Journal of Regulatory Economics (2024) 65:27–63
https://doi.org/10.1007/s11149-023-09467-w

ORIGINAL ARTICLE

A simple way to integrate distributed storage into a wholesale electricity market

Alberto J. Lamadrid[1] · Hao Lu[2] · Timothy D. Mount[3]

Accepted: 23 November 2023 / Published online: 30 December 2023

Abstract

Current plans to decarbonize the electric supply system imply that the generation from wind and solar sources will grow substantially. This growth will increase the uncertainty of system operations due to the inherent variability of these renewable sources, and as a result, more reserve capacity will be required to provide the ramping (flexibility) needed for reliable operations. This paper assumes that all of the increased uncertainty comes from wind farms on the grid, and it shows how distributed storage managed locally by aggregators can provide the ramping needed without introducing a separate market for flexibility. This can be accomplished when the aggregators minimize the expected daily cost of the energy purchased from the grid for their customers by submitting optimal bids into the wholesale market with high and low price thresholds for discharging and charging the storage. This model is illustrated using a stochastic multi-period security constrained optimal power flow together with realistic data for a reduction of the network in the Northeast Power Coordinating

This research was supported in part by the U.S. Department of Energy and the Advanced Research Projects AgencyEnergy ARPA-E PERFORM agreement DE-AR0001277, the Lehigh Faculty Innovation Grant and the National Science Foundation through the EPCN Grant #1809830. The authors are responsible for all conclusions presented, and the views expressed have not been endorsed by the sponsoring agencies.

✉ Alberto J. Lamadrid
ajlamadrid@mit.edu
https://go.lehigh.edu/ajlamadrid

Hao Lu
hl649@cornell.edu

Timothy D. Mount
tdm2@cornell.edu

[1] Department of Economics, Lehigh University, 621 Taylor Street r451, Bethlehem, PA 18015, USA

[2] Independent Electricity System Operator (IESO), 1600-120 Adelaide Street West, Toronto, ON M5H 1T1, Canada

[3] Dyson School of Applied Economics and Management, Warren Hall, Cornell University, Ithaca, NY 14853, USA

Council region of the United States. The results show that the bidding strategy for distributed storage provides ramping to the grid just as effectively as storage managed by a system operator.

Keywords Market design and Intermittent variable renewable energy sources · Markov decision process · Energy storage systems

JEL Classification D40 · L94 · Q48

1 Introduction

The increasing importance of generation from variable renewable energy sources (VRES) on the bulk power system (the grid) and the uncertainty and variability associated with these sources make it more difficult for System Operators (SO) to maintain established standards of operating reliability. More reserve generating capacity is needed to offset the variability of VRES. In addition, VRES in the United States (U.S.) is modifying the typical daily load profile in some regions (EIA, 2022), and wind energy is now the largest source of renewable generation on the grid. The adoption of VRES in the energy sector is part of the transition to a low carbon economy and it is a ubiquitous trend in the U.S. and in many other countries.

An alternative to having more conventional generating capacity committed for reserves is to use Energy Storage Systems (ESS) to respond to the realized levels of VRES. The basic contribution of our paper is to demonstrate that distributed ESS managed locally by aggregators[1] can provide the ramping services needed to maintain operating reliability on the grid without having a separate market for "flexibility." Our empirical application uses thermal storage, that displaces air conditioners on a typical summer day, as the type of distributed ESS analyzed. Government statistics for electricity demand include a component for the provision of *thermal services* (EIA, 2019a), denoted as *deferrable demand* (Jeon et al., 2015). The main advantage of thermal storage is that it has a much lower cost than other distributed ESS technologies, such as batteries (Koohi-Fayegh and Rosen, 2020).

Deferrable demand, by its nature, is located close to load centers, but it can still supply ramping services to the grid. For practical reasons, ESS capacity in the future is likely to include a portfolio of technologies and sizes, including battery storage from electric vehicles as well as thermal storage for winter and summer space conditioning.

This article determines the optimal scheduling of distributed ESS using a stochastic receding (or rolling) horizon (Wu et al., 2012) for two different management situations. The first assumes the ESS is managed centrally by a coordinating SO to minimize the expected operating costs for the grid. The second assumes the ESS is managed locally by aggregators who submit bids into the wholesale market to minimize the expected

[1] We use the term "aggregators" to represent Distributed Energy Resources Aggregators (DERA) who have the responsibility of meeting the energy needs of their customers. In practice, each DERA would participate in the wholesale market like a typical wholesale customer, and manage other local issues such as maintaining a stable power factor. However, we do not discuss the formal structure of the DERA, or how the DERA controls the ESS, in this paper.

cost of the energy purchased from the grid to meet the energy needs of their customers. An essential feature of our analysis is that it uses a stochastic forecast of price for the next 24 h to develop a simple optimal bidding strategy for ESS that determines a balance between shifting load away from the peak and providing ramping to offset the variability of VRES.[2] The strategy discharges ESS when the nodal price is above a high-price threshold, and charges ESS when the nodal price is below a low-price threshold. The difference between the two thresholds is determined by the round-trip efficiency of the ESS (86% in our empirical application).

An important difference between ESS managed centrally by the SO and ESS managed independently is that the State-of-Charge (SOC) of the ESS is not observed by the SO when the ESS is managed independently. Nevertheless, the Federal Energy Regulatory Commission (FERC) has issued Order 841 that requires electricity markets to recognize the physical characteristics of ESS, and to develop procedures for the owners of independent ESS to submit bids that account for these characteristics. This is the major focus in the ongoing development of a new model called the Energy Storage Resource (ESR) by the California Independent System Operator (CAISO) (CAISO, 2022). The ESR allows independently managed ESS to submit bids that are dependent on the SOC into a day-ahead market, and this is an alternative model to the existing Non-Generator Resource (NGR) model that allows bids and offers for charging and discharging energy. The basic rationale for developing the ESR model is that the true marginal costs of charging and discharging ESS are dependent on the SOC (CAISO, 2021). Although an optimum bidding strategy using the ESR model can be derived for a given forecast of prices, the realized bids are dependent on the realized levels of SOC using this strategy. Moreover, offers that are dependent on SOC information add complexities to the market architecture, with increased informational needs in the bidding process and hence in the implementation into real markets. This leads to major analytical and computational complications when evaluating the performance of the ESR bidding strategy (Zheng et al., 2023). However, this is not the place to provide a detailed description or critique of the ESR model. Our objective is to demonstrate that a simple bidding strategy for independently managed ESS, and in our example, distributed ESS managed by aggregators, can provide ramping services to the grid as well as shift load from peak to off-peak periods. It is not necessary to submit information about the SOC into the auction. There are three salient features of our analysis that distinguish it from the ESR model proposed by the CAISO. These are (1) the optimization is conducted each hour to allow for updates in the forecasts of the potential wind generation and the price, and updates in the price/quantity bids submitted by aggregators; (2) the stochastic characteristics of wind generation and price are incorporated explicitly into the optimization; and (3) the optimization is based on the hourly operations for the next 24 h. Even though the optimization solves for a 24-h horizon, updating the forecast of wind every hour means that the forecast for the next hour is always based on the most accurate forecast.[3] Furthermore, allowing

[2] If only a deterministic forecast of price is available, the optimum bidding strategy is to submit a deterministic charging/discharging schedule for ESS that will not provide ramping in response to the realized levels of VRES.

[3] For our estimated time-series model of wind speed, the variance of the forecast for the next hour is roughly one tenth of the variance more than 8 h ahead.

aggregators to modify their bids participating in the real-time balancing makes it feasible for them to monitor the SOC and ensure that the bids for the next hour do not violate the capacity of the ESS, considering the deviation from the day-ahead contract settlements. For example, the quantity discharged in the next hour when the price is above the high-price threshold must be less than both the maximum rate of discharge and the observed SOC in the current hour.

Using a 24-h horizon makes it feasible to determine the optimum balance between the two competing capabilities of ESS, namely, shifting load from peak to off-peak periods versus providing ramping services. The uncertainty of operations in our analysis is represented by different system states (e.g., different levels of potential wind generation) with specified probabilities of occurring. Reserve generating capacity is committed for each hour to ensure that there is sufficient ramping capacity to move from each state in a given hour to all possible states in the next hour. Hence, the largest ramping requirements are for a move from the state with the highest/lowest wind to the next hour's state with the lowest/highest wind. Since spilling wind is allowed, it is often optimum to spill some of the wind in the high-wind states to reduce the cost of ramping.

The rest of this article is organized as follows. In Sect. 2, the related literature on the market design for an SO managing uncertain resources is summarized. Sections 3.1 and 3.2 describe the formal models for ESS managed by an SO, and for ESS managed by aggregators. These two sections are technical and present the theoretical results for each one of the proposed ESS strategies, including a proof for the optimal bidding strategy used by aggregators. Readers who wish to skip these two sections need only know that both ESS strategies use a stochastic multi-period optimization to minimize expected operating costs for the grid. When the SO manages the ESS, the SO can monitor and control the physical capabilities of the ESS to lower costs. When aggregators manage the ESS, the SO faces demand bids that vary with the nodal price and knows nothing about the operating characteristics of the ESS. Section 4 summarizes the numerical results, and shows that the aggregators' bidding strategy for ESS provides ramping services just as well as ESS managed by the SO. In Sect. 5, we offer some concluding recommendations, and in particular, we emphasize how important it is to expose customers to real-time prices in order to get distributed resources to mitigate the uncertainty of VRES effectively.

2 Background and related literature

The value of ESS depends on a complex set of conditions including network location, price spreads, technical characteristics and service applications. In fact, regulation in different jurisdictions affects the integration of these resources. For example, the United States Federal Energy Regulatory Commission (FERC) order 841 and 2222 require Independent System Operators (ISOs) to "establish a participation model for ESS and Distributed Energy Resources (DER) (see e.g., FERC, 2020). Such participation includes the provision of energy and ancillary services alongside conventional resources, clearing rules, tariffs for regional grid operators, and attributes that include, in the case of ESS, a minimum size of 100kW, about twenty times the size of a typical

Tesla Powerwall. For the discussion that follows we consider the terms “distributed storage", “deferrable demand" and “thermal storage" are synonymous in the application.

The majority of system operators in the U.S. use a security constrained unit commitment (SCUC) for operational purposes to jointly determine the dispatches per unit and the aforementioned inflexibilities (see e.g., Zheng et al., 2015, Lamadrid et al., 2019). The recent literature related to the system operator modeling with uncertain VRES can be broadly categorized into three approaches: stochastic programming, probabilistic optimization and robust optimization. Stochastic programming has dimensionality issues (see e.g., Birge and Louveaux, 1997) which are handled using sampling methods to select a subset of possible scenarios by, for example, focusing on the most influential ones (e.g., Morales et al., 2009). Probabilistic optimization allows violation of some specified network constraints within a threshold (e.g., Birge and Louveaux, 1997; Filomena and Lejeune, 2014; Moarefdoost et al., 2016). Robust optimization considers a lower bound on the total social benefits by looking at the worst case realizations before the actual system state is realized (e.g., Bertsimas et al., 2013; Lorca et al., 2016).

The literature studying energy storage from the point of view of individual participants has analyzed threshold policies (see e.g., Secomandi, 2010), with prices for energy generally being positive, agents participating in spot markets, and no transmission constraints. Some works (see e.g., Zhou et al., 2016) solve a discrete-state Bellman dynamic optimization model to determine optimal policies and account for the possibility of negative prices. These models do not consider the participation in scheduling in some existing two-settlements markets (e.g., California Independent System Operator CAISO, the electric reliability council of Texas ERCOT, the New York Independent System Operator NYISO, PJM).

In related models in the literature, (see e.g., Rahimiyan et al., 2014), managers of clusters of demands handle their demand and may have take-or-pay contract with e.g., VRES. Other joint resources setups (see e.g., Zhou et al., 2019), model VRES (e.g., wind farms) with storage systems, proving the optimality of threshold policies and developing heuristics for the participation in the market as a single resource. Here we only assume the aggregator manages the storage capacity. Capacity from VRES is managed via the wholesale market.

The role of storage in supporting the integration of solar power is studied in Schmalensee (2022). The model is a stylized two-period representation of diurnal peak and off-peak (daytime and nighttime) periods, extending work from Boiteux (1952) and Steiner (1957) to account for temporal shifting capabilities from energy storage and abstracting from engineering considerations such as a network and non-convexities from e.g., generators startup. This model assumes inelastic demand, marginal conventional units with non-zero cost in some periods and constant returns to scale among other simplifications. The market structure assumes that scarcity prices are allowed, and therefore the price of electricity may reach the value of lost load (e.g., \$9000/MWh), giving signals and pecuniary incentives to participants able to supply in those periods. This type of ‘energy only’ market could provide revenue streams for suppliers, including storage managers to invest capabilities to supply in the real time process. The aforementioned assumptions allow to have closed form expressions that

may support a merchant model for storage owners. However, there is no proof that the Hessian is positive definite in all cases conditional on the continuous probability density functions for demand shortages and stochastic generation.

