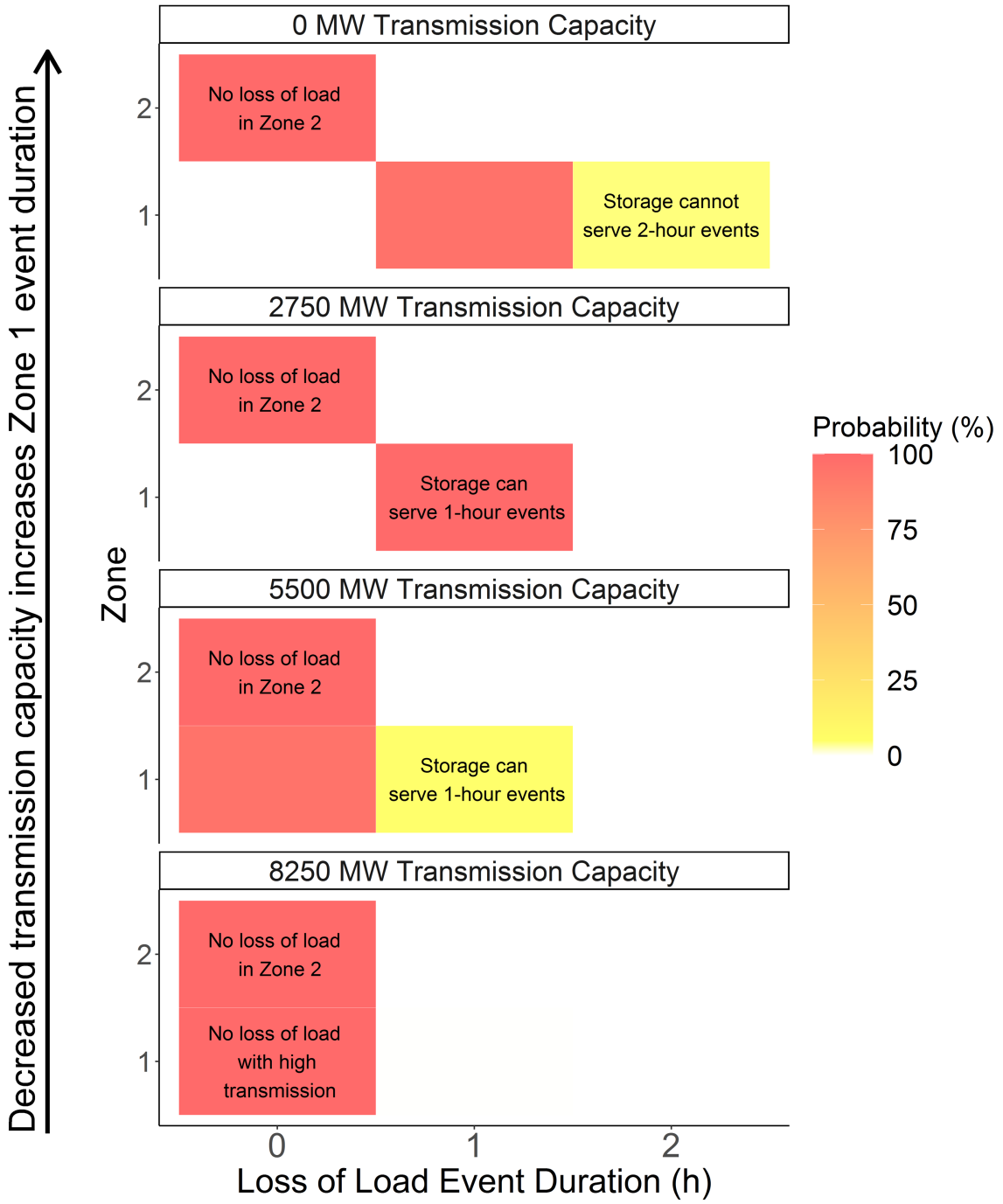

Figure 7.8: Probability of loss-of-load events of each hour duration in each zone for different quantities of transmission capacity between zones. Event durations with 0% probability are white/blank, all non-zero probabilities are shaded yellow, orange, or red as per the legend. 0-hour duration means no loss of load in that zone in the case. ESRs can serve 1-hour events (all of which occur in hour 2) by charging in hour 1, but not 2-hour events. Going from 0 to 2750 MW of transmission capacity eliminates 2-hour events, increasing ESR ELCC.

For the two-hour, two zone case Figure 7.8 exhaustively enumerates all possible LOLP outcomes, and no quantification of ESRs' ability to charge between loss-of-load events is needed. Because LOLP is always higher in hour 2 than hour 1 in the Figure 7.8 case, separating 1-hour events occurring in hour 2 vs. hour 1 is unnecessary. However, note ESRs could not serve hour 1 events if they began with zero state of charge. The starting time period state of charge assumption affects a smaller fraction of hours when evaluating realistic test systems and is ignored for simplicity when extending the event duration screen to 8760 time periods in the next section.

7.4.3 Extension to more zones and time periods with MISO test case

The previous examples considered only two zones with two time periods to demonstrate spatial and temporal criteria relevance. RA assessments of realistic power systems require extension to more zones and time periods. This section quantifies zonal resource ELCCs in a realistic power system using data for a 20-zone representation of MISO. Base case (40% VRE, 18% PRM) EUE is 30 GWh (0.005% of annual load), or 0.4 hours of average hourly load of 78 GWh.

The spatial criterion requires a target RA metric differ across zones. Differences in EUE or another RA metric indicate constrained transmission because excess transmission capacity would be used to equilibrate zonal RA and increase system reliability. This criterion is screened by calculating zonal RA metrics using available data on transmission interfaces and generator locations. Misalignment of zonal loss-of-load events indicates a transmission interface is constrained and should be included. Figure 7.9 shows an example of this approach for the MISO test case using EUE as the target RA metric. Assuming zones with >5% of the maximum zonal EUE (16,000 MWh) cannot be aggregated results in 17 of the 51 lines remaining in the second panel of the figure.

