Curtailing Intermittent Generation in Electrical Systems

Owen Q. Wu Roman Kapuscinski

Stephen M. Ross School of Business, University of Michigan, Ann Arbor, Michigan

March 30, 2013

Energy generation from intermittent renewable sources introduces additional variability into electrical systems, resulting in a higher cost of balancing against the increased variabilities. Ways to balance demand and supply for electricity include using flexible generation resources, storage operations, and curtailing intermittent generation. This paper focuses on the operational and environmental impact of curtailing intermittent generation. We construct a stochastic dynamic optimization model that captures the critical components of the system operating cost and analyze how various generation resources should operate with and without curtailing intermittent generation. We find that the system cost reduction per unit of curtailed energy is consistently significant, and the presence of storage may increase the cost saving per unit of curtailed energy. We also find that curtailing intermittent generation often leads to system emission reductions.

1. Introduction

Intermittent renewable energy sources, such as wind, solar, and tidal power, generate renewable energy at nearly zero marginal cost, and for that reason electrical systems typically use as much intermittent generation as possible to meet the electricity demand. The variability introduced by intermittent generation is absorbed primarily by adjusting the output of conventional resources (e.g., coal or natural gas power plants) or using energy storage. These options, however, are costly. Adjusting the output of conventional resources increases their operating cost; energy storage operations typically incur significant energy losses during energy conversion processes. In contrast, curtailing intermittent generation involves negligible operating cost, e.g., wind power output can be adjusted by pitching the blades of a wind turbine. Thus, curtailing intermittent generation may be considered a third option to manage intermittency. We refer to this type of curtailment as *economic curtailment*, as opposed to the necessary curtailment due to system operating constraints and reliability requirement.

Allowing economic curtailment expands the action space of an electrical system; therefore, the system operating cost should not increase under the optimal decision, but it is important to ask how significant the cost reduction is. Given that intermittent generation incurs nearly zero marginal cost while conventional generation is costly, our prior is that economic curtailment may not bring significant cost savings. Even if economic curtailment significantly reduces cost, the second question is whether it will reduce the environmental benefit of renewable energy and result in more pollution. Our prior is that using less clean energy may be environmentally harmful. Furthermore, curtailing clean energy seems inferior compared to storing it, if energy storage is available. Thus, the third question is how energy storage operations affect the operational and environmental impact of economic curtailment.

The findings of this paper are contrary to our priors. We find that the operating cost reduction per unit of curtailed energy is consistently significant and often exceeds the marginal production cost of conventional resources. Second, minimizing the system operating cost with the curtailment option often reduces emissions. Thus, economic curtailment can be both economically and environmentally beneficial. Third, although energy storage operations reduce the benefit from economic curtailment, it may increase the cost savings per unit of curtailed energy.

2. Literature Review

A vast body of literature is devoted to the integration of intermittent resources into electrical systems. We review the literature that is closely related to our paper.

2.1 Wind Integration Studies and Curtailment Issues

With the rapid growth of wind power in the past decade, utilities and independent system operators (ISOs) have completed a number of wind integration studies, including those by the British electricity system (Gross et al. 2006), New York ISO (GE Energy 2005), California ISO (2007), and Minnesota (EnerNex 2006). These wind integration studies usually simulate a system with various levels of wind energy penetration and evaluate the incremental costs caused by intermittency. Excellent reviews of these studies are conducted by Smith et al. (2007), Ela et al. (2009), and Hart et al. (2012). Recent wind and solar integration studies (GE Energy 2010, EnerNex 2011) expand the analyses by considering integration in large geographical regions. Most of these studies report the amount of wind energy curtailment necessary to meet system constraints and reliability requirement, but do not analyze economic curtailment as a way to save system operating costs.

Economic curtailment of renewable energy is considered counter-effective to meeting the Renewable Portfolio Standards in the U.S. or following the European Union Renewable Energy Directive (2009): "... when dispatching electricity generating installations, transmission system operators shall give priority to generating installations using renewable energy sources. ... Member States shall ensure that appropriate grid and market-related operational measures are taken in order to minimise the curtailment of electricity produced from renewable energy sources." However, more recent studies find that curtailing wind energy has economic benefits. In a three-node system, Ela (2009) shows that when a transmission line is congested, curtailing wind energy may increase the use of inexpensive generation units at the other node, thereby lowering the total cost. Ela and Edelson (2012) provide a case study on the benefit of economic curtailment in the absence of transmission constraints. They find economic curtailment may bring substantial cost savings because it helps relieve physical constraints of generation resources.

Accommodating intermittency also raises environmental concerns. When economic curtailment is not allowed, Bentek Energy (2010) identifies an important cost associated with adjusting conventional resources to accommodate intermittent generation. This adjustment, called cycling, requires extra fuel, leading to extra cost and emissions. Katzenstein and Apt (2009) find that, due to extra emissions from cycling natural gas generators, the carbon dioxide emission reductions are likely to be significantly less than those assumed by policy makers. Xcel Energy (2011) points out that although curtailing wind power reduces cycling, it increases the emissions from fossil generation units that are kept running instead of being shut down.

Despite differing views on wind energy curtailment, market mechanisms for curtailment are being developed. Since 2009, a few ISOs began to allow wind-power producers to provide price offers; see Ela and Edelson (2012) for an overview of this development. In principle, this enhancement means that intermittent resources no longer receive priority and may be economically curtailed. However, because wind-power producers receive production-based subsidies, they typically offer negative prices, i.e., the producers are willing to pay to produce. As a result, wind power effectively receives priority to a significant extent. EnerNex (2011) points out that selective and appropriate use of wind power curtailment is an opportunity for further investigation.

Our work complements these studies by decomposing system operating costs into six components and analyzing the tradeoffs among these costs. We identify situations where economic curtailment can be both economically and environmentally beneficial.

2.2 Advanced Methods for Wind Integration Studies

In a typical electrical system, resources are coordinated by unit commitment (UC) and economic dispatch (ED) programs. The UC program is run every day to determine which generation units should be committed for each hour of the next day, and the ED program is run in real time to determine the output levels of the committed units. Although UC and ED programs involve sophisticated system modeling and optimization techniques, Milligan et al. (2012) point out that wind integration is still a relatively young field and new methodologies are needed. Various advanced methods recently have been proposed and some have been implemented, which we review below.

Look-ahead ED. Current ED programs typically minimize cost over a short time interval (e.g., five minutes). Xie et al. (2011) discuss a look-ahead ED method that optimizes the production schedule over a longer time horizon. A few major ISOs, such as PJM Interconnection and Midwest ISO (MISO), have recently implemented the look-ahead method.

Stochastic UC and rolling planning. Milligan et al. (2012) and Ela et al. (2009) describe the emerging stochastic UC method, which minimizes expected cost over multiple scenarios, and discuss a rolling planning approach, which adjusts UC decisions during the operating day. Weber et al. (2009) use an approximation method to reduce computational complexity: Similar units are grouped and the group's dispatchable capacity is a continuous decision variable. The resulting stochastic programs (typically having two or three stages, each stage three hours or longer) are implemented in a tool called WILMAR. Using this tool, Meibom et al. (2011) evaluate the use of stochastic UC and rolling planning for the Eastern Interconnection. Tuohy et al. (2009) combine WILMAR with a mixed-integer scheduling model and find that stochastic UC reduces system costs by 0.9%.

Sub-hourly analysis. The current hourly-based scheduling programs may be inadequate under high wind penetration levels due to significant sub-hourly wind variations. Statistical analysis in a study by GE Energy (2010) confirms that sub-hourly scheduling is critical. In April 2012, MISO implemented a look-ahead tool that runs every 15 minutes to provide the operator with suggestions for sub-hourly adjustments.

The solution technique in this paper is stochastic dynamic programming (SDP). SDP explicitly models stochasticity and is forward-looking in nature. Therefore, SDP contains the look-ahead ED, stochastic UC, and rolling planning methods. We use 15-minute intervals in SDP to include sub-hourly analysis. To reduce computational complexity, we use an approximation method similar to WILMAR but tailored to the SDP framework.

2.3 Related Work in Capacity and Inventory Management

From a methodological point of view, our work is related to the production-inventory literature that considers capacity adjustment. Rocklin et al. (1984) are among the first to study capacity expansion and contraction under stochastic demand processes. The key tradeoff is between having too much capacity (thus paying unnecessary capacity maintenance cost) and having too little capacity (thus meeting demand at a higher cost). Eberly and Van Mieghem (1997) generalize the problem to include multiple factors (e.g., labor and capital) in the production capacity. Angelus and Porteus (2002) consider the role of inventory in production and capacity adjustment. These production-inventory systems have several similarities to the electrical systems: demand stochastically rises and falls; capacity adjustment corresponds to units' startup and shutdown; having too much or too little capacity increases the cost of meeting demand. On the other hand, electrical systems have some differentiating features: shutting down capacity does not generate a return; storage incurs energy conversion losses and has limited space; importantly, capacity adjustment is gradual and a single capacity decision affects the capacity adjustment process over multiple periods. These features are not considered in the models cited above. In this paper, we model capacity adjustment in electrical systems by introducing pending-up and pending-down capacities. We find the optimal capacity adjustment policy can be characterized by pending capacity targets.

3. The Model

In electrical systems, the important elements for studying the curtailment of intermittent generation are on a time scale of minutes to hours. We use $t \in \{0, 1, \ldots, T\}$ to index time periods, with each period representing a 15-minute interval.

3.1 Production Costs

To approximate a fleet of power generation units with various levels of flexibility, we assume that the generation units have three levels of flexibility: inflexible units, intermediate (semi-flexible) units, and fully flexible units. The latter two are referred to as *flexible resources*.

Inflexible units have the lowest production cost per unit of energy, but generate power at a constant level. Let K^R denote the total capacity (maximum output per period) of inflexible units. Their total output $Q^R \in [0, K^R]$ stays constant once Q^R is chosen prior to $t = 0$. They incur a production cost of $c^R Q^R$ every period, where c^R is the cost per unit of energy produced.

Fully flexible units, also known as *peaking units*, can adjust their output almost instantaneously with negligible adjustment cost, but they have high production cost per unit of energy, denoted as c^P. Producing Q^P units of energy by the peaking units costs $c^P Q^P$.

Intermediate units have limited flexibility and intermediate production cost per unit of energy. Their limited flexibility is reflected in three types of costs (part-load penalty, min-gen penalty, and cycling cost, detailed below) and also in the startup and shutdown processes modeled in §3.2.

To define the aggregate production cost of intermediate units, we start from an individual unit. Let κ denote an intermediate unit's capacity. An intermediate unit normally operates above a minimum generation level $\alpha\kappa$, with $\alpha \in (0, 1)$. The cost of producing $q \in [\alpha\kappa, \kappa]$ units of energy in a period is denoted as $c(q)$, which satisfies the following assumption.

Assumption 1 *(i) $c(q)$ is increasing and convex in q, for $q \in [\alpha\kappa, \kappa]$; (ii) $c(q)/q$ decreases in q, for $q \in [\alpha\kappa, \kappa]$; (iii) $c'(\kappa) < c^P$.*

The convexity in part (i) can be verified in practice and is typically assumed in the literature, e.g., Lu and Shahidehpour (2004) use convex quadratic functions to model the cost of combined-cycle units. Part (ii) assumes that operating at the full load κ is the most efficient and operating at any load below κ results in an increase in the average cost, known as the *part-load penalty*. Part (iii) states that peaking units have a higher marginal cost than the intermediate unit (because $c(q)$ is convex, $c'(\kappa)$ is the highest marginal cost of the intermediate unit).

When the load on an intermediate unit falls below the minimum generation level $\alpha\kappa$, an emergency situation (known as a min-gen event in practice) occurs. We define $c(q) \equiv c(\alpha\kappa)$ for $q \in [0, \alpha\kappa)$, and impose a *min-gen penalty*, p, per unit of output below $\alpha\kappa$. (A min-gen penalty is also used in practical UC and ED programs to approximate the cost associated with equipment damage and unit trips.) Thus, the production cost per period of an intermediate unit can be written as

$$c(q) + (\alpha\kappa - q)^+ p, \qquad \text{for } q \in [0, \kappa]. \tag{1}$$

Because of the part-load and min-gen penalties, operating an intermediate unit at low load is costly and, therefore, during the low-demand periods, it may be desirable to shut down the unit and start it up later. However, cycling an intermediate unit incurs wear and tear costs, and the startup process consumes extra fuel to warm up the turbine. These costs are referred to as *cycling cost*, denoted as $c^s\kappa$ per cycle, where c^s is the cycling cost per unit of capacity per cycle.

We assume the system has many identical intermediate units with a total capacity of K^I, and derive the aggregate production cost. We define *dispatchable capacity* as the capacity that can be instantly used to produce energy. When n intermediate units are up and running, their capacity $K = n\kappa$ is dispatchable, whereas $K^I - K$ is not dispatchable because intermediate units cannot be instantly started up. To produce a given output $Q^I \le K$ by n units, it is optimal to let all n units be equally loaded, because each unit's cost in (1) is convex in q. Therefore, the minimum cost of producing Q^I in a period with dispatchable capacity K is

$$n\,c(Q^I/n) + n(\alpha\kappa - Q^I/n)^+ p \;\equiv\; C(Q^I, K) + (\alpha K - Q^I)^+ p, \tag{2}$$

where we define $C(Q^I, K) \stackrel{\text{def}}{=} n\,c(Q^I/n)$ for $K = n\kappa$. Because $c(q) \equiv c(\alpha\kappa)$ for $q < \alpha\kappa$, we have $C(Q^I, K) \equiv C(\alpha K, K)$ for $Q^I < \alpha K$. Thus, the total production cost in (2) first decreases and then increases in Q^I, with the minimum at $Q^I = \alpha K$. For analytical convenience, we allow $n = K/\kappa$ to be a positive real number and generalize the definition for $C(Q^I, K)$:

$$C(Q^I, K) \stackrel{\text{def}}{=} \frac{K}{\kappa}\, c\Big(\frac{Q^I}{K}\kappa\Big), \qquad \text{for } Q^I \le K,\; K > 0. \tag{3}$$

The following lemma describes the properties of $C(Q^I, K)$. All proofs are in the online supplement.