To the best of our knowledge, our method is the first to combine an (i) optimal control model to quantify the differences between centrally and locally managed ESS; (ii) considering principal-agent misalignment of incentives when the aggregator manages temperature sensitive demand and; (iii) interacting with a detailed optimization model of electric operations with realistic data calibration of the grid. Our stochastic optimization model corresponds to scheduling resources, considering uncertainty explicitly, by a risk-neutral social planner. Our framework makes it feasible to determine an optimum strategy for managing ESS to shift load from peak to off-peak hours as well as to provide ramping services, as is discussed in the following section.

3 Modeling strategy

Here we briefly describe the models for the system operator and the aggregator, with a focus on the salient characteristics and how do they differ from other models.

3.1 The system operator model

3.1.1 Optimal scheduling for the electricity market

This work draws on the stochastic optimization literature, (Arroyo and Galiana, 2005; Powell, 2007; Morales et al., 2009; Pritchard et al., 2010; Bertsimas et al., 2010), with an emphasis on the determination and appraisal of the costs incurred by participants in the system.

Our approach uses a hybrid method, between a stochastic and a robust optimization (RO) program (Delage and Ye, 2010). We model two distinct types of uncertainty faced by the system operator. First, there are a number of events with relatively low impact compared to high impact-low probability (HILP) resiliency events such as hurricanes or earthquakes. These contingencies may affect operations, and if the system is not secure lead to low reliability of the system.[4] Second, there are limitations for the system operator in assessing the characteristics of the stochastic variables, including the probability distribution and support that determine their uncertainty, and the period-to-period variability. This ramping capability is particularly needed to manage the diurnal variation in net demand (total demand-VRES generation).[5] We discretize

[4] We denote this as *aleatoric uncertainty* System operators are mandated to procure enough capacity to maintain a secure operation and the reliability of the bulk power network against these events. We refer to the capacity required for these intra-temporal events as *contingency reserve*. These reserves need to be delivered to provide re-dispatch capabilities and they could be synchronized, or spinning, or resources that can be brought online rapidly.

[5] We denote this as *epistemic uncertainty* Planning over a finite time horizon requires capacity for the ramping between periods, plus deviations to cover possible realizations in future periods. We refer to this inter-temporal capacity as *load-following reserve*. These lower quality reserves are delivered between periods, e.g., ramping reserves, and may be provided by resources not synchronized.

both of these types of uncertainty, and in the case of the inter-temporal events, we operate robustly within the worst possible events. This approach is consistent with a secure operation of the system, using an operating envelope instead of a collection of trajectories as is done in a traditional stochastic program.

The main differences between our approach and previous work can be summarized in four main points. (i) We co-optimize energy and two kinds of ancillary services (contingency and load following). Our model solves optimal amounts of endogenous reserves as part of the variables in the solution set (Lamadrid and Mount, 2012). We use a novel ambiguity robust model (see e.g., Delage and Ye, 2010), instead of a stochastic program or a robust optimization. (ii) We internalize the uncertainty of renewable energy, by assigning specific costs to the changes of dispatch beyond the elastic range of the generators (ER) (Wang and Shahidehpour, 1995). (iii) We treat demand and supply symmetrically, with a valuation of load not served (LNS) at the value of lost load (VOLL). (iv) We determine the management of Energy Storage Systems (ESS) with a multi-period optimization that values the end-of-horizon states. This formulation includes deviations in a range of states to allow both energy shifting and provision of ancillary services for overall system support. With this framework we can model various types of distributed storage and deferrable demand.

For the sake of generality, we describe this model as a Markov Decision Process (MDP), discretizing and indexing periods by $t = 0, 1, \ldots, n^T$. Each generating unit has a vector of unit commitment state and control inputs with linear costs related to the binary decisions $y \in [0, 1]^{n^Y}$. That is, the set of *feasible commitments* has a corresponding set of linear binary constraints $\mathcal{Y} \subset \mathbb{R}^{n_Y}$ with the associated linear function commitment costs $C_Y(y)$. Hence, the problem is a mixed integer program (MIP).

The *states of nature* $\mathcal{S}^t$ are assumed to be finite, containing all the currently available information (e.g., availability of transmission lines, weather events affecting the VRES production). Let $s^t \in \mathcal{S}^t$ denote the state of nature for each time $t = 0, 1, \ldots, n^T$, with $\mathcal{S}^0$ as a singleton. These states of nature evolve following an exogenous, time varying Markov chain. We assume the corresponding transition probability matrices are independent of all potential control actions.

The *state of system resources* are given by $x^t \in \mathbb{R}^{n_X}$ for each time $t = 0, 1, \ldots, n^T$. In the absence of energy storage, the state reflects the output levels for the previous period considering transition constraints (e.g., ramping limits).

The *evolution of system states* is described by a non-recursive linear equation.

$$x^{t+1} = Bu^t, \quad t = 1, \ldots, n^T, \tag{1}$$

where $B \in \mathbb{R}^{n_X \times n_U}$ corresponds to the incidence matrix from all control inputs to the state outputs (including ramping constraints).

The *system dispatches* are given by $u^t \in \mathbb{R}^{n_U}$ for each time $t = 0, 1, \ldots, n^T$. These include power injections (and withdrawals) at each node in the system, as well as nodal reserves (contingency and load following).

The *dispatch constraints* are a set of linear constraints $\mathcal{U}(x^t, y, s^t)$ for each time period $t = 1, \ldots, n^T$ that limit the system dispatches u^t, i.e., power and reserves. These dispatch constraints reflect limitations for the different resources, and typi-

cally depend on e.g., the system state, the power and energy capacities available, the commitment schedule, the network congestion and the state of nature.

We cast the scheduling problem as an MDP with (n^T+1)-stages and a prior feasible commitment stage $y \in \mathcal{Y}$. The *global system state* for this MDP includes the state of resources, x^t, and the state of nature, s^t.

Definition 1 A *feasible (control) action* is given by any feasible dispatch $u^t \in \mathcal{U}(x^t, y, s^t)$ for each time period $t = 1, \ldots, n^T$.

The *evolution of the state of nature* is independent and the time transitions are given by a $S^t \times S^{t+1}$ transition probability matrix Ψ^t for each time period $t = 1, \ldots, n^T$. Let S^t denote the cardinality of $\mathcal{S}^t$. Each column of the transition probability matrix Ψ^t sums to 1.

There is an operation cost incurred at each stage, according to the power injection and reserve vector u^t.

Definition 2 The cost function $C^t(x^t, u^t)$ for a power injection and reserve vector u^t given a system state x^t and commitment y^t for each time period $t = 1, \ldots, n^T$ is assumed to be convex in u^t (Mas-Colell et al., 1995).

Our implementation includes linear, piecewise linear and quadratic cost functions. For simplicity, we assume no costs are incurred at the initial stage $t = 0$, or at the terminal stage $n^T + 1$.

Definition 3 A policy $\pi = (\mu^1, \ldots, \mu^{n^T})$ is a sequence of decision rules such that $\mu^t(x^t, s^t) \in \mathcal{U}(x^t, y, s^t)$ for all x^t, s^t and t

The *cost-to-go function* for a policy π and a state(x^t, s^t) is given by:

$$V_\pi^t(x^t, y, s^t) = C^t(x^t, \mu^t(x^t, s^t)) + \mathbb{E}\left(\sum_{\tau=t+1}^{n^T} C^\tau(x^\tau, \mu^\tau(x^\tau, s^\tau)) \middle| s \right), \quad (2)$$

where we take the expectation with respect to the sequence of states of nature $\{s^\tau\}_{\tau=t=1}^{n^T}$, conditioned on the current state s^t.

Definition 4 A unit commitment schedule $y^* \in \mathcal{Y}$ and a policy π^* are optimal if

$$C^Y(y^*) + V_{\pi^*}^0(x^0, y^*, s^0) = \inf_{\pi, y \in \mathcal{Y}} \left\{ C^Y(y) + V_\pi^0(x^0, y, s^0) \right\}, \quad (3)$$

for all possible initial states (x^0, s^0).

The aforementioned characteristics are important as the amount of stochastic sources of generation increases (Lamadrid and Mount, 2012). Our model integrates demand-side mechanisms for the estimation of consumer surplus using VOLL (Feng et al., 1998), and reflects the economic costs of different rate structures that consumers may face, e.g., real-time pricing, time of use, while maintaining reliability. This in turn would allow the implementation of effective demand response mechanisms as envisioned by Gellings and Smith (1989) and Schweppe.

3.1.2 Uncertainty, energy storage, and ramping

The amount of conventional capacity needed is typically set at the peak system load plus reserve margins, with adjustments for the expected renewable generation. As storage becomes more widely available, in a perfect foresight situation, the optimal management strategy is to buy and store energy during cheap, off-peak periods, and discharge during expensive, peak periods. When uncertainty is considered, deviations from the expected dispatches above (or below) the expected value can be stored (or covered by discharging) in any given period. In this manner, storage capacity can substitute for conventional generation to provide ramping. The economic tradeoff in this case is between the opportunity cost of the energy stored including the inefficiency and losses of storage versus the cost of deploying additional capacity from conventional resources. Additionally, when storage capacity is available, the peak amount of energy purchased from the grid can be reduced, alleviating congestion according to location. If the overall storage capacity is large enough, it endogenizes the peak of the system. In addition, inter-temporal binding constraints affect both conventional capacity and new market participants such as aggregators managing storage capabilities.

The formulation we use has a hierarchical structure, where states with epistemic uncertainty ("intact system states") have higher precedence than states with aleatoric uncertainty ("contingency states") at each period of time t. Let J^t denote the discrete set of states of nature for the epistemic uncertainty in period t.

Consider any two consecutive periods t and $t+1$. Each one of these time periods has a discrete set of epistemic states $j_1 \in J^t$, $j_2 \in J^{t+1}$. Our model guarantees feasibility inside an endogenously determined operating envelope from the states $j_1 \in J^t$, evolving according ψ^t, to the states $j_2 \in J^{t+1}$. Our design follows the need to establish contracts for inter-temporal claims and determine solutions that are *robust* in a range of potential realizations. Hence, we can obtain locational data over the network space and appropriate claims that manage the period to period variability (i.e., load following reserves).

Let K^{tj} denote the discrete set of states of nature for the aleatoric uncertainty at a given state j and time period t. The occurrence of any aleatoric state $k \in K^{tj}$ is therefore *conditional* on a given epistemic state $j \in J^t$. This taxonomy of uncertainties provides a compromise between assuring the security of the system and providing a tractable way to manage the variability form VRES. Let λ^{tijk*} denote the dual variable associated to the balance between supply and demand (i.e., a power balance condition) at period t in location (i.e., bus) i, in epistemic state j and aleatoric state k. Consider the realization of a particular event $jk \in J^t \times K^{tj}$. Supplying an infinitesimal injection at location i in period t would have an unitary cost

$$\lambda^{tijk*} := \frac{\lambda^{tijk}}{\mathbb{P}\{jk\}}, \tag{4}$$

where this expression can only be defined for states of nature jk with strictly positive probabilities, and $\mathbb{P}\{jk\}$ denotes the probability for these states of nature. The system is guaranteed feasibility over the set of specified contingencies. Therefore, the operating conditions in the sets of epistemic and aleatoric states affect the expected prices

that aggregators use. The coordination mechanism is via prices. At the moment of establishing contracts in day ahead markets, it can be proved (Lamadrid et al., 2015) using the first order conditions (FOCs) of the problem and (4) that

$$\lambda^{ti*} = \sum_{j \in J^t} \sum_{k \in K^{tj}} \lambda^{tijk} \tag{5}$$

$$= \mathbb{E}[\lambda^{tijk*}]. \tag{6}$$

In the empirical application, we assume that the aggregators managing the distributed storage have access to the same stochastic forecasts of VRES and price as the SO for the next 24 h. This is in essence how the Australian electricity market has been operating for the past 25 years. The SO uses the forecast to commit reserve capacity for up and down ramping as well as for dispatch, and the aggregators use the forecast to determine an optimal bidding strategy. Using the same forecast is not, however, a requirement, and there is no logical reason why an aggregator could not use an alternative forecast of the price.

Currently, electricity markets in the US still depend on deterministic forecasts for dispatch in day-ahead markets,[6] and various mechanisms are used by the SOs to procure the reserves needed to deal with the uncertainty of VRES. In our evaluation, we incorporate this uncertainty explicitly into the optimization. We show in sections 3.2 and 4 that an aggregator's optimum bidding strategy can provide both load shifting and ramping capabilities to the grid.

3.1.3 Receding horizon optimization

In this section we briefly present our receding horizon control implementation. Please refer to e.g., Mayne (2014); Rossiter (2017) for background information on Model Predictive Control (MPC) and receding horizon optimization. The generality of the MDP formulation makes the characterization of optimal policies for this decision problem difficult. We restrict our attention to the set of all policies such that the decision rules are independent of the system state. That is, we only consider policies that satisfy.

$$\mu^t(x^t, s^t) = u^t(s^t) \in \mathbb{R}^{n_U}, \tag{7}$$

for each time period $t = 1, \ldots, n^T$, each state of nature $s^t \in \mathcal{S}^t$, and all possible system states x^t. Here, $u^t(s^t)$ represents the system dispatch given a realization of a state of nature $s^t \in \mathcal{S}^t$ regardless of its current state x^t, i.e., regardless of the dispatch in the previous time period. This restriction of the policy space is again a choice for numerical tractability. It renders finite the dimension of the policy search space, $\mathbb{R}^{n_U \times S^1} \times \ldots \times \mathbb{R}^{n_U \times S^{n^T}}$.

