Grid Integration of Aggregated Demand Response, Part 2: Modeling Demand Response in a Production Cost Model

Marissa Hummon, David Palchak,
Paul Denholm, and Jennie Jorgenson
National Renewable Energy Laboratory

Daniel J. Olsen, Sila Kiliccote, Nance Matson, Michael Sohn, Cody Rose, Junqiao Dudley, and Sasank Goli
Lawrence Berkeley National Laboratory

Ookie Ma
U.S. Department of Energy

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Technical Report
NREL/TP-6A20-58492
December 2013

Contract No. DE-AC36-08GO28308

NATIONAL RENEWABLE ENERGY LABORATORY

Grid Integration of Aggregated Demand Response, Part 2: Modeling Demand Response in a Production Cost Model

Marissa Hummon, David Palchak,
Paul Denholm, and Jennie Jorgenson
National Renewable Energy Laboratory

Daniel J. Olsen, Sila Kiliccote, Nance Matson,
Michael Sohn, Cody Rose, Junqiao Dudley,
and Sasank Goli
Lawrence Berkeley National Laboratory

Ookie Ma
U.S. Department of Energy

Prepared under Task No. SA12.0384

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

Technical Report
NREL/TP-6A20-58492
December 2013

Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

Foreword

This report is one of a series stemming from the U.S. Department of Energy (DOE) Demand Response and Energy Storage Integration Study. This study is a multi-national-laboratory effort to assess the potential value of demand response (DR) and energy storage to electricity systems with different penetration levels of variable renewable resources and to improve our understanding of associated markets and institutions. This study was originated, sponsored, and managed jointly by the Office of Energy Efficiency and Renewable Energy and the Office of Electricity Delivery and Energy Reliability.

Grid modernization and technological advances are enabling resources, such as DR and energy storage, to support a wider array of electric power system operations. Historically, thermal generators and hydropower in combination with transmission and distribution assets have been adequate to serve customer loads reliably and with sufficient power quality, even as variable renewable generation like wind and solar power become a larger part of the national energy supply. While DR and energy storage can serve as alternatives or complements to traditional power system assets in some applications, their values are not entirely clear. This study seeks to address the extent to which DR and energy storage can provide cost-effective benefits to the grid and to highlight institutions and market rules that facilitate their use.

The project was initiated and informed by the results of two DOE workshops; one on energy storage and the other on DR. The workshops were attended by members of the electric power industry, researchers, and policymakers, and the study design and goals reflect their contributions to the collective thinking of the project team. Additional information and the full series of reports can be found at www.eere.energy.gov/analysis/.

Acknowledgments

The authors would like to thank the following individuals for their valuable input and comments during the analysis and publication process: Mark O'Malley, Brendan Kirby, Elaine Hale, Trieu Mai, Nate Blair, Keith Searight, Erik Ela, Chunlian Jin, Di Wu, and Michael Kintner-Meyer. Any errors or omissions are solely the responsibility of the authors.

Executive Summary

Introduction

Renewable integration studies have evaluated many challenges associated with deploying large amounts of variable wind and solar generation technologies. These studies can evaluate operational impacts associated with variable generation, benefits of improved wind and solar resource forecasting, and trade-offs between institutional changes, including increasing balancing area cooperation and technical changes such as installing new flexible generation. Demand response (DR) resources present a potentially important source of grid flexibility and can aid in integrating variable generation; however, integration analyses have not yet incorporated these resources explicitly into grid simulation models as part of a standard toolkit for resource planners.

Part 1 of this report, "Load Availability Profiles and Constraints for the Western Interconnection," examines the potential for different types of commercial and residential building, municipal, and industrial non-manufacturing loads to provide bulk power system services [1]. Industrial manufacturing loads were examined and quantified by Starke et al. [2]. These include energy, capacity, and operating reserves. Their abilities to provide energy and operating reserve services are represented by a set of hourly availabilities for each service, including some additional resource constraints specific to individual end-use types. The services are defined in Table ES-1. Historically, applications for DR have been limited and used primarily for emergencies and peak shaving. As such, their capabilities are often assessed only for a few select times of the year. This assessment quantifies the DR resource potential throughout the year and can thereby be used in scenarios with substantial deviations of electricity supply from historical norms, such as cases with high levels of wind and solar generation.

Table ES-1. Description and Physical Requirements of the Products Modeled in This Study

Reserves Products		Physical Requirements			
Product Type	**General Description**	**How Fast to Respond**	**Length of Response**	**Time to Fully Respond**	**How Often Called**
Energy	Shed or shift energy consumption over time	5 minutes	≥1 hour	10 minutes	1–2 times per day with 4–8 hour notification
Regulation	Response to random unscheduled deviations in scheduled net load	30 seconds	Energy neutral in 15 minutes	5 minutes	Continuous within specified bid period
Flexibility	Additional load following reserve for large un-forecasted wind/solar ramps	5 minutes	1 hour	20 minutes	Continuous within specified bid period
Contingency	Rapid and immediate response to a loss in supply	1 minute	≤30 minutes	≤10 minutes	≤ Once per day

This report (Part 2) implements DR resources in the commercial production cost model PLEXOS. Production cost models are utility planning tools commonly used in renewable integration analyses because they can mimic many of the near real-time decisions and conditions faced by power system operators. Further, they output a number of useful metrics, such as estimates of different types of operational costs and power plant emissions.

The DR resource assessment detailed in Part 1 has been conducted in a large area across the western United States. In order to explore the methodology, we model a subset of the western United States, the Colorado test system. The Colorado system has been used in several previous studies on the performance and value of storage [3] and concentrating solar power [4]. The work described here demonstrates an approach to modeling DR in production cost models and how its availability to power system operators can reduce operational costs and improve generator utilization.

Approach

Our approach assumes that DR providers submit their capabilities to power system operators, and subsequently are dispatched to minimize the overall cost of generation in a manner similar to conventional generators but including the various operational constraints of the individual DR resources. It also allows DR resources to provide operating reserves in addition to energy, and thereby be fully co-optimized alongside generators in providing bulk power system services.

We assume that DR can provide energy services by either shedding load or shifting its use between different times. In addition to energy, the modeling considers DR resources capable of providing operating reserves. Operating reserves include frequency regulation, contingency reserve, and flexibility (or ramping) reserve. Frequency regulation manages short-term variations in demand due to unpredicted changes in the net of load, wind, and solar generation. Contingency reserve addresses large power plant or transmission line failures. Finally, flexibility reserve is a proposed service that responds to large and unexpected wind and solar ramp events.

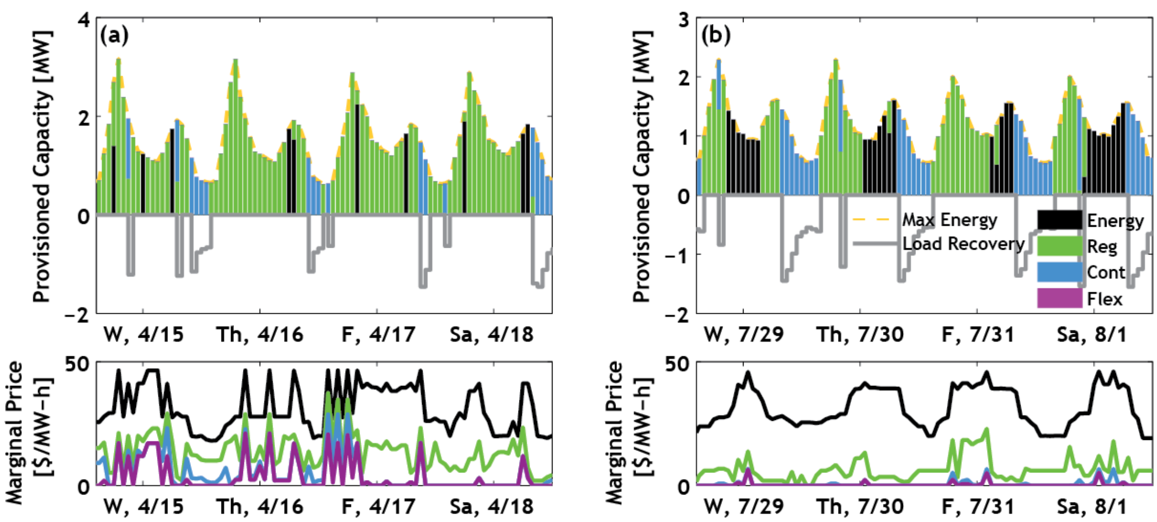

Figure ES-1. Residential water heating DR in Public Service of Colorado (upper): Hourly capacity provision for (a) spring and (b) summer; hourly marginal price of energy and operating reserves (lower)

The production cost model runs a chronological hourly simulation of the power system to assess the operational value of DR. Calculating operational value is approached two ways. First, we compare the total costs of production between scenarios. By making incremental changes to the system, changes to production costs can be attributed to the addition of new resources or differences in operational practices. Changes in production costs represent the total operational value to the system as a whole, irrespective of how value is distributed to different entities. Second, we examine the short-run marginal costs of production. In regions served by an independent system operator or regional transmission organization, marginal costs equate to market clearing prices and indicate the expected revenue that resources would earn as market participants. Outside these regions, marginal costs relate to a vertically integrated utility's avoided cost for providing the associated service.

Results

There are significant variations in the availabilities of different types of DR resources in the Colorado test system. Table ES-2 shows that the average availability for energy and operating reserves as well as the energy available through load shedding during the top 20 load hours (which is a rough estimate for the DR capacity value). The assumed DR resource can provide up to 113 MW of capacity (roughly 2% of peak load) and shift 135 GWh of energy. It can also meet about 33% of the need for frequency regulation, 19% of spinning contingency reserve, and 85% of flexibility reserve. These numbers represent the annual average response that could be achieved for individual grid services but does not include co-optimization. For DR resources that can provide multiple grid services, providing one grid service incurs a lost opportunity for providing another grid service, which is accounted for by constraining the simulation of DR in the production cost model.

Table ES-2. Availability of Demand Response to Provide Energy, Operating Reserves (Regulation, Contingency, and Flexibility) in the Colorado Test System, and Capacity During the Top 20 Hours of Greatest Demand

Demand Response Resource	Energy		Operating Reserves						Capacity
			Mean [MW]			% of Requirement			
	Mean [MW]	Max [GWh]	Reg	Cont	Flex	Reg	Cont	Flex	Mean [MW]
Residential Cooling	10.9	38.8	10.3	10.3	10.3	8.8	2.5	23.9	55.1
Residential Water Heating	1.8	15.7	1.8	1.8	1.8	1.5	0.4	3.4	1.3
Commercial Cooling	2.1	10.7	0.1	1.6	0.2	0.1	0.4	0.6	8.2
Commercial Heating	2.2	3.8	-	-	-	-	-	-	-
Commercial Lighting	-	-	1	3	3	0.8	0.7	5.8	-
Commercial Ventilation	-	-	1.1	3.4	3.4	0.9	0.8	6.6	-
Municipal Pumping	1.7	2.1	-	-	-	-	-	-	2.1
Wastewater Pumping	1.5	1.6	-	-	-	-	-	-	1.5
Outdoor Lighting	1.7	-	23.4	23.4	23.4	20.7	5.8	44.2	-
Refrigerated Warehouses	0.2	0.3	-	-	-	-	-	-	0.3
Agricultural Pumping	17	49.9	0	25.6	0	0	6.3	0	36.6
Data Centers	8	11.7	0	8	0	0	2	0	8
All DR Resources	**45.4**	**49.9**	**37.7**	**77**	**42.1**	**32.9**	**19**	**84.5**	**113**

Implementing these resources into the Colorado test system leads to $7.9 million in total operational savings. This is higher than the $5.4 million in total revenue based on what the resources could earn in a market setting. The difference stems largely from price-suppression effects, as entry of DR into the market results in lower overall market-clearing prices. Most of the revenue comes from the provision of operating reserves, with only 21% coming from energy transactions. This is partly due to the fact that energy provision entails a number of constraints related to the timing of load recovery (which can occur prior to or following a curtailment). For instance, thermal loads must maintain temperature tolerances, limiting the time difference between a load shed and a load recovery and limiting energy price arbitrage value.

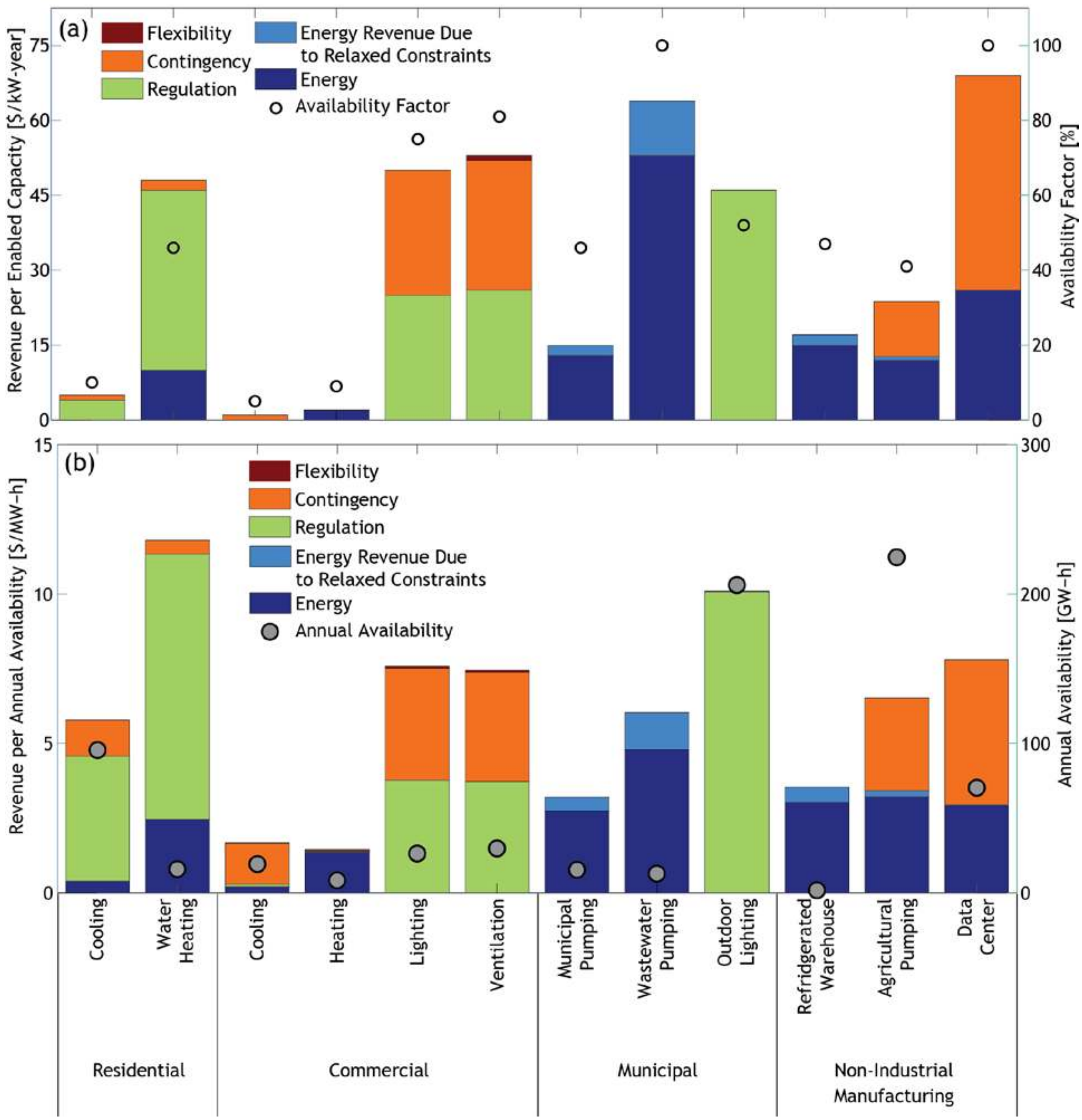

Figure ES-2. Average annual revenue (left axis) from the day-ahead market per (a) total enabled capacity and (b) annual availability for each type of demand response resource in the Colorado test system. Annual DR resource availability (right axis) is expressed as the (a) annual availability factor or as (b) total annual availability

The value of these resources is driven by their availabilities and the coincidence of their availability to times of high production costs. Some resources have a nearly constant available response capacity, while those of others are highly seasonal. For instance, residential cooling availability, during peak demand hours, is much greater than that of residential water heating, yet the average availability of residential cooling is less than that of residential water heating.

The annual availability factor is the sum of the maximum capacity available during each time period divided by the peak available capacity times the number of time intervals (i.e., total hours per year). This is similar to capacity factor, which is a measure of the power plant actual energy production compared to the possible energy production if the plant operated at maximum capacity during all time periods. The value of DR, normalized to dollars per peak enabled capacity ($/kW), closely tracks the availability factor, as shown in Figure ES-2a. Peak-enabled capacity represents the maximum response capability across all hours of the year.

The value of DR can also be measured by the cumulative availability across the year ($/MW-h), as shown in Figure ES-2b. The cumulative availability is the fraction of electricity from an aggregation of end uses that is flexible through DR. This metric reflects the correlation of each resource's availability to times of high market prices for operating reserves as well as its ability to take advantage of large energy price differences across hours of the day. Those resources with lower correlations and more constraints on energy shifting will tend to have lower values on a cumulative availability basis. Figure ES-2 only shows the revenue from each grid service, which neglects other electricity production cost savings such as avoided start-up costs.

Discussion

This study has investigated the abilities of different types of end-use loads to provide bulk power system services, including energy, capacity, and operating reserves in the western United States, and then implemented them in a Colorado test system model. Parts 1 and 2 of this report series provide a framework for incorporating DR in renewable integration analysis. However, additional work is necessary to verify many of the study assumptions.

Thus far, the analysis has been conducted sequentially, rather than iteratively. First, the resource assessment has been conducted without full understanding of the economic value of different resource types. In practice, more valuable resources will have greater incentives to offer their capabilities to the power system. Second, the resulting commitment and dispatch of the DR resources through the production cost modeling has not been tested against detailed examinations of the resources' capabilities. Because we are modeling the aggregate response capabilities of many individual providers, it is not clear how many customers need to be enrolled in order to meet aggregated response utilization or whether the initial assumptions make efficient use of those enrolled.

Finally, the total operational savings found through the production cost modeling ($7.9 million) is comparable in size to what the resources might earn as capacity resources. Capacity market-clearing prices and costs for avoided capacity (i.e., the annual carrying cost of a combustion turbine) are in the range of $77/kW-year [5] and $210/kW-year [6]. We assume that DR can provide 113 MW of capacity, which has a capacity value between $8.7 million and $23.7 million. Future efforts will assess the capacity credit of DR and investigate more thoroughly capacity value alongside operational value.

Table of Contents

List of Figures

List of Tables

1 Introduction

Demand response (DR) is a source of flexibility for the electric power system, which may become more important with increasing use of variable and uncertain renewable generation resources. This is a two-part series of reports. Part 1 of this report series estimates that there are more than 2.7 GW of commercial, residential, municipal, and industrial non-manufacturing DR capacity available for energy and more than 1.7 GW available for frequency regulation reserves, representing an average 0.7% of load and 56% of frequency regulation reserve requirements in the Western Interconnect [1]. Historically, large commercial and industrial customers have participated in interruptible load programs. These resources are typically called upon only during emergencies, rather than integrated dynamically into the power system optimization.