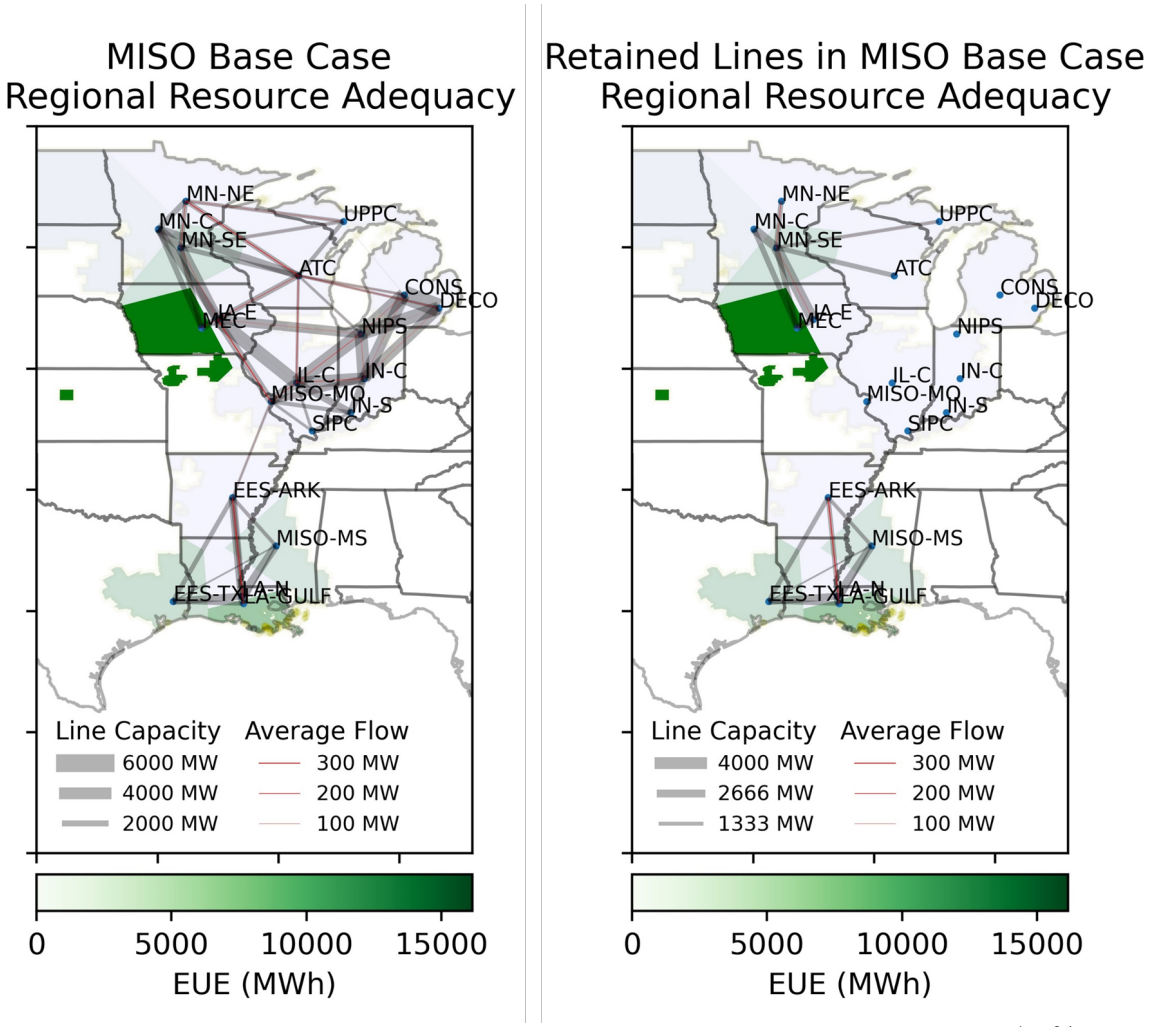

Figure 7.9: Example of spatial criterion evaluation using MISO base case data (40% VRE target). Transmission interfaces between zones without sufficient EUE-based differentiation (EUE<800MWh) are dropped in the right panel. Zones could then be combined into a larger single zone. Using fewer zones decreases computational burden. Average flow is a simple average of all 8760 hours and 1000 Monte Carlo draws.

The temporal criterion depends on the ability of ESRs in a zone with non-zero LOLP to serve loss-of-load events of multiple time durations. An extension of the proposed screen from Figure 7.8 to more than two time periods considers the probability of each duration of expected zonal loss-of-load events on the test system at hourly resolution. The temporal criterion applies when some loss-of-load events exceed the duration of ESR discharge from maximum state of charge. Event duration probabilities are shown for the MISO test system zones below for two cases: base case

transmission and 25% of base case transmission capacity. In the latter case all line capacities are reduced 75%. Ceteris paribus reduced transmission capacity increases LOLH (and EUE) because RA depends more on sufficiency of intrazonal resources.