We use the **M**ATPOWER **O**ptimal **S**cheduling **T**ool MOST framework (Murillo-Sanchez et al., 2013). We apply the model in Jeon et al. (2019) and refer readers to this description and the stylized setup in "Appendix A". This model is a Mixed Integer

[6] This is mainly due to the limitations of the computing capacity needed for stochastic optimization of the dispatch on large networks.

Quadratic Program (MIQP), due to the quadratic function used for the inter-temporal ramping costs. An MIQP problem is difficult to solve and there is no guarantee about the quality of the solutions. Therefore, we implement a Mixed Integer Linear Problem, limiting the costs to piecewise linear, and the ramping costs using an asymmetric absolute value set of constraints. We use a DC OPF approximation. A central issue for using storage efficiently is to determine the optimum balance between shifting load from high-price periods to low-price periods and providing ramping services to mitigate the period-to-period (e.g., hour) variability of generation from renewable sources. With stochastic inputs, this issue has important implications for how energy constraints in the model are imposed on storage capacity. For each hour, the amounts of energy charged/discharged from storage is typically different in the intact system states representing different levels of the stochastic inputs (e.g., the amount of potential wind generation). The intact states are a discretization of the distribution for epistemic uncertainty.

We organize the information for a receding horizon run using finite time discretized profiles. These time profiles describe the potential realization of a given stochastic variable as a percentage of the maximum potential level of that variable. Consider the case of a system planner doing a receding horizon run for N periods. For each settlement of the market ($n = 1, \ldots N$), the planner requires a look-ahead forecast of n^t periods. Algorithm 1 outlines the overall receding horizon setup.

Algorithm 1 Receding Horizon

1: $t \leftarrow t^0$
2: The social planner chooses a time horizon ($T = n^t$) and a number of receding horizon clearing settlements for the optimal scheduling (N)
3: **repeat**
4: A discrete set of possible realizations for stochastic resources (e.g., forecast ranges for wind and demand/load) are provided for the settlement
5: The system operator solves the problem for the period starting in period t and finishing at time $t + n^t$
6: $t \leftarrow t + 1$; Go to Step 4
7: **until** $t = N$, Number of user-specified information updates reached

We assume that only the first period of the horizon is considered binding and dispatched, and all successive hours can be re-dispatched. This provides the features of a look-ahead optimization, allowing for updated forecast inputs. The theoretical properties and establishment of contracts using this mechanism will be the subject of future research.

3.2 The aggregator model

Our aggregator formulation differs from other models in the literature (see e.g., Secomandi, 2010; Zhou et al., 2016; Zhou et al., 2019) in one main aspect. In our model, the aggregator submits bids *before* the price is revealed. Therefore, it is a dispatchable resource (or scheduled loads in the Australian market, AEMO, terms). It is a dispatchable demand when charging, and a dispatchable injection resource when discharging. The system operator determines the equilibrium price at each node using the received bids and offers from all market participants, including the bids from

the aggregators. Other models in the literature decide the participation of aggregators (charge/discharge) *after* observing the spot price. Such models require using a forecast of price from the ISO. Therefore, storage managed in this manner is not dispatchable to the system operator. In comparison, the ability of the system operator to dispatch the resource as we model it can enhance market efficiency, by improving e.g., reserve procurement compared to *ex-post* models in the literature. We assume the aggregator agent is in the money and not a marginal resource.

Consider an aggregator of customers acting as a fiduciary agent. When distributed storage and/or deferrable demand (DD) are managed by aggregators, we posit that their objective is, first, to devise a strategy to submit bids in a way that minimizes the expected cost of purchasing electricity from the grid using a forecast of real-time prices for a given operational horizon, e.g., the next 24 h; and second, to subject strategies to the constraint that all of the energy services demanded by their customers are met. This implies that the storage technology is non-disruptive in the sense that the comfort levels of individual customers are not affected. The basic business plan is that an aggregator promises to lower customers' bills in return for being allowed to manage their DD capacity. Note that the bids and offers that aggregators submit to the system operator are *indistinguishable* from those of traditional participants (e.g., conventional generators, load serving entities). That is, the system operator does *not need to modify* its current structure to accommodate the management of energy storage resources with inter-temporal constraints. The thermal storage case has additional constraints to a general storage problem. Therefore, the optimal strategies devised for the thermal storage problem are part of the feasible set of solutions to the general storage problem, potentially more conservative solutions.

Our formulation has implementation benefits compared to direct participation of thermal storage, as there are minimum size constraints that would preclude smaller individual agents from accessing these markets. To simplify the exposition, in the discussion that follows we assume that there is only one type of DD, thermal storage for space conditioning, but the bidding strategy would be the same if an aggregator managed a portfolio of different storage technologies.

In our model we assume the aggregator does not exercise monopsony power, and therefore cannot alter the prices observed (see e.g., Borenstein et al., 2008). The extension to account for this behavior and mitigate its consequences is a direction of future research. The following two subsections derive the optimum bidding strategy for an aggregator facing (1) deterministic price forecasts, and (2) stochastic price forecasts. Specifying the method to relieve financial constraints, i.e., whether the aggregator or customers should pay for the installation cost of DD is beyond the scope of this article but it is a possible topic for future research.

3.2.1 The deterministic behavior of an aggregator

Consider an aggregator with deterministic forecasts of prices, whose objective is to minimize the cost of procuring electricity from the grid and submit bids and offers to participate in the scheduling mechanism. This aggregator has agreements with individual participants to manage their thermal load. For simplification, this can be modeled as a storage facility with capacity $S_{\max} \in \mathbb{R}^+$ and roundtrip efficiency $\eta \in$

$(0, 1]$, representing potential DD the aggregator manages. The aggregator also has constraints for the maximum charging and discharging capacities that correspond to the thermal demand of the individual participants. The decision variables are the charging $c^t \in \mathbb{R}^+$ and discharging $d^t \in \mathbb{R}^+$ to be done in the period of analysis $t \in \mathcal{T} = \{1, \dots T\}$, (e.g., hours), where T is the number of periods in the horizon (e.g., 24 h). For any $t \in \mathcal{T}$, denote by P^t the random variable for locational price at hour t, $\mathbb{E}[P^t]$ the expected value of the price at hour t and D^t the deferrable demand available at hour t that the aggregator needs to satisfy either directly from the grid or from previously stored energy. The problem is given by (8).

$$\begin{aligned}\min_{c^t, d^t} \Big\{ & \sum_{t \in \mathcal{T}} \mathbb{E}[P^t](D^t + c^t - d^t) \;\; | \\ & g^{\tau_1}(c^t, d^t) = \sum_{t \le \tau_1} (c^t.\eta - d^t) - (S_{\max} - S_0) \le 0, \quad \forall \tau_1 \in \mathcal{T} \backslash \{T\}; \\ & h^{\tau_2}(c^t, d^t) = \sum_{t \le \tau_2} (d^t - c^t.\eta) - S_0 \le 0, \quad \forall \tau_2 \in \mathcal{T} \backslash \{T\}; k(c^t, d^t) \\ & \qquad = \sum_{t \in \mathcal{T}} (d^t - c^t.\eta) = 0; \\ & \qquad 0 \le c^t \le c; \; 0 \le d^t \le \min(d, D^t), \forall t \Big\}. \end{aligned} \tag{8}$$

The first two constraints, $g^{\tau_1}(\cdot)$, $h^{\tau_2}(\cdot)$, include the energy storage capacity, where S_0 is the initial level of storage, accounting for the storage inefficiency, $\eta \in (0, 1]$. Without loss of generality, we assume that the minimum energy storage level is zero. The third constraint is an energy conservation condition over the optimization horizon. This constraint is an expected balance between total charging and discharging over the optimization horizon T (e.g., 24 hours). The last two constraints establish the non-negativity of charging and discharging for all time periods, taking into account the maximum charging (c) and discharging (d) power capacities. We assume that the aggregator cannot sell the stored energy back to the grid, but can reduce the demand for electricity related to thermal services. Thus the discharging is also limited by the level of deferrable demand D^t. This is a linear problem that can be solved to optimality. The solution is to charge during the periods with the lowest expected prices, and discharge in the highest expected price periods, allowing the aggregator to benefit from energy shifting possibilities over the day. This strategy however precludes the possibility of opportunistically charging when prices unexpectedly drop, or avoid unexpected high prices, situations more likely to occur as the penetrations of renewable energy sources increase. Once the aggregator determines the optimal solution of (8), $\{c^{t*}, d^{t*}\}$, it submits a demand schedule for $t \in \{1, \dots T\}$ to the system operator as illustrated in Appendix B, Fig. 5.

3.2.2 The stochastic behavior of an aggregator

Here we start with a motivating two-period model to build the reader's intuition. We then present a generalized model for a horizon $\mathcal{T} = \{1, \dots, T\}$. Relevant proofs for the

aggregator's stochastic behavior are presented in the appendices, including "Appendix E".

Theorem 3.1 *Consider a simple two-period model for an aggregator whose objective is to minimize the cost of procuring electricity. Assume that the aggregator's expected amount of energy is* $S_f \in \mathbb{R}^+$ *by the end of period two. Suppose that either.*

or

1. $\mathbb{E}[P^1] \geq \mathbb{E}[P^2]$
2. $\text{Prob}(P^2 \geq P^1) > 0$,

1. $\mathbb{E}[P^2] \geq \mathbb{E}[P^1]$
2. $\text{Prob}(P^1 \geq P^2) > 0$.

If all the aggregator knows is the expected price for the two periods, the best possible strategy is to purchase with certainty all the energy required in the period with the lowest expected price. That is, the optimal solution has a deterministic change in state of charge. The minimized expected cost assuming $S_0 = 0$ *is* $z_1 = \min(\mathbb{E}[P^1], \mathbb{E}[P^2]) \times S_f$. *The uncertainty in prices for an aggregator, who follows a policy rule obtained assuming deterministic prices, can lead to situations in which the aggregator incurs a cost higher than the expected cost, or higher than* $\min(P^1, P^2).S_f$ *(i.e., the realized minimum cost over the two periods).*

Theorem 3.2 *Assume the conditions in Theorem 3.1. Let* supp *denote the support of a function, and let* f_{X^i} *denote the probability density function of prices* X^i *in period* i. *We assume the aggregators know* f_{X^i}.

An aggregator in a two-period problem whose objective is to minimize the expected energy procurement cost splits the energy purchases between the two periods. Let L^i *denote the low price threshold used for charging in period* i. *The optimal strategy of the problem is*

1. $L^i = L^*, i \in \{1, 2\}$ when $\text{Prob}(X^i \leq L^i) > 0, i \in \{1, 2\}$
2. $L^{i*} \geq \sup \text{supp}(f_{X^i}) = \inf\{L^i : \text{Prob}(P^i \leq L^i) = 1\}$ *when* $f_{X^{-i}}(L^{-i*}) = 0, i \in \{1, 2\}$,

where L^* *denotes the optimal low price threshold,* sup *and* inf *denote the supremum and infimum operators. The objective function value is* $z_2 = \alpha\mathbb{E}[P^1 | P^1 \leq L^*].S_f + (1-\alpha)\mathbb{E}[P^2 | P^2 \leq L^*].S_f$ *for* $\alpha \in [0, 1]$.

Note that, contrarily to Theorem 3.1, the change in state of charge is not deterministic. The aggregator has a non-zero probability of purchasing energy that misses the S_f *target.*

Corollary 3.1 *If* $\mathbb{E}[P^1] = \mathbb{E}[P^2] = \mu$, *then the optimal solution of the problem is* $L^i = L^*, i \in \{1, 2\}$, *such that* $\text{Prob}(P^1 \leq L^*) + \text{Prob}(P^2 \leq L^*) = 1$.

The same logic used to determine the optimal maximum price for charging the storage, L^*, can be used to determine the optimal minimum price for discharging the storage, H^*.

Theorem 3.3 *Consider a generalization of the aggregator model in Theorems 3.1–3.2 and Corollary 3.1, with an optimization horizon* $\mathcal{T} = \{1, \ldots, T\}$. *In each period the aggregator (1) defers an amount c of the demand it manages (charges an amount c)*

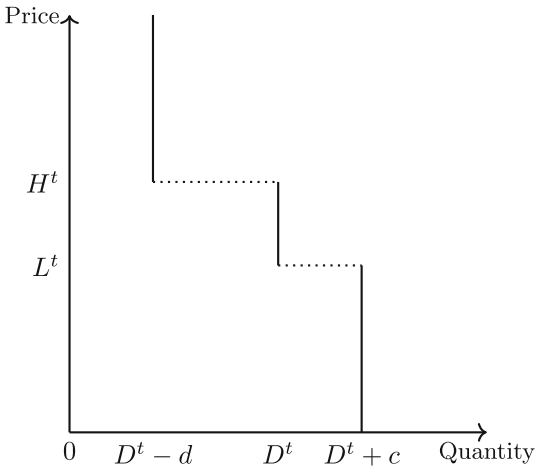


Fig. 1 Bids for stochastic price forecasts

if the price is below or equal to a low threshold L^t; (2) delivers an amount d of the deferrable demand (discharges an amount d) if the price is above a high threshold H^t; (3) neither charges nor discharges if $L^t < P^t \leq H^t$.