Lemma 1 *$C(Q^I, K)$ is increasing in Q^I and K and jointly convex in (Q^I, K).*

The aggregate *part-load penalty* can be expressed as follows. Let $c^I = c(\kappa)/\kappa$ denote the average production cost of the intermediate units operating at full load. From Lemma 1, full-load operation ($Q^I = K$) minimizes the cost of producing Q^I, and part-load operation ($Q^I < K$) leads to inefficiency:

$$\text{Part-load penalty} \;=\; C(Q^I, K) - c^I Q^I, \qquad \text{for } Q^I < K. \tag{4}$$

3.2 Capacity Adjustment

Adjusting the dispatchable intermediate capacity not only incurs cycling cost, but also takes time. Let $K_t \in [0, K^I]$ denote the dispatchable intermediate capacity in period t. We assume that if capacity of size Δ_t^u begins its startup process in period t, then $\gamma^u \Delta_t^u$ becomes dispatchable in period $t+1$, where $\gamma^u \in (0, 1]$ is a constant; the remaining $(1-\gamma^u)\Delta_t^u$ is referred to as *pending-up capacity*. In every following period, a fraction γ^u of the pending-up capacity from the previous period becomes dispatchable. Let R_t^u (R and u stand for "ramp up" or "remain to be up") denote the pending-up capacity in period t before starting up Δ_t^u. Then, R_t^u evolves as follows:

$$R_{t+1}^u = (1-\gamma^u)(R_t^u + \Delta_t^u). \tag{5}$$

Similarly, we assume that if capacity of size Δ_t^d begins to shut down in period t, then $\gamma^d \Delta_t^d$ shuts down in period $t+1$, where $\gamma^d \in (0, 1]$; the remaining $(1-\gamma^d)\Delta_t^d$ is referred to as *pending-down* *capacity* (still dispatchable). In every following period, a fraction γ^d of the pending-down capacity in the previous period shuts down. Let R_t^d denote the pending-down capacity in period t before shutting down Δ_t^d. Then,

$$R_{t+1}^d = (1-\gamma^d)(R_t^d + \Delta_t^d). \tag{6}$$

Note that Δ_t^u and Δ_t^d must be non-negative due to engineering restrictions: Pending-up capacity cannot be shut down, and pending-down capacity cannot be started up. The geometric pattern of the startup process is a reasonable approximation of the actual startup process of combined cycle units described in Henkel et al. (2008). Such an approximation allows us to capture the underlying dynamics of the intermediate units by three state variables instead of a vector of history. Following the dynamics for pending capacities described above, the dispatchable capacity K_t evolves as follows:

$$\begin{aligned} K_{t+1} &= K_t + \gamma^u(R_t^u + \Delta_t^u) - \gamma^d(R_t^d + \Delta_t^d) \\ &= K_{t+1}^o + \gamma^u\Delta_t^u - \gamma^d\Delta_t^d, \end{aligned} \tag{7}$$

where $K_{t+1}^o \stackrel{\text{def}}{=} K_t + \gamma^u R_t^u - \gamma^d R_t^d$ is the dispatchable capacity in period $t+1$ if no new pending capacities are added in period t (i.e., $\Delta_t^u = \Delta_t^d = 0$).

In period t, if the system operator wants to achieve the maximum dispatchable capacity for the next period, it can start up all the remaining non-dispatchable capacity that is not already pending-up, $K^I - K_t - R_t^u$, and the maximum dispatchable capacity in period $t+1$ is

$$K_{t+1}^{\max} \;=\; K_{t+1}^o + \gamma^u(K^I - K_t - R_t^u) \;=\; K_t + \gamma^u(K^I - K_t) - \gamma^d R_t^d. \tag{8}$$

The system operator can initiate the shutdown process on all the dispatchable capacity that is not already pending-down, $K_t - R_t^d$, to achieve the minimum dispatchable capacity for the next period:

$$K_{t+1}^{\min} \;=\; K_{t+1}^o - \gamma^d(K_t - R_t^d) \;=\; (1-\gamma^d)K_t + \gamma^u R_t^u. \tag{9}$$

Because there is no economic reason to initiate startup and shutdown processes at the same time, it is sufficient to use K_{t+1} as a single control variable for capacity, and we have:

$$\Delta_t^u = (K_{t+1} - K_{t+1}^o)^+/\gamma^u, \qquad \Delta_t^d = (K_{t+1}^o - K_{t+1})^+/\gamma^d. \tag{10}$$

Substituting (10) into (5) and (6), we have

$$R_{t+1}^u = (1-\gamma^u)\Big(R_t^u + \frac{(K_{t+1} - K_{t+1}^o)^+}{\gamma^u}\Big), \qquad R_{t+1}^d = (1-\gamma^d)\Big(R_t^d + \frac{(K_{t+1}^o - K_{t+1})^+}{\gamma^d}\Big). \tag{11}$$

In summary, the state of intermediate capacity is described by three variables: dispatchable capacity K_t, pending-up capacity R_t^u, and pending-down capacity R_t^d. In every period t, the system operator decides the dispatchable capacity K_{t+1}, and the pending capacities evolve according to (11).

3.3 System Cost Structure

In an electrical system, when the system operator decides the production for the next 15-minute period, most of the randomness in the demand and wind power for the next period has been resolved. Thus, the demand and wind power can be assumed to be known at the time the production decision is made, and we do not consider the cost related to over- or under-production.

The production cost incurred in period t includes the production cost of the inflexible resource, $c^R Q^R$, and the production cost of the flexible resources (intermediate and peaking units). To produce a total of Q_t in period t using flexible resources, it is best to produce $Q_t \wedge K_t \equiv \min\{Q_t, K_t\}$ by intermediate units and produce $(Q_t - K_t)^+$ by peaking units, because peaking units have a higher marginal production cost than intermediate units. Thus, the total production cost in period t is

$$f(Q_t, K_t) \stackrel{\text{def}}{=} C(Q_t \wedge K_t, K_t) + (\alpha K_t - Q_t)^+ p + (Q_t - K_t)^+ c^P + Q^R c^R, \tag{12}$$

where the first two terms are the production cost of intermediate units. From the discussion after (2), for any given K_t, $f(Q_t, K_t)$ first decreases and then increases in Q_t, with the minimum at $Q = \alpha K$.

For ease of exposition, when starting up intermediate capacity of size $\Delta_{t-1}^u = \frac{(K_t - K_t^o)^+}{\gamma^u}$, we charge the associated cycling cost $\Delta_{t-1}^u c^s$ to period t. Finally, we assume that the curtailment operation itself involves negligible cost (e.g., pitching the blades to curtail wind power simply requires the pitch control to be activated).

3.4 Curtailment Policies for Intermittent Generation

Let $D_t > 0$ denote the demand on flexible resources in period t, which is the total demand minus the inflexible output Q^R. We assume D_t is a deterministic function of a vector of Markovian states, $\mathbf{D}_t$. A simple example is $\mathbf{D}_t = \{t, D_t^r\}$ and $D_t = d(t) + D_t^r$, where the deterministic function $d(t)$ models the predictable component of the demand and D_t^r models the random fluctuations of the demand.

Let $W_t \geq 0$ denote the total wind power in period t if not curtailed. We assume W_t is a deterministic function of a vector of Markovian states, $\mathbf{W}_t$. Elements of $\mathbf{W}_t$ capture predictable and random variations in wind power. We refer to $D_t - W_t$ as the *net demand* on flexible resources.

We study two curtailment policies: the priority dispatch policy and the economic curtailment policy, denoted respectively by superscripts PD and EC in the following analysis.

Under the priority dispatch policy, the system prioritizes wind power over flexible resources.

Wind power is used whenever it can be absorbed by the system; wind power is curtailed only when W_t exceeds D_t. The curtailed energy, denoted as w_t^{PD} (w for 'waste'), is

$$w_t^{\mathrm{PD}} = (W_t - D_t)^+. \tag{13}$$

The production of the flexible resources under the priority dispatch policy is

$$Q_t^{\mathrm{PD}} = D_t - W_t + w_t^{\mathrm{PD}} = (D_t - W_t)^+. \tag{14}$$

Under the economic curtailment policy, the curtailment decision is made jointly with all other decisions to minimize the system operating cost. The minimum production of the flexible resources is Q_t^{PD} and the maximum production is D_t (when all wind power W_t is curtailed):

$$Q_t^{\mathrm{EC}} \in [(D_t - W_t)^+,\ D_t]. \tag{15}$$

The curtailed energy under the economic curtailment policy is

$$w_t^{\mathrm{EC}} \ = \ Q_t^{\mathrm{EC}} + W_t - D_t \ \in \ [(W_t - D_t)^+,\ W_t]. \tag{16}$$

Note that w_t^{PD} defined in (13) is the minimum value of w_t^{EC} in (16).

3.5 Stochastic Dynamic Programs

The problem is formulated as stochastic dynamic programs. In period $t-1$, the system operator observes the capacity state $\mathbf{K}_{t-1} \overset{\mathrm{def}}{=} (K_{t-1}, R_{t-1}^u, R_{t-1}^d)$ and the accurate forecast for the states of demand and wind power in period t, $\mathbf{D}_t$ and $\mathbf{W}_t$ (15-minute ahead forecasts are accurate; see discussions in §3.3). The system operator decides the dispatchable intermediate capacity K_t, which in turn determines the pending capacities R_t^u and R_t^d according to (11):

$$\begin{aligned} R_t^u(K_t, \mathbf{K}_{t-1}) &= (1-\gamma^u)\Big(R_{t-1}^u + \frac{(K_t - K_t^o(\mathbf{K}_{t-1}))^+}{\gamma^u}\Big), \\ R_t^d(K_t, \mathbf{K}_{t-1}) &= (1-\gamma^d)\Big(R_{t-1}^d + \frac{(K_t^o(\mathbf{K}_{t-1}) - K_t)^+}{\gamma^d}\Big), \end{aligned} \tag{17}$$

where $K_t^o(\mathbf{K}_{t-1}) = K_{t-1} + \gamma^u R_{t-1}^u - \gamma^d R_{t-1}^d$. The system operator also decides the flexible generation Q_t according to (14) or in the range given in (15).

Let $\rho \in [0,1]$ be the discount factor. Let V_t^{PD} and V_t^{EC} denote the minimum expected discounted cost from period t onward under the priority dispatch policy and the economic curtailment policy, respectively. The terminal condition is $V_{T+1}^{\mathrm{EC}} = V_{T+1}^{\mathrm{PD}} = 0$. Then, we have, for $t = 0, 1, \ldots, T$,

$$V_t^{\mathrm{PD}}(\mathbf{K}_{t-1},\mathbf{D}_t,\mathbf{W}_t)=\min_{K_t}\big\{f(Q_t,K_t)+\frac{(K_t-K_t^o)^+}{\gamma^u}c^s+\rho\mathsf{E}_t\big[V_{t+1}^{\mathrm{PD}}(\mathbf{K}_t,\mathbf{D}_{t+1},\mathbf{W}_{t+1})\big]\big\}$$
$$s.t.\ K_t\in[K_t^{\min},\ K_t^{\max}],\quad Q_t=(D_t-W_t)^+,\ \text{ and (17)}, \tag{18}$$

$$V_t^{\mathrm{EC}}(\mathbf{K}_{t-1},\mathbf{D}_t,\mathbf{W}_t)=\min_{K_t,Q_t}\big\{f(Q_t,K_t)+\frac{(K_t-K_t^o)^+}{\gamma^u}c^s+\rho\mathsf{E}_t\big[V_{t+1}^{\mathrm{EC}}(\mathbf{K}_t,\mathbf{D}_{t+1},\mathbf{W}_{t+1})\big]\big\}$$
$$s.t.\ K_t\in[K_t^{\min},\ K_t^{\max}],\quad Q_t\in[(D_t-W_t)^+,\ D_t],\ \text{ and (17)}, \tag{19}$$

where $f(Q_t,K_t)$ is defined in (12) and E_t denotes the expectation conditioning on the information about demand and wind power up to period t.

The inflexible units have constant output Q^R, which shifts the demand on flexible resources, D_t, in the above dynamic programs. Prior to $t=0$, the system operator chooses Q^R to minimize the system's expected operating cost:

$$\min_{Q^R\in[0,K^R]} V_0^{\mathrm{PD}}(\mathbf{K}_{-1},\mathbf{D}_0,\mathbf{W}_0;Q^R)\qquad\text{and}\qquad\min_{Q^R\in[0,K^R]} V_0^{\mathrm{EC}}(\mathbf{K}_{-1},\mathbf{D}_0,\mathbf{W}_0;Q^R). \tag{20}$$

For notational simplicity, we suppress Q^R from the arguments in the analysis unrelated to Q^R.

4. Optimal Capacity Adjustment and Economic Curtailment Policy

This section analyzes the structure of the optimal policy for the problem in (19).

4.1 Production and Curtailment under Given Dispatchable Capacity

The system operator chooses both dispatchable intermediate capacity K_t and flexible generation Q_t. In Proposition 1 below, we find the optimal Q_t and the corresponding curtailment w_t under given K_t.

Proposition 1 *Under the economic curtailment policy, for given dispatchable intermediate capacity K_t, the optimal production of flexible resources (intermediate and peaking units) is*

$$Q_t^*(K_t,D_t,W_t)=(D_t-W_t)\vee(\alpha K_t)\wedge D_t, \tag{21}$$

and the corresponding wind power curtailment is $w_t^*=Q_t^*(K_t,D_t,W_t)+W_t-D_t$.

Proposition 1 leads to the following optimal production and curtailment policy:

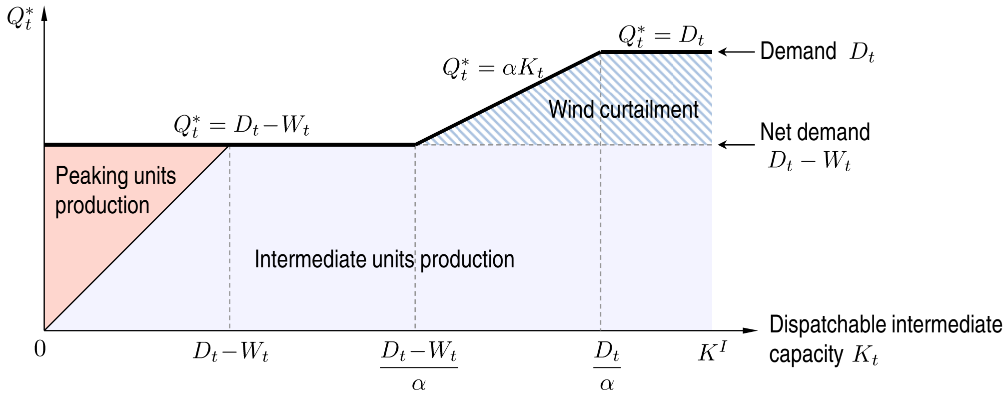
(i) If the net demand on flexible resources is at or above the min-gen level, $D_t-W_t\geq\alpha K_t$, then produce the net demand $Q_t^*=D_t-W_t$ and no wind power is curtailed, $w_t^*=0$.

(ii) If the min-gen level is between the net demand and demand on flexible resources, $D_t-W_t<\alpha K_t<D_t$, then produce $Q_t^*=\alpha K_t$ and partially curtail wind power: $w_t^*=\alpha K_t+W_t-D_t$.

(iii) If the min-gen level is at or above the demand on flexible resources, $\alpha K_t\geq D_t$, then produce the demand $Q_t^*=D_t$ and curtail all wind power: $w_t^*=W_t$.

Figure 1 illustrates the optimal policy under various levels of the dispatchable capacity K_t.