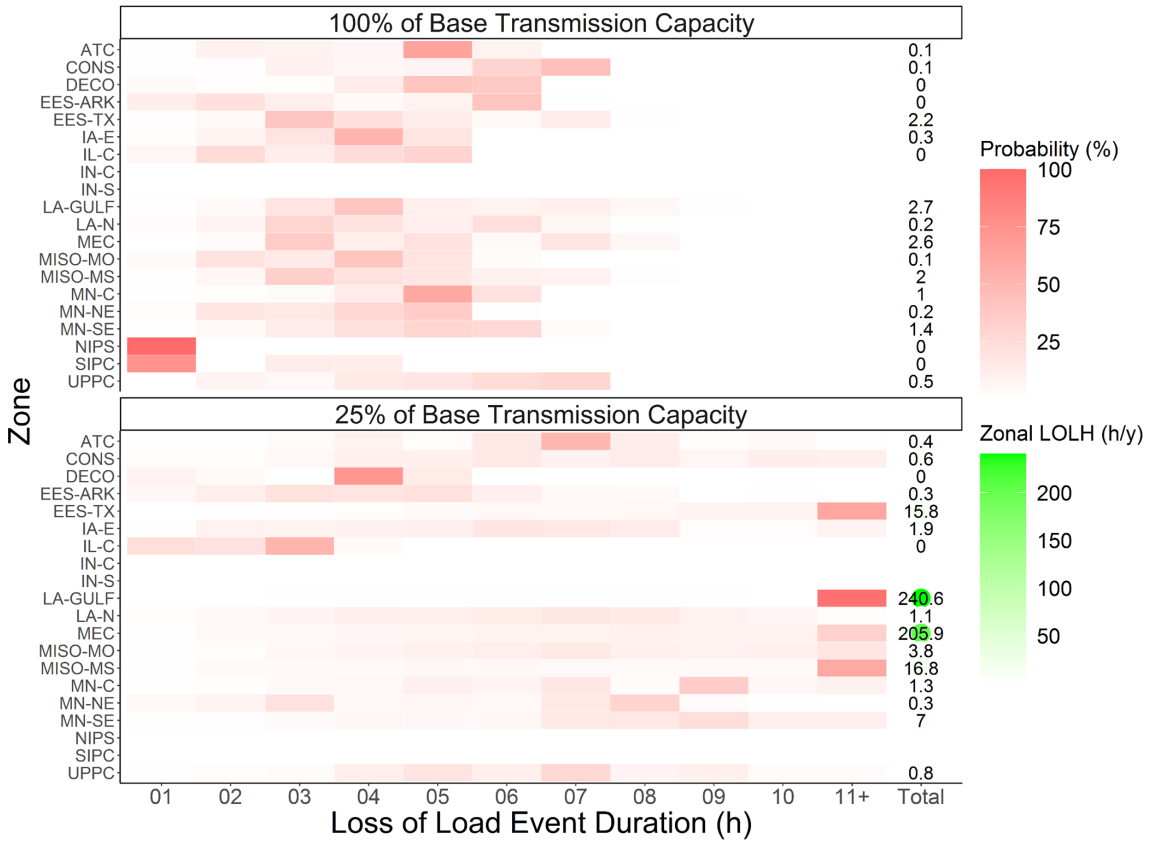

Figure 7.10: Zonal loss-of-load event duration probabilities (red, rectangles) and annual LOLH (green, circles at far right) for base and 25% transmission capacity cases. Probabilities are the percent chance of an event of the labeled duration out of all hourly resolution zonal events. The far right column shows the total zonal LOLH in event hours per year. With increased transmission capacity overall LOLH is lower and shorter duration events are more common; 6-8 hour duration ESRs can serve most events. As transmission capacity is reduced events lengthen and a 6-8 hour duration ESR can no longer fully serve longer-duration events.

Whether the zones meet the temporal criterion then depends on the duration of the ESR considered, shown in more detail in the next section. There are simplifications made in a discrete event-based temporal criterion screen (Figure 7.10): for example, it does not consider the number of hours between consecutive events during which ESRs can recharge to enable serving the next event with sufficient state of charge. Additionally, PRAS dispatches ESRs to minimize unserved

energy; consideration of other objectives in ESRs' dispatch heuristic and uncertainties in dispatch to (co-)optimize other objectives may reduce RA credit. These considerations mean ESRs may not be able to serve shorter duration events than its fully rated capability, so the zonal event duration screen for the temporal criterion is a minimal screen for inclusion.

7.4.4 MISO zonal results

The previous section applies the temporal and spatial criteria to the MISO zonal test system. Screens show as few as 17 of the 51 transmission interfaces can be included and that the temporal criterion does not apply for 6-8 hour duration ESRs with base case transmission capacity, but will apply for shorter duration ESRs or for longer duration ESRs with reduced transmission capacities.

To demonstrate these screens are appropriate in results, ELCCs are calculated for 500MW additions to the MISO test system of four types of resources: a 1-hour ESR, 4-hour ESR, 6-hour ESR, and solar PV. ELCCs are reported as %ELCC (Eq. 7.1). ESR ELCCs decline with decreased transmission capacity, indicating a complementarity benefit between ESRs and transmission. The range of simple average ELCC reductions in the five highest LOLH zones with halved transmission for 1-6 hour ESRs is 25%-71% (Figure 7.11). Solar PV does not show similar ELCC declines with reduced transmission capacity.

Although the spatial criterion suggests some adjacent zones with zero LOLH could be aggregated in the base case, we evaluate resource ELCC for all 20 zones to fully demonstrate results. Each ELCC evaluation uses 1000 Monte Carlo samples of the 8760 hours. ELCCs are reported for a 0.05 p-value evaluating the hypothesis test that the upper bound on ELCC is statistically significantly different than the lower bound after each bisection. Load for ELCC evaluation is added in the same zone as the additional resource. Resource zonal ELCC evaluations for the 1000 Monte Carlo draws take 2-12 minutes of clock time on an Intel Xeon CPU E5-2680 v3 2.50GHz desktop with 12 cores and 64.0 GB memory depending on how many bisections are needed before the upper and lower bound RA metric are no longer statistically significantly different. That each zonal resource evaluation takes this much time for 1000 Monte Carlo draws highlights the importance of applying the criteria screens before including multiple zones in RA.

Figure 7.11 results are for the base case with 0 GW of ESR installed capacity (ICAP) before addition of the 500MW resource for ELCC evaluation. Figure 7.12 also shows the ELCCs of the same resources at 30 GW of 6-hour ESR ICAP; 30 GW is approximately 25% of recent MISO peak load (see Figure 7.2). On the x-axis the transmission capacity between zones is varied by multiplying the base case transmission capacities by a percent scalar; 100% means the base case capacity, 25% would be a 75% reduction in each line's base case capacity, 200% a 100% increase, etc. As suggested by comparing zonal LOLHs for the top and bottom panels in Figure 7.10, changing transmission capacity assuming the same resource portfolio affects overall LOLH: more (less) transmission capacity makes total LOLH increase (decrease) due to increased (decreased) ability of transmission to deliver generation to lower reliability zones. Also as suggested by Figure 7.10, increasing LOLH with decreasing transmission capacity tends to make events longer, which reduces ELCCs of increasingly longer-duration ESR additions. Finally, Figure 7.10 shows five zones with >1.0 LOLH in the base case (100% Transmission Capacity): LA-GULF, MEC, EES-TX, MN-SE, and MISO-MS. While Figure 7.11 and Figure 7.12 include ELCC results for all zones, they highlight the five highest LOLH zones because they are most relevant for increasing RA, similar to how Zone 1 was more relevant than Zone 2 in the example cases in Section 7.4.1-7.4.2.