The emergence of smart grid technologies and new market structures could allow a greater fraction of electricity consumers to participate in DR programs. A key challenge for system planners is to understand the potential operational value of DR. Traditionally, new generation resources are evaluated in grid simulation tools that model the operation of the grid and determine the cost and feasibility of different generation mixes. However, the traditional framework of these modeling tools makes it difficult to evaluate DR resources, and as a result, these resources are often not represented or are represented using simplistic approximations [7].

This report describes the implementation and performance of multiple DR resources in a commercial grid simulation tool. In addition, we demonstrate how this can be used to calculate the value of DR, with an example analysis of DR in a test power system. The resource valuation process uses hourly DR profiles for a large number of appliance and equipment systems in commercial, residential, municipal, and industrial non-manufacturing facilities. These profiles were generated by Lawrence Berkeley National Laboratory and discussed in Part 1 of this report series [1]. These DR profiles were then implemented in the PLEXOS production cost model (PCM) as described in Section 2 of this report. Sections 2.1 and 2.2 introduce production cost modeling and the co-optimization of providing energy and operating reserves. Section 2.3 describes our approach to aggregating DR availability and the implications of that approach on modeling DR in a PCM. Section 3 demonstrates the performance and impact of DR in a test power system, analyzing the impact of all DR on the total production cost and the individual revenue streams for each DR resource providing load shifting/arbitrage and three classes of operating reserve products: regulation, flexibility, and spinning contingency reserves. The appendices have additional data on DR performance. Overall, this analysis demonstrates how DR profiles can be simulated in traditional planning tools to assess the value of DR in the bulk power system.

2 Production Cost Modeling and Demand Response

2.1 Production Cost Modeling: Energy and Operating Reserves

Production cost modeling describes the simulation of grid operation using standard tools used by utilities and system planners. PCMs[1] simulate the commitment (the process of scheduling power plant operation) and dispatch (the actual power output) of the power plant fleet to meet load at least cost while maintaining system security. The simulations are used to evaluate factors, including the operational feasibility of future grid mixes and the cost of system operation.

PCMs are chronological simulations of the grid, where in each time step (typically 1 hour) the objective function of the model is to minimize production cost by dispatching generators in order of variable cost (from lowest to highest) where variable cost includes fuel and operation and maintenance. The model includes a large number of parameters and constraints that affect the dispatch, including power plant efficiency as a function of plant output, plant availability, power plant start-up times, ramp rates, and environmental restrictions. The model may also include the availability of transmission between generators and load centers. A key parameter to the model is ensuring that in each hour, the aggregated set of generators meets the total demand, plus any additional operating reserve requirements. Models optimize against a fixed load profile, and historically, models do not consider the ability to vary load to address generation capacity shortfalls,[2] meet reserve requirements, or reduce the cost of operating the system.

Figure 1 illustrates the result of a PCM for five consecutive days in mid-July for the test system described in Section 3.1. Figure 1a shows the unit commitment of the generation fleet, demonstrating the total online generator availability to meet both the load and the operating reserve requirement. Figure 1b shows the actual economic dispatch of committed units. The figures show how many of the units that are committed to meet the peak demand in the middle of the day must be “backed down” (reduce output) overnight to accommodate the reduced demand. The majority of the variations in load are met by combined cycle (CC) units. These units (along with the coal units) cannot be quickly turned on, so often remain online even during periods of lower demand. Alternatively, higher operational cost combustion turbines (CTs) can be started quickly, so are only committed and dispatched during periods of high demand.

[1] These tools have a number of names, including “production cost” and “security-constrained unit commitment and economic dispatch” models. To realistically model the grid, these tools require extensive generator databases and include transmission constraints and other elements to capture the challenges of reliably operating the electric grid.
[2] Most PCMs place a soft constraint on providing load and reserves to prevent the model from reaching solution infeasibility.

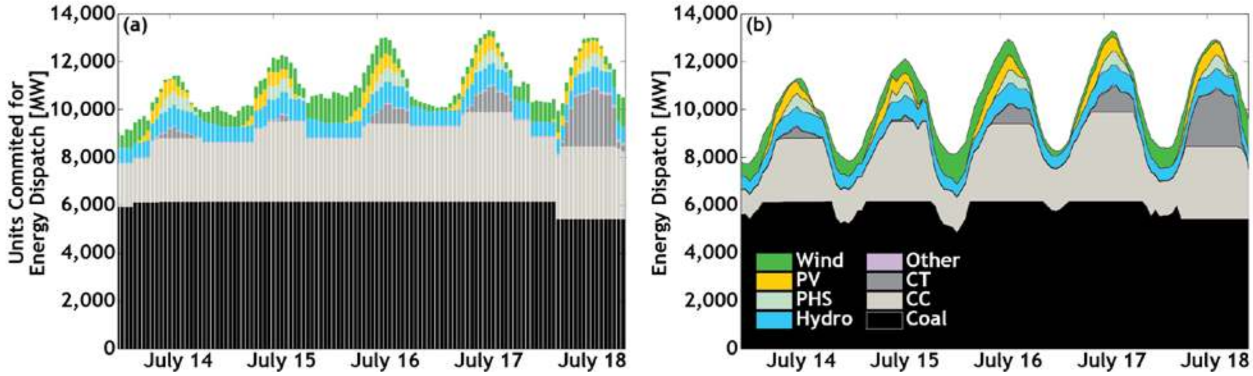

Figure 1. Example of the (a) unit commitment and (b) economic energy dispatch from a production cost model

Coal generators (along with nuclear and geothermal plants) are often referred to as "baseload" units due to their low variable costs. They are dispatched first (at the "bottom" of the dispatch stack) and typically only reduce output during periods of significant reduced demand. Hydro dispatch is performed in a somewhat different manner. While it has essentially zero fuel cost, hydro dispatch also has limited energy availability based on the availability of water resources upstream of the generators and regulation of waterway flow below the generator. Hydro units also have the ability to ramp very quickly in response to variation in load. Hydro is therefore often dispatched as a load-following and peaking plant, while operating under various environmental, recreation, and regulatory constraints of minimum and maximum water levels.[3]

Wind and solar generators do not fit the normal paradigm of baseload or dispatchable resources. These plants act to reduce the net load (load minus solar and wind generation) on the system, reducing output from the marginal (highest cost) plant on the dispatch stack. However, these plants add additional complexity given their variability and uncertainty, which can add to total reserve requirements.

Reserves represent generation capacity that is "held" to ensure reliable system operation. There are a number of types of operating reserves held by system operators and also a large number of names applied to different reserve products depending on the market [8]. In this study, we examine three classes of spinning operating reserve products: regulation, contingency, and flexibility. Spinning reserves refer to reserves provided by generators that are on and producing power at less than full capacity. To hold spinning reserves, a generator commits readily available capacity for power production in real time. Non-spinning reserves, capacity from generators that are not scheduled to be producing power, are also procured in the market at nearly zero cost.

Regulation reserves are designed to meet short-term (seconds to a few minutes) variation in demand due to unpredicted changes in load or variable generation resources, such as solar and wind power plants. Power plants must be equipped with an automatic generator control (AGC) system in order to provide regulation reserves; the AGC receives a regulation control signal and

[3] This is an oversimplification. In some locations large hydro resources allow it to meet a large fraction of load, including baseload demand. In addition, some "run-of-river" hydro plants are not dispatchable and are essentially baseload plants but with an output that varies daily and seasonally.

automatically adjusts the generator output. Regulation reserve requirements vary depending on load characteristics and the contributions from renewable generation.

Contingency reserves address large power plant or transmission line failures and are often based on the single largest contingency that could occur on the system. Contingency reserves are independent of wind and solar penetration, assuming no single wind or solar plant (and associated transmission) becomes the single largest contingency.

Flexibility (sometimes called ramping) reserves are an emerging type of reserves designed to accommodate large, unpredicted changes in generation from solar and wind power plants on time scales greater than the regulation time frame. Flexibility reserve products are under development in two markets, California Independent System Operator (CAISO) and Midwest Independent System Operator (MISO) [9,10]. Detailed discussion of the calculation and provision of reserves can be found in references [11] and [12].

An important consequence of holding operating reserves is that they impose a cost to the system. This cost occurs because of the additional generators are required to be committed and operated at minimum levels, which reduces the efficiency of the system dispatch. An extensive discussion of reserve costs origins and sensitivities is provided by Hummon et al. [12]. The addition of DR providing operating reserves is anticipated to reduce system dispatch inefficiencies.

2.2 Modeled Grid Services Provided by Demand Response

We use a PCM to consider the potential value of DR to provide energy and operating reserves. Part 1 of this report series also assesses the ability of DR to provide firm capacity for the system. The value of capacity is not included in the production cost of energy and therefore is not assessed in this grid-simulation paradigm. DR provides energy services by shedding energy or shifting energy demand over time, generally reducing demand during periods of high prices or moving demand from periods of high to low prices. This can be analogous to energy storage when providing load-leveling or energy arbitrage [3]. We use the term energy to describe the reduction in load and the terms re-charge and pre-charge to describe the replacement of that load either after or before the DR event, respectively. For operating reserves, we consider the three products discussed in Section 3.1.

The attributes of performance for DR providing various bulk power system services are not uniform across regions, nor are DR resources universally allowed to participate in bulk power system services [13]. In order to translate the physical capabilities of different DR strategies into eligibility for participation, we developed a set of generalized product definitions. These products are summarized in Table 1, which include the expected response characteristics of resources seeking to provide each type of service to the system.

Table 1. Description and Physical Requirements of the Products Modeled in This Study

Reserves Products		Physical Requirements			
Product Type	**General Description**	**How Fast to Respond**	**Length of Response**	**Time to Fully Respond**	**How Often Called**
Energy	Shed or shift energy consumption over time	5 minutes	≥1 hour	10 minutes	1–2 times per day with 4–8 hour notification
Regulation	Response to random unscheduled deviations in scheduled net load	30 seconds	Energy neutral in 15 minutes	5 minutes	Continuous within specified bid period
Flexibility	Additional load following reserve for large unforecasted wind/solar ramps	5 minutes	1 hour	20 minutes	Continuous within specified bid period
Contingency	Rapid and immediate response to a loss in supply	1 minute	≤30 minutes	≤10 minutes	≤ Once per day

2.3 Modeling Aggregated Demand Response

2.3.1 Production Cost Modeling Approach

This analysis uses a commercial PCM, PLEXOS,[4] to examine the operation of DR resources when providing both energy and operating reserves. There are three primary approaches to simulating DR: price responsive DR, system optimized load reductions for energy balancing, and system optimized load reduction/recovery for both energy imbalance and operating reserve provision. Several studies implement DR providing energy services using a supply curve to represent price responsive demand [14-16]. Other studies simulated DR energy services by scheduling the flexible portion of demand during the optimization [15,17-21]. Negnevitsky et al. [19] explicitly modeled load recovery, analogous to our "pre-/re-charge," of demand-side resources, as a constraint on the operation of DR in a PCM. Several studies have examined the impact of procuring demand-side resources for operating reserves, including using large interruptible loads to increase reliability (decrease in loss of load probability) at least cost [22,23] or acquiring aggregated DR in the day-ahead market for load-following and regulation operating reserves [24,25]. Other integration studies have included cost-based interruptible load as an aspect that Faria et al. [26] implemented in a co-optimization of DR and distributed generation on a distribution network.

We take the approach of co-optimizing DR to provide energy and operating reserves, as well as constraining the recovery of load allocated for energy shifting, for each DR resource. A single DR resource is represented by an aggregation of the end use over a region; for example, the residential cooling DR hourly time series profile for a large balancing authority (BA) area

[4] PLEXOS is produced by Energy Exemplar. http://energyexemplar.com.

representing most of the load in the State of Colorado represents the hourly sum of all residential cooling units expected to be enrolled in the DR program (outfitted with automatic controls) and operating during each hour. The profile is scaled based on the anticipated call rate (defined as how frequently or length of time the load must reduce demand). For instance, if the residential cooling DR call is typically 15 minutes for a unit, the hourly value is one-quarter of the total capacity available for any 15-minute period. The simulation in this report evaluates the impact on the bulk power system over large areas (such as a utility service territory or balancing area), as opposed to local impacts on the distribution network. DR is constrained in a manner similar to a conventional generation asset but considers the unique aspects of DR, including load recovery (i.e., energy shifting is energy-neutral) and constraints on frequent cycling of loads and time-varying response rates. We are unaware of any previous study that considers large-scale and highly detailed implementation of DR in a commercial PCM.

The PCM objective function remains the same: minimize the total production cost of generating energy to meet the demand in every time interval and hold required reserves to ensure system reliability from the perspective of a central scheduler. DR resources are implemented as virtual generators with time series profiles that define their availability for each grid service. The PCM DR services have small[5] variable operating costs in our model. This allows for an estimate of the potential *value* of various DR resources, which can then be compared to their implementation costs. However, there are significant energy limitations (via constraints limiting hours of operation, starts per day, etc.) and co-optimization constraints that limit the total energy scheduled from DR. DR resource availability by resource type, grid services, and balancing authority were assessed according to Part 1 of this report series [1].

Table 2 summarizes the types of DR considered, categorized by end-use and modeling parameters (operation/constraint similarity), as well as availability for different services.

[5] DR variable operating and maintenance (VO&M) costs are small compared to the fuel cost of conventional generators.

Table 2. Demand Response Resources Modeled in This Report

Product	End-Use Category	Model Category	Energy[6]	Regulation	Contingency	Flex
Commercial Lighting	Commercial	Reserves Only		X	X	X
Commercial Ventilation		Reserves Only		X	X	X
Commercial Cooling		Energy Shifting and Reserves	Shift	X	X	X
Commercial Heating		Energy Shifting and Reserves	Shift	X	X	X
Agricultural Pumping	Industrial Non-Manufacturing	Energy Shifting and Contingency Reserves	Shift		X	
Refrigerated Warehouses		Energy Shifting Only	Shift			
Data Centers		Energy Shifting and Contingency Reserves	Shift		X	
Wastewater Pumping	Municipal	Energy Shifting Only	Shift			
Municipal Pumping		Energy Shifting Only	Shift			
Outdoor Lighting		Reserves Only		X	X	X
Residential Heating	Residential	Energy Shedding and Reserves	Shed	X	X	X
Residential Cooling		Energy Shifting and Reserves	Shift	X	X	X
Residential Water Heating		Energy Shifting and Reserves	Shift	X	X	X

Enabling DR requires equipping end uses with communication and control equipment to allow the load to receive control signals or information from the system operator and act on that information to change their demand. The process of aggregating end-use load response for DR is discussed in Part 1 of this report series. Figure 2 demonstrates the relative terms for end-use capacity: enabled rated, enabled, and available. The enabled rated capacity is the nameplate capacity of each end-use unit times the number of units equipped to receive a control signal. For instance, in residential cooling, the nominal nameplate capacity (k) of a cooling unit is 2.4 kW, and this study estimates participation of approximately 130,000 units (N) in the Colorado test system. Therefore, the enabled rated capacity (R) is

$$R = k_{unit} \cdot N \tag{1}$$

which is equal to 310 MW for residential cooling in the Colorado test system (red, dashed line in Figure 2). The enabled capacity (E) is the maximum response, non-coincident with load, that is

[6] Energy "shift" resources require load recovery, discussed in Section 2.3.2. Energy "shed" resources do not require load recovery.

expected from the enabled rated capacity (R) and is calculated from the duty cycle (δ) and sheddability (σ):

$$E = R \cdot \delta \cdot \sigma \tag{2}$$

For residential cooling, the duty cycle of cooling units during periods of peak demand is estimated to be about 50%. This means that 50% of the cooling units are operating during peak demand. The sheddability is the fraction of the enabled rated capacity that is able to be shut down for a DR event. Residential cooling is assumed to have a sheddability of 0.7 [1]. Thus, the enabled capacity for residential cooling is 109 MW (green, dashed line in Figure 2). The time-varying availability represents the time-varying duty cycle for which residential cooling is inversely proportional to the ambient temperature.

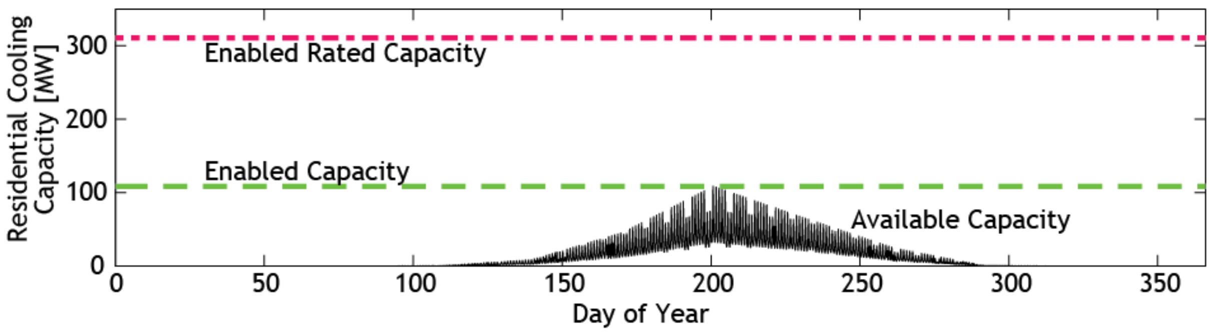

Figure 2. Enabled demand response capacity is defined as the capacity of the end use that is equipped/enrolled to respond to a signal or instruction to reduce load. The available capacity is the fraction of enabled capacity that is "bid" into the wholesale electricity market during each 24-hour optimization window.

Time series profiles of available capacity were generated for each DR resource and for each grid service (energy and operating reserves) that DR resource can provide [1]. Each time series profile of a service is input into PLEXOS, which can then co-optimize all DR resources across all potential services to minimize production cost. The co-optimization is subject to the constraint that the sum of all energy and ancillary service provision during each time interval cannot exceed the maximum availability in that interval. Figure 3 shows three days in August, of commercial cooling availability by grid service in Colorado. These are the input profiles for the PCM. Depending on hourly availability, some combination of energy, contingency, and flexibility reserves can be provided up to 70 MW. Commercial cooling is constrained to only allow energy to be shifted between 6 a.m. and 6 p.m.; thus, the charging profile is zero overnight between 6 p.m. and 6 a.m. The sum of energy and ancillary service provision during each interval cannot exceed the capacity available for the largest product—most often the energy availability. The following sections describe in detail the implementation of these services in the PCM.

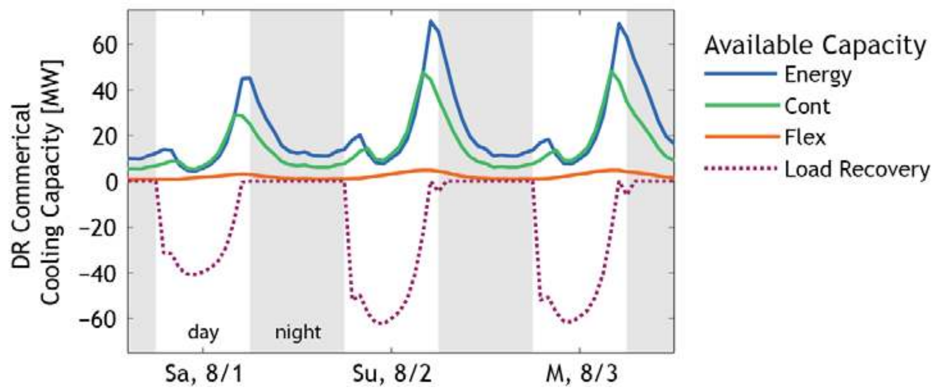

Figure 3. Maximum availability of energy, contingency, and flexibility for demand response from commercial cooling resources in Public Service Colorado on August 1–3

2.3.2 Simulation of Energy Shifting

PLEXOS models the permitted energy shifting scheduling some DR to reduce load during higher cost hours and increase load during lower cost hours. All of the energy shifting DR resources are required to replace the displaced load by increasing the load during other times. This energy replacement is also called load recovery and is analogous to an energy storage device but where the "charging" can occur before or after the discharging (load reduction). DR energy shifting is modeled as a 100% efficient storage device.[7] In other words, an equal amount of charge and discharge energy is accounted for in each DR resource. The charging and energy capacities are constrained by DR profiles. Section 2.3.1 described the DR profiles for energy and operating reserve availability. This section describes the DR profile for charging and the implementation of that in PLEXOS.