Figure 1: Optimal production Q_t^* as a function of K_t for given D_t and W_t



4.2 Capacity Adjustment

Using the optimal production in (21), we express the optimal production cost of the flexible resources as a function of K_t as follows:

$$f(K_t; D_t, W_t) \stackrel{\text{def}}{=} f\big((D_t - W_t) \vee (\alpha K_t) \wedge D_t,\ K_t\big). \tag{22}$$

Then, the problem in (19) becomes

$$V_t^{\text{EC}}(\mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t) = \min_{K_t} \Big\{ f(K_t; D_t, W_t) + \frac{(K_t - K_t^o)^+}{\gamma^u} c^s + \rho \mathsf{E}_t \big[V_{t+1}^{\text{EC}}(\mathbf{K}_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1}) \big] \Big\} \tag{23}$$
$$s.t.\ K_t \in [K_t^{\min},\ K_t^{\max}],\ \text{and (17)}.$$

The last term in the objective in (23) can be written as

$$U_t(K_t; \mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t) \stackrel{\text{def}}{=} \rho \mathsf{E}_t \big[V_{t+1}^{\text{EC}}(K_t,\ R_t^u(K_t, \mathbf{K}_{t-1}),\ R_t^d(K_t, \mathbf{K}_{t-1}), \mathbf{D}_{t+1}, \mathbf{W}_{t+1}) \big],$$

where R_t^u and R_t^d relate to K_t and $\mathbf{K}_{t-1}$ according to (17).

Lemma 2 *(i) $V_t^{\text{EC}}(\mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ is jointly convex in $\mathbf{K}_{t-1} = (K_{t-1}, R_{t-1}^u, R_{t-1}^d)$ for any $\mathbf{D}_t$ and $\mathbf{W}_t$. (ii) The objective function in (23) is convex in K_t. In particular, $f(K_t; D_t, W_t)$ is convex in K_t, and $\frac{(K_t - K_t^o)^+}{\gamma^u} c^s + U_t(K_t; \mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ is convex in K_t.*

Note that the value function is generally not monotone in $\mathbf{K}_t$ and, therefore, the convexity of the objective (part (ii) of the lemma) is not derived from the composition of convex functions. In fact, $U_t(K_t; \mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ is convex in K_t for $K_t \geq K_t^o$ and $K_t \leq K_t^o$, but may have a kink at $K_t = K_t^o$.

The convexity leads to the following optimal policy structure:

Proposition 2 *The optimal capacity adjustment policy is characterized by two pending capacity* *targets, $y^u(K_{t-1}, R^d_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ and $y^d(K_{t-1}, R^u_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$:*

(i) If the pending-up capacity R^u_{t-1} is below the target $y^u(K_{t-1}, R^d_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$, then bring the pending-up capacity to the target by starting up $y^u(K_{t-1}, R^d_{t-1}, \mathbf{D}_t, \mathbf{W}_t) - R^u_{t-1}$ units of capacity;

(ii) If the pending-down capacity R^d_{t-1} is below the target $y^d(K_{t-1}, R^u_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$, then bring the pending-down capacity to the target by initiating the shutdown process on $y^d(K_{t-1}, R^u_{t-1}, \mathbf{D}_t, \mathbf{W}_t) - R^d_{t-1}$ units of capacity.

(iii) Parts (i) and (ii) cannot occur at the same time.

(iv) When neither (i) nor (ii) occurs, it is optimal not to adjust capacity.

We discuss some intuitive features of the optimal policy and the role of economic curtailment. When demand exhibits daily cycles, the dispatchable intermediate capacity typically goes through four phases every day: an expansion phase in the morning, a constant phase in the middle of the day, a downsizing phase in the evening, and a constant phase at night. Economic curtailment reduces the min-gen penalty at night, allowing more intermediate units to stay dispatchable throughout the night. Thus, the cycling cost is reduced, while the part-load penalty increases.

Limited speed of starting up the intermediate capacity also drives the value of economic curtailment. In many electrical systems, intermediate units are unable to follow the increasing net demand in the morning (i.e., $K_t < D_t - W_t$) and thus $D_t - W_t - K_t$ must be produced by peaking units. Curtailing wind power prior to the morning ramp-up period effectively increases the load on the system, allowing more intermediate units to start up early without incurring the min-gen penalty, thereby reducing the peaking cost during the morning ramp-up period.

5. Capacity Adjustment and Curtailment Policy with Storage Operations

5.1 Model for Energy Storage

Energy storage operations involve energy losses that mostly occur during energy conversions. To be stored, electricity must be converted to other forms of energy (storing operation) and converted back to electricity when needed (releasing operation). The energy storage efficiency, denoted by η, measures the proportion of energy recovered after storing and releasing operations. For example, a hydroelectric pumped storage typically has $\eta =$ 70-80%.

In this paper, "storage level" or "inventory level" refer to the amount of energy that the storage can release until empty. Let $\overline{S}$ denote the maximum storage level and S_t denote the inventory level in period t. Storing one unit of energy raises inventory level by η units.

Storage operations also have speed limits. Let $\underline{\lambda}$ denote the maximum amount of energy that can be released per period, and $\overline{\lambda}$ denote the maximum amount of energy that can be stored per period. We assume that the storage can absorb $\overline{\lambda}$ even when it is full (extra energy is wasted). For example, some hydroelectric pumped storage can take energy while releasing water at the same time. Wasting energy via storage is functionally equivalent to curtailment. Under the economic curtailment policy, we assume that the system does not reduce curtailment only to increase the waste via storage.

To model inventory dynamics, let $x_t < 0$ be the amount of energy released from the storage in period t, and $x_t > 0$ be the amount of energy stored. The range of x_t is

$$x_t \in [-\min\{\underline{\lambda}, S_{t-1}\}, \overline{\lambda}].$$

Engineering constraints (e.g., preset pump speeds) may require x_t to take discrete values. The inventory dynamics are described as follows:

$$S_t = \begin{cases} S_{t-1} + x_t, & \text{if } x_t \le 0, \\ \min[\overline{S},\ S_{t-1} + \eta x_t], & \text{if } x_t > 0. \end{cases} \tag{24}$$

5.2 Problem Formulation with Storage

With storage operations, let V_t^{PS} and V_t^{ES} denote the minimum expected discounted cost from period t onward under the priority dispatch and the economic curtailment policy, respectively. The terminal condition is $V_{T+1}^{\mathrm{ES}} = V_{T+1}^{\mathrm{PS}} = 0$. Then, we have

$$\begin{aligned} V_t^{\mathrm{PS}}(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t) = \min_{K_t, x_t} & \big\{ f(Q_t, K_t) + \frac{(K_t - K_t^o)^+}{\gamma^u} c^s + \rho \mathsf{E}_t \big[V_{t+1}^{\mathrm{PS}}(\mathbf{K}_t, S_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1}) \big] \big\} \\ & s.t.\ K_t \in [K_t^{\min},\ K_t^{\max}], \quad x_t \in [-\min\{\underline{\lambda}, S_{t-1}\}, \overline{\lambda}], \\ & Q_t = (D_t + x_t - W_t)^+,\ (24),\ \text{and } (17). \end{aligned} \tag{25}$$

$$\begin{aligned} V_t^{\mathrm{ES}}(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t) = \min_{K_t, Q_t, x_t} & \big\{ f(Q_t, K_t) + \frac{(K_t - K_t^o)^+}{\gamma^u} c^s + \rho \mathsf{E}_t \big[V_{t+1}^{\mathrm{ES}}(\mathbf{K}_t, S_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1}) \big] \big\} \\ & s.t.\ K_t \in [K_t^{\min},\ K_t^{\max}], \quad x_t \in [-\min\{\underline{\lambda}, S_{t-1}\}, \overline{\lambda}], \\ & Q_t \in [(D_t + x_t - W_t)^+,\ D_t + x_t],\ (24),\ \text{and } (17). \end{aligned} \tag{26}$$

Note that when $\overline{\lambda} = \underline{\lambda} = 0$ (no storage case), (25) and (26) reduce to (18) and (19), respectively.

5.3 Optimal Policy under Economic Curtailment

For given storage flow x_t, the demand on the flexible resources becomes $D_t + x_t$. Following from Proposition 1, the optimal production under given dispatchable capacity K_t is

$$Q_t^*(K_t, D_t + x_t, W_t) = (D_t + x_t - W_t) \vee (\alpha K_t) \wedge (D_t + x_t), \tag{27}$$

and the optimal production cost of the flexible resources is

$$f(K_t; D_t + x_t, W_t) = f\big(Q_t^*(K_t, D_t + x_t, W_t), K_t\big).$$

Thus, the problem in (26) becomes

$$\begin{aligned} V_t^{\mathrm{ES}}(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t) = \min_{K_t, x_t} \Big\{ f(K_t; D_t + x_t, W_t) + \frac{(K_t - K_t^o)^+}{\gamma^u} c^s + \rho \mathsf{E}_t \big[V_{t+1}^{\mathrm{ES}}(\mathbf{K}_t, S_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1}) \big] \Big\} \\ s.t.\ K_t \in [K_t^{\min},\ K_t^{\max}], \quad x_t \in [-\min\{\underline{\lambda}, S_{t-1}\},\ \overline{\lambda}],\ (24),\ \text{and}\ (17). \end{aligned}$$

Proposition 3 *(i) $V_t^{\mathrm{ES}}(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ is decreasing in S_{t-1}.*

(ii) $V_t^{\mathrm{ES}}(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ is jointly convex in $(\mathbf{K}_{t-1}, S_{t-1})$ for any $\mathbf{D}_t$ and $\mathbf{W}_t$.

(iii) For any given state $(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$, we have $V^{\mathrm{PD}} \geq V^{\mathrm{EC}} \geq V^{\mathrm{ES}}$ and $V^{\mathrm{PD}} \geq V^{\mathrm{PS}} \geq V^{\mathrm{ES}}$.

With storage operations, the optimal capacity adjustment policy has a structure similar to that in Proposition 2. If the pending-up capacity is below a target, it is raised to the target; furthermore, the target is independent of the pending-up capacity. A similar structure holds for the pending-down capacity. However, the optimal storage operations cannot be characterized by an inventory-independent target. Proposition 3(iii) formalizes the intuition that storage operations lower the operating cost, and that economic curtailment results in a lower operating cost than the priority dispatch policy. The magnitude of these differences is examined in §6.

6. Numerical Analysis

The goal of our numerical analysis is to explore the implications of our theoretical model and address the series of research questions regarding the impact of economic curtailment. We use data from MISO, but we do not intend to assess the performance of MISO.

6.1 Data and Model Parameters

6.1.1 Load and Wind Power

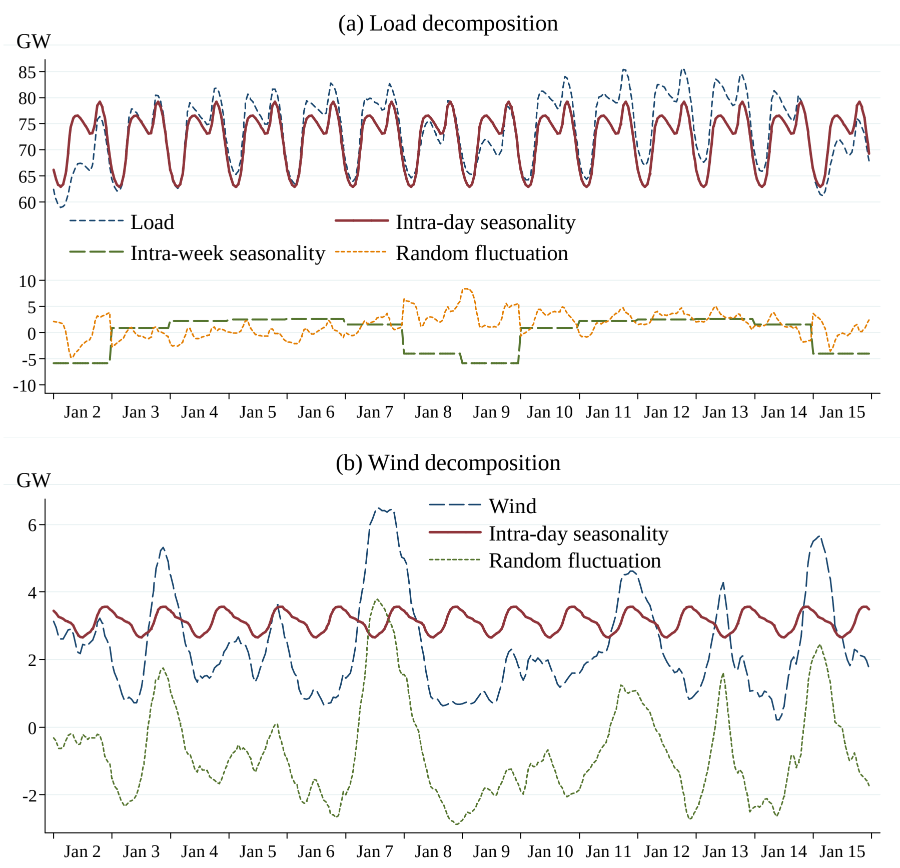
Load and wind power consist of predictable and random components (see §3.4). We model the dynamics of these components in a typical winter season based on the data from MISO, available at https://www.midwestiso.org/Library/MarketReports. Using hourly load data over eight weeks from Jan. 2 to Feb. 26, 2011, and using an ordinary least-squared regression with day-of-week and hour-of-day dummies, we decompose the load into intra-week, intra-day, and random components.

Figure 2(a) shows the first two weeks of load data and its components. The intra-week component captures a weekday-weekend effect and is not considered in our analysis to ensure computational tractability of SDPs. The intra-day component (solid curve in Figure 2(a)) represents the predictable load variability. The random component is modeled as a Markov chain described shortly. Figure 2(b) shows MISO wind power, its intra-day seasonality, and the random component. Note that wind power variations are mostly contributed by the random component.

Because each period in SDP is a 15-minute interval, we linearly interpolate three values between adjacent hourly values of the predictable component. The predictable component of the 15-minute load varies between 15.73 and 19.81 GWh and averages at 18.05 GWh over the eight-week period. The predictable component of 15-minute wind energy varies between 0.664 and 0.891 GWh and averages at 0.781 GWh. The average wind energy penetration is 4.3% (= 0.781/18.05). We observe that evening hours typically have low load but high wind power, in terms of their predictable components.

The random components of load and wind power are generally uncorrelated and we model them independently. We assume the random component of the 15-minute load, denoted as D_t^r, follows an

Figure 2: Load and wind power in MISO's footprint: January 2-15, 2011



AR(1) process, $D^r_{t+1} = \beta_1 D^r_t + \varepsilon_{1t}$, where β_1 is the autoregressive coefficient and ε_{1t} is independent normal random variables with zero mean and standard deviation σ_1. We use the maximum likelihood method to find parameter estimators $\hat{\beta}_1 = 0.988$ and $\hat{\sigma}_1 = 0.170$ GWh. Similarly, the random component of 15-minute wind power is assumed to follow $W^r_{t+1} = \beta_2 W^r_t + \varepsilon_{2t}$, and the parameter estimators are $\hat{\beta}_2 = 0.995$ and $\hat{\sigma}_2 = 0.038$ GWh.

When the average wind penetration increases k times, we assume the predictable component is multiplied by k. The standard deviation of the random component increases k times if the random components of the existing and added wind power are perfectly correlated, or $\sqrt{k}$ times if they are independent. The realistic case is likely in between and we assume that the standard deviation of the random component increases $k^{0.75}$ times.