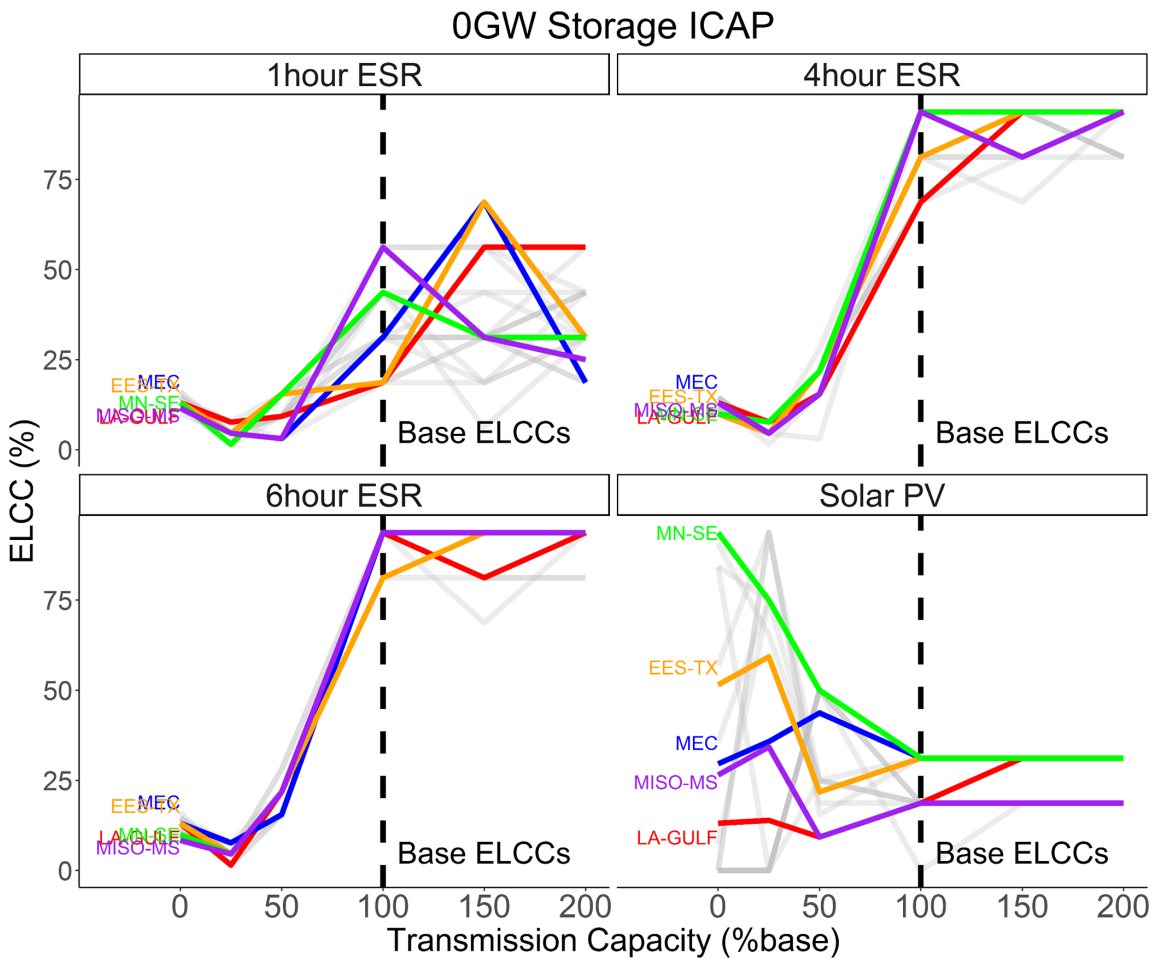

Figure 7.11: Zonal ELCCs as a function of increasing transmission capacity with 0GW of installed ESR capacity. The five zones with highest LOLH are highlighted and labeled.

The ESR complementarity benefit is most prominently displayed in Figure 7.11 for 4-hour and 6-hour duration ESRs as transmission capacity increases from 50% of base case to base case (100%) capacity: 4-hour ESR ELCCs increase from a simple average of 19% to 86% in the five highlighted zones, and 6-hour ESR ELCCs from 21% to 92%. Increased transmission capacity reduces the expected duration of loss-of-load events in the five zones, meaning the temporal criterion applies and ESRs are able to serve an higher percent of loss-of-load events. 1-hour ESRs also see an increase in ELCC from 9% to 34%; the effect is smaller because most loss-of-load events are longer than one hour. Reducing transmission capacity by half thus reduces ELCCs of 1-6 hours ESRs by 25-71%. This effect does not apply for the generator: simple average solar PV ELCCs in the five zones

increase from 26% in the base case to 27% for the same halved transmission capacity.

Increasing transmission capacity above base case values shortens average event duration, but 6-hour duration ESRs ELCCs are already saturated because events are rarely longer than 6 hours with base case transmission capacity. Shorter average event duration due to increased transmission capacity increases 1-hour ESR average ELCCs from 34% to 51% (+17%) and 4-hour ESRs from 86 to 89% (+3%) in the five highlighted zones with 50% more transmission capacity than the base case. However, doubling of transmission capacity (+100%) does not increase average 1-hour ESR ELCC more than a 50% increase because with high levels of transmission capacity deliverability no longer constrains resources in one zone from reducing EUE in another. Average solar PV ELCCs are unaffected by increasing transmission capacity. Increasing transmission capacity increases the ELCCs of ESRs when it reduces length of loss-of-load events to shorter than the ESRs' duration. This complementarity benefit does not apply if event length or ESR duration are too short, and does not apply to generation since generators cannot temporally shift load to increase use of available transmission capacity at earlier times.

Figure 7.12 shows the same 500MW addition resource ELCCs for a "high existing ESR" case with 30 GW of existing 6h ESR ICAP (MISO-wide PRM is adjusted to remain 18%). A comparison of resource ELCCs with increased ESR ICAP shows solar PV is complemented by ESR installation (higher solar PV ELCC at higher ESR penetration), but ESRs exhibit declining marginal capacity credit (lower ESR ELCC at higher ESR penetration).