An illustration of the capacity available for generation and charging is shown in Figure 4. The total load profile is defined by the solid black line. The fraction of the end use that is available for DR activities is shaded dark grey, and the fraction of the end use that is fixed and not available for DR is shaded light blue. Figure 4b zooms in on the fraction of end-use load available for energy and the corresponding capacity available for DR charging (shown in pink). DR can "peak shift," providing a reduced load during periods of high demand and increased load during periods of low demand. Charging for heating and cooling loads is subject to the additional constraint that it must be completed during the daytime hours (see Table 3).

[7] DR pre-charge and re-charge may result in storage efficiencies greater than or less than 100%. Further research is needed to determine the appropriate modeling parameters for each type of DR and may vary by climate, season, time of day, and implementation.

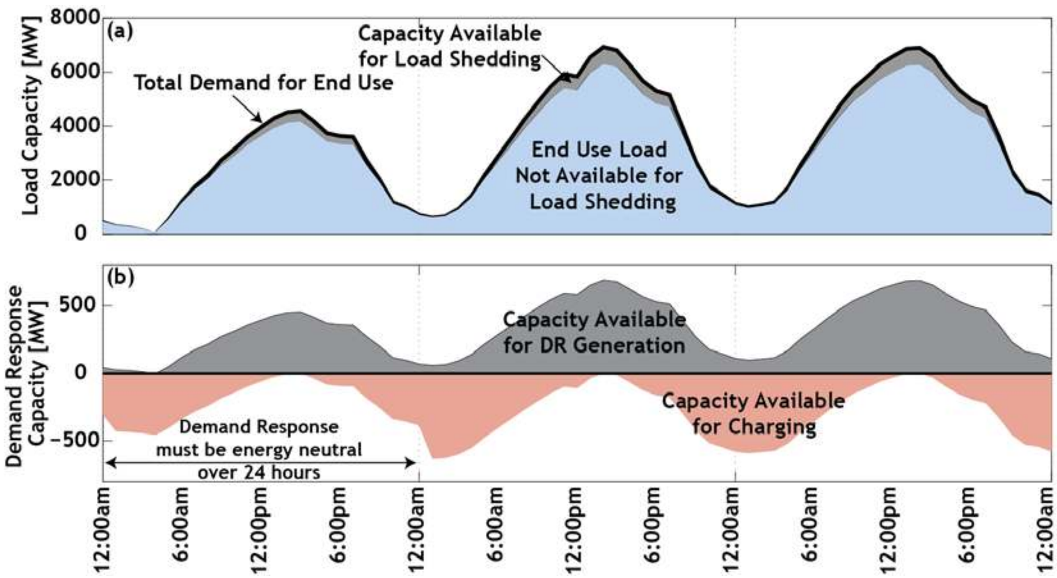

Figure 4. Example of the (a) total load from an illustrative end use and the portion of that load available for demand response activities and (b) the constrained availability of DR capacity for charging and generation

The time series of capacity available for generation is assigned to the discharge of the storage device. Similarly, the time series of capacity available for charging is assigned to the load of the storage device. In this configuration, DR is essentially a storage device with limitations of when and how much the device can provide energy to the system or draw energy from the system.

The operational constraints associated with energy shifting are listed by type of DR in Table 3. Energy scheduling is constrained by both the maximum number of hours of operation per day and by the schedule for charging. The entire energy-shifting DR is energy neutral over 24 hours. This means that the system must either pre-charge or re-charge the DR resource. While some DR resources may be capable of shifting energy over longer time periods (greater than 24 hours), the modeling horizon in this study is 24 hours. Decision complexity (and thus computation time) increases with the length of the optimization horizon. In addition to the challenges associated with modeling longer duration load shifting, Sioshansi et al. [27] has demonstrated that most of the value of load shifting appears to occur on the diurnal cycle.

Several DR resources have soft constraints for energy operation with penalty prices for constraint violations. Soft constraints are presented in PCMs as a dollar value on the violation of that constraint. For instance, municipal pumping resources are constrained to “operate” once per day. If this were a hard constraint, there would never be an occasion when that DR resource may operate more than once per day. As a soft constraint, the cost of violating the constraint is part of the cost minimization function and may occur when it is the least-cost option of meeting load. The soft constraint with the least penalty will be violated first. For this reason, it is important that all soft constraint costs are set relative to each other. The penalties are paid to the DR resource and are part of the total revenue for DR.[8]

Table 3. Demand Response Resource Energy Operation Restrictions, Penalties, and Costs

DR Resource	Operation Restrictions			Penalties		Costs
	Charging Hours (within 24 hours)	Max Hours per Day	Max Starts per Day	Max Hour Shed (per hour)	Max Starts (per start)	Variable Operating & Maintenance ($/MWh)
Residential Heating	N/A	1	1	Strictly enforced		
Commercial Cooling	6 a.m.–6 p.m.					
Commercial Heating	3 a.m.–7 p.m.					
Residential Cooling	6 a.m.–6 p.m.					
Data Centers	Any	4		Strictly enforced		
Residential Water Heating	Any					
Wastewater Pumping	Any	3	1	$20.00	$50.00	
Agricultural Pumping	Any	8	1	$20.00	$50.00	$2
Municipal Pumping	Any	2	1	$50.00	$100.00	
Refrigerated Warehouses	Any	4	1	Strictly enforced		

2.3.3 Co-Optimization of Energy and Reserves

As discussed in Section 2.3.1, our approach to modeling DR in PLEXOS is to enforce energy shifting and charging for DR providing energy services and to co-optimize the DR capacity for energy shifting with provision of operating reserves. The capability to provide operating reserves is largely dependent on the response rate of the load. The response rate (along with other characteristics) of DR changes with the mechanisms that control load. The response rate varies across DR resources and is defined differently than conventional fuel-burning thermal generators. Ramp rates set the maximum change in power output per minute and the capacity available for operating reserves. For instance, if a generator has a ramp rate of 5 MW/minute, the maximum regulation reserve provision is 25 MW because regulation services have to be fully met within 5 minutes. Cooling and heating DR uses both changes to the thermostat set-point or direct load control to activate and "ramp" the load shed [28]. DR from pumping applications often involves human action to change the load consumption of the system [29]. The differences between DR service activation and generators necessitated that we model DR resources with time-varying and service-varying response time.

[8] If the penalty scales with the unit of the violation (e.g., number of hours) rather than the size of the DR capacity, it will be evaluated in future DR modeling studies.

Table 4 defines the "fast" and "slow" time lapse to full load shed for each DR resource. Each DR resource could either have a fast actuation or a slow actuation, yet the total capacity available to respond is constrained by the resource and not the response mechanism. The fast response time is assigned to the contingency and regulation reserve capacity and the slow response time is assigned to the energy and flexibility reserve capacity. The grouping of energy/flexibility and contingency/regulation resources is based on the response times needed for the product (see Section 2.2). The PCM co-optimizes both sets of responses and capacities, while maintaining the differing response rates. We convert the time to full load shed to a ramp rate by taking the capacity time series for the service type (e.g., contingency reserve) divided by the time to full load shed. Our implementation in PLEXOS uses two separate generation objects[9] with different response rates. One generator represents the energy shifting and flexibility reserve (slow response services), while its counterpart provides spinning contingency and regulation reserves (fast response services). While the provision of each grid service from DR is limited by the individual profile for each grid service, a single constraint overlaps all four grid services to enforce that the sum of all DR capacity providing energy and operating reserves cannot exceed the maximum of any one grid service profile.

Table 4. Time Lapse to Full Load Shed for Each Demand Response Resource

Product	Time Lapse to Full Load Shed	
	Contingency and Regulation	Energy Shifting and Flexibility
Residential Heating	1 min	15 min
Commercial Cooling	1 min	15 min
Commercial Heating	1 min	15 min
Residential Cooling	1 min	15 min
Data Centers	1 min	15 min
Residential Water Heating	30 sec	30 sec
Wastewater Pumping	1 min	5 min
Agricultural Pumping	1 min	1 min
Municipal Pumping	1 min	5 min
Refrigerated Warehouses	1 min	5 min
Commercial Lighting	30 sec	30 sec
Commercial Ventilation	1 min	15 min
Outdoor Lighting	40 sec	40 sec

[9] In PLEXOS, a generator cannot have separate (unrelated) ramp rates for different services.

3 Results

3.1 Colorado Test System

To evaluate the impact of DR on an electric power system, we developed a test case composed of two balancing areas largely in the State of Colorado. The test system is described extensively in three sources [3,4,12]. The Colorado test system consists of two balancing areas [PSCo and Western Area Colorado Missouri (WACM)] using data derived from the database established by the Western Electricity Coordinating Council (WECC) Transmission Expansion Policy Planning Committee (TEPPC) model and other publicly available datasets. Transmission is modeled zonally, without transmission limits within each BA area. Projected generation and loads were derived from the TEPPC 2020 scenario [30]. Hourly load profiles were based on 2006 data and scaled to match the projected TEPPC 2020 annual load. Hourly solar and wind power generation profiles are time synchronized to the load profiles for the year 2006.

The system peaks in the summer with a 2020 coincident peak demand of 13.7 GW and annual demand of 79.0 TWh. A total of 201 thermal and hydro generators are included in the test system, with total capacities listed in Table 5. We adjusted the conventional generator mix to ensure the available capacity (after outages) was always at least 9% greater than demand by adding a total of 1,450 MW (690 MW of combustion turbines and 760 MW of CC units). This adjustment was necessary in part because the simulated system does not include contracted capacity from surrounding regions or any capacity contribution from solar and wind resources. The base case of the test system assumes a wind and solar penetration of 16% on an energy basis. For comparison, Colorado received about 11% of its electricity from wind in 2012 [31].[10] Photovoltaic (PV) profiles were generated using the System Advisor Model (SAM) [32] with 2006 meteorology. Wind data was derived from the Western Wind and Solar Integration Study (WWSIS) dataset [33].[11] Discrete wind and solar plants were added from the WWSIS datasets until the installed capacity produced the targeted energy penetration.[12]

[10] Colorado generated 6,045 GWh from wind in 2012 compared to total generation of 53,594 GWh. EIA “Electric Power Monthly with Data for December 2012,” February 2012. See [30].
[11] All generation profiles were adjusted to be time synchronized with 2020, which is a leap year.
[12] The sites were chosen based on capacity factor and do not necessarily reflect existing or planned locations for wind and solar plants.

Table 5. Test System Generator Capacity in 2020

	System Capacity (MW)
Coal	6,178
Combined Cycle	3,724
Gas Turbine/Gas Steam	4,045
Hydro	773
Pumped Storage	560
Wind	3,347 (10.7 TWh)
Solar PV	878 (1.8 TWh)
Demand Response	293
Other[a]	513
Total	15,793

[a] Includes oil- and gas-fired internal combustion generators.

Fuel prices were derived from the TEPPC 2020 database. Coal prices were $1.42/MMBtu for all plants. Natural gas prices varied by month and ranged from $3.90/MMBtu to $4.20/MMBtu, with an average of $4.10/MMBtu [34]. No constraints or costs were applied to carbon or other emissions.

We generated hourly requirements for contingency, regulation, and flexibility reserves.[13] Contingency reserves are based on the single largest unit (an 810-MW coal plant) and allocated with 451 MW to PSCo and 359 MW to WACM; 50% is met by spinning units.[14] Regulation and flexibility reserve[15] requirements vary over time based on the statistical variability of load, wind, and PV, with the methodology described in detail by Ibanez et al. [11]. Hummon et al. [12] describes the application of the methodology to the test system.

The sum of the total operating reserves (met by spinning units) averages 582 MW, which corresponds to about 6.4% of average load. Table 6 summarizes the general characteristics of the three modeled reserve services. Reserves were modeled as "soft constraints," meaning the system was allowed to not meet requirements if the cost of provision exceeded the threshold value shown in Table 6. These penalties were chosen to be high enough so that the least-cost decision is likely to be starting a new unit to provide reserves. Unserved load was also modeled as a soft constraint, with a penalty price of $10,000/MWh.

[13] For additional discussion of these reserves (especially flexibility reserves, which is not yet a well-defined market product), see Ela et al. [8].
[14] The PSCo and WACM balancing areas are part of the Rocky Mountain Reserve group, which shares contingency reserves based on these values.
[15] For these services only the "upward" reserve requirements were evaluated. The need for downward reserves becomes of greater importance at high renewable penetration when conventional thermal generators are operated at or near their minimum generation points for more hours of the year. Future work will evaluate the cost and price of separate up and down reserve products in these scenarios.

Table 6. Summary of Operating Reserves in the Base Case of Test System

Operating Reserve Service	System Drivers	Time to Respond [minutes]	Requirement (% of load) Mean (min/max)	Penalty [$/MW-h]
Regulation	PV, wind, load	5	1.33 (1.00/1.71)	9,500
Contingency	Largest generator	10	4.54 (2.97/5.95)	9,000
Flexibility	PV, wind	20	0.64 (0.13/1.07)	8,500

The availability and constraints of individual generators providing reserves are major drivers for the cost of providing reserves and therefore for the value of DR-providing reserves. Not all generators are capable of providing regulation reserves based on operational practice or lack of necessary equipment to follow a regulation signal. The results reported are for a system that only allows a subset of generators to provide regulation and flexibility. We based our assumptions on the PLEXOS database established for the CAISO's *33% Renewable Integration Study* [35]. This dataset assigns regulation capability to a subset of plants, which is about 60% of total capacity within California (as measured by their ramp rate). Similarly, we allowed only 60% of all dispatchable generators (i.e., coal, gas CC, dispatchable hydro, and pumped storage) to provide regulation.[16] Based on feedback from various utilities and system operators, we further restricted CTs from providing regulation. We allow all dispatchable plants (including CTs) to provide flexibility and contingency reserves.

An additional cost was assigned to plants providing regulation, associated with additional wear and tear and heat rate degradation associated with non-steady-state operation. This is functionally equivalent to a generator regulation "bid cost" in restructured markets, discussed in PJM Manual 15: Cost Development Guidelines [36]. The assumed regulation costs, by unit type, are provided in Table 7. We did not apply a regulation bid cost to DR resources that were capable of providing regulation. Essentially this makes DR "first in line" for providing regulation services, with the only limitation being co-optimization of the DR resource for load shedding/recovery.

Table 7. Assumed Additional Operating Cost for Units Providing Frequency Regulation Service

Generator Type	Cost ($/MW-h)
Supercritical Coal	15
Subcritical Coal	10
Combined Cycle	6
Gas/Oil Steam	4
Hydro	2
Pumped Storage	2

The PLEXOS simulations performed in this analysis used day-ahead scheduling with a 48-hour optimization window, rolling forward in 24-hour increments. The extra 24 hours in the unit commitment horizon (for a full 48-hour window) were necessary to properly commit the

[16] In practice, there may be far fewer generators that provide regulation services, where information is proprietary, and which may affect the marginal price of regulation.

generators with high start-up costs and the dispatch of energy storage. All scenarios were run for one chronological year using PLEXOS version 6.207 R08, using the Xpress-MP 23.01.05 solver, with the model performance relative gap set to 0.5%.

3.2 Base Case Results (No DR Resources)

Figure 5 summarizes the energy and reserves prices produced in the test system in the base case without added DR. Figure 5a shows a price duration curve (PDC) for energy, with an average energy price of $32/MWh. Any value associated with load shifting will occur by reducing demand during periods of highest price (on the left side of the PDC) and increasing demand during periods of lower price (on the right side of the PDC) [12]. Operating reserve price duration curves for the base system are provided in Figure 5b. The average price of regulation reserves was $14.7/MW-h. For comparison, the average market clearing price for regulation in 2011 was $11.8/MW-h in New York Independent System Operator (NYISO), $10.8/MW-h in MISO, and $16.1/MW-h in CAISO [37].

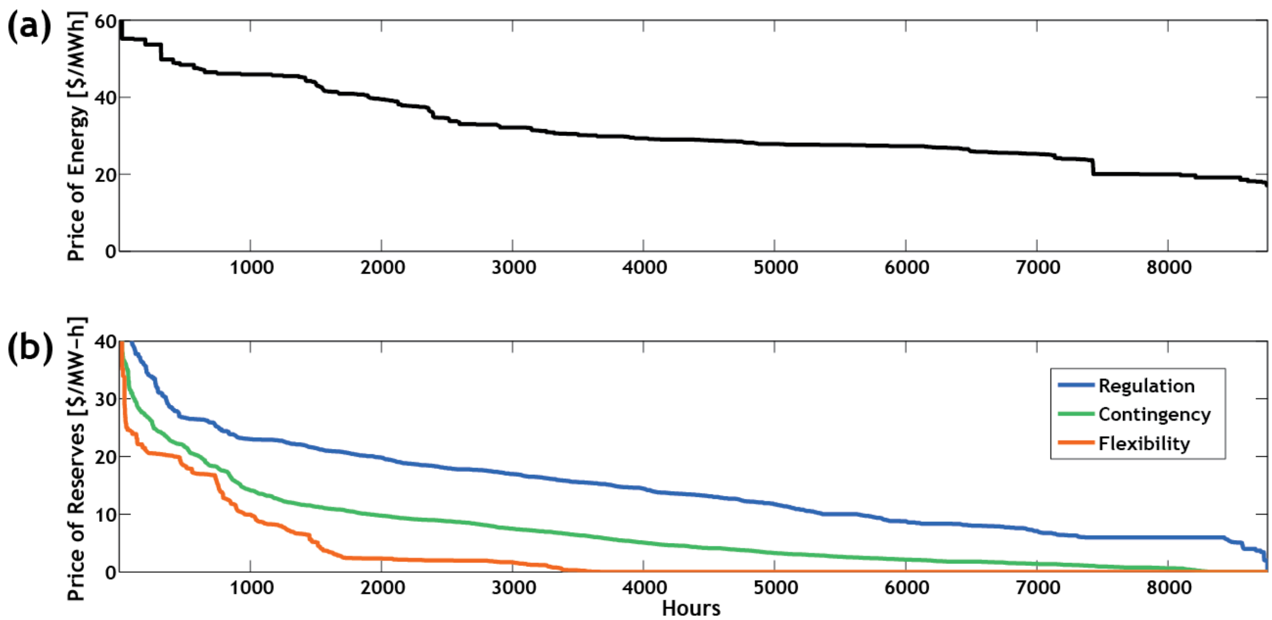

Figure 5. System price duration curve for (a) energy and (b) operating reserves in the base case of the Colorado test system

The average price of spinning reserves in the base system across the two balancing areas simulated in the test system was $7.0/MW-h.[17] The values can be compared to 2011 average market clearing prices of $7.4/MW-h in NYISO, $2.8/MW-h in MISO, and $7.2/MW-h in CAISO. Of note is the large number of hours where the price of spinning reserves is close to zero, which is often observed in the clearing price for spinning reserves in wholesale markets. For example, in 2011, the clearing price for spinning contingency reserves in both MISO and CAISO was less than $1/MW-h for over 2,000 hours.