6.1.2 Conventional Resources

We consider a generation fleet consisting of 40 GW of inflexible capacity provided by nuclear power plants, 25 GW of intermediate capacity provided by either coal or natural gas combined-cycle (NGCC) units, and natural gas peaking capacity that is sufficient to meet the peak demand. Table 1 lists the operating parameters of these resources and energy storage. Table 1 also shows CO_2 and NOx emission rates, which are useful for analyzing the implications for emissions.

The rest of this subsection substantiates the operating parameters listed in Table 1. Cycling cost consists of startup fuel cost and wear-and-tear cost. GE Energy (2012) shows that wear-and-tear cost per cycle is 2 to 7 times as high as the startup fuel cost; we use 4.5 times in our analysis. According to Wu and Bennett (2010), starting up a 1000-MW NGCC unit requires about 10,000 MBtu of energy. At natural gas price \$3.81 per MBtu, this implies a startup fuel cost of \$38.10 per MW. Thus, the cycling cost is $\$38.10 \times (4.5+1) = \$209.55$ per MW per cycle. Starting up a 520-MW coal unit requires about 41,000 MBtu. At coal price \$2.24 per MBtu, this translates into a startup fuel cost of \$176.62 per MW and a cycling cost of \$971.41 per MW per cycle. The above fuel prices are from the 2013 fuel price forecast by Energy Information Administration (2012).

The estimation of the aggregate production cost function is provided in the online supplement. The min-gen level of a unit depends on its technical specifications. For each unit in the MISO market in January 2011, we calculate the ratio of economic minimum output over economic maximum output. We find the median ratio of 224 steam turbines is 0.52. Thus, we assume the min-gen level is 50% of the capacity. We set the min-gen penalty to be \$1000 per MWh for NGCC units and \$2000 per MWh for coal units. Our numerical experiments show that for the penalty values between \$1000

Table 1: Operating cost parameters and emission rates

	Coal intermediate capacity	NGCC intermediate capacity
Cycling cost, c^s ($ per MW per cycle)	971.41	209.55
Aggregate production cost, $C(Q,K)$ ($)	$2.56\frac{Q^2}{K} + 17.55Q + 2.56K$	$9.16\frac{Q^2}{K} + 11.66Q + 9.16K$
Min-gen level as a percentage of dispatchable capacity, α	50%	50%
Min-gen penalty, p ($ per MWh)	2000	1000
Ramp-up rate, γ^u	15%	30%
Ramp-down rate, γ^d	30%	60%
Carbon dioxide (CO_2) emission (lb. per MBtu)	215	117
Nitrogen oxides (NOx) emission (lb. per MBtu)	0.05	0.0073
Nitrogen oxides (NOx) emission during startup (lb. per MBtu)	0.05	0.07

Peaking units emission: same as NGCC	Inflexible (nuclear) units emission: zero
Peaking cost: c^P = $40.20 per MWh	Inflexible (nuclear) units variable cost: c^R = $20 per MWh
Maximum storage level $\overline{S} = 10$ GWh	Round-trip efficiency $\eta = 75\%$
Maximum release rate: $\underline{\lambda} = 2$ GW	Maximum absorbtion rate: $\overline{\lambda} = 2.667$ GW

and $4000 per MWh, the system operating cost under the priority dispatch policy is insensitive to the min-gen penalty, and that min-gen events do not occur under economic curtailment policy for the set of instances we tested.

Ihle (2003) reports that typical 500-MW coal units built in the 1980s to early 1990s have ramp-up rates 3 to 7 MW per minute or, on average, 1% of capacity per minute. Thus, for a 15-minute period, we assume $\gamma^u = 0.15$. The shutdown process is faster and we assume $\gamma^d = 0.3$. NGCC units have faster startup and shutdown processes; we assume $\gamma^u = 0.3$ and $\gamma^d = 0.6$ for NGCC units.

The emission rates (not during startup) in Table 1 are from Black & Veatch Corporation (2012). The CO_2 emission rates are for power plants without carbon capture and sequestration systems, while the NOx emission rates are for power plants with selective catalytic reduction (SCR, a NOx reduction system). Because the SCR system is not fully operative during the startup of a NGCC unit, the NOx emission rate during startup is higher and varies widely across different units. We use a conservative estimate of 0.07 lb. of NOx per MBtu during NGCC startup.

Peaking units (natural gas combustion turbines) have the same emission rates as NGCC units after startup. Although peaking units have a higher NOx emission rate during startup, the startup is very quick (assumed to be instantaneous in this paper), resulting in negligible emissions during startup. The processes of running a nuclear power plant generates no CO_2 or NOx.

At natural gas price \$3.81 per MBtu and heat rate 10.55 MBtu per MWh (Brinkman 2012), peaking cost is c^P = \$40.20 per MWh (= 3.81×10.55). We assume the variable cost of inflexible units is c^R = \$20 per MWh.

The storage we consider is a large hydroelectric pumped storage with a maximum inventory level $\overline{S} = 10$ GWh and round-trip efficiency $\eta = 75\%$. The maximum inventory level change is 0.5 GWh per period. Thus, the maximum amount of energy that can be released every 15-minute period is $\underline{\lambda} = 0.5$ GWh, and the maximum load the storage can create is $\overline{\lambda} = 0.5/\eta = 0.667$ GWh per period.

6.1.3 Discretization and Problem Size

To keep computation time manageable, we discretize the state space of the random load D_t^r into 7 levels, evenly spaced between -2.2 and 2.2 GWh (this range covers four standard deviations of the stationary distribution of D_t^r). At each level, the random load either stays at the same level in the next period or transits to an adjacent level. The transition probabilities are set to match the conditional mean and variance implied by the random load model in §6.1.1. For the random component of the wind power, we choose 13 levels, evenly spaced between -0.664 and 0.664 GWh (this ensures that wind power is always non-negative after the predictable component is added). The transition probabilities are also set to match the conditional mean and variance. As wind penetration increases, this range expands to $-0.664k^{0.75}$ and $0.664k^{0.75}$ GWh, and the transition probabilities are set to reflect the increased variability.

The 25 GW intermediate capacity is discretized into n^I levels: $0, \delta, 2\delta, \ldots, (n^I - 1)\delta = 25$ GW. It can be shown that the capacity vector $\mathbf{K}_t = (K_t, R_t^u, R_t^d)$ can have $n^I(n^I + 1)(n^I + 2)/6$ possible states. In our analysis, we set $n^I = 17$, which gives a total of 969 capacity states.

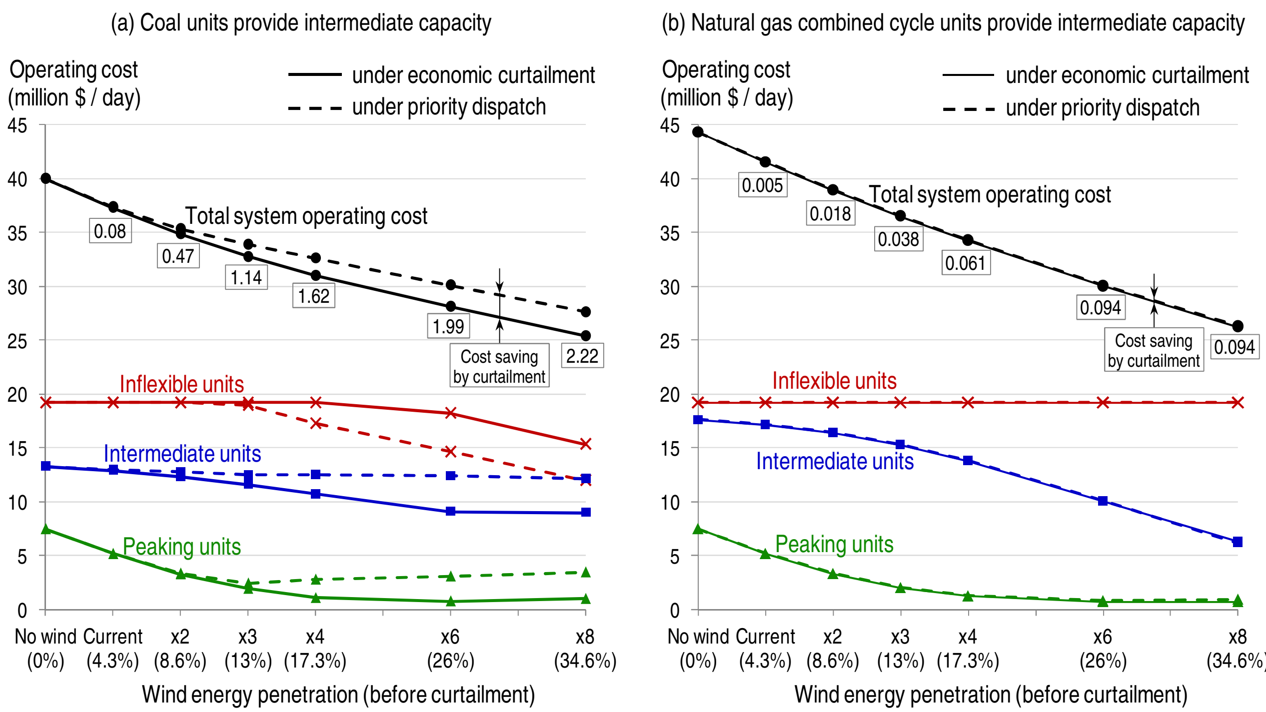
We discretize the 10 GWh of inventory space into 41 levels; each step is 0.25 GWh. We have 96 periods each day; the period index is also a state that determines the predictable components of the wind power and load. In total, the dynamic program is defined on $969 \times 41 \times 7 \times 13 \times 96 = 347.1$ million states. The finite-horizon discounted objective is used in problem formulations for analytical convenience. Setting the discount factor to one and dividing the total cost by the number of periods gives the average cost. Our numerical analysis uses the infinite-horizon average cost criterion, which allows us to measure the system's performance using a single number, the average cost. The algorithm we implement is the value iteration with average cost criteria (Puterman 1994, §8.5).

6.2 Impact of Economic Curtailment on Operating Costs

We first examine the operating cost under priority dispatch policy and coal intermediate capacity, shown as dashed curves in Figure 3(a). The system operating cost declines nonlinearly in wind penetration: From no wind to 4.3% wind penetration, the system operating cost declines at a rate of \$34.8 per MWh of wind energy; from 26% to 34.6% wind penetration, it declines at only \$16.7 per MWh of wind energy. This difference is because at low wind penetration, wind energy mainly displaces expensive peaking output, whereas at medium to high wind penetration, it mostly displaces inflexible generation. The displacement of inflexible generation ensures adequate minimum load on the intermediate units so they can operate without excessive cycling cost or min-gen penalty.

The operating cost under economic curtailment policy is shown as solid curves in Figure 3(a). The system operating cost reduction brought by economic curtailment increases from \$0.08 million per day at the current wind penetration to \$2.22 million per day at 34.6% wind penetration. Under economic curtailment, wind energy displaces more output from intermediate and peaking units than under priority dispatch. This displacement is because economic curtailment reduces the variability facing the system and, therefore, a lower amount of flexible generation is needed and more demand

Figure 3: Operating cost of generation resources: economic curtailment vs. priority dispatch



can be met by the cheaper, inflexible generation.

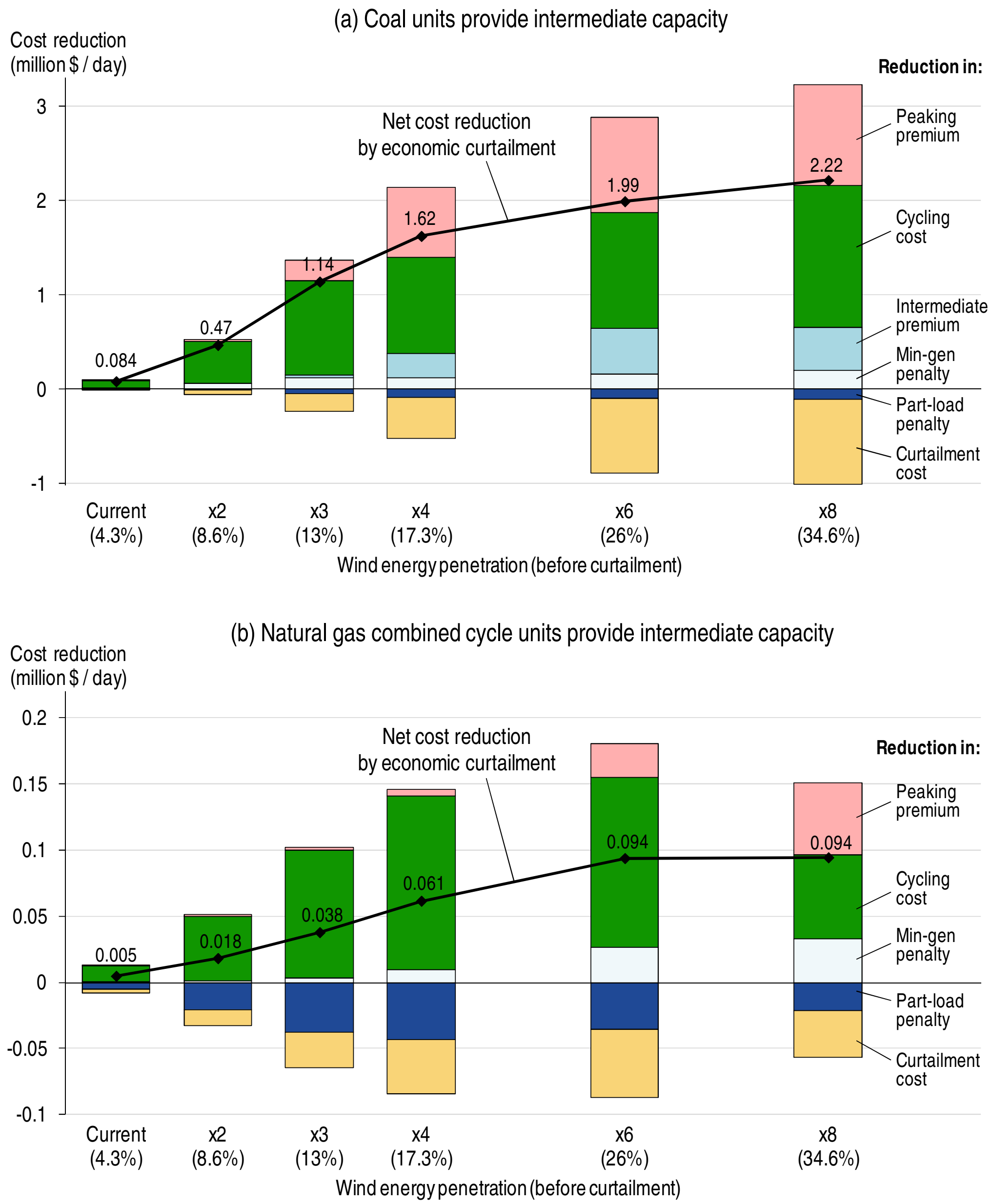
Flexibility of the intermediate capacity is a substitute for economic curtailment. Figure 3(b) shows that if NGCC units replace coal units as intermediate capacity, the system operating cost reduction by economic curtailment is much smaller. Within the wind penetration levels we examined, economic curtailment does not affect inflexible generation and slightly affects the operating costs of NGCC and peaking units.