Lemma 3.1 *If the storage capacity bounds are never binding from period t to the end of the horizon T, the optimal low threshold prices are the same for those periods. In particular, if the storage capacity bounds are never binding throughout the optimization horizon, the optimal low threshold prices are the same for all periods.*

Corollary 3.2 *If the storage capacity bounds are never binding throughout the optimization horizon, for any hour t, the price arbitrage between the low thresholds and the high threshold must be large enough to compensate for the round-trip inefficiency of storage, $L^t = \eta H^t$.*

Once an aggregator determines the optimal threshold prices $\{L^{t*}, H^{t*}\}$, a set of hourly bids are submitted to the wholesale auction. These bids are price responsive and correspond to $(D^t + c)$ if the energy prices are below the low threshold prices L^{t*}, $(D^t - d)$ when the energy prices are above the high threshold prices H^{t*}, and D^t when the energy prices are between the threshold prices.

The implied form of the demand curve for managing storage is illustrated in Fig. 1. The figure shows an hour where the deferrable demand of that hour is greater than the maximum rate of discharging, $(D^t - d) > 0$. Let S^t be the amount of energy stored at the end of period t. Using the optimal strategy characterized in the theorems above, the actual purchase of energy from the grid is $(D^t + c - d) > 0$. For each hour, $\min\{d, D^t, S^{t-1}\}$ is the upper limit on the amount discharged, and $\min\{c, (S_{\max} - S^{t-1})\}$ is the upper limit on the amount charged. We assume implicitly that the energy capacity of DD storage $S_{\max}$ is, by design, greater than the maximum value of D^t. The values c and d are fixed for computational ease. However note that these parameters are related to technical characteristics of the overall aggregated thermal demand.

"Appendix F" characterizes the calibration process for the probability density function of an aggregator used in the simulations. The charging/discharging strategy in

Theorem 3.3, Lemma 3.1 and Corollary 3.2 decreases the cost for an aggregator, and additionally provides implicit flexibility in the form of ramping services for the system operator. For power systems with high penetration of VRES, this flexibility is particularly valuable. In general, the realized prices for energy will typically be high when the generation from renewables is lower than forecasted and low when the generation from renewables is higher than forecasted. Thus an aggregator submitting a high threshold price for discharging storage and a low threshold price for charging storage will reduce purchases from the grid if the price is high enough, and increase purchases if the price is low enough. In this way, a self-interested aggregator will provide ramping services to the system operator *even though* no instructions to do this are given.

4 Numerical illustration

We explain our model using four cases that compare the costs of serving a given demand profile for a 24-h peak period under different regimes. (i) Case 1: Base case. (ii) Case 2: Case 1 + 16 GW of New Wind Capacity at 16 locations; (iii) Case 3: Case 2 + 17 GWh of DD Storage at 5 load centers managed by the system operator. (iv) Case 4: Case 3 with DD Storage managed by Aggregators (deterministic and stochastic price forecasts). The wind capacity in Cases 2–4 represents approximately 14% of the peak system load. Case 1 is treated as a benchmark for a system with no uncertainty from Potential Wind Generation (PWG) other than the uncertainty of the standard contingencies, i.e., $n - 1$ reliability.

For the two cases with storage, the results in Case 3 using the centralized management of storage by a system operator, as in U.S. FERC order 841, are compared with the results in Case 4 using distributed management by aggregators, as in U.S. FERC order 2222, who submit bids into the wholesale market. Most demand is covered by purchasing electric energy from the grid, but the deferrable demand for space cooling can be met by either purchasing electric energy and/or discharging thermal storage. In other words, the delivery of some cooling services can be decoupled from the purchase of the electric energy needed. Thermal energy can be stored by, e.g., producing ice at night when wind generation is high and electricity is inexpensive, and cooling services can be delivered when needed by e.g., melting the ice in the afternoon when electricity is more expensive. This non-disruptive delivery of cooling services substitutes discharging thermal storage for air conditioning, and it reduces the peak load and the amount of conventional generating capacity needed for system adequacy. An analogous logic applies to heating services. The details of the underlying network and data used are available in "Appendix G".

4.1 Optimizing for a fixed horizon

The first analysis considers a fixed 24-h horizon that is consistent with the day-ahead markets currently implemented in several systems (e.g., NYISO). Using MOST, we simulate a system with more stringent requirements for optimality. We perform an

(a) No Deferrable Demand (b) Deferrable Demand Centrally Managed

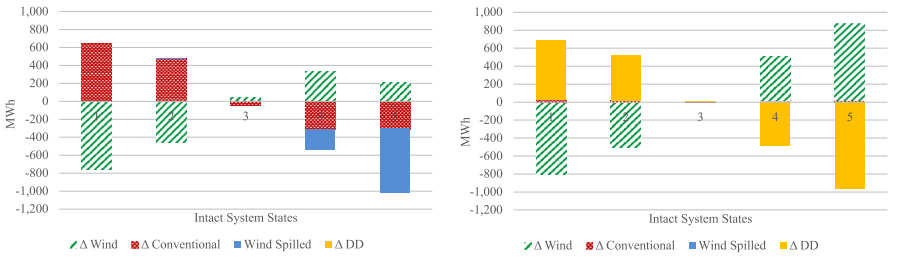
Fig. 2 Mitigating the variability of wind generation with and without storage

analysis of the dispatches in the different states of the system for the fixed horizon optimization to illustrate the tradeoffs incurred under alternative management regimes. We model both epistemic and aleatoric uncertainty. For each hour of the day there are 15 possible system states. The probabilities of each state occurring are known and remain the same for each hour, by construction. Five are epistemic intact system states representing a discretization of systemic uncertainty, and determining the period to period variability (see section 3.1). Each one of these five intact system states has a hierarchical set of aleatoric contingency states, in this application two per intact state. These intact states correspond to a different potential level for the variable renewable energy resources and the demand. Figure 2 shows the changes in dispatches for the five system states considered for two selected cases at the peak hour of the system (hour 13). Each intact probability state is associated with a different PWG realization, and is numbered from 1 to 5. These five states are ranked from the lowest PWG (State 1) to the highest PWG (State 5). The variables with a Δ correspond to the positive or negative deviations from the weighted average dispatch. Therefore our results illustrate the combined effect of epistemic and aleatoric uncertainty. The height of the bar indicates the positive or negative deviation from the weighted average dispatch over all of these states.

For each hour, the level of demand by customers is the same for all five states and it can be met by conventional generation, wind generation and, in Case 3, by discharging storage (Deferrable Demand (DD) in Fig. 2b). It is also possible to shed load but this does not happen in Fig. 2.

In Case 2, Fig. 2a, there is no storage and hence conventional generation has to adjust to accommodate the different levels of PWG. The cost of the reserves needed for ramping means that it is sometimes optimal to spill PWG generation. In fact, most of the additional PWG in the high PWG states is spilled, and the levels of conventional generation in States 2–5 are quite similar because 453 MW and 935 MW of PWG are spilled in States 4 and 5, respectively. The maximum conventional generation in State 1 is 56.9 GW, and the range of dispatch across the five states is 728 MW, which is less than the amount of PWG spilled in State 5. Total energy supplied is very similar in all cases, and there is no load shedding at the optimal solution.

Figure 2b shows the different levels of dispatch in Case 3 when storage is managed centrally by the system operator. The storage capacity accommodates the full range of PWG, discharging in States 1 and 2 and charging in States 3 and 4. None of the



(a) Deferrable Demand with Deterministic Price Forecasts

(b) Deferrable Demand with Stochastic Price Forecasts

Fig. 3 Hierarchical storage management by aggregators

PWG is spilled and conventional generation is roughly the same in all five states (a range of only 33 MW). Spilling less PWG in Case 3 implies that fuel costs are lower in the peak hour than they are for Case 2. More importantly, the maximum dispatch of conventional generation is now 54.9 GW, which is 2 GW lower than the maximum in Case 2. This implies a reduction in the amount of installed capacity needed for adequacy and a corresponding reduction in capital costs. Since the full range of PWG is used in Case 3, the range of energy purchased from the grid is now 1652 MWh, compared to only 141 MWh in Case 2. This highlights the flexibility provided to the system by distributed storage.

Figure 3 shows the different levels of dispatch in Case 4 when the storage is managed by aggregators who submit bids into the market using either deterministic or stochastic forecasts of the price of electricity. Figure 3a shows the deviations from the weighted average dispatch when the aggregators use deterministic price forecasts and submit a deterministic charge/discharge schedule. Although discharging storage reduces the maximum conventional generation to 55.4 GW, 1.6 GW lower than the maximum in Case 2 when there is no storage, conventional generation is the only way to accommodate the different levels of PWG. The results in Figs. 3a and 2a (Case 2) are similar. In Fig. 3a, the range of conventional generation is 951 MW, compared to 728MW in Case 2. Moreover, less of the PWG is spilled (231MW and 712MW in States 4 and 5, respectively) than the amounts in Case 2. Overall, if aggregators use deterministic price forecasts, the potential benefits to the system that could be provided by the flexibility of storage are wasted. Generally, storage managed by an aggregator using deterministic bids serves as a bridge between a situation in which no storage is available, and the case in which storage is optimally managed by the system operator. Figure 3b shows the deviations from the weighted average dispatch when the aggregators use stochastic price forecasts and submit bids with price thresholds for charging and discharging storage. The results in this instance are almost identical to Case 3 when the storage is managed by the SO. None of the PWG is spilled and the range of conventional generation is still relatively small (82 MW). The maximum conventional generation is 55.1 GW, only 0.2 GW higher than Case 3. The fact that the storage managed by aggregators does provide flexibility to the grid implies that the shadow prices are sufficiently negatively correlated with the level of PWG to provide the signals needed to trigger charging and discharging effectively.

Even though the results in Figs. 2b and 3b are similar in terms of the composition of dispatch in the different system states, they are derived using two fundamentally different approaches. In Fig. 2b, the SO has visibility of the full system and perfect knowledge about everything except the actual level of PWG. Consequently, the SO can ensure that the expected dispatch does not violate the physical limits of storage over the 24-h horizon. In contrast, an aggregator only has access to a price forecast, and with a fixed horizon, only one set of the bids is submitted for 24 h ahead. Since the SO does not know the physical limits of storage when it is managed by aggregators, it is likely that the SO's optimal dispatch plan will violate these limits. Moreover, we assume that the aggregator has knowledge of the probability distributions for the prices, f_{X^i}. Mismatches the information available to the SO and the aggregator can lead to divergent situations between the central management and the distributed management. This problem is exacerbated by the characteristics of VRES, and in particular PWG. Even with knowledge of the underlying price distributions, the high positive autocorrelation of the residuals implies that forecasting errors are persistent, and as a result, on unexpectedly high/low wind days, the SO will tend to charge/discharge storage more than the physical limits allow. In this situation, an aggregator would have to ignore a price signal to, for example, charge the energy storage if the capacity is already full.

The situation in Case 3 corresponds to having thermal storage that is fully dispatchable by the SO, and the ramping provided by storage is treated the same way as it is for conventional generation. By contrast, when aggregators submit bids with price thresholds in Case 4, the storage represents price-responsive load in a two-sided market with demand that is no longer perfectly inelastic. The fact that the performance of the two cases in Figs. 2b and 3b are so similar is an encouraging sign. As the number of Distributed Energy Resources (DER) increase and become more varied, it is simply unrealistic to expect an SO to manage this complexity effectively. Establishing the interface between demand and supply at substations with DER managed locally by aggregators[7] and the bulk power grid managed by the SO, is a more promising structure for the wholesale market, as in FERC order 2222.

Nevertheless, the problem of aggregators being unable to respond to price signals when storage limits are binding still remains. One way to address this issue is to use a receding horizon optimization and allow aggregators to adjust their hourly bids. This assumes that the aggregators will also have access to updated price forecasts for the next periods (e.g., 24 h) so that they can determine an optimum plan for their storage. There is, however, a more important justification for using a receding horizon optimization. The forecasts of PWG can also be updated, and given the statistical properties of these forecasts, the range of PWG levels in the intact system states will be smaller for 1-h ahead than it is, for example, in Figs. 2 and 3 when the peak hour is 13 h ahead. Consequently, the ramping requirements will be substantially lower, and having more accurate forecasts of PWG is always an effective way to reduce operating costs.

In the next subsection, we depart from our current theme of operating within current market designs and illustrate how a receding horizon optimization can improve

[7] Or managed by third-party aggregators that coordinate with a Distribution System Operator (DSO).

operations on the grid and lower costs. This is done by comparing the daily system costs for the four different cases using both a fixed horizon and a receding horizon optimization. In Case 4, the aggregators have access to stochastic forecasts of the price.

4.2 Optimizing for a 1-h ahead receding horizon

Table 1 presents a selected set of physical metrics at the optimum levels of operation for the four cases using a fixed 24-h horizon, and the corresponding four cases using a receding-horizon. The receding horizon results are based on 24 separate optimizations using updated 24-h ahead forecasts of PWG and price for each hour. We assume that the next day is exactly the same as the current day to make the results for the fixed and receding horizons easier to compare, and for the same reason, the realized PWG is assumed to be the same as the forecasted PWG.