[17] As with regulation reserves, this excludes hours of extremely high prices driven by internal model penalties.

The cost of flexibility reserves was very low due to the relatively slow response rate (20 minutes compared to 10 minutes for spinning reserves and 5 minutes for regulation) and a small overall requirement. However, the actual use of flexibility reserves in real-time dispatch could be more costly than our model captures. This reserve service has yet to be implemented in a restructured market, and our assumptions regarding requirements and use may be substantially different from a flexibility product actually implemented by utilities and system operators. Additional analysis of the performance of flexibility reserve capacity deployed in sub-hourly dispatch will be required to fully evaluate the potential benefits in terms of system cost savings by allowing demand to provide this service.

3.3 Demand Response Resources in the Test System

As discussed in Section 2.3.1, the capacity available for each DR resource, for each grid service, was assessed independently. Table 8 summarizes the availability of each DR resource to provide energy on an annual basis as well as during peak demand hours (top 1%) in the Colorado test system. Table 9 summarizes the availability of each DR resource to provide operating reserves. In addition to available capacity, Table 9 includes the availability of DR resources in terms of the average percentage of hourly reserve requirement that could be met by the resource. This calculation does not take into account the mutual exclusivity of the resource capacity.

Table 8. Availability of Demand Response to Provide Energy in the Colorado Test System on an Annual Basis as Well as in the Top 20 Hours of Greatest Demand

Demand Response Resources Providing Energy[a]	Annual			Top 20 Load Hours	
	Capacity (mean/min/max) [MW][b]	Annual Available Energy with Constraints [GWh][c]	Annual Hours Available[d]	Capacity (mean/min/max) [MW]	% of Load (mean/min/max)
Residential Cooling	10.9 / 0 / 108.7	38.8	5,390	55.1 / 41.3 / 72.2	0.4 / 0.3 / 0.5
Residential Water Heating	1.8 / 0.5 / 3.9	15.7	8,784	1.3 / 1.1 / 1.6	0 / 0 / 0
Commercial Cooling	2.1 / 0 / 46.4	10.7	4,338	8.2 / 4.2 / 12.9	0.1 / 0 / 0.1
Commercial Heating	2.2 / 0 / 25.5	3.8	8,390	0 / 0 / 0	0 / 0 / 0
Municipal Pumping	1.7 / 0.4 / 3.8	2.1	8,784	2.1 / 1.7 / 2.7	0 / 0 / 0
Wastewater Pumping	1.5 / 1.5 / 1.5	1.6	8,784	1.5 / 1.5 / 1.5	0 / 0 / 0
Refrigerated Warehouses	0.2 / 0 / 0.4	0.3	8,685	0.3 / 0.3 / 0.4	0 / 0 / 0
Agricultural Pumping	17 / 1.7 / 41.2	49.9	8,784	36.6 / 32 / 40	0.3 / 0.2 / 0.3
Data Centers	8 / 8 / 8	11.7	8,784	8 / 8 / 8	0.1 / 0.1 / 0.1
Total[e]	**45.4 / 14.5 / 227.8**	**134.6**	**8,784**	**113.1 / 91.8 / 137.3**	**0.8 / 0.7 / 1**

[a] In the test system, residential heating is primarily sourced from natural gas, not electricity. We include the methods of modeling residential heating because it is modeled in the western interconnect. See Ref. [1].
[b] Peak megawatt is the maximum capacity available for that resource (across all products), except for agricultural pumping where the maximum capacity available for contingency is 61.9 MW.
[c] Annual gigawatt-hours of energy availability is calculated by finding the maximum energy available for each day constrained by the number of allowable hours. This expresses the maximum energy shifting potential; the production cost model optimizes that energy shifting and will always be less than this quantity.
[d] The year analyzed (2020) is a leap year, hence 8,784 hours.
[e] Total numbers are less than the sum of the end uses due to non-coincident availability.

Table 9. Availability of Demand Response for Operating Reserves (Regulation, Contingency, and Flexibility) in the Colorado Test System

Demand Response Resource	Mean Capacity [MW]			Mean % of Reserve Requirement Available from Demand Response			Hours of Average or Higher Availability
	Reg	Cont	Flex	Reg	Cont	Flex	
Residential Cooling	10.3	10.3	10.3	8.8	2.5	23.9	2,656
Residential Water Heating	1.8	1.8	1.8	1.5	0.4	3.4	4,334
Commercial Cooling	0.1	1.6	0.2	0.1	0.4	0.6	2,052
Commercial Lighting	1.0	3.0	3.0	0.8	0.7	5.8	3,945
Commercial Ventilation	1.1	3.4	3.4	0.9	0.8	6.6	4,890
Outdoor Lighting	23.4	23.4	23.4	20.7	5.8	44.2	4,329
Agricultural Pumping	0.0	25.6	0.0	0.0	6.3	0.0	0
Data Centers	0.0	8.0	0.0	0.0	2.0	0.0	0
Total[a]	**37.7**	**77.0**	**42.1**	**32.9**	**19.0**	**84.5**	**4,985**

[a] Total numbers are less than the sum of the end uses due to non-coincident availability.

Table 8 and Table 9 show that the total capacity available for energy and operating reserves is not distributed equally over the DR resource categories. Some resources have a nearly constant capacity available, while others are highly seasonal. For instance, residential cooling capacity, during peak demand hours, is 50 times greater than residential water heating, yet the annual expected energy from residential cooling is less than 3 times greater than residential water heating. The annual availability factor (AF) is the sum of the maximum capacity available during each time period divided by the peak available capacity times the number of time intervals (i.e., total hours per year). This is similar to capacity factor, which is a measure of the power plant actual energy production compared to the possible energy production if the plant operated at maximum capacity during all time periods. The "energy" AF for a DR resource is calculated based on the probable energy production (with operation constraints such as maximum number of hours per day; see Table 3) divided by the sum of the peak available capacity over all hours of the year. Table 10 provides the availability factors for energy and for all services. AF does not capture the time varying value of DR but does convey the difference between seasonal and non-seasonal resources.

Table 10. Demand Response Availability Factor for Energy and All Services

Demand Response Resource	Energy Availability Factor [%]	Total Availability Factor [%]	Demand Response Resource	Energy Availability Factor [%]	Total Availability Factor [%]
Residential Cooling	4	10	Municipal Pumping	6	46
Residential Water Heating	46	46	Wastewater Pumping	13	100
Commercial Cooling	3	5	Outdoor Lighting	0	52
Commercial Heating	2	9	Refrigerated Warehouses	10	47
Commercial Lighting	0	75	Agricultural Pumping	9	41
Commercial Ventilation	0	81	Data Centers	17	100

The time series profiles provide the PCM with the upper bound of availability for each service, and we impose a constraint such that the total energy and reserve provision could not exceed the availability of any one service. On average, about 0.5% of Colorado test system load is available for DR, with a minimum of about 0.1% occurring during early evening hours in the winter and a maximum of about 2.3% occurring during late afternoon hours in the summer.

While we did not perform a detailed capacity credit analysis, some insights can be gained by examining the availability of DR during periods of highest demand. Figure 6 shows the hourly system load (right axis) and the maximum hourly capacity available from DR (left axis) in the four categories: municipal, industrial non-manufacturing, commercial, and residential loads (see Section 2.3 for the DR resources in each category). During the top 1% of demand hours, about 115 MW of DR was available to shed or shift load, which is about 2.5 times the average DR availability.

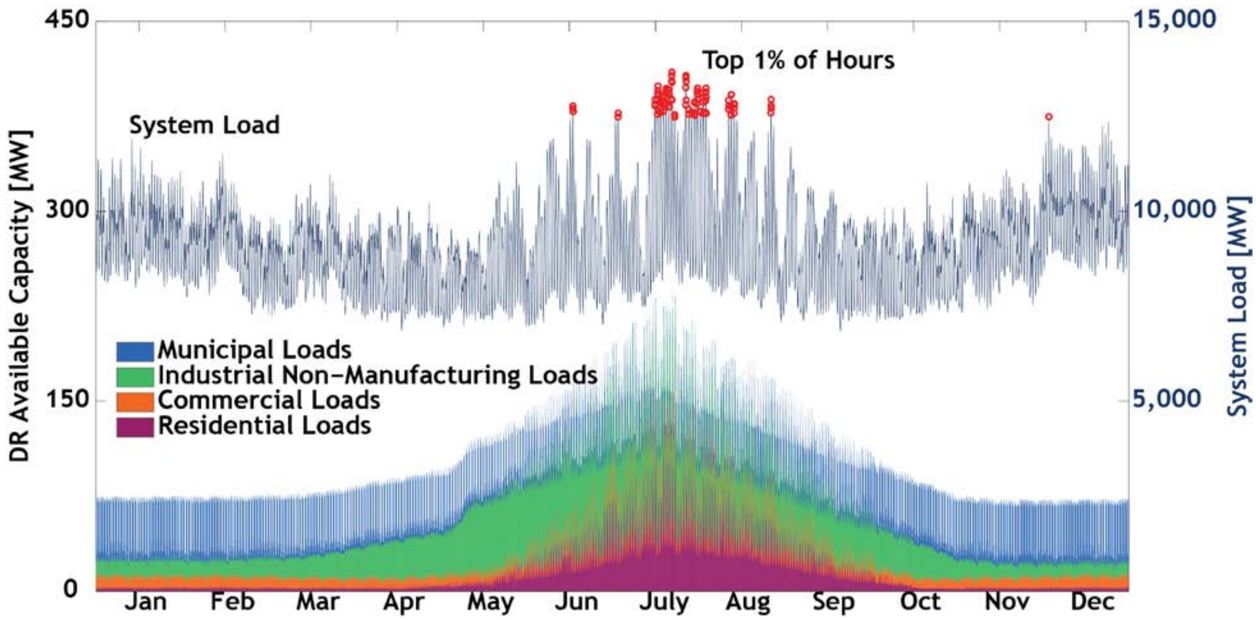

Figure 6. Comparison of the (left axis) hourly capacity available from DR by sector and (right axis) hourly system load DR availability is the maximum available from each DR resource across the energy and ancillary service markets. The top 1% of hours, measured by system load, are marked with red circles.

3.4 Examples of Demand Response Performance

Three examples of DR are examined in this section: commercial lighting, wastewater pumping, and residential water heating. For each resource, we examine the allocation of capacity via two figures. The first figure shows the annual daily average provision of the DR resource across energy and operating reserves as well as the annual hourly average provision. The second figure shows two snapshots of hourly energy and ancillary service provision data, from spring and summer. Additional examples of DR performance can be found in the appendix.

3.4.1 Commercial Lighting

Commercial lighting (shown in Figure 7 and Figure 8) is able to hold flexibility, regulation, and contingency reserves but cannot shift energy. Capacity available for flexibility reserves is the largest during every interval and therefore sets the total capacity of commercial lighting DR that can provide operating reserves. Commercial lighting is cyclic over the day and week (see Figure 7), decreasing overnight and on the weekends. The provision for regulation is near its full regulation availability at all times; the remaining capacity is primarily provisioned for contingency reserves, with intermittent provision of flexibility reserves, the lowest price service.

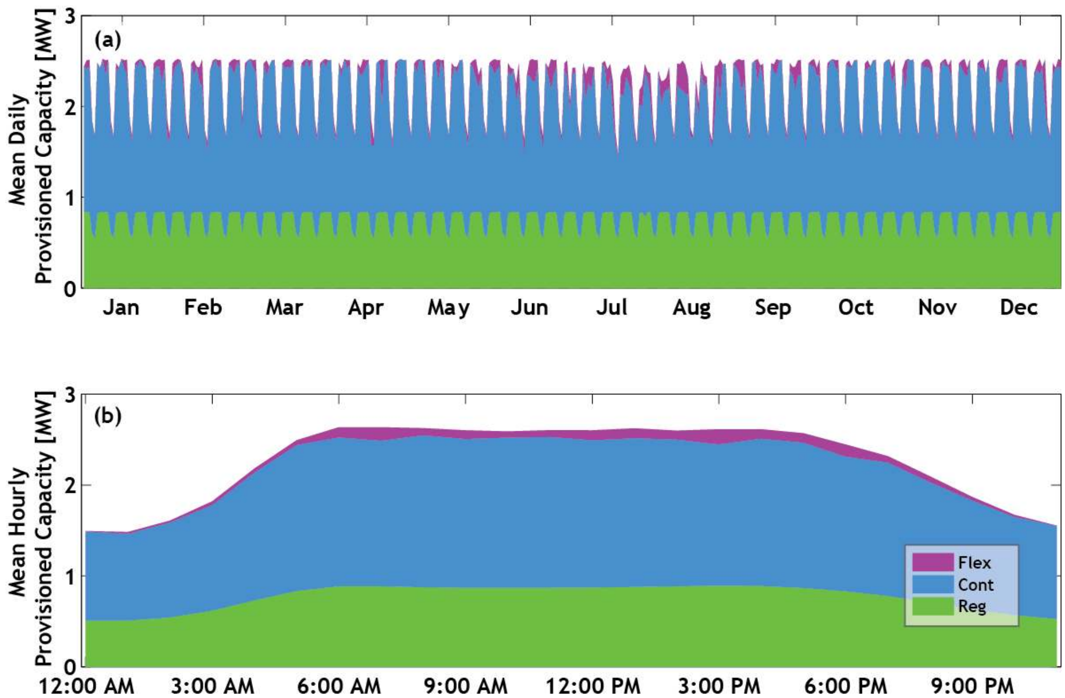

Figure 7. Commercial lighting DR in PSCo: Mean (a) daily and (b) hourly provision of capacity

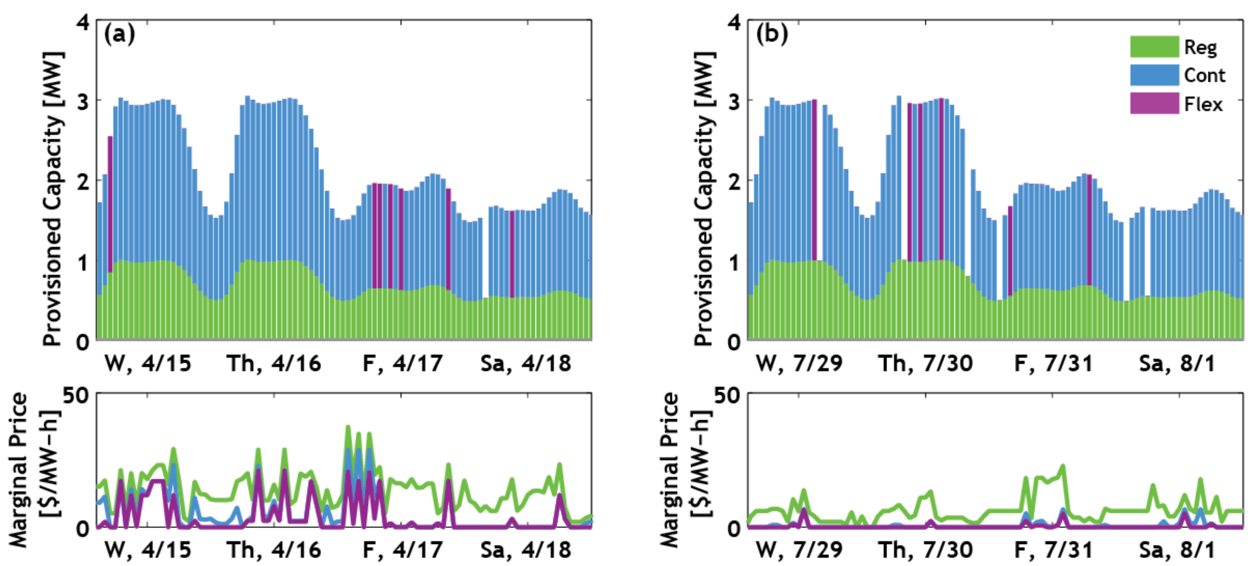

Figure 8. Commercial lighting DR in PSCo (upper): Hourly capacity provision for (a) spring and (b) summer; hourly marginal price of operating reserves (lower)

3.4.2 Municipal Wastewater Pumping

Wastewater pumping can provide energy shifting but is assumed not able to provide operating reserves. When many pumps are aggregated together, there is the potential to offer operating reserves, but individual facilities are assumed to be too risk-averse to participate in calls of that frequency. Part 1 of this report series has a more in-depth discussion on the controllability and acceptability factors for wastewater pumping [1]. Wastewater pumping has a constant available capacity, as can be observed in Figure 9a and Figure 9b. Similar to many of the other DR resources that can shift energy, Figure 9b shows that the average day shifts energy away from the early evening hours to the late evening and early morning hours.

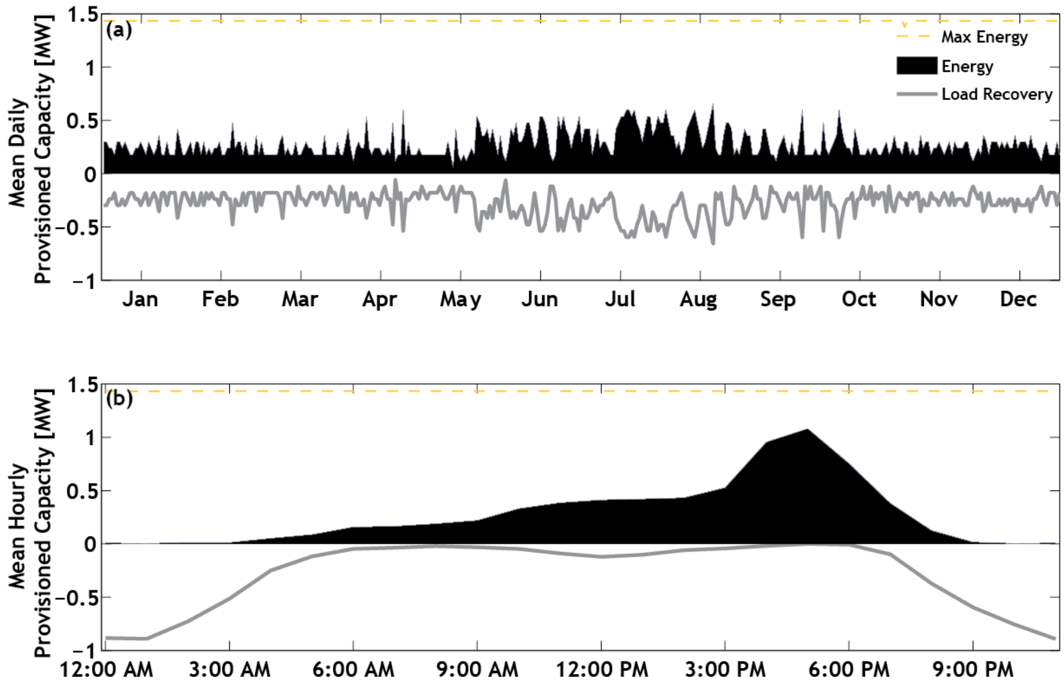

Figure 9. Wastewater pumping DR in PSCo: Mean (a) daily and (b) hourly provision of capacity

The discrepancy between the maximum energy and the energy utilization (black area) in Figure 9a is explained by the constraints placed on the usage of wastewater pumping for DR. There is a $50 violation for scheduling the aggregated set of pumps to turn off more than once a day and a $20 violation for every 1 hour of utilization over 3 hours. Figure 10 shows the operation of these units in the (a) spring and (b) summer. Similar to municipal pumping, the violations are more frequent in July when energy prices are higher.