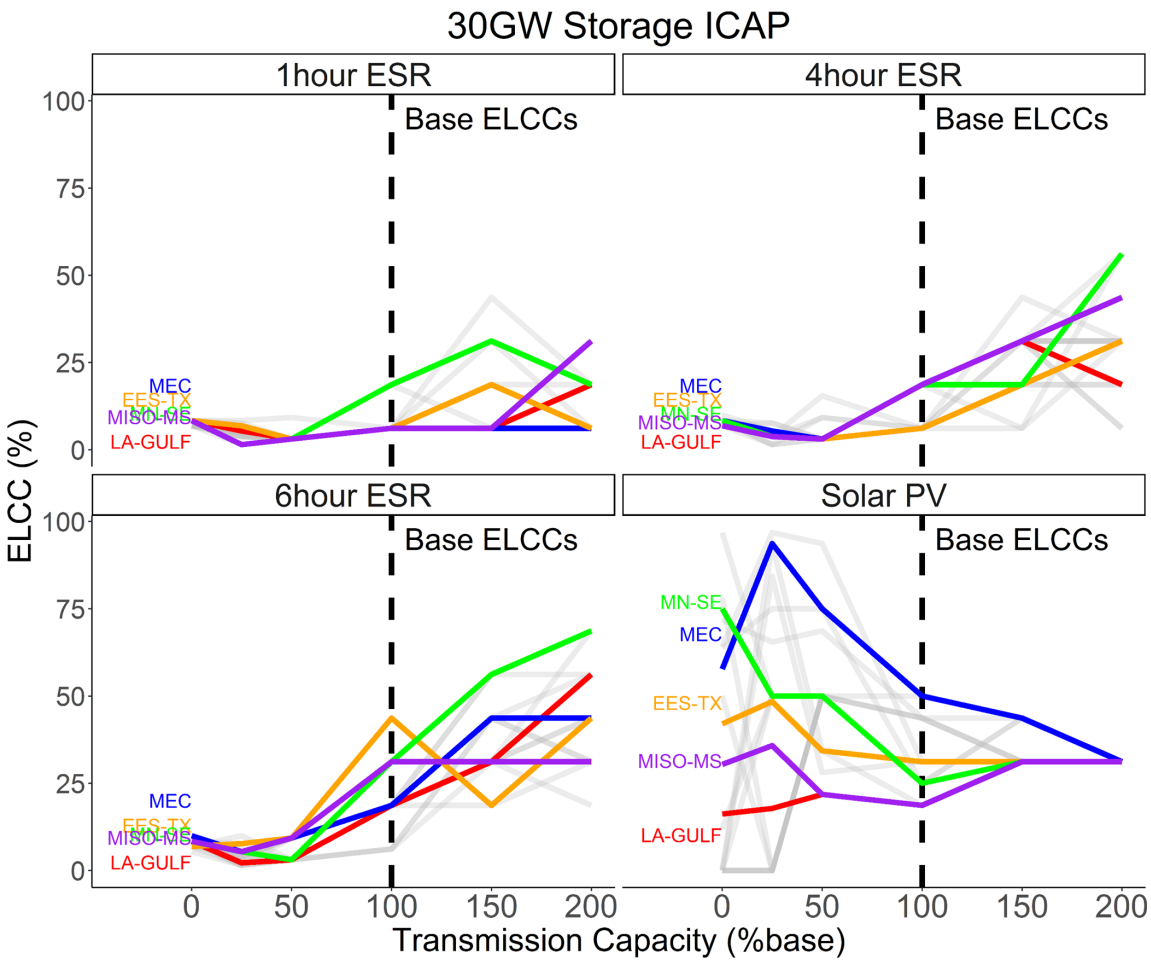

Figure 7.12: Zonal ELCCs as a function of increasing transmission capacity with 30GW of installed 6h ESR capacity. The five zones with highest LOLH are highlighted and labeled.

Average transmission base case ELCCs in the five highest LOLH zones are 9%, 16%, and 29% for 1-,4-,and 6-hour ESRs, respectively, compared to 34%, 86%, and 92% in the no existing ESR base case (Figure 7.11). Reduced transmission base case ELCCs quantify declining marginal capacity credit of ESRs with ESR penetration. Average solar PV ELCC is 29%, higher than the 26% in the no existing ESR base case. Higher ELCC means solar PV is complementary with ESRs, supporting previous research conclusions [12, 35]. Increased ESR ICAP shifts complementarity with transmission to higher installation capacity: in the 30GW ESR ICAP case, doubling the base case transmission capacity increases average 1-,4-, and 6-hour ESR ELCCs from 9%, 16%, and 29% to 16% (+7%), 36% (+20%), and 49% (+20%), respectively. More transmission capacity is needed

to realize complementarity benefits than in the no existing ESR case because a system with high more ESR contribution to an equal RA target has relatively less generation. This makes the system more dependent on its transmission to charge ESRs from available, potentially variable generation in different zones. Transmission constraints bind more often, increasing EUE, until higher quantities of transmission capacity than in the no existing ESR case, so more transmission capacity is needed to maximize complementarity benefit with ESRs.

Case results show that while zonal resource ELCC is system-dependent, the spatial and temporal criteria apply for ESRs of sub-6 hour duration on a realistic test system. Reducing transmission capacity reduces the ELCC of increasingly long (6+ hour) duration ESRs. As increased transmission capacity reduces some loss-of-load event durations below an ESR's duration its ELCC increases. This indicates an ESR-transmission complementarity benefit in zonal RA.

7.5 Conclusion and Policy Implications

We develop a two-part screening criteria for including transmission interfaces in RA models with ESRs. The two criteria are different levels of zonal reliability, indicating a constrained transmission interface in reliability-relevant hours, and ESR discharge durations shorter than some loss-of-load event durations, indicating the ESR's need to shift as much energy in time as possible to maximize its reliability contribution (ELCC). Together these "spatial" and "temporal" conditions indicate transmission constrains an ESR's ability to charge to serve reliability events by discharging for its full duration at full rated capacity. This indicates more transmission capacity will complement the ESR, increasing its ability to charge for later reliability-relevant discharge and thus increase its ELCC. The resulting complementarity benefit is demonstrated in model cases by increased ESR ELCC with increased transmission capacity when moving more energy in space and time increases reliability. The same complementarity benefit does not apply for generators because they do not increase use of additional transmission capacity by charging for later discharge.

Results emphasize the importance of considering transmission interfaces in resource adequacy with ESRs when ESRs will charge from resources external to the ESR's zone for later discharge

when transmission capacity is constrained. Under these conditions ESRs increase use of incremental transmission capacity for reliability. This means incorporation of transmission capacity in RA endogenizes a benefit traditionally considered in transmission planning or proposed approaches for ESRs substituting for transmission [36].