The first four columns in Table 1 show the results for the fixed-horizon optimization. The availability of PWG in Case 2 replaces over 15% of the conventional generation in Case 1. However, the uncertainty of PWG also requires increases in the amounts of reserves for up and down ramping, δ_{+}^{ti}, δ_{-}^{ti}, and contingency reserves, r_{+}^{ti}, r_{-}^{ti}, by just over 50%, from 60 GW/day in Case 1 to 91 GW/day in Case 2. These reported amounts are the sums of the 24 hly commitments of reserves. Even though ramping increases in Case 2, the sum of the maximum commitment for each generating unit over all system states and all hours (The last row in Table 1), decreases by more than 5 GW, roughly one third of installed wind capacity. Consequently, the capital costs of the installed conventional capacity needed for adequacy is also reduced.

Adding storage in Case 3 improves operations on the grid compared to Case 2 by (1) spilling less PWG so that an additional 2.3 GWh/day of wind energy is dispatched; (2) providing ramping services that displace 33 GW/day of reserves capacity; and (3) reducing the maximum commitment of conventional generation by an additional 2 GW due to shifting some load to off-peak hours. The results for Case 4, when the storage is managed by aggregators, are very similar to Case 3 when the SO manages everything.

The last four columns of Table 1 show the results using a receding horizon, and these results are compared with the corresponding cases using a fixed horizon. The differences for Case 1 are trivial because the initial amounts of wind capacity are negligible and this is the main source of uncertainty. For Case 2, the improved forecasts of PWG using a receding horizon lead to a 20% reduction in ramping requirements.

However, comparing Case 3 with the corresponding case using a fixed horizon, the increase in wind generation is small, the reductions in ramping are modest, and the reduction in conventional capacity is small.

Doing the same comparison with Case 4, the results are mixed. Wind generation is slightly higher, but the total ramping and the conventional capacity are also slightly higher. However, using a receding horizon avoids violations of the energy storage limits. The overall conclusion is that the improvements associated with using a receding horizon optimization are most apparent, particularly for ramping, in Case 2 when there

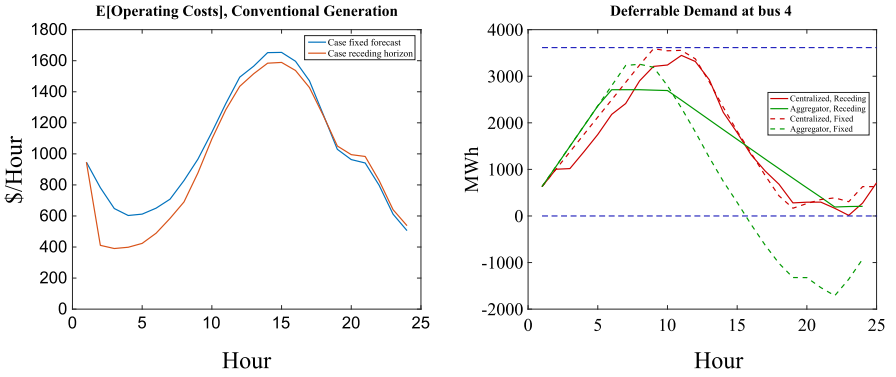
Table 1 Summary of the optimum operating levels for the fixed and receding horizons

		Fixed horizon				Receding horizon			
		Case 1	Case 2	Case 3	Case 4 [a]	Case 1	Case 2	Case 3	Case 4
1. E[Wind Gen.]	(MWh)	689	206,117	208,433	209,362	719	211,769	214,492	214,449
2. E[Conventional Gen.]	(MWh)	1,268,793	1,063,375	1,063,570	1,058,375	1,268,764	1,057,719	1,059,379	1,058,302
3. LF Up Reserve δ_{+}^{ti}	(MW)	22,030	35,049	25,084	25,756	22,060	28,363	23,234	25,438
4. LF Down Reserve δ_{-}^{ti}	(MW)	20,360	31,072	23,324	20,954	20,390	25,674	21,378	21,751
5. Contingent Res. r_{+}^{ti}, r_{-}^{ti}	(MW)	18,087	25,038	10,309	10,785	18,777	18,433	7,113	10,878
Conv. Gen., Max Intact States	(MW)	62,100	56,985	54,962	55,052	62,100	56,502	54,486	55,566
Conv. Gen., Max	(MW)	63,078	57,857	55,842	55,913	63,078	57,535	55,511	56,614

[a] Dispatch of thermal storage does not comply with physical limits

Table 2 A summary of the operating costs for the fixed and receding horizons

	Fixed horizon				Receding horizon			
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
Composition of wholesale costs ($1000/day)								
E[generation cost]	30,947	22,871	19,565	19,575	31,992	21,845	19,324	19,559
E[ramp wear cost]	2	198	28	29	5	29	5	5
LF ramp-up reserve cost	234	1,161	363	376	272	690	294	327
LF ramp-down reserve cost	204	387	239	219	244	304	167	186
Contingency reserve cost	88	122	50	53	94	92	36	54
E[total operating cost]	31,475	24,739	20,245	20,252	32,607	22,960	19,826	20,131



(a) Expected Operating Costs for Case 2 (b) Expected Energy Management, bus 4

Fig. 4 Fixed versus receding horizon examples

is no storage. This conclusion is confirmed by inspecting the cost results summarized in Table 2.

Table 2 presents results for the major components of the expected daily operating costs using the same layout as Table 1. The differences in total costs among the different cases are more pronounced than the differences between the fixed and receding horizon. Installing more wind capacity in Case 2 shows the largest reductions in costs of 22% and 29% for the fixed and receding horizon, respectively. The reductions in fuel costs caused by displacing fossil generation are much larger than the increases in ramping costs compared to Case 1. Additional reductions of 14% and 10% are obtained by adding storage in Case 3 for the fixed and receding horizons, respectively (the corresponding reductions for Case 4 are 14% and 9%). Although these reductions in cost from adding storage are modest, the main economic benefits from storage come from reducing the peak load and the amount of conventional capacity needed for adequacy, but the associated reductions in capital costs are not part of the operating costs in Table 2.

Figure 4a shows the expected hourly operating costs using a fixed (blue line) versus a receding (orange line) horizon for Case 2 when there is no storage. The lower hourly costs using the receding horizon are mainly caused by dispatching more wind generation in the early morning hours when PWG is abundant. This additional wind generation displaces conventional generation because there is no storage in Case 2. This has implications for the aggregator's economic viability: an aggregator participating in an energy-only market is able to perform close to the benchmark of a Ramsey central planner.

Figure 4b shows the expected hourly levels of stored energy at a specific bus close to Boston (Bus 4) for Cases 3 and 4 using a fixed and receding horizon. The levels for Case 3, when storage is managed by an SO, are similar for both the fixed and receding horizon. Storage is almost fully charged at night and then almost fully discharged during the day. The results when the storage is managed by aggregators in Case 4 are different. With a fixed horizon, storage is fully charged at night but the amount discharged during the day is much larger than the storage capacity. In other words, it is not a feasible plan and no discharging would be possible after Hour 15. However, with a receding horizon, the stored energy is always within the capacity limits of storage. In this example, however, the maximum level of stored energy is well below the capacity limit and this probably contributes to making the operating costs in Table 2 higher for Case 4 than they are for Case 3 when storage is managed by an SO.

Hence, in situations as illustrated in Fig. 4b in which the threshold prices for discharging are not high enough to prevent excessive discharging, the system operator requests the aggregator to perform dispatches that are not possible physically (out of the feasible solution region for the aggregator). The receding horizon optimization updates the bids for each period according to the most updated forecasts. Therefore, the strategy precludes exceeding the capacity limits on the deferrable demand.

5 Conclusion

This article proposes a two-sided market in which aggregators manage ESS, in the form of deferrable demand for space cooling, thermal storage, and submit bids into the wholesale market for purchasing energy from the grid for a given planning horizon (e.g., 24 h). This structure establishes the market interface at the substation level, aggregating potentially millions of individual loads. The bids by aggregators would have similar characteristics to the bids by wholesale customers, and in this respect, the market structure is simple. Even though the aggregators only participate in the energy market, they still provide ramping capabilities to the system operator. The overall result is that the variability of VRES can be accommodated when the aggregators face real-time prices, [8] paving the way, for example, for the implementation of order 2222 from FERC. Negative prices may arise when there are "excess" VRES, and this provides a clear incentive for aggregators to charge their storage. Similarly, high prices when the wind generation is less than expected trigger discharging and a reduction in purchases

[8] To avoid paying excessively high prices, aggregators should be hedged by, for example, holding collar derivatives that specify a "floor" and a "ceiling" for prices determined in a separate market or by bilateral negotiations.

from the grid. Ideally, distributed storage could smooth out the realized generation from conventional units and increase their average capacity factors. This would be a valuable benefit for conventional generators because their earnings in the wholesale market tend to fall when there are high penetrations of renewable generation.

An empirical application demonstrates how distributed storage managed by an SO or by aggregators can reduce the conventional capacity needed for ramping, and also increase the amount of VRES dispatched.

Currently, there are two competing proposals for managing Distributed Energy Resources (DER). The first is to extend the logic of nodal pricing from the high-voltage grid to distribution systems, so that all customers can participate in the wholesale market. The second, which we favor, is to have DER managed locally by aggregators who submit bids into the wholesale market and work on behalf of their customers to reduce the expected cost of their energy purchases from the grid.

There is an important qualification that underlies our conclusions. Our empirical application assumes implicitly that the optimum results in all system states are in equilibrium with a unit power factor. However, with rooftop solar, for example, local voltage problems will occur whenever clouds pass overhead. In our two-sided market, these voltage problems should be managed locally through the installation of equipment such as smart inverters that can respond rapidly to voltage problems. A simple market mechanism that provides the incentives for maintaining a stable power factor already exists for wholesale customers. These customers pay a penalty whenever their power factor falls outside a specified range. In contrast, in a highly disaggregated market, local voltage problems on the distribution systems will be the responsibility of the DSO.

In summary, our results show how the local control of DERs by aggregators, as opposed to centralized control by an SO, can manage the increasing complexity of DERs effectively. From the perspective of an SO, it would be preferable to have a few wholesale customers with stable power factors in our two-sided markets than to manage all of the local voltage problems caused by thousands of retail customers. This does, however, leave open the questions of how should aggregators bill their customers and how should they control the ESS. These are promising topics for future research. Although state regulators would lose their jurisdiction over the new wholesale customers, they could still treat the customers of an aggregator as retail customers and regulate the rates that they pay. The combination of aggregators submitting bids into the energy market and having equipment to manage local voltage problems automatically is practical, and possibly cost effective, and it is consistent with the concept of “grid edge intelligence."

Acknowledgements We thank Wooyoung Jeon and Jung Youn Mo for their inputs to this work. We also thank Daniel Munoz-Alvarez, Ray D. Zimmerman, Carlos E. Murillo-Sanchez, Robert J. Thomas, Michael Crew and other participants at the Eastern and Western Rutgers Conferences organized by the Center for Research in Regulated Industries at the Rutgers Business School-Newark and New Brunswick for their comments and input.

Author Contributions AL, HL and TM wrote the main manuscript text. All authors reviewed the manuscript.

Funding 'Open Access funding provided by the MIT Libraries'

Declarations

Conflict of interest The authors declare no competing interests.

Appendix A: Interaction between aggregators and system operator

The full formulation for the system operator model and the economic analysis and counterfactual studies are presented in Murillo-Sanchez et al. (2013), Lamadrid et al. (2019) and Jeon et al. (2019). Table 3 presents the nomenclature for a stylized presentation of the system operator model.