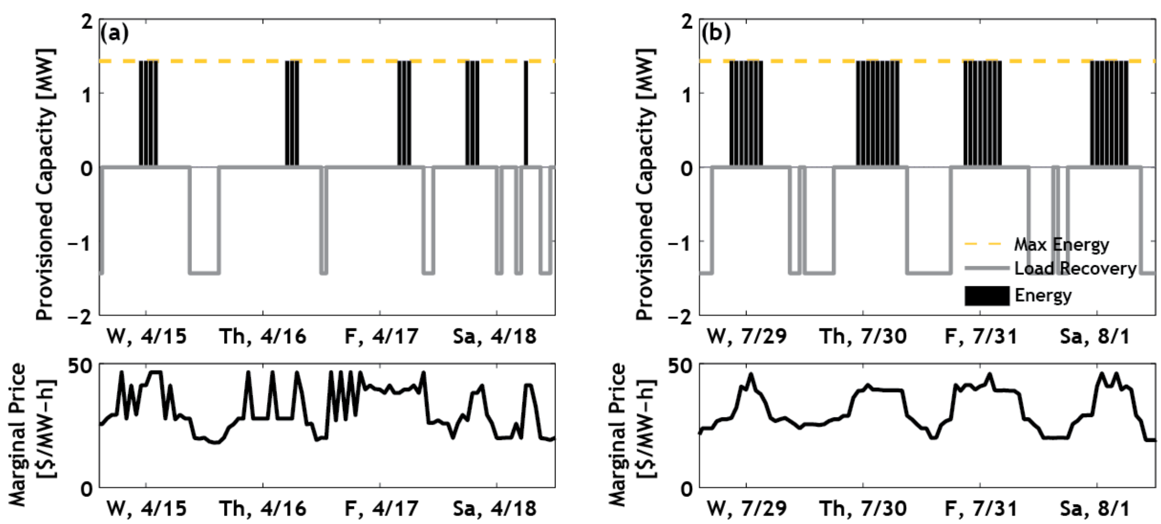

Figure 10. Wastewater pumping DR in PSCo (upper): Hourly capacity provision for (a) spring and (b) summer; hourly marginal price of energy and operating reserves (lower)

3.4.3 Residential Water Heating

Residential water heating can provide both energy shifting and reserves services. Figure 11a shows a seasonal availability pattern, with greater overall demand during the winter. The maximum energy and operating reserves available from water heating peak twice a day (morning and evening). The generation obtained from varying the electrical load of a water heater must be shifted to other hours within the day, just as with other energy shifting DR resources. The fraction of load from water heaters offered for DR is such that the temperature of the water heater will stay within its operating range. Similar to residential cooling DR, residential water heating can hold contingency during charging. The shifted energy is not restricted to a set of hours, and thus can be optimized to the hours with the lowest price (i.e., overnight) (see Figure 11b).

Regulation dominates the allocation of reserve provision because it is the most valuable product but also because the maximum available capacity of residential water heating for regulation is 4.5 times greater than the capacity available for contingency. Regulation is energy-neutral over 15–20 minutes; thus, there is a high potential to use water heater capacity for regulation reserves without disturbing the operation of the water heater.

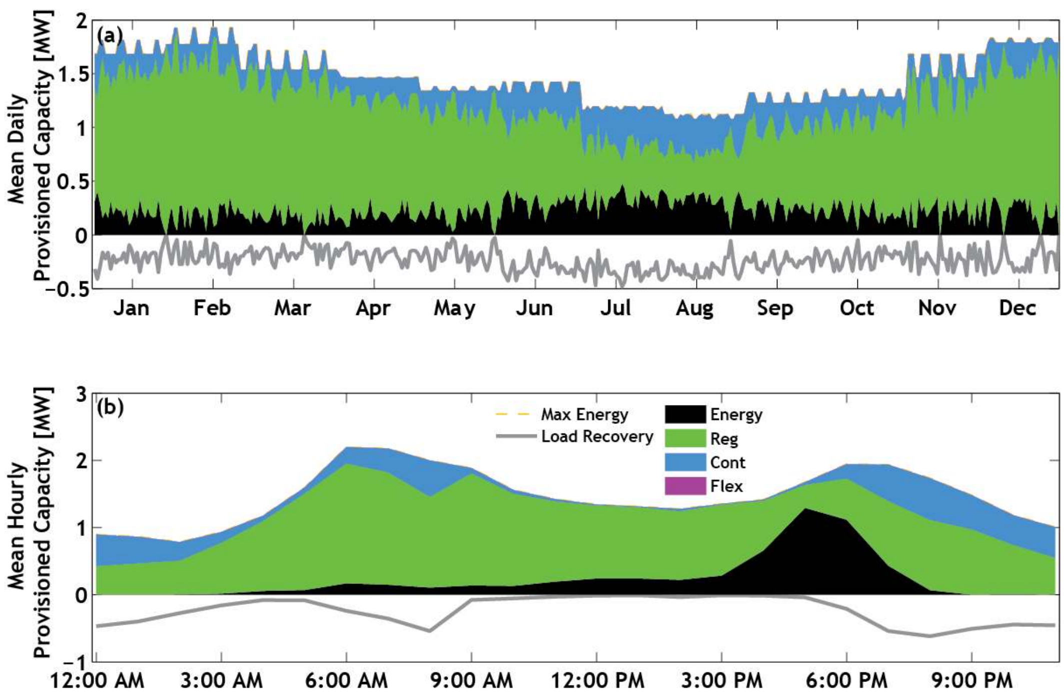

Figure 11. Residential water heating DR in PSCo: Mean (a) daily and (b) hourly provision of capacity

Figure 12 shows the hourly allocation of capacity in residential water heating during two 4-day periods in the spring and summer, as well as the marginal price of energy and operating reserves. In the summer, the residential water heating DR resource is typically used for energy between the hours of 1 p.m. and 8 p.m. and shifts that energy to the late evening and overnight hours when the price of energy is lowest. This is similar to the operation of an energy storage device arbitraging on-/off-peak energy prices. It also holds capacity for contingency events while recovering energy. During all other hours, the residential water heating DR capacity is held for regulation.

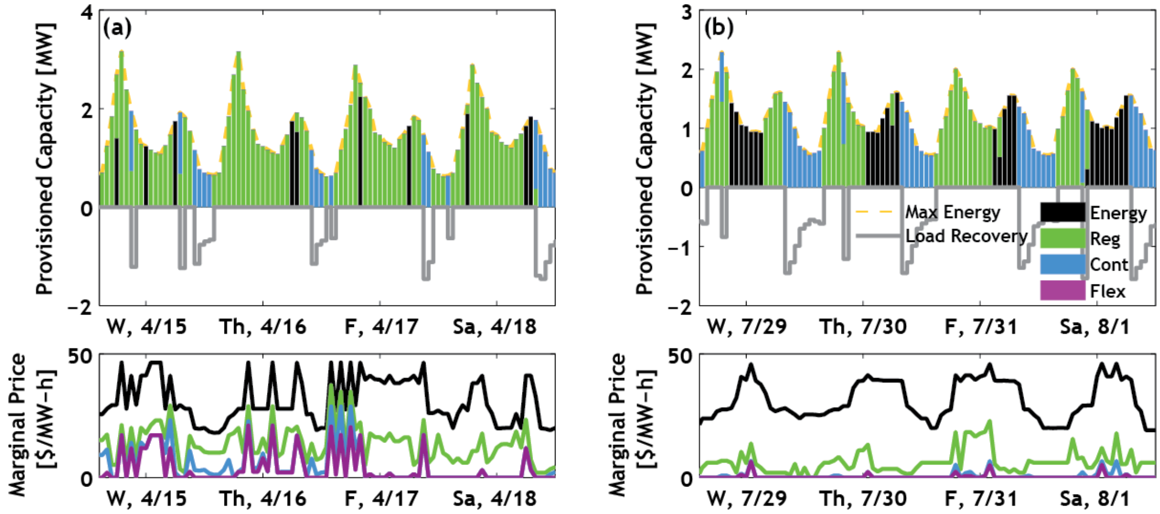

Figure 12. Residential water heating DR in PSCo (upper): Hourly capacity provision for (a) spring and (b) summer; hourly marginal price of energy and operating reserves (lower)

3.5 Operational Value of Demand Response

3.5.1 System Value

The operational value of DR is derived from two sources. First, when shifting energy, it can reduce use of the highest cost generation units. Second, while providing reserves it can reduce the use of less efficient partially loaded thermal generators, as well as the variable cost associated with providing regulation services from conventional generators.

Figure 13 shows the impact of DR energy shifting on mean daily load patterns for spring and summer. In the spring, DR usually shifts load from daytime to overnight hours. In the summer, DR primarily reduces system load between 1 p.m. and 8 p.m., while load recovery occurs overnight and into the early morning hours. The shift in load is much more dramatic in the summer, swinging between a load shed of 100 MW to a load recovery of 50 MW.

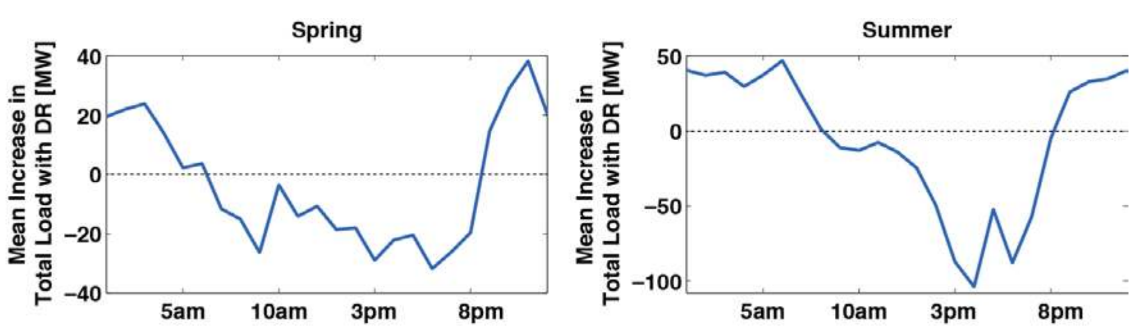

Figure 13. Mean increase in the daily load profile for spring and summer when DR provides energy. Negative numbers are a reduction in total system load, positive numbers are an increase in load.

Figure 14 shows the distribution of reserves provided by the different types of generation. On an annual basis, DR provides about 27% of regulation reserves and 8% of contingency reserves. DR provides very little flexibility reserves; the sum of all DR providing reserves is limited by the largest grid service availability, and generally all of the DR capacity is allocated to energy, regulation reserves, and contingency reserves.

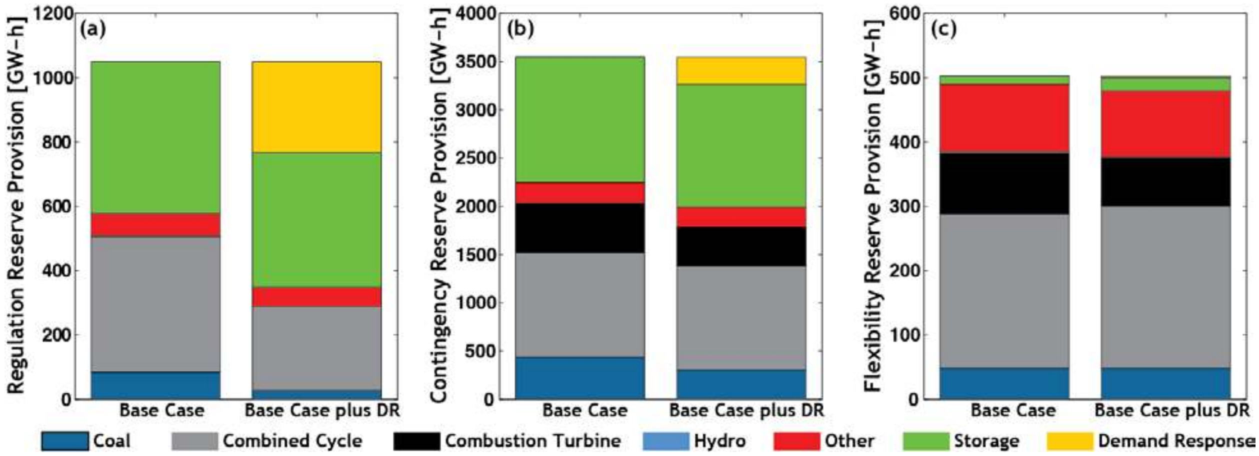

Figure 14. Reserve provision (a) regulation, (b) contingency, and (c) flexibility for the base case and when DR resources are available in the system for reserves

The overall impact of the services provided by 293 MW of peak available DR is summarized in Table 11. DR acts to reduce generation from the highest cost gas-fired generation and shift this generation to lower cost resources. In particular, DR reduced the use of the lowest efficiency power plants (combustion turbines) by about 10%.

Table 11. Comparison of the Test System Energy Results When Demand Response Resources are Available

	Without Demand Response	With Demand Response	Increase (Absolute / %)
Generation (GWh)			
Coal	45,981	46,111	130 / 0.3%
Gas Combined Cycle	14,741	14,637	-104 / -0.7%
Gas Combustion Turbine	1,199	1,075	-124 / -10.3%
Hydro	3,792	3,792	0 / 0%
Pumped Hydro Storage (PHS)	1,037	1,013	-24 / -2.3%
Wind[a]	10,705	10,705	0 / 0%
PV	1,834	1,834	0 / 0%
Other	92	92	0 / 0%
Demand Response	-	117	-
Total Generation (GWh)[b]	**79,381**	**79,376**	**-5 / 0%**
Fuel Use (1,000 MMBTU)			
Coal	486,467	487,870	1,403 / 0.3%
Gas	128,473	126,195	-2,277 / -1.8%
Total Fuel Use	**614,939**	**614,065**	**-874 / -0.1%**

[a] Neither wind nor PV experienced curtailment in these cases.
[b] The difference in generation is associated with the additional losses that occur from changes in the use of the pumped hydro storage.

The shift in fuel use resulting from DR deployment translates into annual fuel savings of about $7 million, or 0.6%. Table 12 summarizes the changes between the base case and system with DR in the four cost categories tracked in the PLEXOS model. The dominant source of benefit is associated with reduced fuel use; however, there is also a significant savings associated with avoided regulation costs. This cost is associated with decreased non-steady state operation of conventional generators providing regulation.

The addition of 293 MW of DR in the test system reduced total production cost by about 0.6%. Because DR provides a large fraction of the total regulation requirement it also reduces the total costs associated with regulation "bid" costs by about 37%. This result will change if DR resources bid in an equivalent wear and tear cost associated with DR providing regulation services.

Table 12. Production Costs for Base Case and Base Case With Demand Response Providing Energy and Operating Reserves

Production Cost [M$]	Base Case	Base Case with DR	Decrease in Cost With DR (M$ / % of Base)
Fuel Cost	1,215.0	1,208.0	7.0 / 0.6%
Variable O&M Cost	151.8	152.2	-0.4 / -0.3%
Start and Shutdown Cost	58.4	58.7	-0.4 / -0.6%
Regulation Reserve Bid Price	4.5	2.9	1.7 / 36.8%
Total Generation Cost	**1,429.7**	**1,421.8**	**7.9 / 0.6%**

3.5.2 Value of Individual DR Products

The value of DR provided in Table 12 represents the aggregated value of all DR products. Overall, dividing the $7.9 million in production cost savings by the non-coincident enabled DR capacity available, 293 MW, yields a value of $26.91/kW-year of DR available capacity. An alternative metric is the value of DR in terms of the annual availability, which is calculated by finding the sum of the maximum hourly availability across all DR resources. For the Colorado test system, the annual availability is 725 GW-h, and the value of the availability is $10.90/MW-h. However, this does not provide any insight into the relative value of different DR products or the value in a market setting. Furthermore, DR resources can have both operational value and capacity value. The PCM determines only the operational value. Determining the operational value of individual DR products would require separate simulations, each comparing individual or combinations of various DR. However, some insight can be gained by examining the provision of each service and the corresponding price.

Table 13 summarizes the market revenue for each DR product, under the assumption that the marginal energy and reserve prices produced by PLEXOS represent the market-clearing price in a restructured market.[18] The revenue of each DR resource is calculated by taking the hourly energy schedule or reserve provision times the hourly marginal price for each service. The revenue due to energy also accounts for the increase in load at another time; thus, it is the revenue from buying energy when the marginal price is low and selling it back to the grid when the marginal price of energy is higher. All DR resources have a no-cost bid for operating reserves.

[18] This implicitly assumes that the marginal prices generated by a PCM are equal to the marginal prices generated in a market setting. This is an important and potentially significant limitation when comparing the value of storage in a vertically integrated utility and a restructured market. PCMs do not typically include generator bidding and other factors that could drive market prices much higher. The results presented here are unlikely to represent the true difference between DR value in a market and non-market setting. However, they do represent some of the general challenges associated with value capture by DR associated with generator starts and price suppression.

Table 13. Annual Demand Response Energy and Reserve Provision (GWh or GW-h) and Revenue (M$)

DR Resources	Energy	Regulation	Contingency	Flexibility	Total
	Scheduled/Revenue (GWh/M$)	Provision/Revenue (GW-h/M$)			Revenue [M$]
Residential Cooling	22.4 / 0.037	48.5 / 0.400	23.4 / 0.115	0.1 / 0.000	0.553
Residential Water Heating	2.6 / 0.038	10.3 / 0.139	2.7 / 0.007	0 / 0.000	0.185
Commercial Cooling	6.5 / 0.004	0.2 / 0.002	5.5 / 0.026	0.2 / 0.001	0.032
Commercial Heating	0.7 / 0.011	-	-	-	0.012
Commercial Lighting	-	8.6 / 0.099	16.3 / 0.098	0.8 / 0.002	0.199
Commercial Ventilation	-	9.8 / 0.111	18.6 / 0.109	1 / 0.002	0.222
Municipal Pumping	1.7 / 0.042	-	-	-	0.042
Wastewater Pumping	2.5 / 0.062	-	-	-	0.062
Outdoor Lighting	-	204.6 / 2.073	0.8 / 0.005	-	2.078
Refrigerated Warehouses	0.3 / 0.005	-	-	-	0.005
Agricultural Pumping	68.9 / 0.723	-	155.1 / 0.695	-	1.418
Data Center	11.3 / 0.207	-	59 / 0.342	-	0.548
Total DR	**116.8 / 1.129**	**282.1 / 2.824**	**281.5 / 1.398**	**2.1 / 0.005**	**5.355**

Table 13 also shows the annual allocation of each DR resource across energy and operating reserves, expressed as either gigawatt-hours for energy or the provision of a gigawatt of reserves capacity for a scheduled hour (GW-h). Overall, DR capacity is primarily allocated for reserves; 17% of the allocated capacity provides energy, 41% provides regulation, 41% provides contingency, and less than 1% for flexibility. Outdoor lighting and residential cooling make up 90% of the DR provision for regulation. Agricultural pumping and data centers make up 76% of the DR provision for contingency.[19] Regulation services cost about twice as much as contingency; thus, the revenue for DR from regulation is about 53% of the total revenue for DR.