Results are policy relevant to planning areas with zonal transmission constraints currently using or considering ELCC-based methods for assigning ESR capacity credit. This minimally includes all planning areas with centralized capacity markets, resource-level capacity credits, and zonal clearing prices. While RA models will not include the topological detail in nodal production cost models, in practice representing even one major limiting transmission interface can be important. For example, Mills et al. split PJM into Western and Eastern zones to better reproduce historical prices [37], and California Independent System Operator (CAISO) wholesale prices are commonly geographically differentiated into northern and southern areas (denoted NP15 and SP15, respectively) in analyses [38]. To underscore the importance of including a major limiting interface two zonal capacity-weighted clearing prices for each of the three centralized United States capacity markets (PJM, ISO-NE, NYISO) since their inception are shown in Figure 7.13.

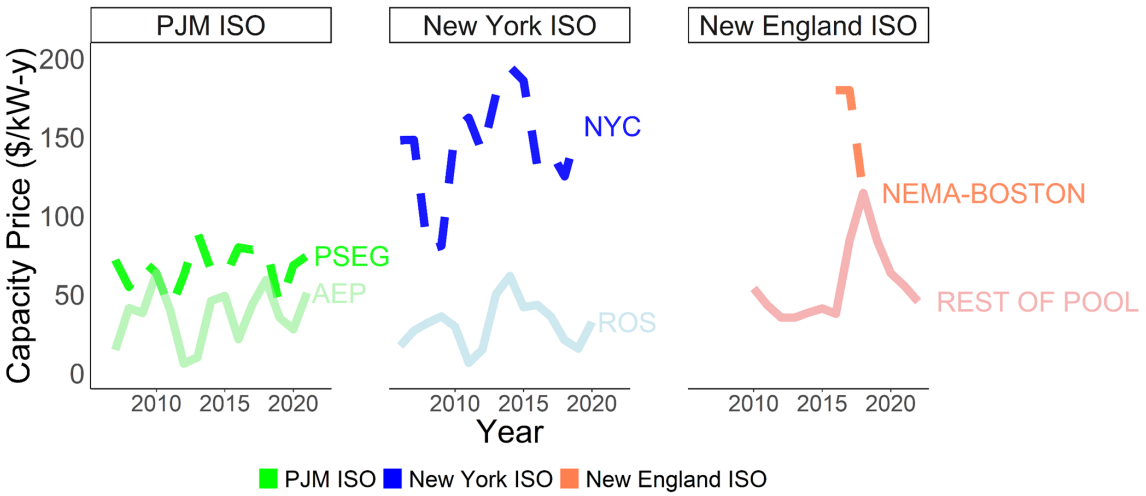

Figure 7.13: Annual capacity clearing prices for a more and less expensive zone in each of the three major United States ISOs with centralized capacity auctions. NYISO runs 6-month winter and summer auctions; values shown are summer clearing prices. ISO-NE only reports zonal prices when different from the "Rest of Pool" clearing price. The NYISO Rest of State ("ROS") line is renamed the New York Control Area (NYCA) clearing price from 2016-present. PSEG is Public Service Electric and Gas (NJ), AEP is American Electric Power (OH), NEMA is Northeastern Massachusetts, NYC is New York City. All three markets have more than two different zonal clearing prices in some auctions.

Different zonal clearing prices in today's markets suggests limits on inter-zone deliverability, so the spatial criterion applies. If ESR capacity additions are shorter duration than some loss-of-load events, the temporal criterion applies and considering ESR-transmission complementarity in zonal RA may materially affect ELCC. For example, if loss-of-load events are more likely in NYISO's New York City (NYC) zone[9] than upstate but are short enough to be served by an ESR, this could be identified as a lower cost alternative to additional, rarely used transmission capacity or a conventional generator in the resource adequacy process without requiring a separate transmission or non-wires alternative study process. Using the two criteria will directly compare the ELCC contribution of a ESR in NYC to an ESR in other zones, a generator in NYC, or additional transmission deliverability. The NYC ESR may be preferable when it can use available existing transmission capacity in earlier hours to charge for discharge in later hours with non-zero LOLP when transmission constrains deliverability.

Results also underscore the general importance of considering transmission investment alongside

[9]New York City is NYISO Zone J, but we refer to it here as NYC.

generation and ESRs in planning. Research identifies considerable benefits in better coordinating transmission investment decisions with other resources using probability-weighted scenarios for future loads and generation resources [39, 40]. The importance of co-optimizing transmission planning increases with forecast future increases in variable and energy-limited resource penetration [14]. However, transmission and generation planning are commonly done in parallel, even for the high-voltage bulk system. Separate tracks for ESRs to be valued in transmission ("storage as a transmission asset", SATA) and generation planning creates potential for under- or double-counting ESRs' value. SATA assumes ESRs substitute for additional transmission under some circumstances; we show ESRs are more accurately thought of as complementing transmission by shifting load to increase use of existing transmission capacity. The broader point is ESRs are neither transmission nor a generator, so trying to fit them into old constructs developed for those resources instead of better valuing services like resource adequacy can be quixotic. Compensating ESRs through rate-of-return constructs historically applied to transmission may have benefits in increasing projects' revenue certainty, but misses an opportunity to take a step toward more holistically evaluating resource contributions in providing services. This general point also applies to hybrid resources made up of ESRs and generators, demand-side flexibility, and other resources that may not look like transmission, variable, nor conventional generation for resource adequacy purposes in coming years. Hybrid resources, which may be configured with an interconnection limit that optimizes use of existing transmission capacity with ESR charging from otherwise curtailed variable generation [29], particularly underscore the point that market participants with financial incentive will consider complementarity between transmission, ESRs, and generation in planning.

Future research can consider additional dispatch heuristics for ESRs, better incorporating uncertainty in ESR dispatch decisions and objectives other than capacity value maximization with an EUE-based target metric.

References

[1] J. Blake, M. Gambrel, S. Elstein, and M. Hecht, "Joint Petition of the North American Electric Reliability Corporation and ReliabilityFirst Corporation for Approval of Proposed Regional Reliability Standard BAL-502-RF-03," 2017. https://www.nerc.com/FilingsOrders/us/NERC%20Filings%20to%20FERC%20DL/Final%20BAL-502-RF-03%20Petition%209.7.17.pdf.