Next, we identify the drivers of the cost saving of economic curtailment. The change in the system operating cost can be decomposed into six components. The first three are related to the limited flexibility of the intermediate units: cycling cost change, part-load penalty change, and min-gen penalty change (refer to §3.1); the remaining three are related to the change in the sources of energy: $c^R\Delta\overline{Q}^R$, $c^I\Delta\overline{Q}^I$, and $c^P\Delta\overline{Q}^P$, where $\Delta\overline{Q}^R$, $\Delta\overline{Q}^I$, and $\Delta\overline{Q}^P$ denote the changes in the daily average output of the inflexible, intermediate, and peaking capacity, respectively. Let $\Delta\overline{w}$ denote the change in the average wind energy curtailment. Flow balance implies that $\Delta\overline{Q}^I = -\Delta\overline{Q}^P - \Delta\overline{Q}^R + \Delta\overline{w}$. If economic curtailment reduces peaking output ($\Delta\overline{Q}^P < 0$), intermediate output will increase and the net cost change is $(c^P - c^I)\Delta\overline{Q}^P$, referred to as *peaking premium change*. Similarly, if economic curtailment increases inflexible output ($\Delta\overline{Q}^R > 0$), intermediate output will decrease and the net cost change is $(c^R - c^I)\Delta\overline{Q}^R$, referred to as *intermediate premium change*. Economic curtailment ($\Delta\overline{w} > 0$) increases intermediate output, resulting in cost referred to as *wind curtailment cost*. Thus, we can rewrite $c^P\Delta\overline{Q}^P + c^R\Delta\overline{Q}^R + c^I\Delta\overline{Q}^I$ as

$$\begin{aligned} & c^P\Delta\overline{Q}^P + c^R\Delta\overline{Q}^R + c^I\big(-\Delta\overline{Q}^P - \Delta\overline{Q}^R + \Delta\overline{w}\big) \\ = \; & \underbrace{(c^P - c^I)\Delta\overline{Q}^P}_{\text{Peaking premium change}} + \underbrace{(c^R - c^I)\Delta\overline{Q}^R}_{\text{Intermediate premium change}} + \underbrace{c^I\Delta\overline{w}}_{\text{Wind curtailment cost change}} . \end{aligned}$$

The changes in the cost components described above are shown in Figure 4. We see cost savings in four components: peaking premium, cycling cost, intermediate premium, and min-gen penalty, represented by the four stacked bars above the horizontal axis. The stacked bars under the axis correspond to the increase in part-load penalty and wind curtailment cost. For both types of intermediate capacity, the two dominant drivers of the total cost reduction are cycling cost and peaking premium. The cycling cost reduction plays a major role when the wind penetration level is relatively low, whereas the contribution from the peaking premium reduction becomes more prominent as the wind penetration increases. With NGCC intermediate capacity and 34.6% wind penetration, the cycling cost reduction is less significant, because the cycling cost under priority dispatch in this case

Figure 4: Operating cost reductions by economic curtailment



is less significant: It is optimal to displace intermediate generation and fully use inflexible capacity (see Figure 3(b)).

Finally, we examine the amount of curtailment and the cost saving per MWh of economic curtailment, shown in Table 2. Under priority dispatch, wind power is curtailed only when the net demand on flexible resources is negative. Thus, the curtailment is either zero or very small, except for the case of NGCC intermediate capacity and 34.6% wind penetration. Under economic curtailment, up to 7.8% of wind energy may be curtailed. An important result shown in Table 2 is that the cost saving per MWh of curtailed energy is consistently significant. For all the cases we examined, the cost saving is at least \$42 per MWh of curtailed energy. This significant cost saving can be explained by two factors. First, curtailing wind power during an hour may reduce operating costs for several following hours. For many electrical systems, intermediate capacity cannot keep up with the rising demand in the morning, requiring expensive peaking units to fill the gap. Curtailing 1 MW of wind power prior to the morning hours effectively increases the load by 1 MW, allowing 2 MW of inter-

Table 2: Economic curtailment (EC) vs. priority dispatch (PD), without energy storage

Wind energy penetration:	Current (4.3%)	×2 (8.6%)	×3 (13%)	×4 (17.3%)	×6 (26%)	×8 (34.6%)
(a) Coal units provide intermediate capacity						
Curtailment under PD (% of wind power)	0	0	0.001%	0.004%	0.034%	0.11%
Curtailment under EC (% of wind power)	0.29%	1.41%	3.64%	6.38%	7.80%	6.80%
Total cost under PD (mil \$/day)	37.40	35.34	33.89	32.62	30.12	27.61
Total cost under EC (mil \$/day)	37.32	34.87	32.75	31.00	28.13	25.40
Cost saving by EC (mil \$/day)	0.084	0.47	1.14	1.62	1.99	2.22
Cost saving per MWh of EC (\$/MWh)	380.4	222.9	139.3	85.0	57.1	55.2
Total CO_2 emission under PD (10^3 t/day)	636.0	584.9	537.0	542.7	524.9	506.0
Total CO_2 emission under EC (10^3 t/day)	635.7	584.7	532.8	480.6	400.6	398.2
CO_2 reduction by EC (10^3 t/day)	0.28	0.27	4.3	62.2	124.3	107.8
CO_2 reduction per MWh of EC (t/MWh)	1.26	0.13	0.52	3.26	3.56	2.69
Total NOx emission under PD (t/day)	135.6	128.0	119.1	119.6	114.7	109.4
Total NOx emission under EC (t/day)	135.5	128.1	119.3	109.2	91.3	90.2
NOx reduction by EC (t/day)	0.06	-0.08	-0.23	10.4	23.4	19.3
NOx reduction per MWh of EC (kg/MWh)	0.26	-0.036	-0.028	0.55	0.67	0.48
(b) NGCC units provide intermediate capacity						
Curtailment under PD (% of wind power)	0	0	0.003%	0.084%	1.43%	5.53%
Curtailment under EC (% of wind power)	0.13%	0.27%	0.40%	0.54%	1.80%	5.73%
Total cost under PD (mil \$/day)	41.54	38.95	36.56	34.31	30.09	26.33
Total cost under EC (mil \$/day)	41.53	38.93	36.52	34.25	30.00	26.24
Cost saving by EC (mil \$/day)	0.005	0.018	0.038	0.061	0.094	0.094
Cost saving per MWh of EC (\$/MWh)	49.2	45.2	42.4	44.6	55.2	80.6
Total CO_2 emission under PD (10^3 t/day)	311.2	274.1	238.8	205.2	143.5	91.8
Total CO_2 emission under EC (10^3 t/day)	311.3	274.5	239.4	206.0	144.0	91.7
CO_2 reduction by EC (10^3 t/day)	-0.08	-0.32	-0.62	-0.77	-0.53	0.13
CO_2 reduction per MWh of EC (t/MWh)	-0.85	-0.81	-0.70	-0.56	-0.31	0.11
Total NOx emission under PD (t/day)	19.44	17.25	15.27	13.44	9.90	6.58
Total NOx emission under EC (t/day)	19.43	17.20	15.18	13.31	9.76	6.49
NOx reduction by EC (t/day)	0.011	0.046	0.093	0.13	0.14	0.094
NOx reduction per MWh of EC (kg/MWh)	0.12	0.12	0.10	0.094	0.083	0.081

mediate capacity (recall $\alpha = 50\%$) to start earlier, reducing peaking output by 2 MW over several morning hours. Second, significant cycling cost is saved by having an option to curtail wind power. With no curtailment option, when strong wind is likely to occur at night, much of the intermediate capacity needs to be shut down beforehand to reduce the expected min-gen penalty. Curtailing wind power can relieve the system from min-gen penalty in real time and, therefore, some intermediate capacity can be kept up and running overnight.

6.3 Impact of Economic Curtailment on Emissions

We next study the environmental impact of economic curtailment, focusing on CO_2 and NOx emissions from power generation (we do not consider life-cycle emissions of power plants).

The cost changes in Figure 4 are linked to emission changes in the following ways. First, coal units have a cost advantage but an emission disadvantage over natural gas peaking units. Thus, with coal intermediate capacity, peaking premium reduction actually increases emissions; with NGCC intermediate capacity, peaking premium reduction corresponds to emission reduction. Second, fuel consumption during startup leads to emissions; wear-and-tear cost and min-gen penalty do not result in emissions. Third, because inflexible (nuclear) capacity has zero emissions, a reduced intermediate premium translates into emission reductions. Fourth, increased part-load penalty and wind curtailment correspond to increased emissions.

The relative strengths of these four forces determine the net effect on emissions, as reported in Table 2. With coal intermediate units, economic curtailment of wind power reduces CO_2 emission by about 270 tons per day at low wind penetration and more than 100,000 tons per day at high wind penetration. The CO_2 reduction per MWh of economic curtailment ranges from 0.13 to 3.56 tons (cf. a typical coal unit emits about 1 ton of CO_2 per MWh). At high wind penetration, economic curtailment also reduces NOx emission by 0.5 to 0.7 kgs per MWh of curtailment (cf. a typical coal unit emits 0.23 kgs of NOx per MWh), but at wind penetration levels 8.6% and 13%, economic curtailment slightly increases NOx emission by about 0.03 kgs per MWh of curtailment.

In contrast, when NGCC units provide intermediate capacity, economic curtailment results in an increase in CO_2 emission (except for 34.6% wind penetration), because more natural gas is consumed with the increase in part-load penalty and wind curtailment. NGCC units also differ from coal units in that NGCC units emit much more NOx during the startup process than during normal operations (see Table 1). Economic curtailment reduces the cycling of NGCC units, and thus reduces NOx emission for all cases in Table 2(b).

Although the emission impact of economic curtailment is mixed, economic curtailment can be both economically and environmentally beneficial. These benefits are strongest when coal units provide intermediate capacity under high wind penetration: Economic curtailment brings a cost saving of about \$2 million per day, reduces CO_2 emission by over 100,000 tons per day, and reduces NOx emission by about 20 tons per day for the system we consider.

6.4 Interaction of Economic Curtailment and Storage Operations

In this section, we examine how storage affects the operational and environmental impact of economic curtailment. We optimize storage operations jointly with the flexible resources and compute the effects of economic curtailment. To facilitate comparison, we keep the inflexible generation the same as in the case without storage.

With storage operations, economic curtailment brings smaller cost savings, as indicated in Table 3.

Table 3: Economic curtailment (EC) vs. priority dispatch (PD), with energy storage

(a) Coal units provide intermediate capacity

Wind energy penetration:	Current (4.3%)	×2 (8.6%)	×3 (13%)	×4 (17.3%)	×6 (26%)	×8 (34.6%)
Curtailment under PD (% of wind power)	0	0	0	0.001%	0.012%	0.053%
Curtailment under EC (% of wind power)	0.023%	0.45%	1.85%	3.92%	5.42%	4.99%
Total cost under PD (mil \$/day)	37.24	34.90	33.19	31.92	29.34	26.79
Total cost under EC (mil \$/day)	37.23	34.75	32.59	30.81	27.90	25.16
Cost saving by EC (mil \$/day)	0.008	0.16	0.59	1.11	1.44	1.63
Cost saving per MWh of EC (\$/MWh)	461.0	234.9	142.4	94.3	59.0	55.0
$\frac{\text{Cost saving by EC with storage}}{\text{Cost saving by EC without storage}}$	9.6%	33.6%	51.8%	68.3%	72.0%	73.5%
Total CO_2 emission under PD (10^3 t/day)	640.9	589.5	541.7	547.3	529.9	511.8
Total CO_2 emission under EC (10^3 t/day)	640.8	589.3	534.7	478.7	395.1	392.7
CO_2 reduction by EC (10^3 t/day)	0.033	0.15	7.0	68.6	134.8	119.1
CO_2 reduction per MWh of EC (t/MWh)	1.88	0.22	1.69	5.84	5.54	4.02
CO_2 reduction by storage under PD (10^3 t/day)	-4.8	-4.5	-4.7	-4.5	-5.0	-5.8
CO_2 reduction by storage under EC (10^3 t/day)	-5.1	-4.6	-1.9	1.9	5.5	5.5
Total NOx emission under PD (t/day)	137.4	130.0	121.3	121.8	117.2	112.3
Total NOx emission under EC (t/day)	137.4	130.0	120.4	109.2	90.4	89.3
NOx reduction by EC (t/day)	0.007	0.001	0.91	12.6	26.8	23.0
NOx reduction per MWh of EC (kg/MWh)	0.41	0.001	0.22	1.07	1.10	0.78
NOx reduction by storage under PD (t/day)	-1.8	-2.0	-2.3	-2.2	-2.5	-2.9
NOx reduction by storage under EC (t/day)	-1.9	-1.9	-1.1	-0.1	0.9	0.8

(b) NGCC units provide intermediate capacity

Wind energy penetration:	Current (4.3%)	×2 (8.6%)	×3 (13%)	×4 (17.3%)	×6 (26%)	×8 (34.6%)
Curtailment under PD (% of wind power)	0	0	0	0.023%	0.81%	4.37%
Curtailment under EC (% of wind power)	0.011%	0.073%	0.15%	0.25%	1.05%	4.49%
Total cost under PD (mil \$/day)	41.49	38.87	36.44	34.17	29.87	26.03
Total cost under EC (mil \$/day)	41.49	38.87	36.43	34.14	29.82	25.99
Cost saving by EC (mil \$/day)	0.0005	0.004	0.013	0.026	0.047	0.041
Cost saving per MWh of EC (\$/MWh)	55.7	40.2	38.0	38.7	44.6	54.5
$\frac{\text{Cost saving by EC with storage}}{\text{Cost saving by EC without storage}}$	9.9%	24.5%	34.6%	42.8%	50.5%	43.3%
Total CO_2 emission under PD (10^3 t/day)	310.8	273.8	238.7	205.2	142.3	88.7
Total CO_2 emission under EC (10^3 t/day)	310.8	273.9	239.0	205.7	142.9	89.0
CO_2 reduction by EC (10^3 t/day)	-0.010	-0.11	-0.31	-0.52	-0.62	-0.33
CO_2 reduction per MWh of EC (t/MWh)	-1.21	-0.96	-0.89	-0.76	-0.59	-0.43
CO_2 reduction by storage under PD (10^3 t/day)	0.43	0.37	0.14	0.06	1.18	3.09
CO_2 reduction by storage under EC (10^3 t/day)	0.50	0.58	0.45	0.32	1.08	2.64
Total NOx emission under PD (t/day)	19.39	17.14	15.09	13.21	9.63	6.28
Total NOx emission under EC (t/day)	19.39	17.12	15.05	13.14	9.54	6.22
NOx reduction by EC (t/day)	0.001	0.013	0.037	0.068	0.096	0.067
NOx reduction per MWh of EC (kg/MWh)	0.15	0.12	0.11	0.10	0.090	0.089
NOx reduction by storage under PD (t/day)	0.05	0.11	0.18	0.23	0.27	0.30
NOx reduction by storage under EC (t/day)	0.04	0.08	0.12	0.17	0.23	0.27

At 4.3% wind penetration, the storage facility is able to store most of the excessive wind energy that would otherwise be curtailed and, thus, the cost saving by economic curtailment drops to about 10% of the cost saving without storage. As wind penetration increases, the need to balance the extra variability rises. The cost saving by economic curtailment increases to about 70% of the saving without storage under coal intermediate capacity, and to about 50% under NGCC units.