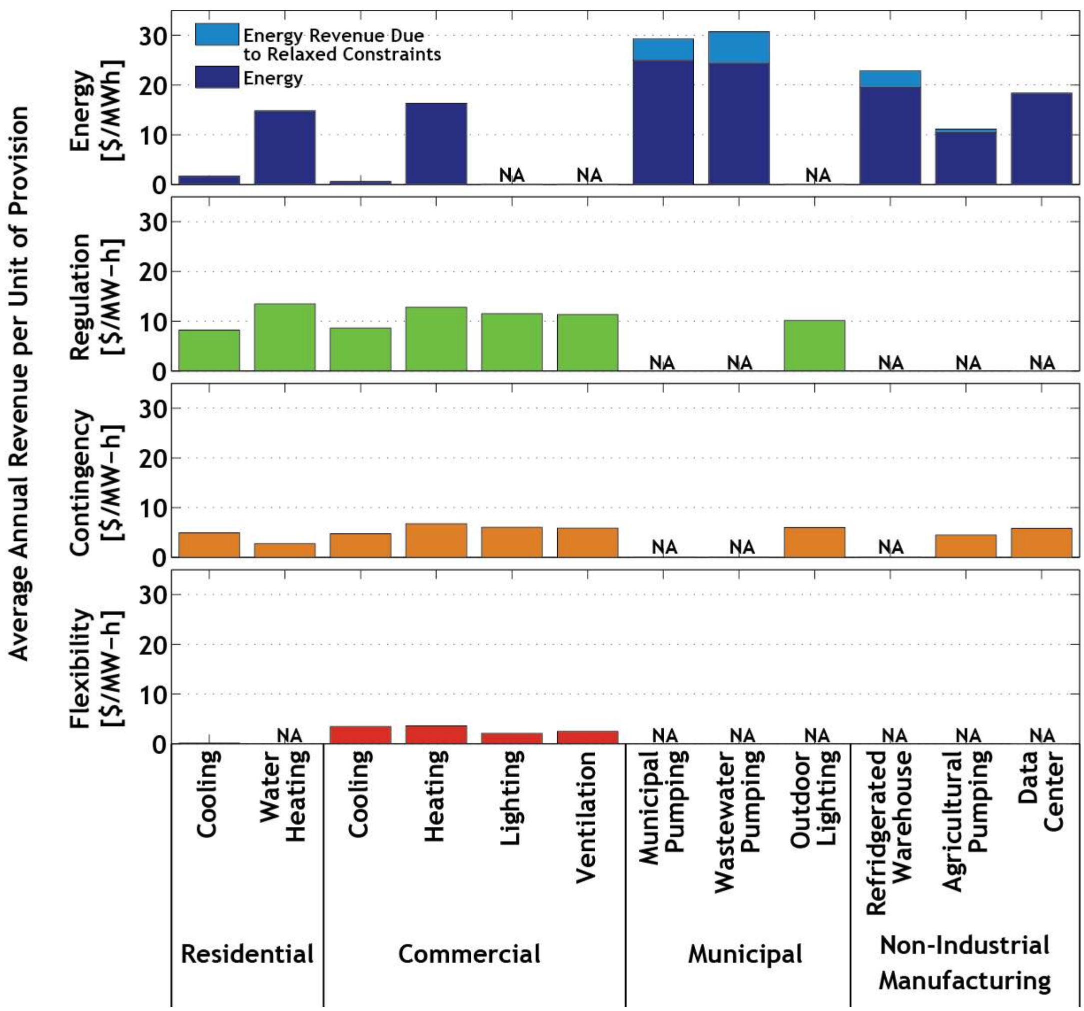

Figure 15. Annual average revenue per unit of generation or reserves for each DR resource

Figure 15 shows the annual average revenue for each grid service by each DR resource. The revenue per unit of reserve provision is fairly constant across DR resources, despite that the availability of DR for reserves is highly time varying. The variance in the value of energy from DR is much larger. The cooling resources are arbitraged within daytime hours, limiting the net

[19] Large industrial loads were not considered in this study but have historically been used to provide spinning contingency services.

revenue. Commercial heating resources are also arbitraged within daytime hours; however, the difference between the average highest and lowest price hour during the daytime in the winter is $19.0/MWh while the summer difference is $11.8/MWh. The municipal and non-industrial manufacturing resources are not constrained by the time of day that the generation must be recharged. The average daily difference between the highest and lowest priced hours, across all seasons, is $20.0/MWh.

The total revenue of $5.4 million is less than the $7.9 million in production cost savings largely due to the price-suppression effects of DR on the reserves prices.[20] This is demonstrated in Table 14, where the addition of DR reduces the mean price of regulation by about 23% and the mean price of contingency reserve by about 20%. This presents a significant challenge to DR in market environments as demonstrated previously by Denholm et al. [3]. See the appendix for price duration curves for regulation, contingency, and flexibility reserves.

Table 14. Marginal Reserve and Energy Price When Demand Response Resources are Available in the Test System

Service	System without DR	System with DR	DR Contribution to Annual Reserve and Energy Requirement
	Median/Mean ($/MW-h)		
Regulation	13.4 / 14.72	8.14 / 11.31	282.1 GW-h / 26.9%
Contingency	4.34 / 7	2.54 / 5.62	281.5 GW-h / 7.9%
Flexibility	0 / 3.27	0 / 2.75	2.1 GW-h / 0.4%
Energy	28.96 / 31.97	28.98 / 31.39	116.8 GWh / 0.1%

Figure 16a translates the total revenue for the aggregated DR products into a value per kilowatt of enabled capacity. It includes the various grid service components of this value. In addition to the energy and reserves categories, it includes a value associated with penalties—several types of DR resource were permitted to exceed their daily operation constraints by incurring a penalty cost (see Table 3 for operation constraints and penalties). Agricultural pumping and wastewater pumping both exceeded the constrained maximum available energy provision, which increases their revenue from energy by about 7% and 35%, respectively.

[20] In addition to price suppression, DR has the potential to reduce start costs that would typically not be captured in a market environment. This is observed and discussed previously by Denholm et al. (2013).

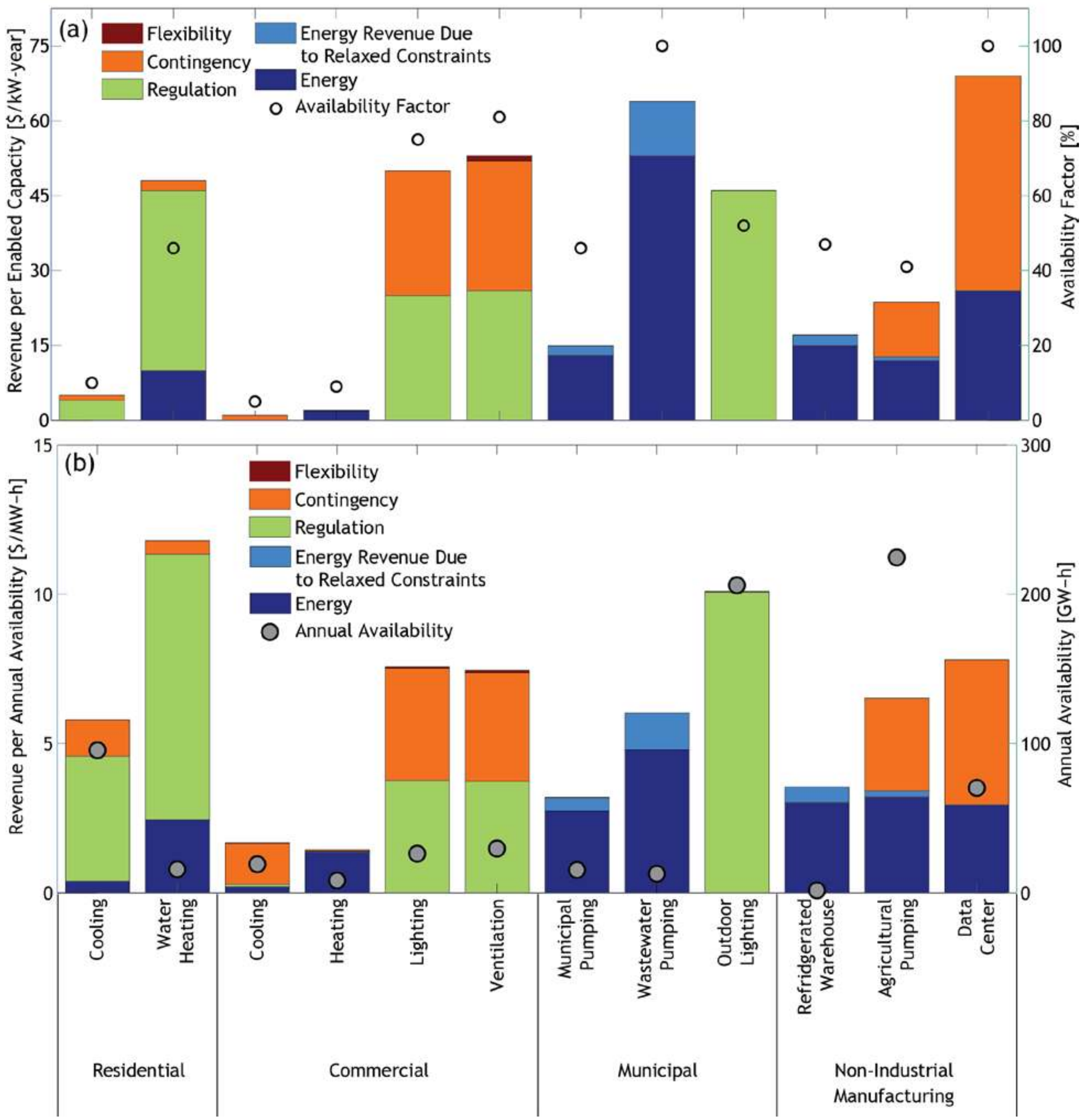

Figure 16. Average annual revenue (left axis) from the day-ahead market per (a) total enabled capacity and (b) annual availability for each type of demand response resource in the Colorado test system. Annual DR resource availability (right axis) is expressed as the (a) annual availability factor or as (b) total annual availability.

The overall value per kilowatt of each service (Figure 16a) is largely determined by the availability factor, which relates to the enabled capacity discussed in Section 2.3.2 and illustrated in Figure 2. The scaling factor applied to create the enabled capacity effectively "derates" each DR resource by a different amount. This produces the large difference between certain products, such as residential cooling and water heating. Dividing by the enabled rated capacity, which removes this scaling factor would yield a different set of revenue per kilowatt values. For instance, the enabled capacity of residential cooling is 109 MW and the enabled rated capacity is

310 MW (see discussion is Section 2.3.2), while the enabled capacity of residential water heaters is 4 MW and the enabled rated capacity is 223 MW [see Eq. (1), using a duty cycle of 7% and a sheddability of 25%]. This is illustrated in Figure 17, where we show the available capacity relative to the enabled rated capacity. The revenue per enabled rated capacity for residential cooling and water heating is \$3.1/kW-year and \$0.7/kW-year, respectively. Alternately, utilities often measure DR participation in terms of number of customers or devices that are enabled as opposed to the capacity enabled. The 310 MW of residential cooling is captured in approximately 130,000 units, while the 223 MW of residential water heating is captured in approximately 50,000 units. This yields a value per unit-year of \$7.4/unit-year and \$3.3/unit-year for residential cooling and water heating, respectively.

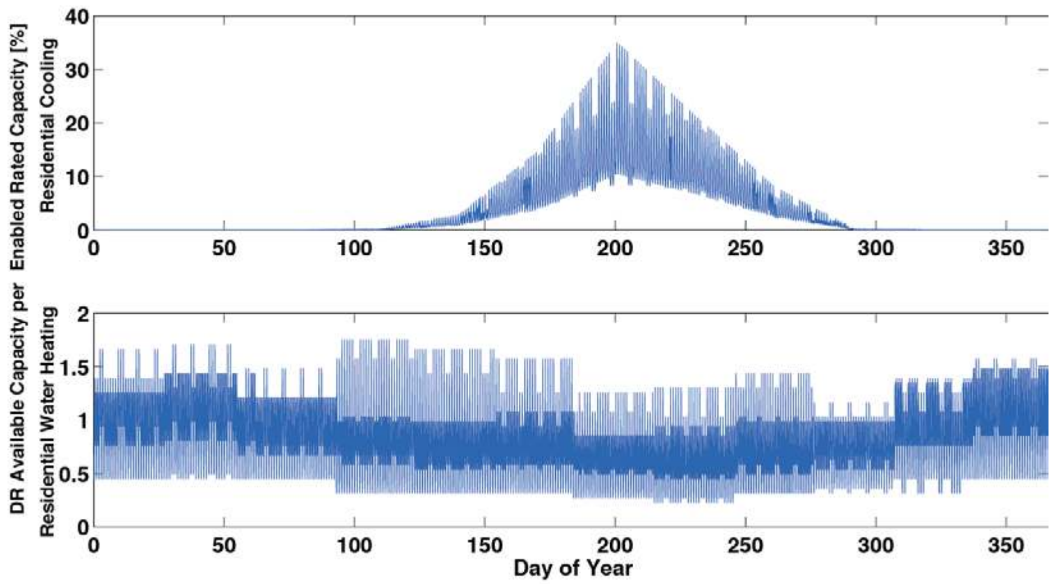

Figure 17. Illustration of the DR available capacity, scaled by the enabled rated capacity for (a) residential cooling and (b) residential water heating in the Colorado test system

Figure 16b shows the total annual revenue per annual availability. The annual availability is the sum of the maximum hourly availability across all of the grid services for each DR resource. The annual availability is the fraction of electricity from an aggregation of end uses that is flexible through DR. This metric reflects the correlation of each resource's availability to times of high market prices for operating reserves as well as its ability to take advantage of large energy price differences across hours of the day. Those resources with lower correlations and more constraints on energy shifting will tend to have lower values on a cumulative availability basis. Dividing by the annual availability instead of the enabled capacity increases the value of residential cooling relative to water heating. This is because cooling resources have a strong seasonal dependence that correlates well with peak demand and high energy and operating reserve prices, while water heating is relatively constant throughout the year (see Figure 17).

Finally, it should be emphasized that these values only consider the operational value of the DR device and do not consider capacity value. Many existing DR programs focus primarily on the capacity value of DR. For example, Xcel Energy offers \$40/year for each residential cooling unit with a Saver Switch, which is used to curtail cooling loads during critical peak periods [38].

Wisconsin offers $32/year ($8/month for June, July, and August) for direct load control of residential air conditioning and $24/year ($2/month) for direct load control of residential electric water heating [39]. The difference in capacity value payment is due to the time-varying availability of that capacity: a single residential cooling unit provides more capacity during peak load hours than a single residential electric water heater.

4 Conclusion

We demonstrated a method for modeling an energy-limited DR resource that both explicitly accounts for the constraints on the operation DR and co-optimizes DR capacity between providing energy and operating reserves. This approach enables DR to participate in multiple grid service markets, while price responsive demand curves are limited to providing one grid service. Each DR resource model was customized using capacity and ramping availability profiles and operating constraints in Part 1 of this report series [1]. Generation from variable and uncertain resources continues to grow, increasing the operating reserves necessary to meet intra-hour variability and multi-hour forecast error for net load. This report demonstrates that DR resources can provide a significant fraction of operating reserves and reduce peak demand.

Our approach yielded production cost savings in the Colorado test system of about $27/kW-year for the 293 MW of non-coincident peak DR capacity, or an average value of about $11/MW-h for the 725 GW-h of annual availability. Most of this production cost savings came from DR displacing natural-gas-fired combustion turbines from providing peak energy and contingency reserves. The revenue per kilowatt of enabled DR capacity (annual peak available capacity) varies significantly across the resources from less than $1/kW-year to more than $65/kW-year. Across all DR resources, only 20% of the revenue came from the energy market, while more than 50% of revenue came from the regulation reserve market and the remainder from the contingency reserve market. This revenue calculation did not include capacity value, which may be of significantly higher value for many DR resources.

In order to assess the value of DR we used peak-enabled capacity. However, this may not fully reflect the value of these resources because it may not capture the cost of enabling the capacity. For some DR resources, the capital costs of enabling DR scales with the number of end-use loads enabled, not necessarily with the capacity of DR enabled. The value of DR on a per-unit basis can be significantly different from the value per kilowatt of available capacity. For instance, in the Colorado test system with a 15% renewable penetration, residential cooling and water heating have a value per enabled kilowatt of $5/kW-year and $47/kW-year, respectively. However, on a per-unit basis, the value is $7.4/unit-year and $3.3/unit-year, respectively. Again, these values only reflect their operation in the Colorado test system and neglect the capacity value of DR.

Additional analysis is needed to understand the impact of renewables on several aspects of DR operation and value. The provision of down reserves from DR was not evaluated in this study and might become more important at higher penetrations of variable renewable generation and thus present a source of revenue for DR. Greater wind and solar penetration may also increase the actual deployment of regulating and flexibility reserves, changing the operation of DR. This analysis only included day-head hourly modeling. Sub-hourly analysis of DR performance is needed to evaluate the impact of the DR control signal on end-use performance. For instance, under real-time operating conditions should DR be allowed to exceed its scheduled performance if the price of energy is higher than expected? Our research is the first step toward evaluating the economic potential of DR. A more thoroughly assessment of the potential should be pursued by iteratively modeling various DR penetrations to inform the incremental value of additional DR. Finally, this analysis points to the need to consider the capacity value of DR. Past work in storage analysis has demonstrated considerable capacity value, and existing programs are essentially paying for DR to provide firm capacity during peak operating periods.

In summary, DR can provide a significant source of operating reserves when co-optimized between availability for reserves and energy. Using operating constraints, such as maximum operating time and requiring pre-/re-charge for energy use from DR, yields realistic DR performance without the use of a price responsive supply curve. These modeling methods can be incorporated into renewable penetration integration studies as a source of flexibility for the system.

References

[1] Olsen, D. J.; Kiliccote, S.; Matson, N.; Sohn, M.; Rose, C.; Dudley, J.; Goli, S.; Hummon, M.; Palchak, D.; Denholm, P.; Jorgenson, J.; Ma, O. (September 2013). *Grid Integration of Aggregated Demand Response, Part 1: Load Availability Profiles and Constraints for the Western Interconnection*. LBNL-6417E. Berkeley, CA: Lawrence Berkeley National Laboratory.

[2] Starke, M.; Alkadi, N.; Ma, O. (2013). *Assessment of Industrial Load for Demand Response across U.S. Regions of the Western Interconnect*. Technical report ORNL/TM-2013/407. Oak Ridge, TN: Oak Ridge National Laboratory.

[3] Denholm, P.; Jorgenson, J.; Hummon, M.; Jenkin, T.; Palchak, D.; Kirby, B.; Ma, O.; O'Malley, M. (2013). *Value of Energy Storage for Grid Applications*. TP-6A20-58465. Golden, CO: National Renewable Energy Laboratory.

[4] Denholm, P.; Hummon, M. (2012). *Simulating the Value of Concentrating Solar Power with Thermal Energy Storage in a Production Cost Model*. TP-6A20-56731. Golden, CO: National Renewable Energy Laboratory.

[5] PSCo. (October 2011). *2011 Electric Resource Plan. Volume II Technical Appendix.* Denver, CO: Public Service Company of Colorado.

[6] CAISO. (April 2012). *2011 Annual Report on Market Issues and Performance*. Technical report. Folsom, CA: California Independent System Operator.

[7] Holttinen, H. (2013). *Wind RP16. Task 25. WIND INTEGRATION STUDIES*. Technical report. Washington, DC: International Enery Agency (IEA).

[8] Ela, E.; Milligan, M.; Kirby, B. (2011). *Operating Reserves and Variable Generation*. NREL/TP-5500-51978. Golden, CO: National Renewable Energy Laboratory.

[9] Navid, N. (2012). "Challenges and Opportunities to Bring Prices to Devices." *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*.