[2] The North American Electric Reliability Corporation, "2020 Long-Term Reliability Assessment," December, 2020. https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_LTRA_2020.pdf.

[3] L. L. Garver, "Effective Load Carrying Capability of Generating Units," *IEEE Transactions on Power Systems*, vol. PAS-85, no. 8, pp. 910–919, 1966. doi:https://doi.org/10.1109/TPAS.1966.291652.

[4] MISO, "Planning Year 2020-2021 Wind Capacity Credit," December 2013, 2019. doi:https://cdn.misoenergy.org/2020%20Wind%20&%20Solar%20Capacity%20Credit%20Report408144.pdf.

[5] C. Mbae, D. Ogolla, and J. Mbebe, "PJM Interconnection L.L.C., Docket No. ER21-278-000 Effective Load Carrying Capability Construct," vol. 2, no. 3, pp. 16–34, 2020. https://www.pjm.com/directory/etariff/FercDockets/5832/20201030-er21-278-000.pdf.

[6] N. Schlag, "Capacity and Reliability Planning in the Era of Decarbonization," *E3*, p. 1–23, August 2020. https://www.ethree.com/wp-content/uploads/2020/08/E3-Practical-Application-of-ELCC.pdf.

[7] A. D. Mills and R. H. Wiser, "Changes in the economic value of photovoltaic generation at high penetration levels: A pilot case study of California," *Conference Record of the IEEE Photovoltaic Specialists Conference*, Part 2, 2012. doi:https://doi.org/10.1109/PVSC-Vol2.2013.6656763.

[8] M. Milligan, B. Frew, E. Ibanez, J. Kiviluoma, H. Holttinen, and L. Söder, "Capacity value assessments of wind power," *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 6, no. 1, pp. 1–15, 2017. doi:https://doi.org/10.1002/wene.226.

[9] K. Parks, "Declining Capacity Credit for Energy Storage and Demand Response With Increased Penetration," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4542–4546, 2019. doi:https://doi.org/10.1109/TPWRS.2019.2915805.

[10] V. Sivaram and S. Kann, "Solar power needs a more ambitious cost target," *Nature Energy*, vol. 1, no. 4, pp. 7–9, 2016. doi:https://doi.org/10.1038/nenergy.2016.36.

[11] D. S. Mallapragada, N. A. Sepulveda, and J. D. Jenkins, "Long-run system value of battery energy storage in future grids with increasing wind and solar generation," *Applied Energy*, vol. 275, p. 115390, July 2020. doi:https://doi.org/10.1016/j.apenergy.2020.115390.

[12] P. L. Denholm and R. M. Margolis, "The Potential for Energy Storage to Provide Peaking Capacity in California Under Increased Penetration of Solar Photovoltaics," March, 2018. http://www.osti.gov/servlets/purl/1427348/.

[13] A. Mileva, J. Johnston, J. H. Nelson, and D. M. Kammen, "Power system balancing for deep decarbonization of the electricity sector," *Applied Energy*, vol. 162, pp. 1001–1009, 2016. doi:https://doi.org/10.1016/j.apenergy.2015.10.180.

[14] A. Bloom, J. Novacheck, G. Brinkman, J. McCalley, A. L. Figueroa-Acevedo, A. Jahanbani-Ardakani, H. Nosair, A. Venkatraman, J. Caspary, D. Osborn, and J. Lau, "The Value of Increased HVDC Capacity Between Eastern and Western U.S. Grids: The Interconnections Seam Study: Preprint," October 2020. doi:https://doi.org/10.2172/1696787.

[15] A. Figueroa-Acevedo, A. Jahanbani Ardakani, H. Nosair, A. Venkatraman, J. D. McCalley, A. Bloom, D. Osborn, J. Caspary, J. Okullo, J. Bakke, and H. Scribner, "Design and Valuation of High-Capacity HVDC Macrogrid Transmission for the Continental US," *IEEE Transactions on Power Systems*. 2020. doi:https://doi.org/10.1109/TPWRS.2020.2970865.

[16] L. Reed, M. Granger Morgan, P. Vaishnav, and D. E. Armanios, "Converting existing transmission corridors to HVDC is an overlooked option for increasing transmission capacity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 28, pp. 13879–13884, 2019. doi:https://doi.org/10.1073/pnas.1905656116.

[17] C. Byers and A. Botterud, "Additional capacity value from synergy of variable renewable energy and energy storage," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 1106–1109, 2020. doi:https://doi.org/10.1109/TSTE.2019.2940421.

[18] E. Tomasson and L. Soder, "Multi-area generation adequacy and capacity credit in power system analysis," *2017 IEEE Innovative Smart Grid Technologies - Asia: Smart Grid for Smart Community, ISGT-Asia 2017*, pp. 1–6, 2018. doi:https://doi.org/10.1109/ISGT-Asia.2017.8378349.

[19] S. H. Tindemans, M. Woolf, and G. Strbac, "Capacity value of interconnection between two systems," *arXiv*, 2018. doi:https://doi.org/10.1109/PESGM40551.2019.8973865.

[20] S. Murphy, F. Sowell, and J. Apt, "A time-dependent model of generator failures and recoveries captures correlated events and quantifies temperature dependence," *Applied Energy*, vol. 253, p. 113513, July 2019. doi:https://doi.org/10.1016/j.apenergy.2019.113513.

[21] S. Murphy, L. Lavin, and J. Apt, "Resource adequacy implications of temperature-dependent electric generator availability," *Applied Energy*, 2020. doi:https://doi.org/10.1016/j.apenergy.2019.114424.

[22] A. D. Mills and P. Rodriguez, "A simple and fast algorithm for estimating the capacity credit of solar and storage," *Energy*, vol. 210, p. 118587, 2020. doi:https://doi.org/10.1016/j.energy.2020.118587.

[23] F. D. Munoz and A. D. Mills, "Endogenous Assessment of the Capacity Value of Solar PV in Generation Investment Planning Studies," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1574–1585, 2015. doi:https://doi.org/10.1109/TSTE.2015.2456019.