The substitution effect between storage and economic curtailment is expected. However, since storage reduces both the amount of economic curtailment and the corresponding cost saving, it is unclear how the cost saving per MWh of economic curtailment will change. Comparing Tables 3 and 2 indicates that the cost saving per MWh of economic curtailment may be higher when storage is present than if storage is absent—this occurs at most wind penetration levels under coal intermediate capacity, and at low wind penetration levels under NGCC units. Importantly, in all cases, the cost savings per MWh of economic curtailment remain significant in the presence of storage. Furthermore, the emission reduction (if positive) per MWh of economic curtailment in Table 3 is higher than the emission reduction without storage in Table 2.

The environmental impact of storage is also affected by economic curtailment. Storage impacts emissions in three ways: First, storage allows more wind energy to be used and thus reduces emissions. Second, storage reduces the peaking cost while increasing the use of intermediate units, which leads to more (less) emissions under coal (NGCC) intermediate capacity. Third, the energy conversion losses during storage operations increase emissions. The net effect of storage on emissions depends on the relative strengths of these three factors.

Table 3(a) shows that when coal units provide intermediate capacity, under priority dispatch, storage results in a net increase in CO_2 and NOx emissions, because priority dispatch allows most wind power to be used and the first factor above does not contribute much to emission reduction. In contrast, under economic curtailment, the environmental effect of storage changes from negative at low wind penetration to positive at high wind penetration, because storage significantly reduces the amount of curtailed wind power at high wind penetration, lowering the overall emissions.

Table 3(b) shows that when NGCC units provide intermediate capacity, storage operations reduce emissions in all cases, because the first two factors described above dominate the third.

6.5 Deterministic Optimization vs. Stochastic Dynamic Optimization

The analysis in this paper is based on SDP. In this subsection, we examine the value of economic curtailment under deterministic optimization used in practice and illustrate the benefit of increased frequency of deterministic optimization.

The deterministic optimization procedure is performed on a rolling horizon basis and governed by three time-related parameters: $T_P \geq T_I \geq T_L$. The length of the planning-horizon is T_P periods (say t through $t+T_P-1$), but the decisions are implemented only for the first T_I periods (t through $t+T_I-1$), after which the planning-horizon is forwarded by T_I periods. Because the optimization takes time in practice and the results need to be communicated, the optimization for periods t to $t+T_P-1$ is run T_L periods earlier (at time $t-T_L$), based on the information available then.

It is non-trivial to evaluate the average operating cost of this deterministic history-dependent policy. We sweep through all possible deterministic decisions and evaluate the long-run average cost under all stochastic paths of wind and load processes.

We study two deterministic policies, labeled A and B, and compare them to stochastic policy C. Policy A corresponds to the typical day-ahead UC process without intra-day re-optimization: $T_P = T_I = 96$ periods (24 hours) and $T_L = 48$ periods (12 hours). Policy B involves four intra-day re-optimizations, and each optimization uses forecasts two hours ahead of the planning horizon: $T_P = 96$ periods (24 hours), $T_I = 24$ periods (6 hours), and $T_L = 8$ periods (2 hours). To keep computation time manageable, we keep the inflexible generation at the same level as in Figure 3 and do not consider energy storage. Table 4 reports the results.

Table 4 shows that economic curtailment brings more cost savings under deterministic optimization than under SDP. Intuitively, curtailment serves as a recourse of the deterministic policy optimized every T_I periods, but this recourse is not as valuable under SDP because the stochastic policy adjusts decisions in response to the wind and load conditions. Thus, if economic curtailment is implemented with the current deterministic optimization, the cost saving will be higher than that estimated using SDP.

Table 4 also shows that when both economic curtailment and intra-day re-optimization (deterministic policy B) are implemented, the system operating cost is nearly optimal under low wind penetration levels. Furthermore, comparing the cost saving by economic curtailment under A and B reveals that intra-day re-optimization improves the cost saving of economic curtailment at low wind penetration levels. At 26% and 34.6% wind penetration, the gap between policies B and C increases to about $0.3 million per day for both types of intermediate capacity. These results suggest

Table 4: Deterministic optimization vs. stochastic dynamic optimization

All numbers are measured in million dollars per day.

Wind energy penetration:	Current (4.3%)	×2 (8.6%)	×3 (13%)	×4 (17.3%)	×6 (26%)	×8 (34.6%)
(a) Coal units provide intermediate capacity						
Cost under EC and deterministic policy A	38.14	36.70	35.19	33.79	31.01	28.37
Cost under EC and deterministic policy B	37.32	34.87	32.84	31.22	28.41	25.72
Cost under EC and stochastic policy C	37.32	34.87	32.75	31.00	28.13	25.40
Cost saving by EC under A	0.22	1.52	4.18	5.06	8.84	13.09
Cost saving by EC under B	0.36	2.20	4.74	5.40	7.25	9.20
Cost saving by EC under C	0.084	0.47	1.14	1.62	1.99	2.22
(b) NGCC units provide intermediate capacity						
Cost under EC and deterministic policy A	41.61	39.15	36.85	34.73	30.76	27.10
Cost under EC and deterministic policy B	41.53	38.95	36.59	34.39	30.30	26.58
Cost under EC and stochastic policy C	41.53	38.93	36.52	34.25	30.00	26.24
Cost saving by EC under A	0.10	0.55	1.68	3.57	8.09	11.56
Cost saving by EC under B	0.17	0.90	1.86	2.58	3.44	3.94
Cost saving by EC under C	0.005	0.018	0.038	0.061	0.094	0.094

that intra-day re-optimizations, coupled with economic curtailment policy, perform well under the current low wind penetration levels, but more advanced algorithms will be needed to manage more intermittent generation resources in the future.

7. Conclusion

Generating energy from intermittent renewable sources requires no fuel and thus involves negligible marginal production cost and no pollution. Because of these "obvious" economic and environmental benefits, intermittent generation is typically prioritized over fossil generation and is curtailed only for system operating constraints and reliability concerns. This paper shows that, in addition to the physical reasons, there are economic and environmental reasons to curtail intermittent generation, which casts a shadow on the motivation for policies that prioritize intermittent renewable energy.

We evaluate the tradeoffs involved in economic curtailment and identify why it can be beneficial. Curtailing intermittent generation during low-demand periods helps relieve min-gen events and reduce the cycling cost of intermediate units. Curtailment allows intermediate units to start up earlier in the morning, increasing the capacity to meet the rising demand and reducing the peaking cost. In addition, the flexibility provided by economic curtailment may allow the use of more inflexible capacity, thereby reducing the system's operating cost. Our numerical analysis shows that the cost saving per unit of curtailed energy is consistently significant.

Although curtailment reduces the use of clean energy, we find that it also reduces CO_2 emission under coal intermediate capacity. This reduction is driven by reduced cycling of coal units and increased use of inflexible (nuclear) capacity. However, under NGCC intermediate capacity, curtailment increases CO_2 emission, driven by increased fuel consumption due to part-load and curtailment. On the other hand, curtailment lowers NOx emission because a significant amount of NOx emitted during the cycling of NGCC units can be avoided by curtailment. Under coal intermediate capacity, curtailment typically reduces NOx emission, but may slightly increase NOx emission due to the increased use of coal units.

We examine how the benefits of curtailment are affected by the presence of energy storage. When the storage operations are jointly optimized with other resources, the need for curtailment does not disappear. Interestingly, the cost and emission reductions per unit of curtailed energy remain significant and are often higher than without storage.

Finally, we discuss the challenges in implementing economic curtailment. First, although some ISOs began to allow economic curtailment, production-based subsidies effectively grant wind energy priority to a significant extent. This paper indicates that the priority dispatch policy is inefficient from both cost saving and emission reduction perspectives. Thus, the design of subsidies for wind energy should facilitate economic curtailment and avoid unintended consequences of production-based subsidies. The second challenge is to incorporate cost of pollution in the objectives of electrical systems. Our results show that operating cost minimization is not perfectly aligned with emission reduction. If all emissions are priced and reflected in the marginal cost of fossil units, economic curtailment would be most effective. Third, implementing economic curtailment will influence the revenue streams of wind-power producers and other conventional generators. New market settlement schemes would be required to make sure no market participant is adversely affected by a policy that improves the overall system efficiency.

References

Angelus, A., E. L. Porteus 2002. Simultaneous capacity and production management of short-life-cycle, produce-to-stock goods under stochastic demand. *Management Science* **48**(3) 399–413.

Bentek Energy 2010. How less became more: Wind, power and unintended consequences in the Colorado energy market. Available at http://www.bentekenergy.com/WindCoalandGasStudy.aspx.

Black & Veatch Corporation 2012. Cost and performance data for power generation technologies. Prepared for National Renewable Energy Laboratory. Available at http://bv.com/docs/reports-studies/nrel-cost-report.pdf.

Brinkman, G. 2012. Emission impact of fossil fuel unit cycling. In Western Wind and Solar Integration Study Phase 2. Available at http://wind.nrel.gov/public/WWIS/Emissions.pdf.

California ISO 2007. Integration of renewable resources. Available at http://www.caiso.com.

Eberly, J. C., J. A. Van Mieghem 1997. Multi-factor dynamic investment under uncertainty. *Journal of Economic Theory* **75**(2) 345–387.

Ela, E. 2009. Using economics to determine the efficient curtailment of wind energy. National Renewable Energy Laboratory, NREL/TP-550-45071.

Ela, E., D. Edelson 2012. Participation of wind power in LMP-based energy markets. *IEEE Transactions on Sustainable Energy* **3**(4) 777–783.

Ela, E., M. Milligan, B. Parsons, D. Lew, D. Corbus 2009. The evolution of wind power integration studies: Past, present, and future. *Proceedings of Power & Energy Society General Meeting, 2009.*

Energy Information Administration 2012. Short-term energy outlook data file: May12_base.xls.

Available at http://www.eia.gov/forecasts/steo/outlook.cfm.

EnerNex 2006. 2006 Minnesota wind integration study. Prepared for Minnesota Public Utilities Commission.

EnerNex 2011. Eastern wind integration and transmission study. National Renewable Energy Laboratory, NREL/SR-5500-47078. Available at http://www.nrel.gov/wind/publications.html.

European Union 2009. Directive 2009/28/EC of the European Parliament and of the Council on the promotion of the use of energy from renewable sources.

GE Energy 2005. The effects of integrating wind power on transmission system planning, reliability, and operations. Report on phase 2: System performance evaluation. Prepared for the New York State Energy Research and Development Authority.

GE Energy 2010. Western wind and solar integration study. National Renewable Energy Laboratory, NREL/SR-550-47434. Available at http://www.nrel.gov/wind/publications.html.

GE Energy 2012. Analysis of cycling costs in Western Wind and Solar Integration Study. National Renewable Energy Laboratory, NREL/SR-5500-54864.

Gross, R., P. Heptonstall, D. Anderson, T. Green, M. Leach, J. Skea 2006. The costs and impacts of intermittency: An assessment of the evidence on the costs and impacts of intermittent generation on the British electricity network. The UK Energy Research Centre.

Hart, E. K., E. D. Stoutenburg, M. Z. Jacobson 2012. The potential of intermittent renewables to meet electric power demand: Current methods and emerging analytical techniques. *Proceedings of the IEEE* **100**(2) 322–334.

Henkel, N., E. Schmid, E. Gobrecht 2008. Operational flexibility enhancements of combined cycle power plants. Available at http://www.energy.siemens.com.

Ihle, J. 2003. Coal-wind integration. Platts Research & Consulting Renewable Power Service.

Katzenstein, W., J. Apt 2009. Air emissions due to wind and solar power. *Environmental Science and Technology* **43**(2) 253–258.

Lu, B., M. Shahidehpour 2004. Short-term scheduling of combined cycle units. *IEEE Transactions on Power Systems* **19**(3) 1616–1625.

Meibom, P., H. V. Larsen, R. Barth, H. Brand, A. Tuohy, E. Ela 2011. Advanced unit commitment strategies in the United States Eastern Interconnection. National Renewable Energy Laboratory, NREL/SR-5500-49988.

Milligan, M., E. Ela, D. Lew, D. Corbus, Y. Wan, B. Hodge, B. Kirby 2012. Operational analysis and methods for wind integration studies. *IEEE Transactions on Sustainable Energy* **3**(4) 612–619.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* John Wiley & Sons. New York, NY.

Rocklin, S. M., A. Kashper, G. C. Varvaloucas 1984. Capacity expansion/contraction of a facility with demand augmentation dynamics. *Operations Research* **32**(1) 133–147.

Smith, J. C., M. Milligan, E. A. DeMeo, B. Parsons 2007. Utility wind integration and operating impact state of the art. *IEEE Transactions on Power Systems* **22**(3) 900–908.

Tuohy, A., P. Meibom, E. Denny, M. O'Malley 2009. Unit commitment for systems with significant wind penetration. *IEEE Transactions on Power Systems* **24**(2) 592–601.

Weber, C., P. Meibom, R. Barth, H. Brand 2009. WILMAR: A stochastic programming tool to analyze the large-scale integration of wind energy. *In* J. Kallrath, P. M. Pardalos, S. Rebennack and M. Scheidt eds. *Optimization in the Energy Industry.* Springer-Verlag. Chapter 19.

Wu, O. Q., N. D. Bennett 2010. Sustainability at Detroit Edison: Using natural gas as a transition fuel. Case 1-429-143, William Davidson Institute, University of Michigan.

Xcel Energy 2011. Wind induced coal plant cycling costs and the implications of wind curtailment for public service company of Colorado. Available at http://www.xcelenergy.com.

Xie, L., P. M. S. Carvalho, L. A. F. M. Ferreira, J. Liu, B. H. Krogh, N. Popli, M. D. Ilić 2011. Wind integration in power systems: Operational challenges and possible solutions. *Proceedings of the IEEE* **99**(1) 214–232.