[10] Xu, L.; Tretheway, D. (2012). *Flexible Ramping Products Draft Final Proposal.* Technical report. Folsom, CA: California Independent System Operator.

[11] Ibanez, E.; Brinkman, G.; Hummon, M.; Lew, D. (2012). "A Solar Reserve Methodology for Renewable Energy Integration Studies Based on Sub-hourly Variability Analysis." *2nd Annual International Workshop on Integration of Solar Power into Power Systems Conference, 12-13 November 2012, Lisbon, Portugal*.

[12] Hummon, M.; Denholm, P.; Jorgenson, J.; Kirby, B.; Ma, O.; Palchak, D. (2013). *Fundamental Drivers of Operating Reserve Cost in Electric Power Systems.* TP-6A20-58491. Golden, CO: National Renewable Energy Laboratory.

[13] Cappers, P.; MacDonald, J.; Goldman, C. (2013). *Market and Policy Barriers for Demand Response Providing Ancillary Services in U.S. Markets.* LBNL-6155E. Berkeley, CA: Lawrence Berkeley National Laboratory.

[14] Conejo, A. J.; Morales, J. M.; Baringo, L. (2010). "Real-Time Demand Response Model." *IEEE Transactions on Smart Grid* (1:3); pp. 236–242.

[15] Dietrich, K.; Latorre, J. M.; Olmos, L.; Ramos, A. (2012). "Demand Response in an Isolated System With High Wind Integration." *IEEE Transactions on Power Systems* (27:1); pp. 20–29.

[16] Oldewurtel, F.; Ulbig, A.; Parisio, A.; Andersson, G.; Morari, M.(2010). "Reducing Peak Electricity Demand in Building Climate Control Using Real-Time Pricing and Model Predictive Control." *2010 49th IEEE Conference on Decision and Control (CDC 2010),* pp. 1927–1932.

[17] Alizadeh, M.; Wang, Z.; Scaglione, A.; Chen, C.; Kishore, S. (2012). "On the Market Effects of Queueing Energy Requests as an Alternative to Storing Electricity." *2012 IEEE Power & Energy Society General Meeting. New Energy Horizons - Opportunities and Challenges.*

[18] Chen, C.; Kishore, S.; Wang, Z.; Alizadeh, M.; Scaglione, A. (2012). "A Cournot Game Analysis on Market Effects of Queuing Energy Request as Demand Response. *2012 IEEE Power & Energy Society General Meeting. New Energy Horizons - Opportunities and Challenges.*

[19] Negnevitsky, M.; Nguyen, D. T.; de Groot, M. (2012). "Load Recovery in Demand Response Scheduling." *Power and Energy Society General Meeting, 2012 IEEE,* pp. 1-8

[20] Nottrott, A.; Kleissl, J.; Washorn, B. "Energy Dispatch Schedule Optimization and Cost Benefit Analysis for Grid-Connected, Photovoltaic-Battery Storage Systems." *Renewable Energy* (55); pp. 230–240.

[21] Parvania, M.; Fotuhi-Firuzabad, M. (2010). "Demand Response Scheduling by Stochastic SCUC." *IEEE Transactions on Smart Grid* (1:1); pp. 89–98.

[22] Aalami, H. A.; Moghaddam, M. P.; et al. (2010). "Demand Response Modeling Considering Interruptible/Curtailable Loads and Capacity Market Programs." *Applied Energy* (87:1); pp. 243–250.

[23] Reid, E.; Gerber, S.; Adib, P. (2009). "Integration of Demand Response Into Wholesale Electricity Markets." *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES.*

[24] Callaway, D. S. (2009). "Tapping the Energy Storage Potential in Electric Loads to Deliver Load Following and Regulation, With Application to Wind Energy." *Energy Conversion and Management* (50:5); pp. 1389–1400.

[25] Hobbs, B. F.; Hu, M.-C.; Inon, J. G.; Stoft, S. E.; Bhavaraju, M. P. (2007). "A Dynamic Analysis of a Demand Curve-Based Capacity Market Proposal: The PJM Reliability Pricing Model." *IEEE Transactions on Power Systems* (22:1); pp. 3–14.

[26] Faria, P.; Vale, Z.; Soares, T.; Morais, H. (2012). "Energy and Reserve Provision Dispatch Considering Distributed Generation and Demand Response." *3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe).*

[27] Sioshansi, R.; Denholm, P. (2010). *The Value of Concentrating Solar Power and Thermal Energy Storage.* TP-6A20-45833. Golden, CO: National Renewable Energy Laboratory.

[28] Piette, M. A.; Kiliccote, S.; Dudley, J. H. (2012). "Field Demonstration of Automated Demand Response for Both Winter and Summer Events in Large Buildings in the Pacific Northwest." *Energy Efficiency* (6); pp. 671–684.

[29] Marks, G.; Wilcox, E.; Olsen, D.; Goli, S. (2013). *Opportunities for Demand Response in California Agricultural Irrigation: A Scoping Study*. LBNL-6108E. Berkeley, CA: Lawrence Berkeley National Laboratory.

[30] TEPPC. (September 2011). *TEPPC 2010 Study Program 10-Year Regional Transmission Plan*. Technical report. Salt Lake City, UT: Western Electricity Coordinating Council Transmission Expansion Planning Policy Committee.

[31] EIA. (June 2013). *Electric Power Monthly with Data.* Technical report. Washington, DC: Energy Information Administration.

[32] Gilman, P.; Dobos, A. (2012). *System Advisor Model, SAM 2011.12.2: General Description*. TP-6A20-53437. Golden, CO: National Renewable Energy Laboratory.

[33] GE. (2010). *Western Wind and Solar Integration Study*. SR-550-47434. Golden, CO: National Renewable Energy Laboratory.

[34] EIA. (December 2012). *Electric Power Monthly with Data*. Technical report. Washington, DC: Energy Information Administration.

[35] CAISO. (2011). *Summary of Preliminary Results of 33% Renewable Integration Study -- 2010*. Technical report. CPUC LTPP Docket No. R.10-05-006. San Francisco, CA: California Public Utility Commission.

[36] PJM. (2012). *PJM Manual 15: Cost Development Guidelines Rev. 20.* Technical report. Norristown, PA: PJM.

[37] EIA. (June 2013). *Natural Gas Monthly: Office of Oil, Gas, and Coal Supply*. Technical report. Washington, DC: Energy Information Administration.

[38] Xcel Saver Switch. Accessed September 13, 2013: http://www.xcelenergy.com/Save_Money_&_Energy/Find_a_Rebate/Saver's_Switch_for_Residences_-_CO.

[39] WPS. "Wisconsin Public Service Cool Credits Direct Load Control Program." Accessed December 10, 2013: http://www.wisconsinpublicservice.com/home/cool_credits.aspx.

Appendix

Additional Examples of Demand Response Performance in the Test System

Each DR resource has a separate time series that defines the capacity available for energy shifting, charging, and each type of ancillary service. This section demonstrates the annual daily average availability and performance, as well as two snapshots (spring and summer) of hourly performance for each DR resource. The DR resources are organized by their modeling characteristics: thermal resources, pumping resources, and resources that only provide operating reserves. In most cases we present data for the DR resource installed in the PSCo balancing area. Section 3 explores the aggregate effects of DR-providing energy and operating reserves in PSCo and WACM.

Residential Cooling

DR from residential cooling end use peaks in the summer and is zero from October to April (Figure 18a). The load shed (in order to provide generation) must be recovered in the form of pre- or post-increased cooling of the residence between 6 a.m. and 6 p.m. (time is reported in PST). The energy use is a smaller fraction of total provision than regulation and is about equal to contingency. Two factors influence this co-optimization. First, in the PCM, the energy must be recovered (net-zero over a 24-hour period), while capacity held for reserves does not need to be "paid back." Second, the response time to full load shed is 1 minute for contingency and regulation and 15 minutes for energy. Therefore, there is more capacity available for contingency and regulation reserve provision than there is available for energy.

The average daily profile (Figure 18b) shows that the residential cooling DR is optimally dispatched in a pre-cooling pattern. This resource, as well as all other resources that provide contingency as a service, are allowed to hold contingency reserves during charging because the additional load could be shed in a contingency event. The residential cooling DR also holds only a tiny fraction of its capacity for flexibility reserves. This is likely due to the fact that flexibility reserves in the test system can be provisioned by generators at no opportunity cost (and therefore has zero marginal price) 40% of the time. This reduces the incentive for other zero marginal cost resources to provide flexibility capacity.

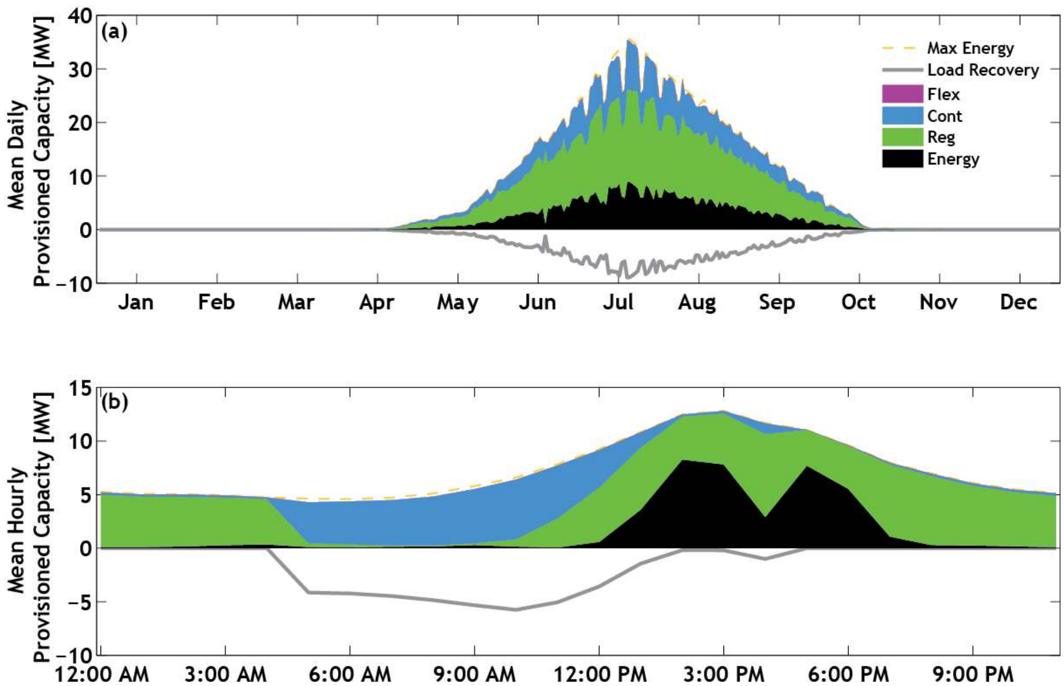

Figure 18. Residential cooling demand response in PSCo: Mean (a) daily and (b) hourly provision of capacity

Figure 19 shows the hourly allocation of capacity for the residential cooling DR in the spring (at the tail of the available capacity window) and summer (two orders of magnitude greater than spring, at the peak of the available capacity window), as well as the marginal price of energy and operating reserves (lower panels). In the summer the residential cooling DR resource is usually used for energy between the hours of 1 p.m. and 5 p.m. and shifts that energy to the late evening and early morning hours (when the price of energy is lower). It also holds capacity for contingency events while recovering load. During all other hours the residential cooling DR capacity is held for regulation.

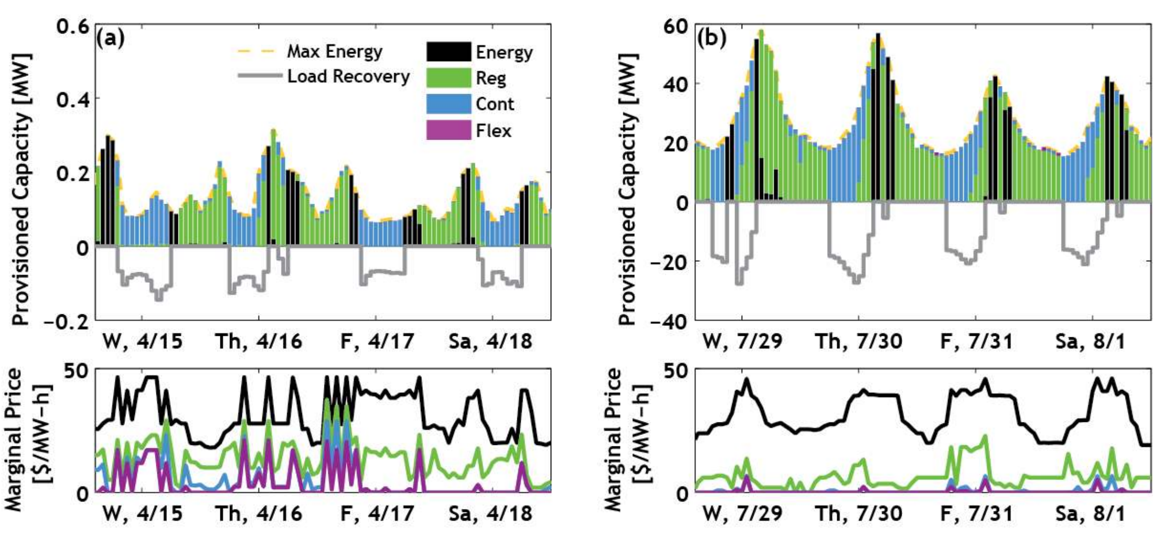

Figure 19. Residential cooling demand response in PSCo: (upper) Hourly capacity provision for (a) spring and (b) summer and (lower) hourly marginal price of energy and operating reserves

Residential Heating

Residential heating is the only thermal process that does not accrue shedded load that must be recharged by the system. It is also the only DR resource that is not available in the test system because residential heating is primarily sourced from natural gas rather than electricity. In areas where the residential heating is sourced from the grid, the available energy shedding capacity is available for one hour per day. The data shown in Figure 20 is from the northwest United States. The seasonal availability follows the expected pattern, going to zero over the summer months, with slightly more capacity available on the weekends. The energy schedule is largely driven by price because there is no charging. The total capacity available for regulation provision is less than the maximum availability for energy; when residential heating is not used for energy it is nearly always provisioned for regulation. The capacity availability for regulation is less than the capacity available for energy.

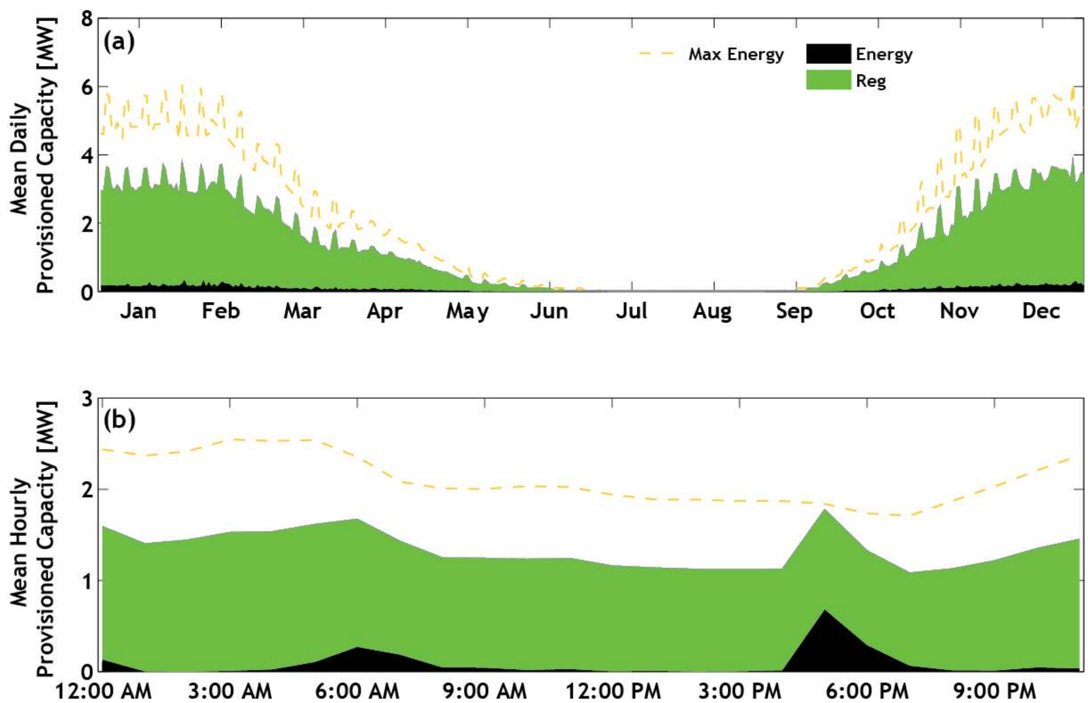

Figure 20. Residential heating demand response in the Northwest: Average (a) daily and (b) hourly provision of capacity; the daily capacity is zero between mid-June and late-August, when residences do not use electricity for heating

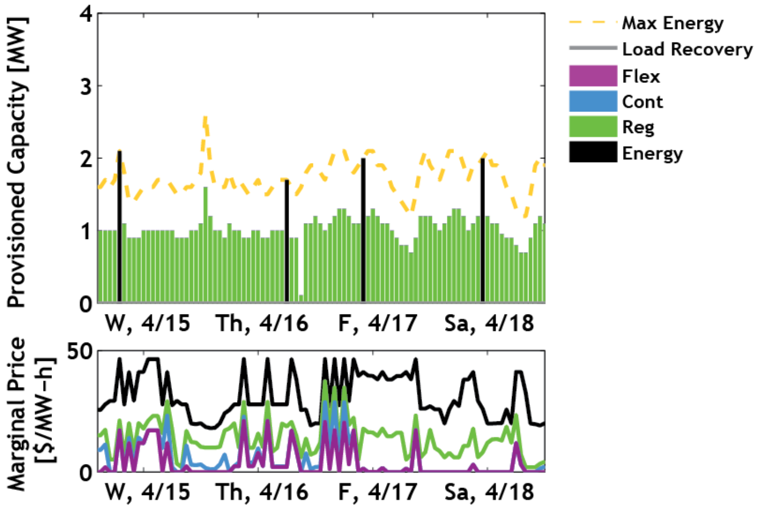

Figure 21. Residential heating demand response in the Northwest: (upper) Hourly capacity provision for April 15–April 18 and (lower) hourly marginal price of energy and operating reserves. There is no residential heating capacity available for demand response in the summer.

Commercial Cooling

DR originates from commercial cooling peaks in the summer and is zero from October to April (Figure 22a). The load shed for energy must be recovered in the form of pre- or post-cooling of the commercial building between 6 a.m. and 6 p.m. The energy use is a smaller fraction of total provision than contingency. This is primarily driven by the operation constraint that the energy must be recovered, while capacity held for reserves does not need to be re-charged.

The average daily profile (Figure 22b) shows that the optimized commercial cooling operates through pre-cooling. The resource is allowed to hold contingency reserves when recovering load, similar to other energy shifting resources. The commercial cooling DR holds only a tiny fraction of its total capacity for regulation and flexibility reserves. This is because the capacity available for regulation from commercial cooling is about one-tenth the capacity available for contingency, and flexibility is largely covered by other generators at no opportunity cost to the system.