[24] M. P. Evans, S. H. Tindemans, and D. Angeli, "Minimising unserved energy using heterogeneous storage units," *arXiv*, vol. 34, no. 5, pp. 3647–3656, 2018. doi:10.1109/TPWRS.2019.2910388.

[25] E. Ibanez and M. Milligan, "Comparing Resource Adequacy Metrics," September, 2014. https://www.nrel.gov/docs/fy14osti/61017.pdf.

[26] K. Carden and N. Wintermantel, "The Economic Ramifications of Resource Adequacy White Paper," pp. 1–88, 2013. http://www.naruc.org/grants/Documents/EconomicsofResourceAdequacyWhitePaper_Astrape_Final.pdf.

[27] R. Sioshansi, S. H. Madaeni, and P. Denholm, "A dynamic programming approach to estimate the capacity value of energy storage," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 395–403, 2014. doi:https://doi.org/10.1109/TPWRS.2013.2279839.

[28] M. J. Fisher and J. Apt, "Emissions and Economics of Behind-the-Meter Electricity Storage," *Environmental Science & Technology*, vol. 51, pp. 1094–1101, Feb 2017. doi:https://doi.org/10.1021/acs.est.6b03536.

[29] W. Gorman, A. Mills, M. Bolinger, R. Wiser, N. G. Singhal, E. Ela, and E. O'Shaughnessy, "Motivations and options for deploying hybrid generator-plus-battery projects within the bulk power system," *Electricity Journal*, vol. 33, no. 5, p. 106739, 2020. doi:https://doi.org/10.1016/j.tej.2020.106739.

[30] R. Lueken and J. Apt, "The effects of bulk electricity storage on the PJM market," *Energy Systems*, vol. 5, no. 4, pp. 677–704, 2014. doi:https://doi.org/10.1007/s12667-014-0123-7.

[31] E. S. Hittinger and I. M. Azevedo, "Bulk energy storage increases united states electricity system emissions," *Environmental Science and Technology*, vol. 49, no. 5, pp. 3203–3210, 2015. doi:https://doi.org/10.1021/es505027p.

[32] G. He, Q. Chen, P. Moutis, S. Kar, and J. F. Whitacre, "An intertemporal decision framework for electrochemical energy storage management," *Nature Energy*, vol. 3, no. 5, pp. 404–412, 2018. doi:https://doi.org/10.1038/s41560-018-0129-9.

[33] J. Bakke, M. Boese, A. Figueroa-Acevedo, B. Heath, Y. Li, J. Okullo, A. J. Prabhakar, and C.-h. Tsai, "Renewable Integration Impact Assessment : The MISO Experience," no. Figure 2, pp. 47–50, 2019. https://www.misoenergy.org/stakeholder-engagement/committees/renewable-integration-impact-assessment/.

[34] L. o. L. E. W. Group, "2021/22 PY Planning Reserve Margin and Local Reliability Requirement- Draft Results," 2020. https://cdn.misoenergy.org/20200908LOLEWGItem032021-22PYPRMLRRResults472186.pdf.

[35] A. Mills and R. Wiser, "Strategies for Mitigating the Reduction in Economic Value of Variable Generation with Increasing Penetration Levels," *Lbnl*, vol. 36, no. 2, pp. 119–127, 2014. doi:https://doi.org/10.1016/j.exphem.2007.09.002.

[36] "Redrawing the Network Map: Energy Storage as Virtual Transmission." https://poweringpastcoal.org/insights/energy-security/the-inevitable-decline-of-austra-.

[37] A. Mills, R. Wiser, D. Millstein, J. P. Carvallo, W. Gorman, J. Seel, and S. Jeong, "The impact of wind, solar, and other factors on the decline in wholesale power prices in the United States,"

Applied Energy, p. 116266, November 2020.
doi:https://doi.org/10.1016/j.apenergy.2020.116266.

[38] M. T. Craig, P. Jaramillo, B. M. Hodge, N. J. Williams, and E. Severnini, "A retrospective analysis of the market price response to distributed photovoltaic generation in California," *Energy Policy*, vol. 121, pp. 394–403, July 2018.
doi:https://doi.org/10.1016/j.enpol.2018.05.061.

[39] E. Spyrou, J. L. Ho, B. F. Hobbs, R. M. Johnson, and J. D. McCalley, "What are the Benefits of Co-Optimizing Transmission and Generation Investment? Eastern Interconnection Case Study," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4265–4277, 2017.
doi:https://doi.org/10.1109/TPWRS.2017.2660249.

[40] V. Krishnan, J. Ho, B. F. Hobbs, A. L. Liu, J. D. McCalley, M. Shahidehpour, and Q. P. Zheng, *Co-optimization of electricity transmission and generation resources for planning and policy analysis: review of concepts and modeling approaches*, vol. 7. 2016.
doi:https://doi.org/10.1007/s12667-015-0158-4.

A MISO and RIIA Data Appendix

Zones in strikethrough in Table A-1 are excluded in results because we have load but not renewable profile data for these zones. Their generation resources and transmission capacity are modeled for RA metric calculations.

Table A-1: Zonal load data. Zones in strikethrough are excluded in the 20-zone MISO representation used in cases in the main text results.

Full Zone Name	Abbreviation	Average Hourly Load (MW)
~~Associated Electric Cooperative, Inc.~~	~~AECIZ~~	~~3,328~~
American Transmission Company	ATC	7,437
~~Corn Belt Power Cooperative~~	~~CBPC-NIPCO~~	~~521~~
Consumers Energy	CONS	5,412
DTE Energy (Detroit Edison Company)	DECO	6,085
Entergy Arkansas	EES-ARK	4,043
Entergy Texas	EES-TX	2,520
Iowa East	IA-E	3,150
Illinois Central	IL-C	5,368
Indiana Central	IN-C	4,307
Indiana South	IN-S	730
Louisiana Gulf	LA-GULF	3,897
Louisiana North	LA-N	5,530
MidAmerican Energy Company	MEC	4,768
MISO Missouri	MISO-MO	6,342
MISO Mississippi	MISO-MS	2,511
Minnesota Central	MN-C	1,691
Minnesota Northeast	MN-NE	965
Minnesota Southeast	MN-SE	7,206
Northern Indiana Power Service	NIPS	1,485
Southern Illinois Power Company	SIPC	54
Upper Peninsula Power Company	UPPC	121