Table 3 Nomenclature associated to stylized system operator model

Functions and variables	
$V_u(\cdot)$	Upper level value function for aggregator and supply-side participants
$V_\delta(\cdot)$	Upper level cost function for quadratic ramping adjustments, supply-side participants
$V_{\text{oth}}(\cdot)$	Upper level value function for other variables including demand-side participants
$f_{\text{uc}}(\cdot)$	Upper level cost function for binary variables, supply-side participants
$V_l(\cdot)$	Lower level value function for aggregator participants using proxies to system variables (e.g., nodal prices)
$g_{\text{bal}}(\cdot)$	Function for balancing supply and demand
$g_\Delta(\cdot)$	Functions for balancing economic, financial and physical variables
a	Vector of energy decisions for energy participation, aggregator
x	Vector of state system resources
u	Vector of system dispatches, supply-side participants
s	Vector of states of nature
x_{oth}	Vector for all other variables for both supply and demand side participants (e.g., reserves, curtailable demands, minimum operating times)
y	Vector of binary decisions (e.g., startup and shutdown states)
$\langle\cdot\rangle'$	Vectors of proxies used by aggregators to determine their participation strategy

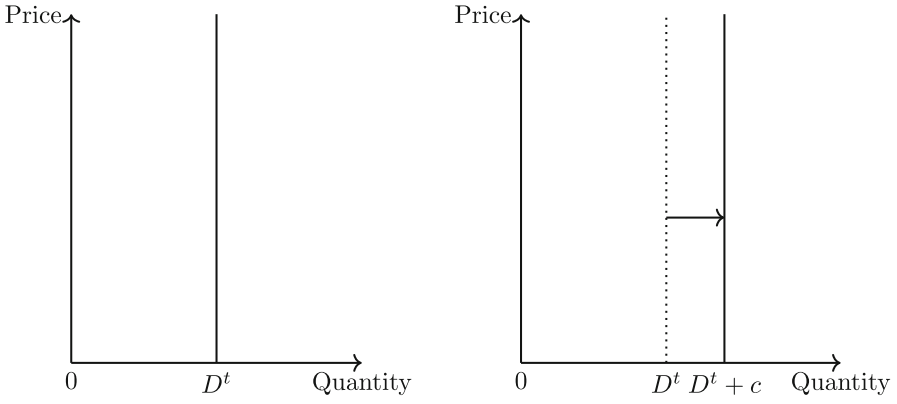
Consider a stylized formulation of the system operator problem accounting for the bi-level nature of the aggregator decisions, a. Namely,

$$\begin{aligned}
\min \quad & V_u(a, x, u, s) + V_\delta(u) + V_{\text{oth}}(x_{\text{oth}}, s) + f_{\text{uc}}(y, s) && \text{(MOST)}\\
\text{s.t.} \quad & a \in \arg\min_{a'} V_l(a', x', u', s', x'_{\text{oth}}, y', \lambda'), && \text{(9a)}\\
& g_{\text{bal}}(a, x, u, s, x_{\text{oth}}) = 0, \quad (\lambda) && \text{(9b)}\\
& g_\Delta(a, x, u, s, x_{\text{oth}}, y) \leq 0, && \text{(9c)}\\
& y \in \{0, 1\}, && \text{(9d)}
\end{aligned}$$

where the objective function (MOST) is a Mixed Integer Linear Program (MILP) that maximizes social welfare as outlined in section 3.1. The aggregator solves a problem (9a) as outlined in section 3.2.2 and section E, using priors to determine, e.g., the probability density functions of energy prices, λ', at their participating nodes. The way these priors are obtained can vary from market to market (e.g., the system operator can provide non-binding forecasts). The system operator balances supply and demand as per (9b), and the dual variable is presented in parenthesis, after the considerations in (O'Neill et al., 2005; Kuang et al., 2019).. This market is cleared using a supply function equilibria with a uniform price at each node (locational marginal prices). All other inequality constraints, including both economic, financial and physical limits, for the system operator are represented by (9c) (e.g., energy and reserve contracts, ramping limits, reserve amounts, power).

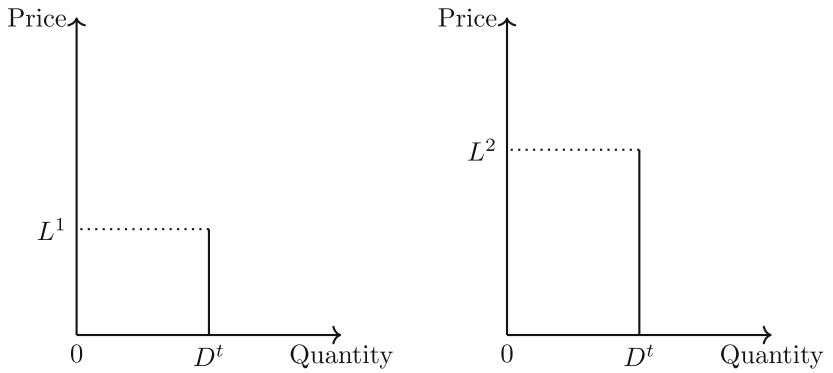
Appendix B: Proof of Theorem 3.1

Proof Due to the cost minimizing behavior we impose, the aggregator in a two-period problem buys the energy needed in the period with the lowest expected energy price, hence minimizing the expected procurement cost.



(a) Hour t without charging/discharging (b) Hour t charging c

Fig. 5 Demand curves for deterministic price behavior



(a) Period 1 (b) Period 2

Fig. 6 Demand curves with threshold prices L^1 and L^2

These demand curves are completely inelastic, as shown if Fig. 5. For any $t \in \{1, \ldots T\}$, Fig. 5a shows the case for a moderate price, with $c^{t*} = d^{t*} = 0$, and Fig. 5b illustrates a case for a low price, with $c^{t*} = c, d^{t*} = 0$.

Once prices are stochastic, there is a probability $p > 0$ that the aggregator pays a price above the expected price in a given period, even when the expected price for that period is the lowest of the two periods. Let P^1 and P^2 be the stochastic prices faced by an aggregator in periods one and two respectively. Let $f_{X^i}, i \in \{1, 2\}$ denote the known the probability density functions for the prices in the two periods. A possible strategy is to specify low threshold prices for each time period, $L^i, i \in \{1, 2\}$. The aggregator purchases an amount S_f if the realized price is below the threshold as illustrated in Fig. 6.

In general we can assume that the objective for an aggregator with deterministic price forecasts is $J_d = \min \mathbb{E}[P^t](D^t + c^t - d^t) = \min \mathbb{E}[P^t * (D^t + c^t - d^t)]$ whereas the objective function with the stochastic price forecasts satisfies $J_s = \min \mathbb{E}[P^t * (D^t(P^t) + c^t(P^t) - d^t(P^t))] \leq J_d$. J_s is a relaxation of J_d, and therefore the feasible set for J_s is at least as large as that of J_d. □

Appendix C: Proof of Theorem 3.2

Proof The problem for an expected cost minimizing aggregator is given by (10).

$$\min_{L^1, L^2} \Big\{ \mathbb{E}[P^1 | P^1 \leq L^1].S_f.\text{Prob}(P^1 \leq L^1) + \mathbb{E}[P^2 | P^2 \leq L^2].S_f.\text{Prob}(P^2 \leq L^2) | \\ \text{Prob}(P^1 \leq L^1) + \text{Prob}(P^2 \leq L^2) = 1 \Big\}, \tag{10}$$

where the constraint in (10) means that the aggregator is expected to buy a quantity S_f over the two periods.

1. For $f_{X^i}(L^{i*}) > 0, i \in \{1,2\}$, then $\text{Prob}(P^i \leq L^{i*}) > 0, i \in \{1,2\}$, and the aggregator can make purchases in both periods. Because $\mathbb{E}[P^i | P^i \leq L^i] = \int_{-\infty}^{L^i} P^i f_{X^i}(P^i) dP^i / \text{Prob}(P^i \leq L^i)$, and $S_f > 0$, the objective function is equivalent to (11).

$$\min_{L^1, L^2} \left\{ \int_{-\infty}^{L^1} P^1 f_{X^1}(P^1) dP^1 + \int_{-\infty}^{L^2} P^2 f_{X^2}(P^2) dP^2 | \int_{-\infty}^{L^1} f_{X^1}(P^1) dP^1 + \int_{-\infty}^{L^2} f_{X^2}(P^2) dP^2 = 1 \right\}. \tag{11}$$

We solve this problem using Lagrangean relaxation, leading to (12).

$$\mathcal{L}(L^1, L^2, \lambda) = \int_{-\infty}^{L^1} P^1 f_{X^1}(P^1) dP^1 + \int_{-\infty}^{L^2} P^2 f_{X^2}(P^2) dP^2 + \lambda \left(1 - \int_{-\infty}^{L^1} f_{X^1}(P^1) dP^1 - \int_{-\infty}^{L^2} f_{X^2}(P^2) dP^2 \right). \tag{12}$$

The first order conditions (FOCs) for the problem are given by (13).

$$\begin{aligned} (L^i) \frac{\partial \mathcal{L}}{\partial L^i} \equiv \quad & L^i f_{X^i}(L^i) - \lambda f_{X^i}(L^i) && \leq 0, \quad i \in \{1,2\} \\ (\lambda) \frac{\partial \mathcal{L}}{\partial \lambda} \equiv \quad & 1 - \int_{-\infty}^{L^1} f_{X^1}(P^1) dP^1 - \int_{-\infty}^{L^2} f_{X^2}(P^2) dP^2 && = 0. \end{aligned} \tag{13}$$

2. For $f_{X^1}(L^{1*}) = 0$, L^{1*} is out of the range of P^1. Formally, $L^{1*} \leq \inf \text{supp}(f_{X^1}) = \sup\{L^1 : \text{Prob}(P^1 \leq L^1) = 0\}$; the (λ) FOC implies $\int_{-\infty}^{L_{2*}} f_{X^2}(P^2) dP^2 = 1$ or $L^{2*} \geq \sup \text{supp}(f_{X^2}) = \inf\{L^2 : \text{Prob}(P^2 \leq L^2) = 1\}$. In such case, the objective value is $\mathbb{E}[P^2 | P^2 \leq L^{2*}].S_f = \mathbb{E}[P^2]$. Therefore it is optimal for the aggregator to purchase all the energy in period 2, given the high prices expected for period 1. The analysis for the case where $f_{X^2}(L^{2*}) = 0$ is analogous and in that case, the objective value is $\mathbb{E}[P^1 | P^1 \leq L^{1*}].S_f = \mathbb{E}[P^1]$
3. If $\text{Prob}(X^i \leq L^i) > 0, i \in \{1,2\}$, the aggregator makes purchases in both periods. For $0 < f_{X^i}(L^i) < 1, i \in \{1,2\}$, the aggregator purchases energy in period i, implying $L^i = \lambda$. Hence, setting $L^i = L^*, i \in \{1,2\}$ so that $\text{Prob}(P^1 \leq L^*) + \text{Prob}(P^2 \leq L^*) = 1$ is the optimal strategy. In such case, the objective value is $\alpha \mathbb{E}[P^1 | P^1 \leq L^1].S_f. + (1-\alpha)\mathbb{E}[P^2 | P^2 \leq L^2].S_f$ for some $\alpha \in (0,1)$ □

Appendix D: Proof of Corollary 3.1

Proof If $\mathbb{E}[P^1] = \mathbb{E}[P^2]$ then the objective value of the local optimal solution satisfying $L^i = L^*, i \in \{1,2\}$, such that $\text{Prob}(P^1 \leq L^*) + \text{Prob}(P^2 \leq L^*) = 1$ is $z_2 = \alpha \mathbb{E}[P^1 | P^1 \leq L^1].S_f. + (1-\alpha)\mathbb{E}[P^2 | P^2 \leq L^2].S_f \quad \alpha \in (0,1)$. Thus, since

$\mathbb{E}[P^i|P^i \leq L^*] \leq \mathbb{E}[P^i], i \in \{1,2\}$ then $z_2 \leq \alpha\mathbb{E}[P^1].S_f. + (1-\alpha)\mathbb{E}[P^2].S_f = S_f\mu = z_1$, where z_1 is the objective value given by the first case above. A useful inference from the results is that the aggregator's optimal thresholds for the two periods should be equal. Otherwise, the aggregator could always reduce the expected procurement cost by lowering the higher threshold and raising the lower threshold accordingly. Using this strategy, the optimal expected cost is lower than $\mathbb{E}[P^i], i \in \{1,2\}$, the purchase price is capped at L^* and therefore it is less risky than buying in the period with the lowest expected energy price. □

Appendix E: Stochastic behavior of an aggregator, multiple periods

We formulate the problem for an aggregator in the day-ahead energy market (first stage) by combining the optimal strategies outlined in section 3.2 to minimize the expected cost of serving deferrable demand over certain periods, and using the differences in prices over a optimization horizon (e.g., the 24 h of a day).

Proof Proof of Theorem 3.3

The optimization problem is given by (14).

$$\min_{L^t,H^t} \Big\{ \begin{aligned} v(L^t,H^t) &= \sum_{t\in\mathcal{T}} \big(D^t.\mathbb{E}[P^t] + c.\mathbb{E}[P^t|P^t \leq L^t]\,\mathrm{Prob}(P^t \leq L^t)\big) \\ &\quad - \sum_{t\in\mathcal{T}} \min\{D^t,d\}\mathbb{E}[P^t|P^t > H^t]\,\mathrm{Prob}(P^t > H^t)| \\ g^{\tau_1}(L^t,H^t) &= \sum_{t\leq\tau_1} \big(c.\,\mathrm{Prob}(P^t \leq L^t).\eta - \min\{D^t,d\}\,\mathrm{Prob}(P^t > H^t)\big) \\ &\quad - (S_{\max} - S_0) \leq 0, \quad \forall\tau_1 \in \mathcal{T}\backslash\{T\} \\ h^{\tau_2}(L^t,H^t) &= \sum_{t\leq\tau_2} \big(\min\{D^t,d\}\,\mathrm{Prob}(P^t > H^t) - c.\,\mathrm{Prob}(P^t \leq L^t).\eta\big) \\ &\quad - S_0 \leq 0, \quad \forall\tau_2 \in \mathcal{T}\backslash\{T\} \\ k(L^t,H^t) &= \sum_{t\in\mathcal{T}} \big(\min\{D^t,d\}\,\mathrm{Prob}(P^t > H^t) - c.\,\mathrm{Prob}(P^t \leq L^t).\eta\big) = 0 \Big\}. \end{aligned} \tag{14}$$

The set of constraints $g^{\tau_1}(L^t,H^t) \leq 0$ and $h^{\tau_2}(L^t,H^t) \leq 0$ are on the upper and lower bounds of the storage capacity respectively at the end of each period. The constraint $k(L^t,H^t) = 0$ is an expected balance between total charging and discharging over the optimization horizon T (e.g., 24 h). This is equivalent to a transversality condition stating that the final level of energy stored should be equal to the initial one, thus avoiding assuming that the energy S_0 is free. Note that the storage capacity constraints for the final hour h^T and g^T would be weaker than $k(L^t,H^t) = 0$ and therefore are not included.