Online Supplement for "Curtailing Intermittent Generation in Electrical Systems"

Owen Q. Wu and Roman Kapuscinski

A. Estimation of Aggregate Production Cost Functions in Table 1

The average production cost of an individual unit, $c(q)/q$, can be calculated as average heat rate (MBtu/MWh) $\times$ fuel price (\$/MBtu), where the average heat rate measures the amount of energy input needed to generate one MWh of electricity. Lew et al. (2011) illustrates that the average heat rate is typically convex and downward sloping in the output level, and the slope is close to zero when the output level is near full capacity. We assume that the production cost of an individual unit with capacity κ can be approximated by a quadratic function:

$$c(q) = aq^2 + bq + c, \qquad \text{with } a > 0,\ c = a\kappa^2. \tag{A.1}$$

The average cost, $\overline{c}(q) \equiv \frac{c(q)}{q} = aq + b + \frac{c}{q}$, is convex and decreasing in q, and $\overline{c}'(\kappa) = a - \frac{c}{\kappa^2} = 0$, which are consistent with the aforementioned properties of the average heat rate. Then, using (A.1) and (3) in the paper, we write the aggregate production cost as

$$C(Q, K) = \frac{K}{\kappa}\Big(a\frac{Q^2\kappa^2}{K^2} + b\frac{Q\kappa}{K} + a\kappa^2\Big) \;=\; a\kappa\frac{Q^2}{K} + bQ + a\kappa K. \tag{A.2}$$

Thus, we need to estimate only parameters $a\kappa$ and b. We provide a method to estimate these two parameters below. For an individual unit, let $\overline{c}_1$ and $\overline{c}_{0.5}$ denote the average production cost (\$/MWh) at 100% and 50% load, respectively. Then,

$$\begin{aligned} \overline{c}(\kappa) &= a\kappa \;+\; b \;+\; \frac{a\kappa^2}{\kappa} \;=\; 2a\kappa + b \;=\; \overline{c}_1, \\ \overline{c}(0.5\kappa) &= 0.5a\kappa + b + \frac{a\kappa^2}{0.5\kappa} \;=\; 2.5a\kappa + b \;=\; \overline{c}_{0.5}. \end{aligned}$$

Solving these two equations gives

$$a\kappa = 2(\overline{c}_{0.5} - \overline{c}_1), \qquad b = 5\overline{c}_1 - 4\overline{c}_{0.5}. \tag{A.3}$$

Brinkman (2012) provides the average heat rate of units by types. For coal units, the average heat rate at full and 50% load is 10.12 MBtu/MWh and 10.69 MBtu/MWh, respectively; for NGCC units, the average heat rate at full and 50% load is 7.87 MBtu/MWh and 9.07 MBtu/MWh, respectively.

Using the fuel price forecast from Energy Information Administration (2012), we have

$$\begin{aligned} &\text{Coal:} && \overline{c}_1 = 2.24 \times 10.12 = \$22.67/\text{MWh}, && \overline{c}_{0.5} = 2.24 \times 10.69 = \$23.95/\text{MWh}, \\ &\text{NGCC:} && \overline{c}_1 = 3.81 \times 7.87 = \$29.98/\text{MWh}, && \overline{c}_{0.5} = 3.81 \times 9.07 = \$34.56/\text{MWh}. \end{aligned}$$

The above parameters determine the values of $a\kappa$ and b in (A.3), which in turn determine the aggregate cost function in (A.2). This gives the aggregate cost functions listed in Table 1.

B. Proofs

Proof of Lemma 1. By the definition in (3) and Assumption 1(i), $C(Q^I, K)$ is increasing in Q^I.

Rewrite (3) as $C(Q^I, K) = \dfrac{c(Q^I\kappa/K)}{Q^I\kappa/K}Q^I$. Because $\dfrac{c(q)}{q}$ decreases in q due to Assumption 1(ii), we see that $C(Q^I, K)$ increases in K.

To show that $C(Q^I, K)$ is jointly convex in (Q^I, K), we arbitrarily choose two points (Q_i, K_i) with $Q_i \in [0, K_i]$, $i = 1, 2$, and we have

$$\begin{aligned}
C(Q_1, K_1) + C(Q_2, K_2) &= \tfrac{K_1}{\kappa} c\big(\tfrac{Q_1}{K_1}\kappa\big) + \tfrac{K_2}{\kappa} c\big(\tfrac{Q_2}{K_2}\kappa\big) \\
&= \tfrac{K_1+K_2}{\kappa}\big[\tfrac{K_1}{K_1+K_2} c\big(\tfrac{Q_1}{K_1}\kappa\big) + \tfrac{K_2}{K_1+K_2} c\big(\tfrac{Q_2}{K_2}\kappa\big)\big] \\
&\geq \tfrac{K_1+K_2}{\kappa} c\big(\tfrac{K_1}{K_1+K_2}\tfrac{Q_1}{K_1}\kappa + \tfrac{K_2}{K_1+K_2}\tfrac{Q_2}{K_2}\kappa\big) \\
&= \tfrac{K_1+K_2}{\kappa} c\big(\tfrac{Q_1+Q_2}{K_1+K_2}\kappa\big) \\
&= 2C\big(\tfrac{Q_1+Q_2}{2}, \tfrac{K_1+K_2}{2}\big),
\end{aligned}$$

where the inequality is due to the convexity of $c(q)$. ∎

Proof of Proposition 1. In the problem in (19), for any given dispatchable capacity K_t, we can find the optimal production of flexible resources by solving

$$\min_{Q_t} \big\{ f(Q_t, K_t) : \ Q_t \in [(D_t - W_t)^+, \ D_t] \big\}.$$

Because $f(Q_t, K_t)$ is decreasing in Q_t for $Q_t \leq \alpha K_t$, and then increasing in Q_t for $Q_t > \alpha K_t$, $f(Q_t, K_t)$ is minimized at αK_t, and the optimal production is to produce αK_t or as close as possible. Therefore, we have

$$Q_t^*(K_t, D_t, W_t) = (D_t - W_t) \vee (\alpha K_t) \wedge D_t.$$

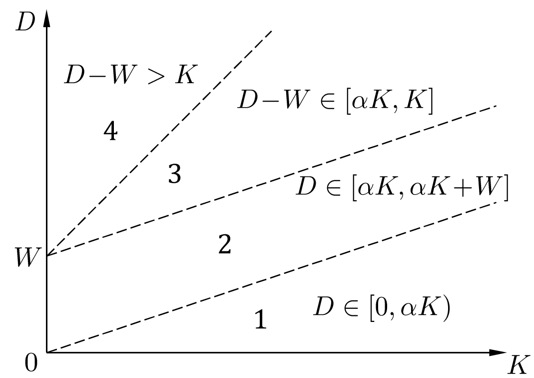
The amount of curtailed wind power is $W_t - (D_t - Q_t^*(K_t, D_t, W_t))$, i.e., the difference between the available wind power W_t and the used wind power $D_t - Q_t^*(K_t, D_t, W_t)$. ∎

Proof of Lemma 2. Lemma 2 (ii) states that $f(K; D, W) = f\big((D - W) \vee (\alpha K) \wedge D, \ K\big)$ defined in (22) is convex in K. Here, we first prove a more general result: $f(K; D, W)$ is jointly convex in (K, D). This result is useful for proving this lemma and other propositions. Using the definition $f(Q, K) = C(Q \wedge K, K) + (\alpha K - Q)^+ p + (Q - K)^+ c^P + Q^R c^R$ in (12), we have

$$f(K; D, W) - Q^R c^R = \begin{cases} C(\alpha K, K) + (\alpha K - D)p, & \text{Region 1: } K \in [\frac{D}{\alpha}, K^I], \\ C(\alpha K, K), & \text{Region 2: } K \in [\frac{D-W}{\alpha}, \frac{D}{\alpha}), \\ C(D - W, K), & \text{Region 3: } K \in [D - W, \frac{D-W}{\alpha}), \\ C(K, K) + (D - W - K)c^P, & \text{Region 4: } K \in [0, D - W). \end{cases} \tag{A.4}$$

The above four regions are illustrated in Figure A.1 below.

Figure A.1: Regions for $f(K, D, W)$



Note that $f(K; D, W)$ is jointly convex in (K, D) within each of the four regions, because $C(Q, K)$ is jointly convex in (Q, K). Thus, we need to prove the convexity across region boundaries.

- Across regions 1 and 2, we can write $f(K; D, W) = C(\alpha K, K) + (\alpha K - D)^+ p$, which is jointly convex in (K, D) because $(\alpha K - D)^+$ is jointly convex in (K, D).
- Across regions 2 and 3 but excluding the area with $D < W$, we can write

$$f(K; D, W) = \max\{C(\alpha K, K),\ C(D - W, K)\},$$

  which is jointly convex in (K, D) because the maximum of two convex functions is convex.
- To prove $f(K; D, W)$ is convex across regions 3 and 4, we define auxiliary functions:

$$\widetilde{C}(Q, K) \stackrel{\text{def}}{=} \frac{K}{\kappa}\widetilde{c}\Big(\frac{Q}{K}\kappa\Big), \qquad \widetilde{c}(q) \stackrel{\text{def}}{=} \begin{cases} c(q), & q \in [0, \kappa], \\ c(\kappa) + (q - \kappa)c^P, & q > \kappa. \end{cases}$$

  Then, $\widetilde{C}(Q, K)$ and $C(Q, K)$ defined in (3) have the following relation:

$$\widetilde{C}(Q, K) = \begin{cases} C(Q, K), & Q \in [0, K], \\ C(K, K) + (Q - K)c^P, & Q > K. \end{cases}$$

  Lemma 1 proves that $C(Q, K)$ is jointly convex due to the convexity of $c(q)$. Following the same lines of proof and the fact that $\widetilde{c}(q)$ is convex under Assumption 1, we see that $\widetilde{C}(Q, K)$ is jointly convex in (Q, K). Because $f(K; D, W) = \widetilde{C}(D - W, K)$ in regions 3 and 4, we conclude that $f(K; D, W)$ is jointly convex in (K, D) across regions 3 and 4.

For any two points (K_1, D_1) and (K_2, D_2), the values of $f(K; D, W)$ on the segment between these two points can be written as $f\big(xK_1 + (1 - x)K_2;\ xD_1 + (1 - x)D_2, W\big) \equiv g(x)$, where $x \in [0, 1]$. The convexity of $f(K; D, W)$ in (K, D) across adjacent regions ensures that $g(x)$ is convex in x for $x \in [0, 1]$. Because the two points are arbitrary, $f(K; D, W)$ is jointly convex in (K, D) across all four regions.

Next, note that the problem in (23) is equivalent to an alternative formulation using startup capacity Δ^u_{t-1} and shutdown capacity Δ^d_{t-1} as decision variables:

$$V^{\mathrm{EC}}_t(\mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t) = \min_{\Delta^u_{t-1}, \Delta^d_{t-1}} \Big\{ f(K_t; D_t, W_t) + \Delta^u_{t-1} c^s + \rho \mathsf{E}_t[V^{\mathrm{EC}}_{t+1}(\mathbf{K}_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})] \Big\} \tag{A.5}$$

$$\Delta^u_{t-1} \in [0, K^I - K_{t-1} - R^u_{t-1}], \qquad \Delta^d_{t-1} \in [0, K_{t-1} - R^d_{t-1}], \tag{A.6}$$

$$K_t = K_{t-1} + \gamma^u (R^u_{t-1} + \Delta^u_{t-1}) - \gamma^d (R^d_{t-1} + \Delta^d_{t-1}), \tag{A.7}$$

$$R^u_t = (1 - \gamma^u)(R^u_{t-1} + \Delta^u_{t-1}), \tag{A.8}$$

$$R^d_t = (1 - \gamma^d)(R^d_{t-1} + \Delta^d_{t-1}). \tag{A.9}$$

We discussed in §3.2 that there is no economic reason to initiate startup and shutdown processes at the same time, and thus, the optimal solution has the property $\Delta^u_t \cdot \Delta^d_t = 0$.[1] Using this property, we can reduce the formulation in (A.5)-(A.9) to that in (23).

We now prove the convexity of the value function by induction. The terminal value function V^{EC}_{t+1} is assumed to be zero. Suppose that $V^{\mathrm{EC}}_{t+1}(\mathbf{K}_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})$ is convex in $\mathbf{K}_t$ for any given $\mathbf{D}_{t+1}$ and $\mathbf{W}_{t+1}$. Then, the objective function in (A.5) is convex in $\mathbf{K}_t$. Equations (A.7)-(A.9) show that $\mathbf{K}_t$ is a linear function of $(\mathbf{K}_{t-1}, \Delta^u_{t-1}, \Delta^d_{t-1})$. Hence, the objective function in (A.5) is jointly convex in $(\mathbf{K}_{t-1}, \Delta^u_{t-1}, \Delta^d_{t-1})$ on a closed convex set defined as

$$\big\{(\mathbf{K}_{t-1}, \Delta^u_{t-1}, \Delta^d_{t-1}) : K_{t-1} \in [0, K^I],\ R^u_{t-1} \in [0, K^I - K_{t-1}],\ R^d_{t-1} \in [0, K_{t-1}],\ \text{and (A.6)}\big\}.$$

By the theorem on convexity preservation under minimization (Heyman and Sobel 1984, p. 525), we conclude that $V^{\mathrm{EC}}_t(\mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ is convex in $\mathbf{K}_{t-1}$.

---

[1] A formal proof for $\Delta^u_t \Delta^d_t = 0$ is as follows. For any feasible policy with $\Delta^u_t > 0$ and $\Delta^d_t > 0$, we consider a revised policy that differs only in the decisions in period t and period $t+1$:

$$\widetilde{\Delta}^u_t = \Delta^u_t - \varepsilon/\gamma^u, \qquad \widetilde{\Delta}^d_t = \Delta^d_t - \varepsilon/\gamma^d, \qquad \text{with } \varepsilon \in (0, \min\{\gamma^u \Delta^u_t, \gamma^d \Delta^d_t\}), \tag{A.10}$$

$$\widetilde{\Delta}^u_{t+1} = \Delta^u_{t+1} + \frac{1-\gamma^u}{\gamma^u}\varepsilon, \qquad \widetilde{\Delta}^d_{t+1} = \Delta^d_{t+1} + \frac{1-\gamma^d}{\gamma^d}\varepsilon. \tag{A.11}$$

Under the above new policy, with the decisions in (A.10), the dispatchable capacity in period $t+1$ does not change compared to the original policy, while the pending capacities decline:

$$\begin{aligned}
\widetilde{K}_{t+1} &= K_t + \gamma^u(R^u_t + \Delta^u_t) - \varepsilon - \gamma^d(R^d_t + \Delta^d_t) + \varepsilon \ = \ K_{t+1}, \\
\widetilde{R}^u_{t+1} &= (1-\gamma^u)(R^u_t + \widetilde{\Delta}^u_t) = R^u_{t+1} - \tfrac{1-\gamma^u}{\gamma^u}\varepsilon < R^u_{t+1}, \\
\widetilde{R}^d_{t+1} &= (1-\gamma^d)(R^d_t + \widetilde{\Delta}^d_t) = R^d_{t+1} - \tfrac{1-\gamma^d}{\gamma^d}\varepsilon < R^d_{t+1}.
\end{aligned}$$

Furthermore, in period $t+2$, we have $\widetilde{\mathbf{K}}_{t+2} = \mathbf{K}_{t+2}$, because the decisions in (A.11) make the pending capacities equal to those in the original policy:

$$\widetilde{R}^u_{t+1} + \widetilde{\Delta}^u_{t+1} = R^u_{t+1} + \Delta^u_{t+1}, \qquad \widetilde{R}^d_{t+1} + \widetilde{\Delta}^d_{t+1} = R^d_{t+1} + \Delta^d_{t+1}.$$

The capacity dynamics of the new policy after $t+1$ are the same as the original policy.

Because the dispatchable capacity process under the two policies are the same, the production costs are the same. The capacity adjustment costs are different: The new policy saves $c^s \varepsilon/\gamma^u$ in period t but incurs an extra cost of $c^s \frac{1-\gamma^u}{\gamma^u}\varepsilon$ in period $t+1$. The net saving is $c^s \varepsilon(\frac{1}{\gamma^u} - \rho\frac{1-\gamma^u}{\gamma^u}) > 0$. Thus, the new policy is strictly better than the original policy. As long as $\Delta^u_t \Delta^d_t > 0$, we can make strict improvement, and thus, the optimal policy must have $\Delta^u_t \Delta^d_t = 0$.