Let f_{X^t} denote the probability density function of the energy price at hour t. This problem can be expressed as (15). The dual variables are indicated in parentheses on the right.

$$\begin{aligned}
\min_{L^t,H^t} v(L^t,H^t) &= \sum_{t\in\mathcal{T}}\left(\int_{-\infty}^{+\infty} P^t f_{X^t}(P^t)dP^t.D^t + c\int_{-\infty}^{L^t} P^t f_{X^t}(P^t)dP^t\right)\\
&\quad -\sum_{t\in\mathcal{T}}\min\{D^t,d\}\int_{H^t}^{+\infty} P^t f_{X^t}(P^t)dP^t\\
&\text{st.}\\
g^{\tau_1}(L^t,H^t) &= \sum_{t\le\tau_1}\left(c\int_{-\infty}^{L^t} f_{X^t}(P^t)dP^t.\eta - \min\{D^t,d\}\int_{H^t}^{+\infty} f_{X^t}(P^t)dP^t\right)\\
&\quad -(S_{\max}-S_0)\\
&\le 0,\quad \forall\tau_1\in\mathcal{T}\backslash\{T\} \qquad (\mu^{\tau_1})\\
h^{\tau_2}(L^t,H^t) &= \sum_{t\le\tau_2}\left(\min\{D^t,d\}\int_{H^t}^{+\infty} f_{X^t}(P^t)dP^t - c\int_{-\infty}^{L^t} f_{X^t}(P^t)dP^t.\eta\right) - S_0\\
&\le 0,\quad \forall\tau_2\in\mathcal{T}\backslash\{T\} \qquad (\lambda^{\tau_2})\\
k(L^t,H^t) &= \sum_{t\in\mathcal{T}}\left(\min\{D^t,d\}\int_{H^t}^{+\infty} f_{X^t}(P^t)dP^t - c\int_{-\infty}^{L^t} f_{X^t}(P^t)dP^t.\eta\right) = 0. \qquad (\gamma)
\end{aligned} \tag{15}$$

The Karush-Kuhn-Tucker (KKT) conditions can be expressed as (16).

$$\begin{aligned}
(L^t)\quad & \frac{\partial\mathcal{L}}{\partial L^t} \equiv && \frac{\partial v}{\partial L^t} + \sum_{t\le\tau_1}\mu^{\tau_1}\frac{\partial g^{\tau_1}}{\partial L^t} + \sum_{t\le\tau_2}\lambda^{\tau_2}\frac{\partial h^{\tau_2}}{\partial L^t} + \gamma\frac{\partial k}{\partial L^t} = 0,\\
& && \forall\tau_1\in\mathcal{T}\backslash\{T\},\quad \forall\tau_2\in\mathcal{T}\backslash\{T\}\\
(H^t)\quad & \frac{\partial\mathcal{L}}{\partial H^t} \equiv && \frac{\partial v}{\partial H^t} + \sum_{t\le\tau_1}\mu^{\tau_1}\frac{\partial g^{\tau_1}}{\partial H^t} + \sum_{t\le\tau_2}\lambda^{\tau_2}\frac{\partial h^{\tau_2}}{\partial H^t} + \gamma\frac{\partial k}{\partial H^t} = 0,\\
& && \forall\tau_1\in\mathcal{T}\backslash\{T\},\quad \forall\tau_2\in\mathcal{T}\backslash\{T\}\\
(g^{\tau_1})\quad & g^{\tau_1}(L^t,H^t) && \le 0,\quad \forall\tau_1\in\mathcal{T}\backslash\{T\}\\
(h^{\tau_2})\quad & h^{\tau_2}(L^t,H^t) && \le 0,\quad \forall\tau_2\in\mathcal{T}\backslash\{T\}\\
(k)\quad & k(L^t,H^t) && = 0\\
(cg^{\tau_1})\quad & \mu^{\tau_1}g^{\tau_1}(L^t,H^t) && = 0,\ \mu^{\tau_1}\ge 0\ \ \forall\tau_1\in\mathcal{T}\backslash\{T\}\\
(ch^{\tau_2})\quad & \lambda^{\tau_2}h^{\tau_2}(L^t,H^t) && = 0,\ \lambda^{\tau_2}\ge 0\ \ \forall\tau_2\in\mathcal{T}\backslash\{T\}.
\end{aligned} \tag{16}$$

For the condition (L^t), the FOC implies:

$$\begin{aligned}
(L^t 1)\quad \frac{\partial\mathcal{L}}{\partial L^t} \equiv\ & c.\eta.f_{X^t}(L^t)\sum_{t\le\tau_1\le T}\mu^{\tau_1} - c.\eta.f_{X^t}(L^t)\sum_{t\le\tau_2\le T}\lambda^{\tau_2} - \gamma.c.\eta.f_{X^t}(L^t)\\
& + c.L^t f_{X^t}(L^t) = 0,
\end{aligned} \tag{17}$$

which can be simplified to

$$L^t f_{X^t}(L^t) + \eta . f_{X^t}(L^t) \left(\sum_{t \le \tau_1 \le T} \mu^{\tau_1} - \sum_{t \le \tau_2 \le T} \lambda^{\tau_2} - \gamma \right) = 0, \tag{18}$$

and for any hour t, if $f_{X^t}(L^t) \neq 0$ in optimality, then

$$L^t + \eta \left(\sum_{t \le \tau_1 \le T} \mu^{\tau_1} - \sum_{t \le \tau_2 \le T} \lambda^{\tau_2} - \gamma \right) = 0, \tag{19}$$

and therefore

$$(L^{t'}) \quad L^t = -\eta \left(\sum_{t \le \tau_1 \le T} \mu^{\tau_1} - \sum_{t \le \tau_2 \le T} \lambda^{\tau_2} - \gamma \right). \tag{20}$$

By the complementary slackness condition (cg^{τ_1}), if $g^{\tau_1}(L^t, H^t) < 0$, when the upper bound of the storage is not binding at the end of hour τ_1, this implies $\mu^{\tau_1} = 0$. Similarly, by the complementary slackness condition (ch^{τ_2}), $\lambda^{\tau_2} h^{\tau_2}(L^t, H^t) < 0$, when the lower bound of the storage is not binding at the end of hour τ_2. This implies $\lambda^{\tau_2} = 0$. Thus, for any hour t, if $\mu^t = \mu^{t+1} = \cdots = \mu^{T-1} = \lambda^t = \cdots = \lambda^{T-1} = 0$, then $L^t = \eta\gamma$, implying that $L^t = L^{t+1} = \cdots = L^T$. This establishes the result. □

Proof Proof of Lemma 3.1 For the condition (H^t),

$$\begin{aligned}(H^t 1) \quad \frac{\partial \mathcal{L}}{\partial H^t} \equiv \quad & \min\{D^t, d\}.H^t f_{X^t}(H^t) + \min\{D^t, d\}.f_{X^t}(H^t) \sum_{t \le \tau_1 \le T} \mu^{\tau_1} \\ & - \min\{D^t, d\}.f_{X^t}(H^t) \sum_{t \le \tau_2 \le T} \lambda^{\tau_2} - \gamma \min\{D^t, d\}.f_{X^t}(H^t) = 0,\end{aligned} \tag{21}$$

and for any hour t, if $f_{X^t}(H^t) \neq 0$ in optimality, then

$$H^t + \left(\sum_{t \le \tau_1 \le T} \mu^{\tau_1} - \sum_{t \le \tau_2 \le T} \lambda^{\tau_2} - \gamma \right) = 0. \tag{22}$$

and therefore

$$(H^{t'}) \quad H^t = -\left(\sum_{t \le i \le T} \mu^{\tau_1} - \sum_{t \le \tau_2 \le T} \lambda^{\tau_2} - \gamma \right). \tag{23}$$

Analogous to the analysis for L^t using the complementary slackness conditions, if $\mu^t = \mu^{t+1} = \cdots = \mu^{T-1} = \lambda^t = \cdots = \lambda^{T-1} = 0$, then $H^t = H^{t+1} = \cdots = H^T = \gamma$. This establishes the result. □

Analogous to the analysis for L^t using the complementary slackness conditions, if $\mu^t = \mu^{t+1} = \cdots = \mu^{T-1} = \lambda^t = \cdots = \lambda^{T-1} = 0$, then $H^t = H^{t+1} = \cdots = H^t = \gamma$. This implies the following.

Proof Proof of Corollary 3.2 If the upper bound of the storage capacity is binding at the end of period τ_1, $g^{\tau_1}(L^t, H^t) = 0$, and $\mu^{\tau_1} > 0$. From $(H^{t'})$, the optimal discharging thresholds for period τ_1 and all periods before τ_1 are lowered by μ^{τ_1}. The charging thresholds for periods $k \leq \tau_1$ are also lowered, according to $L^t = \eta H^t$.

Similarly, if the lower bound of the storage capacity is binding at the end of period τ_2, $h^{\tau_2}(L^t, H^t) = 0$, and $\lambda^{\tau_2} > 0$. From $(L^{t'})$, the optimal charging threshold L^{τ_2} is increased at period τ_2 and for all periods before τ_2 by $\eta\lambda^{\tau_2}$. The discharging thresholds for periods $k \leq \tau_2$ are also increased, according to $L^t = \eta H^t$. □

In practice, it is possible that the storage is charged in one system state and discharged in another for the same hour. This capability implies that DD can provide ramping services even though there is no formal market for ramping. Consequently, the benefits of the aggregator's strategy for managing storage are not limited to minimizing the expected cost of meeting the DD requirements. The bid strategy also provides the flexibility needed to deliver ramping services to the system operator even though the nodal price of energy is the only market signal. Note that in this market there are non-convexities due to the unit commitment problem that can affect the prices. This fact is beyond the control of aggregators.

Appendix F: Parameters for the aggregator proxies

We assume that the forecast of prices follows a shifted lognormal distribution for each hour. Further, we assume that the participation in the market does not change the distribution of the price forecast on which the bids and offers are based.

Consider a $Z \sim \log N(\mu, \sigma^2)$. The probability density function (PDF) of z is given by

$$f_{Z^i}(z, \mu, \sigma) = \frac{1}{z\sigma\sqrt{2\pi}} e^{-\frac{(\ln z - \mu)^2}{2\sigma^2}}, z > 0. \tag{24}$$

Let $X = g(Z) = Z + P_{\min}$ denote a shifted lognormal distribution. Then $Z = g^{-1}(X) = X - P_{\min}$. The pdf for X is given by

$$f_{X^i}(x, \mu, \sigma, P_{\min}) = \frac{1}{(x - P_{\min})\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x - P_{\min}) - \mu)^2}{2\sigma^2}}, x > P_{\min}. \tag{25}$$

The Cumulative Distribution Function (CDF) of X is given by

$$F_{X^i}(x, \mu, \sigma, P_{\min}) = F_{Z^i}(x - P_{\min}, \mu, \sigma), x > P_{\min}. \tag{26}$$

The first two moments of this distribution are

$$\mathbb{E}[X] = \mathbb{E}[Z] + P_{\min} = e^{\mu + \frac{\sigma^2}{2}} + P_{\min}$$
$$\mathbb{V}[X] = \mathbb{V}[Z] = (\mathbb{E}[X] - P_{\min})^2 (e^{\sigma^2} - 1). \quad (27)$$

Therefore, the mean and variance can be calibrated by the expectation and variance of the prices, $\mathbb{E}[P]$, $\mathbb{V}[P]$ as follows

$$\mu = \ln(\mathbb{E}[P] - P_{\min}) - \frac{\sigma^2}{2}$$
$$\sigma^2 = \ln(1 + \frac{\mathbb{V}[P]}{(\mathbb{E}[X] - P_{\min})^2}). \quad (28)$$

We also consider other distributions (truncated normal, triangular). A normal distribution is attractive, as it provides a simple closed form. However the implicit assumption is that prices are distributed symmetrically around the mean. In our simulation cases, this is not the case.

The triangular distribution is given by three parameters. The lower limit, the higher limit and the mode. We omit the results from using a triangular distribution for brevity.

Appendix G: Model calibration

Our calibration is based on publicly available sources, including the information for the electricity network, the characteristics of the installed generating units, the model attributes for the stochastic resources and load, and the energy and power properties for deferrable demand. The stochastic resources (e.g., wind, solar, wave energy; demand) are modeled as Markovian processes with discretized probability distributions over a finite number of states (in our application five per time period), and each of these states allows for contingencies (equipment failures). We summarize the input here and highlight the main differences with respect to our previous work (Jeon et al., 2015, section 4).

G.1 Network and generation fleet

The test network is a 36-bus reduction of a New England and New York centric version of the North Eastern Power Coordinating Council (NPCC) network (Allen et al., 2008). The information for the generating units corresponds to reported data to the U.S. Energy Information Administration (EIA, 2011), complemented with data provided by Energy Visuals (EVFR, 2012) for over 690 generators, including the location of the units, and the initial estimated fuel costs. We internalize the ramping costs incurred by conventional generators when accommodating the variability and uncertainty of stochastic resources. These costs are consistent with the information in Lew et al. (2013) and the estimation of ramping effects by Cullen (2013), and they are specified by fuel type. Changing the dispatch of conventional generation to

provide ramping services reduces their efficiency and causes damages that are accrued over time (Kumar et al., 2012; Moarefdoost et al., 2016). The ramping costs include lower thermal performance, e.g., heat rate degradation for thermal generating units; equipment damage, e.g., creep damage, increases in equipment forced outage rates (EFOR); and higher operating and maintenance costs (O&M). The system operator manages aleatoric uncertainty by scheduling enough contingency reserve capacity to cover (by potentially re-dispatching) intra-temporal equipment failure contingencies. This means that if the system losses any element randomly, it can continue operating, or what is called $n - 1$ reliability in electric systems. Each period of time has its own contingency set. The system operator also schedules robust inter-temporal load following reserves for the worst-case ramp necessary, in the empirical application for the five states described in "Appendix G.2". In practice, reserve capacity can be used for covering both contingency and load following events. The amount of down ramping reserves determines how much potential wind generation has to be spilled when e.g., wind speeds are unexpectedly high.

G.2 Modeling the uncertainty of VRES

We limit our attention to modeling wind energy sources. This follows two main research considerations: (1) the negatively correlated diurnal patterns between the demand for thermal services and the availability of the VRES modeled (2) the potential for capacity additions of VRES in the region of analysis, the Northeastern United States (EIA, 2019b). We estimate time-series models using hourly data to forecast temperature, wind speed and load for specified locations. These equations are estimated in two steps, first for temperature, and then for wind speed and load, using the estimated temperature as an input, as Auto-Regressive Moving Average (ARMAX) models with exogenous variables. We use the estimated variances and covariances of the white-noise residuals for these estimated equations to generate random sequences of multivariate normal residuals. We specify for any day and starting hour and then calculate the deterministic forecasts for temperature. For this we use the previously estimated model for a 24-h ahead forecast. This model is used to do deterministic forecasts of load, and a Monte Carlo simulation of wind speed for 1,000 realizations. The obtained wind speeds are transformed to the equivalent potential wind generation (PWG) obtained from a multi-turbine modeling approach. The details of the statistical fit and the Monte Carlo simulation are presented in Jeon et al. (2015). One of the specifications of our model is that the uncertainty of PWG is Markovian, and discretized in a finite number of states. We rank the simulated PWG and group them into a finite number of bins (five in what follows). We build the profiles and transition probability matrices based on simulations for August 8th, 2006. We create inputs for 24 receding horizon settlements ($N = 24$) for successive periods based on the Monte Carlo econometric model, each one with a 24-h horizon $n^t = 24$. The transition probability matrices are consistent with respect to the first settlement period. That is, they represent a conditional forecast. This assumption bounds extreme possible realizations out of the forecasts available at the beginning of the day. But the substantial persistence observed in the transitions between periods supports this modeling decision. The two

more important features of the PWG profiles are (1) they exhibit substantial persistence because the residuals of the forecasts are highly positively auto-correlated; and (2) the range between the highest and lowest PWG increases over the 24-h horizon, because the forecasting accuracy deteriorates over time. Some demand may not be served, particularly in the rare contingency states. However, shedding load is expensive, and the specified VOLL are $10,000/MWh for urban areas and $5000/MWh for rural areas.

G.3 Deferrable demand model

The specifications of deferrable demand (DD) consider only thermal storage for space cooling because air conditioning is the main cause of the annual peak load. Reducing this peak reduces the amount of installed generating capacity needed for generation adequacy. The ARMAX model of load distinguishes between temperature-sensitive and non-temperature-sensitive load. Some customers have thermal storage but most do not. The energy capacity of thermal storage is 17 GWh corresponding to 1/15 of the total daily amount of electricity used for space cooling that is potentially deferrable (temperature-sensitive demand in the peak day). The optimal management of storage determines when to charge (usually at night) and when to discharge (usually during peak load periods) the storage. This storage is allocated to five load centers in proportion to the load at each center. The technical characteristics of storage are based on the products described by Evapco (EVAPCO, 2007) and Calmac (Hunt et al., 2010). The hourly ice building power rate is 12% and the hourly ice melting power rate is 16.7% of the total storage capacity, but rates can vary in practice with the number of chillers installed. The specified efficiency of 86% is based on an average Energy Efficiency Ratio (EER) of 8.8 for thermal storage compared to an EER of 10.2 for a conventional air conditioner.

References

Allen, E., Lang, J., & Ilic, M. (2008). A combined equivalenced-electric, economic, and market representation of the northeastern power coordinating council U.S. electric power system. *IEEE Transactions on Power Systems, 23*(3), 896–907.

Arroyo, J., & Galiana, F. (2005). Energy and reserve pricing in security and network-constrained electricity markets. *IEEE Transactions on Power Systems, 20*(2), 634–643.

Bertsimas, D., Iancu, D. A., & Parrilo, P. A. (2010). Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research, 35*(2), 363–394.

Bertsimas, D., Litvinov, E., Sun, X., Zhao, J., & Zheng, T. (2013). Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems, 28*(1), 52–63.

Birge, J., & Louveaux, F. (1997). *Introduction to stochastic programming. Springer series in operations research series*. Springer.

Boiteux, M. (1952). Report of the Louvain meeting, September 12–14, 1951. *Econometrica, 20*(2), 306–332.

Borenstein, S., Bushnell, J., Knittel, C. R., & Wolfram, C. (2008). Inefficiencies and market power in financial arbitrage: A study of California's electricity markets. *The Journal of Industrial Economics, 56*(2), 347–378.

CAISO. (2021). Energy storage enhancements: Straw proposal, 12/9/2022. Technical report, CAISO

CAISO. (2022). Energy storage enhancements: Final proposal, 10/27/2022. Technical report, CAISO.

Cullen, J. A. (2013). Dynamic response to environmental regulation in the electricity industry. Technical report, Washington University in St. Louis.
Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research, 58*, 595–612.
EIA. (2011). Voluntary reporting of greenhouse gases program fuel carbon dioxide emission coefficients. Technical report, Energy Information Administration.
EIA, R. S. (2019). The northeast leads the country in net-metered wind electricity generation capacity. Technical report, EIA.
EIA, M. W. (2019a). North Dakota, Montana, and Wyoming consume the most residential energy per capita. Technical report, EIA.
EIA, N. V. (2022). Small-scale solar is changing hourly utility electricity demand in New England. Technical report, EIA.
EVAPCO. (2007). Thermal ice storage—application and design guide. Technical report, EVAPCO, Inc.
EVFR. (2012). Firstrate. Technical report, Energy Visuals, Inc.
Feng, Z., Ajjarapu, V., & Maratukulam, D. (1998). A practical minimum load shedding strategy to mitigate voltage collapse. *IEEE Transactions on Power Systems, 13*(4), 1285–1290.
FERC. (2020). Order no. 2222. Technical report, Federal Energy Regulatory Commission.
Filomena, T., & Lejeune, M. (2014). Warm-start heuristic for stochastic portfolio optimization with fixed and proportional transaction costs. *Journal of Optimization Theory and Applications, 161*(1), 308–329.
Gellings, C., & Smith, W. (1989). Integrating demand-side management into utility planning. *Proceedings of the IEEE, 77*(6), 908–918.
Hunt, M., Heinemeier, K., Hoeschele, M., & Weitzel, E. (2010). HVAC energy efficiency maintenance study. Technical report, CALMAC.
Jeon, W., Lamadrid, A. J., Mo, J., & Mount, T. D. (2015). Using deferrable demand in a smart grid to reduce the cost of electricity for customers. *Journal of Regulatory Economics, 47*(3), 239–272.
Jeon, W., Lamadrid, A. J., & Mount, T. D. (2019). The economic value of distributed storage at different locations on an electric grid. *The Energy Journal, 40*(4), 165–190.
Koohi-Fayegh, S., & Rosen, M. (2020). A review of energy storage types, applications and recent developments. *Journal of Energy Storage, 27*, 101047.
Kuang, X., Lamadrid, A. J., & Zuluaga, L. F. (2019). Pricing in non-convex markets with quadratic deliverability costs. *Energy Economics, 80*, 123–131.
Kumar, N., Besuner, P. M., Lefton, S. A., Agan, D. D., & Hilleman, D. D. (2012). Power plant cycling costs. Technical report, Intertek APTECH.
Lamadrid, A. J., & Mount, T. D. (2012). Ancillary services in systems with high penetrations of renewable energy sources, the case of ramping. *Energy Economics, 34*(6), 1959–1971.
Lamadrid, A. J., Muñoz-Álvarez, D., Murillo-Sánchez, C. E., Zimmerman, R. D., Shin, H., & Thomas, R. J. (2019). Using the Matpower Optimal Scheduling Tool to test power system operation methodologies under uncertainty. *IEEE Transactions on Sustainable Energy, 10*(3), 1280–1289.
Lamadrid, A., Shawhan, D., Murillo-Sanchez, C., Zimmerman, R., Zhu, Y., Tylavsky, D., Kindle, A., & Dar, Z. (2015). Stochastically optimized, carbon-reducing dispatch of storage, generation, and loads. *IEEE Transactions on Power Systems, 30*(2), 1064–1075.
Lew, D., Brinkman, G., Kumar, N., Lefton, S., Jordan, G., & Venkataraman, S. (2013). Finding flexibility: Cycling the conventional fleet. *IEEE Power and Energy Magazine, 11*(6), 20–32.
Lorca, Á., Sun, X. A., Litvinov, E., & Zheng, T. (2016). Multistage adaptive robust optimization for the unit commitment problem. *Operations Research, 64*(1), 32–51.
Mas-Colell, A., Whinston, M., & Green, J. (1995). *Microeconomic theory*. Oxford: Oxford University Press.
Mayne, D. Q. (2014). Model predictive control: Recent developments and future promise. *Automatica, 50*(12), 2967–2986.
Moarefdoost, M. M., Lamadrid, A. J., & Zuluaga, L. F. (2016). A robust model for the ramp-constrained economic dispatch problem with uncertain renewable energy. *Energy Economics, 56*, 310–325.
Morales, J. M., Conejo, A. J., & Perez-Ruiz, J. (2009). Economic valuation of reserves in power systems with high penetration of wind power. *IEEE Transactions on Power Systems, 24*(2), 900–910.
Murillo-Sanchez, C., Zimmerman, R., Anderson, C., & Thomas, R. (2013). Secure planning and operations of systems with stochastic sources, energy storage, and active demand. *IEEE Transactions on Smart Grid, 4*(4), 2220–2229.

O'Neill, R. P., Sotkiewicz, P. M., Hobbs, B. F., Rothkopf, M. H., & Stewart, W. R. (2005). Efficient market-clearing prices in markets with nonconvexities. *European Journal of Operational Research, 164*(1), 269–285.

Powell, W. (2007). *Approximate dynamic programming: Solving the curses of dimensionality. Wiley series in probability and statistics*. Wiley.

Pritchard, G., Zakeri, G., & Philpott, A. (2010). A single-settlement, energy-only electric power market for unpredictable and intermittent participants. *Operations Research, 58*(4–part–2), 1210–1219.

Rahimiyan, M., Baringo, L., & Conejo, A. J. (2014). Energy management of a cluster of interconnected price-responsive demands. *IEEE Transactions on Power Systems, 29*(2), 645–655.

Rossiter, J. A. (2017). *Model-based predictive control: A practical approach*. CRC Press.

Schmalensee, R. (2022). Competitive energy storage and the duck curve. *The Energy Journal, 43*(2), 1–16.

Secomandi, N. (2010). Optimal commodity trading with a capacitated storage asset. *Management Science, 56*(3), 449–467.

Steiner, P. O. (1957). Peak loads and efficient pricing. *The Quarterly Journal of Economics, 71*(4), 585–610.

Wang, C., & Shahidehpour, S. (1995). Optimal generation scheduling with ramping costs. *IEEE Transactions on Power Systems, 10*(1), 60–67.

Wu, O. Q., Wang, D. D., & Qin, Z. (2012). Seasonal energy storage operations with limited flexibility: The price-adjusted rolling intrinsic policy. *Manufacturing and Service Operations Management, 14*(3), 455–471.

Zheng, N., Qin, X., Wu, D., Murtaugh, G., & Xu, B. (2023). Energy storage state-of-charge market model. *IEEE Transactions on Energy Markets, Policy, and Regulation, 1*(1), 11–22.

Zheng, Q. P., Wang, J., & Liu, A. L. (2015). Stochastic optimization for unit commitment; a review. *IEEE Transactions on Power Systems, 30*(4), 1913–1924.

Zhou, Y. H., Scheller-Wolf, A., Secomandi, N., & Smith, S. (2016). Electricity trading and negative prices: Storage vs. disposal. *Management Science, 62*(3), 880–898.

Zhou, Y. H., Scheller-Wolf, A., Secomandi, N., & Smith, S. (2019). Managing wind-based electricity generation in the presence of storage and transmission capacity. *Production and Operations Management, 28*(4), 970–989.