Finally, we prove that the objective in (23) is convex in K_t. We write

$$V_t^{\mathrm{EC}}(\mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t) = \min_{K_t} \big\{ f(K_t; D_t, W_t) + F_t(K_t; \mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t) : K_t \in [K_t^{\min},\ K_t^{\max}] \big\}, \quad \text{(A.12)}$$

where $F_t(K_t; \mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ can be written as

$$F_t(K_t; \mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t) = \min_{\Delta_{t-1}^u} \Big\{ \Delta_{t-1}^u c^s + \rho \mathsf{E}_t[V_{t+1}^{\mathrm{EC}}(K_t, R_t^u, R_t^d, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})] \Big\} \quad \text{(A.13)}$$

$$\Delta_{t-1}^u \in [0, K^I - K_{t-1} - R_{t-1}^u], \quad \text{(A.14)}$$

$$\Delta_{t-1}^u \geq \frac{K_t - K_t^o}{\gamma^u}, \qquad \Delta_{t-1}^u \leq \frac{K_t - K_t^{\min}}{\gamma^u}, \quad \text{(A.15)}$$

$$R_t^u = (1 - \gamma^u)(R_{t-1}^u + \Delta_{t-1}^u), \quad \text{(A.16)}$$

$$R_t^d = (1 - \gamma^d)\frac{1}{\gamma^d}(K_{t-1} + \gamma^u(R_{t-1}^u + \Delta_{t-1}^u) - K_t), \quad \text{(A.17)}$$

where the inequalities in (A.15) are derived from the constraint $\Delta_{t-1}^d \in [0, K_{t-1} - R_{t-1}^d]$, and (A.17) follows from (A.7) and (A.9).

Because $V_{t+1}^{\mathrm{EC}}(\mathbf{K}_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})$ is convex in $\mathbf{K}_t = (K_t, R_t^u, R_t^d)$ for any given $\mathbf{D}_{t+1}$ and $\mathbf{W}_{t+1}$, and because (A.16)-(A.17) express linear relations between (R_t^u, R_t^d) and (K_t, Δ_{t-1}^u), the objective function in (A.13) is jointly convex in (K_t, Δ_{t-1}^u) on a closed convex set defined by $K_t \in [K_t^{\min},\ K_t^{\max}]$, (A.14) and (A.15). Therefore, $F_t(K_t; \mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ is convex in K_t (Heyman and Sobel 1984).

Using the property of $\Delta_{t-1}^u \Delta_{t-1}^d = 0$, we know that the minimizer of (A.13) is $\Delta_{t-1}^u = \frac{(K_t - K_t^o)^+}{\gamma^u}$. Thus, $F_t(K_t; \mathbf{K}_{t-1}, \mathbf{D}_t, \mathbf{W}_t) = \frac{(K_t - K_t^o)^+}{\gamma^u} c^s + \rho \mathsf{E}_t\big[V_{t+1}^{\mathrm{EC}}(K_t,\ R_t^u(K_t, \mathbf{K}_{t-1}),\ R_t^d(K_t, \mathbf{K}_{t-1}), \mathbf{D}_{t+1}, \mathbf{W}_{t+1})\big]$ is convex in K_t. ∎

Proof of Proposition 2. Consider minimizing the objective in (23) over two separate regions: $K_t \in [K_t^{\min}, K_t^o]$ and $K_t \in [K_t^o,\ K_t^{\max}]$. The corresponding problems are:

$$\min_{K_t \in [K_t^{\min}, K_t^o]} f(K_t; D_t, W_t) + \rho \mathsf{E}_t\big[V_{t+1}^{\mathrm{EC}}(\mathbf{K}_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})\big] \quad \text{(A.18)}$$

$$s.t. \quad R_t^u = (1 - \gamma^u) R_{t-1}^u, \qquad R_t^d = (1 - \gamma^d)\Big(R_{t-1}^d + \frac{K_t^o - K_t}{\gamma^d}\Big),$$

$$\min_{K_t \in [K_t^o,\ K_t^{\max}]} f(K_t; D_t, W_t) + \frac{K_t - K_t^o}{\gamma^u} c^s + \rho \mathsf{E}_t\big[V_{t+1}^{\mathrm{EC}}(\mathbf{K}_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})\big] \quad \text{(A.19)}$$

$$s.t. \quad R_t^d = (1 - \gamma^d) R_{t-1}^d, \qquad R_t^u = (1 - \gamma^u)\Big(R_{t-1}^u + \frac{K_t - K_t^o}{\gamma^u}\Big).$$

For the problems in (A.18) and (A.19), we change the decision variables to the pending-down capacity *after* shutting down Δ_{t-1}^d and the pending-up capacity *after* starting up Δ_{t-1}^u, respectively:

$$y^d = R_{t-1}^d + \Delta_{t-1}^d = R_{t-1}^d + \frac{K_t^o - K_t}{\gamma^d} = \frac{K_{t-1} + \gamma^u R_{t-1}^u - K_t}{\gamma^d},$$

$$y^u = R_{t-1}^u + \Delta_{t-1}^u = R_{t-1}^u + \frac{K_t - K_t^o}{\gamma^u} = \frac{K_t - K_{t-1} + \gamma^d R_{t-1}^d}{\gamma^u}.$$

Then, $K_t \in [K_t^{\min},\ K_t^o]$ is equivalent to $y^d \in [R_{t-1}^d, K_{t-1}]$, and $K_t \in [K_t^o,\ K_t^{\max}]$ is equivalent to

$y^u \in [R^u_{t-1}, K^I - K_{t-1}]$. The problems in (A.18) and (A.19) are equivalent to:

$$\min_{y^d \in [R^d_{t-1}, K_{t-1}]} \quad F(K_{t-1}, R^u_{t-1}, y^d, \mathbf{D}_t, \mathbf{W}_t), \tag{A.20}$$

$$\min_{y^u \in [R^u_{t-1}, K^I - K_{t-1}]} F(K_{t-1}, y^u, R^d_{t-1}, \mathbf{D}_t, \mathbf{W}_t) + (y^u - R^u_{t-1})c^s, \tag{A.21}$$

where,

$$\begin{aligned} F(K_{t-1}, y^u, y^d, \mathbf{D}_t, \mathbf{W}_t) \stackrel{\text{def}}{=} & f\big(K_{t-1} + \gamma^u y^u - \gamma^d y^d;\ D_t, W_t\big) \\ & + \rho \mathsf{E}_t\Big[V^{\text{EC}}_{t+1}\big(K_{t-1} + \gamma^u y^u - \gamma^d y^d,\ (1-\gamma^u) y^u,\ (1-\gamma^d) y^d, \mathbf{D}_{t+1}, \mathbf{W}_{t+1}\big)\Big]. \end{aligned} \tag{A.22}$$

The problems in (A.20) and (A.21) imply that the optimal policy has the following structures:

If the pending-down capacity R^d_{t-1} is below a target level defined as

$$y^d(K_{t-1}, R^u_{t-1}, \mathbf{D}_t, \mathbf{W}_t) \stackrel{\text{def}}{=} \inf \operatorname*{arg\,min}_{y^d \in [0, K_{t-1}]} F(K_{t-1}, R^u_{t-1}, y^d, \mathbf{D}_t, \mathbf{W}_t),$$

then it is optimal to bring the pending-down capacity up to the target.

If the pending-up capacity R^u_{t-1} is below a target level defined as

$$y^u(K_{t-1}, R^d_{t-1}, \mathbf{D}_t, \mathbf{W}_t) \stackrel{\text{def}}{=} \inf \operatorname*{arg\,min}_{y^u \in [0, K^I - K_{t-1}]} F(K_{t-1}, y^u, R^d_{t-1}, \mathbf{D}_t, \mathbf{W}_t) + y^u c^s,$$

then it is optimal to bring the pending-up capacity up to the target.

As shown in Lemma 2 (ii), the objective in (23) is convex in K_t. Thus, if the problem in (A.18) has a minimizer $K^*_t < K^o_t$, then K^*_t minimizes the objective in (23) over the entire feasible region $K_t \in [K^{\min}_t, K^{\max}_t]$. Similarly, when solving (A.19), if a minimizer $K^*_t > K^o_t$ exists, it is also the global minimizer. If neither of the above two situations occur, then $K^*_t = K^o_t$ is the optimal solution, i.e., do not adjust capacity. ∎

Proof of Proposition 3. (i) We prove by induction that $V^{\text{ES}}_t(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ decreases in S_{t-1}. The terminal value function V^{ES}_{T+1} is assumed to be zero. Suppose that $V^{\text{ES}}_{t+1}(\mathbf{K}_t, S_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})$ decreases in S_t. Consider two inventory levels: $S^a_{t-1} < S^b_{t-1}$. Starting from state $(\mathbf{K}_{t-1}, S^a_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$, denote the optimal decision as (K^*_t, x^*_t), the resulting inventory as S^a_t, and the resulting capacity states as $\mathbf{K}^*_t$. If the system starts from $(\mathbf{K}_{t-1}, S^b_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$, the decision (K^*_t, x^*_t) remains feasible because $x^*_t \in [-\min\{\underline{\lambda}, S^a_{t-1}\}, \overline{\lambda}] \subseteq [-\min\{\underline{\lambda}, S^b_{t-1}\}, \overline{\lambda}]$, and $S^a_t \le S^b_t$ according to (24).We have

$$\begin{aligned} V^{\text{ES}}_t(\mathbf{K}_{t-1}, S^a_{t-1}, \mathbf{D}_t, \mathbf{W}_t) &= f(K^*_t; D_t + x^*_t, W_t) + \frac{(K^*_t - K^{*o}_t)^+}{\gamma^u} c^s + \rho \mathsf{E}_t\big[V^{\text{ES}}_{t+1}(\mathbf{K}^*_t, S^a_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})\big]\big\} \\ &\ge f(K^*_t; D_t + x^*_t, W_t) + \frac{(K^*_t - K^{*o}_t)^+}{\gamma^u} c^s + \rho \mathsf{E}_t\big[V^{\text{ES}}_{t+1}(\mathbf{K}^*_t, S^b_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})\big]\big\} \\ &\ge V^{\text{ES}}_t(\mathbf{K}_{t-1}, S^b_{t-1}, \mathbf{D}_t, \mathbf{W}_t), \end{aligned}$$

where the first inequality is due to the induction assumption and $S^a_t \le S^b_t$, and the last inequality follows from the feasibility of (K^*_t, x^*_t).

(ii) We formulate the problem using alternative decision variables Δ^u_{t-1} and Δ^d_{t-1} as in the proof of Lemma 2. Additionally, we use alternative decision variables x^u_t, x^d_t, and x^w_t, which respectively represent the energy flows for raising inventory, lowering inventory, and being wasted. Then, the problem in (26) can be written as:

$$
\begin{aligned}
V^{\mathrm{ES}}_t(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t) = \min_{\Delta^u_{t-1}, \Delta^d_{t-1}, x^u_t, x^d_t, x^w_t} \Big\{ & f(K_t; D_t + x_t, W_t) + \Delta^u_{t-1} c^s \\
& + \rho \mathsf{E}_t[V^{\mathrm{ES}}_{t+1}(\mathbf{K}_t, S_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})] \Big\}
\end{aligned}
\tag{A.23}
$$

$$
\begin{aligned}
&\Delta^u_{t-1} \in [0, K^I - K_{t-1} - R^u_{t-1}], && \Delta^d_{t-1} \in [0, K_{t-1} - R^d_{t-1}], \\
&x^u_t \in [0, \min\{\overline{\lambda},\ (\overline{S} - S_{t-1})/\eta\}], && x^d_t \in [0, \min\{\underline{\lambda}, S_{t-1}\}], \\
&x^w_t \in [0, (\overline{\lambda} - (\overline{S} - S_{t-1})/\eta)^+], && \\
&x_t = x^u_t - x^d_t + x^w_t, && S_t = S_{t-1} + \eta x^u_t - x^d_t, \\
&R^u_t = (1 - \gamma^u)(R^u_{t-1} + \Delta^u_{t-1}), && R^d_t = (1 - \gamma^d)(R^d_{t-1} + \Delta^d_{t-1}), \\
&K_t = K_{t-1} + \gamma^u(R^u_{t-1} + \Delta^u_{t-1}) - \gamma^d(R^d_{t-1} + \Delta^d_{t-1}). &&
\end{aligned}
$$

The optimal solution has the property that $x^u_t \cdot x^d_t = 0$, because if $x^u_t \cdot x^d_t > 0$, reducing x^u_t and x^d_t by the same amount does not change x_t, but raises S_t, thereby improving the objective due to the monotonicity in part (i). Similarly, the optimal solution has the property that $x^d_t \cdot x^w_t = 0$. These two properties, together with $\Delta^u_{t-1} \cdot \Delta^d_{t-1} = 0$, indicate that the above alternative formulation is equivalent to the original formulation in (26).

In the proof of Lemma 2, we proved $f(K_t; D_t, W_t)$ is jointly convex in (K_t, D_t). Now, we prove the convexity of the value function by induction. The terminal value function V^{ES}_{T+1} is assumed to be zero. Suppose $V^{\mathrm{ES}}_{t+1}(\mathbf{K}_t, S_t, \mathbf{D}_{t+1}, \mathbf{W}_{t+1})$ is jointly convex in $(\mathbf{K}_t, S_t)$ for any given $\mathbf{D}_{t+1}$ and $\mathbf{W}_{t+1}$. Then, because all constraints of the problem in (A.23) are linear in $(\mathbf{K}_{t-1}, S_{t-1}, \Delta^u_{t-1}, \Delta^d_{t-1}, x^u_t, x^d_t, x^w_t)$, the objective function in (A.23) is jointly convex in $(\mathbf{K}_{t-1}, S_{t-1}, \Delta^u_{t-1}, \Delta^d_{t-1}, x^u_t, x^d_t, x^w_t)$ on a closed convex set. Therefore, $V_t(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$ is convex in $(\mathbf{K}_{t-1}, S_{t-1})$.

(iii) For any given state $(\mathbf{K}_{t-1}, S_{t-1}, \mathbf{D}_t, \mathbf{W}_t)$, by definition of the value functions in (18), (19), (25), and (26), we see that $V^{\mathrm{PD}} \geq V^{\mathrm{EC}}$ and $V^{\mathrm{PS}} \geq V^{\mathrm{ES}}$. Furthermore, the optimal policy for (18) is a feasible policy for (25), and the optimal policy for (19) is a feasible policy for (26). Hence, we have $V^{\mathrm{EC}} \geq V^{\mathrm{ES}}$ and $V^{\mathrm{PD}} \geq V^{\mathrm{PS}}$. ∎

Additional References (see also the references in the paper)

Heyman, D. P., M. J. Sobel 1984. *Stochastic Models in Operations Research*. Vol. 2. McGraw-Hill, New York.

Lew, D., G. Brinkman, M. Milligan, S. Lefton, D. Piwko 2011. How does wind affect coal? Cycling, emissions, and costs. National Renewable Energy Laboratory, NREL/PR-5500-51579.