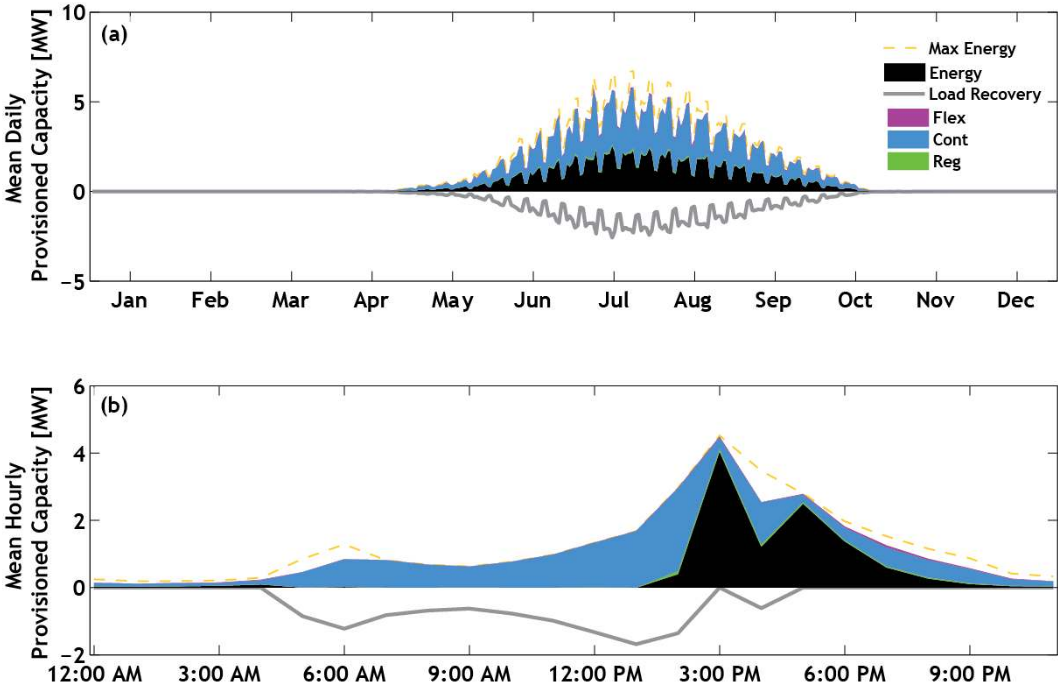

Figure 22. Commercial cooling demand response in PSCo: Mean (a) daily and (b) hourly provision of capacity

Figure 23 shows the hourly allocation of capacity for the commercial cooling DR in the spring (at the tail end of the available capacity window) and summer (two orders of magnitude greater than spring, at the peak of the available capacity window), as well as the marginal price of energy and operating reserves (lower panels). In the summer the commercial cooling DR resource is typically used for energy between the hours of 1 p.m. and 5 p.m. and shifts that energy to the late evening and early morning hours when the price of energy is lower. It also holds capacity for contingency reserve while paying back energy and during all other hours when the load shedding is less than the maximum energy available.

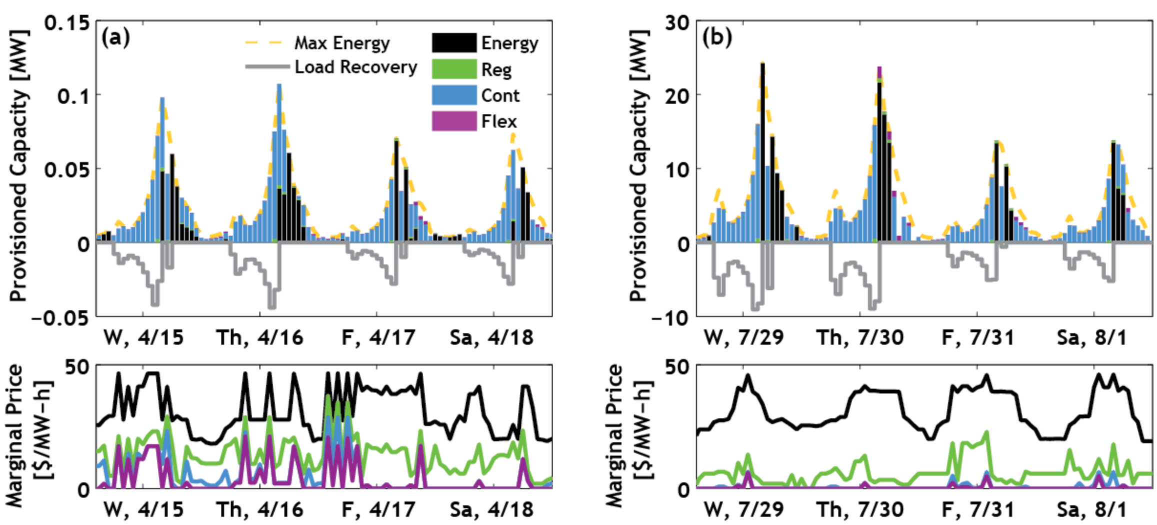

Figure 23. Commercial cooling demand response in PSCo: (upper) Hourly capacity provision for (a) spring and (b) summer and (lower) hourly marginal price of energy and operating reserves

Commercial Heating

Commercial heating DR is seasonally available, peaking in the winter, and is not available from June through September. It can provide energy shifting as well as contingency, flexibility, and regulation, although the operating reserves are less than 0.1 MW in all time periods and are therefore considered negligible and ignored for our simulations. WACM is used as the example here because there was not an adequate commercial electric heating resource identified for the PSCo region. Charging is constrained from 3 a.m. through 6 p.m. to be congruent with the heating patterns in a typical commercial setting, which decreases the opportunities for energy arbitrage. Charging constraints could depend on the types of buildings and vary by region and occupant behavior. Commercial heating typically runs for 2.5 hours per day and recovers load over 3.5 hours.

The average daily profile is shown in Figure 24b. The middle of the day offers minimal to no opportunities for energy shifting because of the direct effect on the comfort of the occupants. The combination of the average day maximum energy constraint shown in Figure 24b, along with the charging constraints, explains the relatively infrequent provision of capacity represented in Figure 25. There are only a small number of hours that arbitrage makes economic sense given the constrained number of capable provision and charging hours. Figure 25 (lower) also shows the marginal price of energy and operating reserves, which is the driver of arbitrage for the resource.

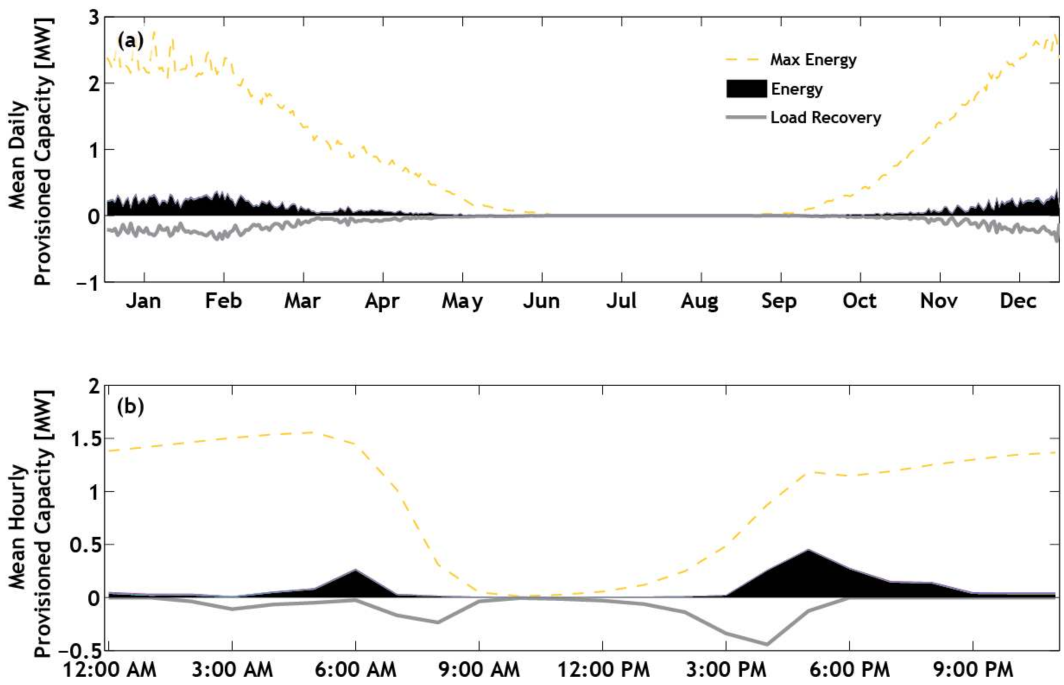

Figure 24. Commercial heating demand response in WACM: Mean (a) daily and (b) hourly provision of capacity

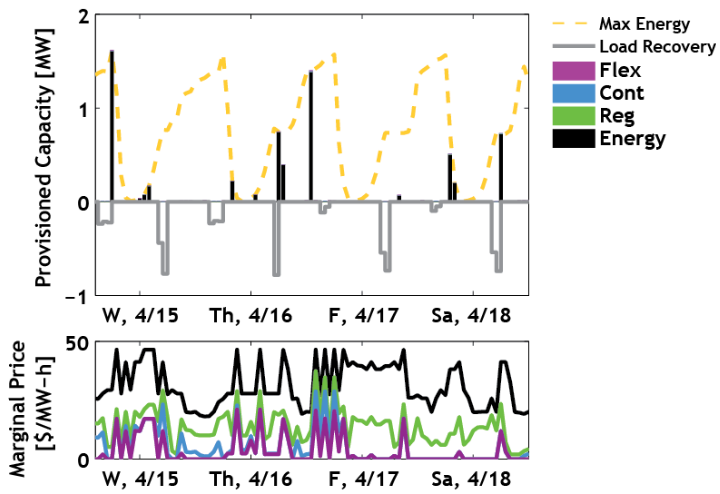

Figure 25. Commercial heating demand response in WACM: (upper) Hourly capacity provision for spring and (lower) hourly marginal price of energy and operating reserves (x-axis tick marks 12 p.m.)

Data Centers

Data centers are capable of providing contingency and energy shifting with a relatively predictable load and therefore a very uniform provision capability. The data centers are able to hold contingency reserves as much as the load profiles will allow, but the energy shifting has the added restriction that it can shed load for a maximum of 4 hours in a 24-hour period. This time restriction is in place because a large part of the load for a data center is cooling for the servers. Limiting the amount of time that the load can be shed keeps the temperature drift within an acceptable range. The ramping capability is the same as other thermal DR resources, such as commercial and residential cooling, with contingency reserve able to ramp to its maximum in one minute through direct load control (e.g., fans and pumps) and thermostat changes, which take 15 minutes to reach their maximum energy shifting capability. Figure 26 shows the annual provision of energy and contingency as (a) daily averages and (b) the mean hourly provision. While the availability of contingency and energy is equal for this resource in PSCo, the time restrictions and the required charging cause the average energy to be much lower than its ability to hold contingency reserve capacity, as can be seen by the discrepancy between the services in Figure 26.

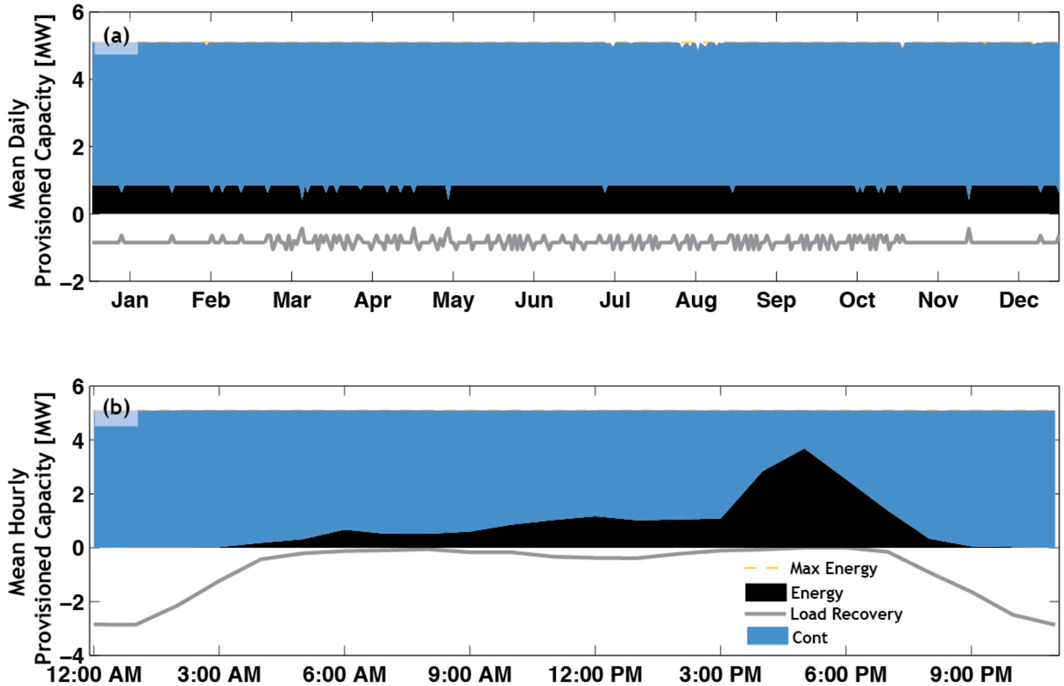

Figure 26. Data center demand response in PSCo: Mean (a) daily and (b) hourly provision of capacity

Figure 27 shows the DR operation of the data center over a 3-day period for the spring and summer, as well as the marginal prices for energy and all operating reserves in those periods. It can be gathered from the following figures that energy is often shed during the late afternoon/early evening, when prices are typically highest, and that charging is taking place during the night, when prices for energy are typically lowest. The energy shifting capability was utilized for the full four hour daily maximum 99% of the days in our annual simulation.

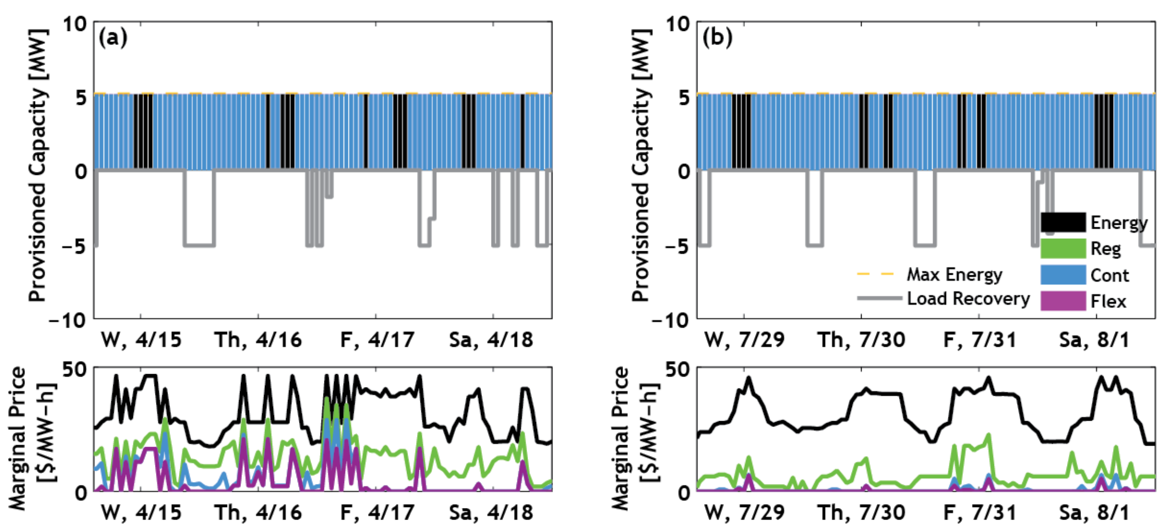

Figure 27. Data center demand response in PSCo: (upper) Hourly capacity provision for (a) spring and (b) summer and (lower) hourly marginal price of energy and operating reserves

Agricultural Pumping

DR from agricultural pumping is different than the rest of the resources because it can provide more contingency than energy shifting. This is a reflection of the participation rate that is expected from agricultural pumps, which is lower than the less-committing contingency reserve [1]. The resource is largely seasonal, as seen in Figure 28a, with the bulk of the capacity available in summer due to the growing season cycle. There are no restrictions on the time of day that energy can be shifted, which helps to explain the tendency for energy to be shed during the high priced hours in the middle of the day, as seen in Figure 28b.

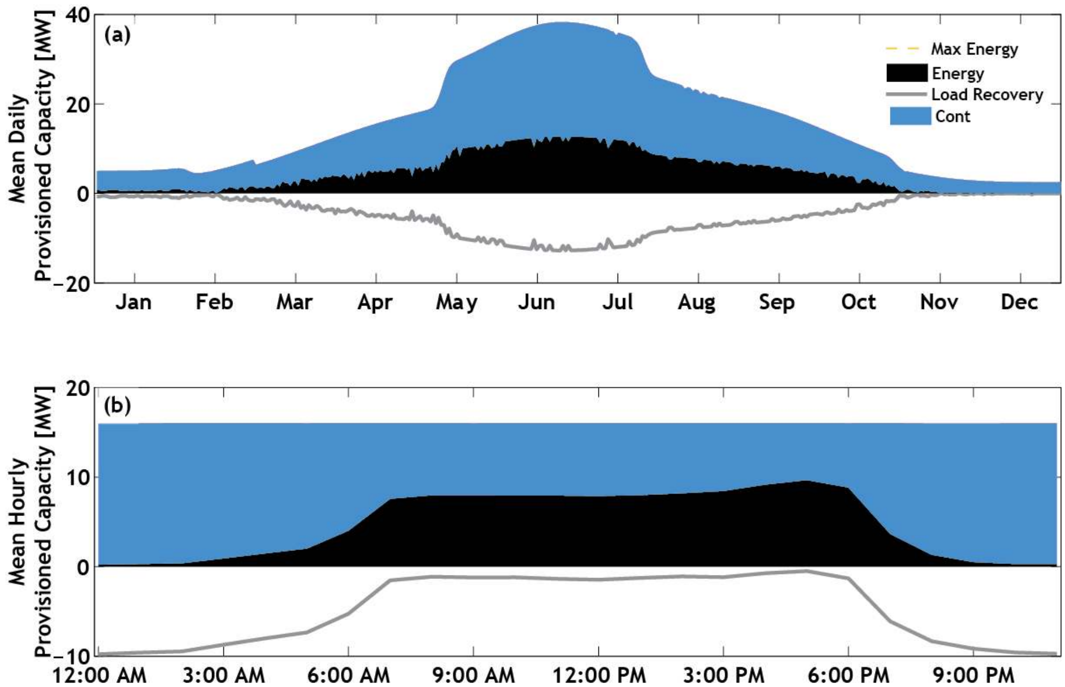

Figure 28. Agricultural pumping demand response in PSCo: Mean (a) daily and (b) hourly provision of capacity

Although there are no restrictions on the time of day that energy can be shifted, there are three other restrictions that limit the capabilities of agricultural pumping to shift energy freely. One is a $10 fee for every call, which likely requires a person physically turning off the agricultural watering pumps. The pumps are also limited by a maximum start of once per day, with a $50 penalty for a violation, as well as a limit of 8 hours of load shed per day, with a violation fee of $20 per hour violated. Both of these “soft constraints” refer only to the energy allocation of the resource and are violated during the simulation incurring penalties to the optimization. A start limitation can be a deterrent to frequent switching, which could lead to decreased effectiveness or increased wear for a mechanical machine like an agricultural pump. The operation of the agricultural pumping DR resource proved to be sensitive to the magnitude of the penalty, although a more thorough understanding is needed of the pricing of the penalty relative to other system costs. Figure 29 shows that energy shifting is utilized more than the 8-hour limit and breaks the start violation multiple times. It also shows the expected result—that charging is happening during lower price hours and energy is being shed during high price hours.

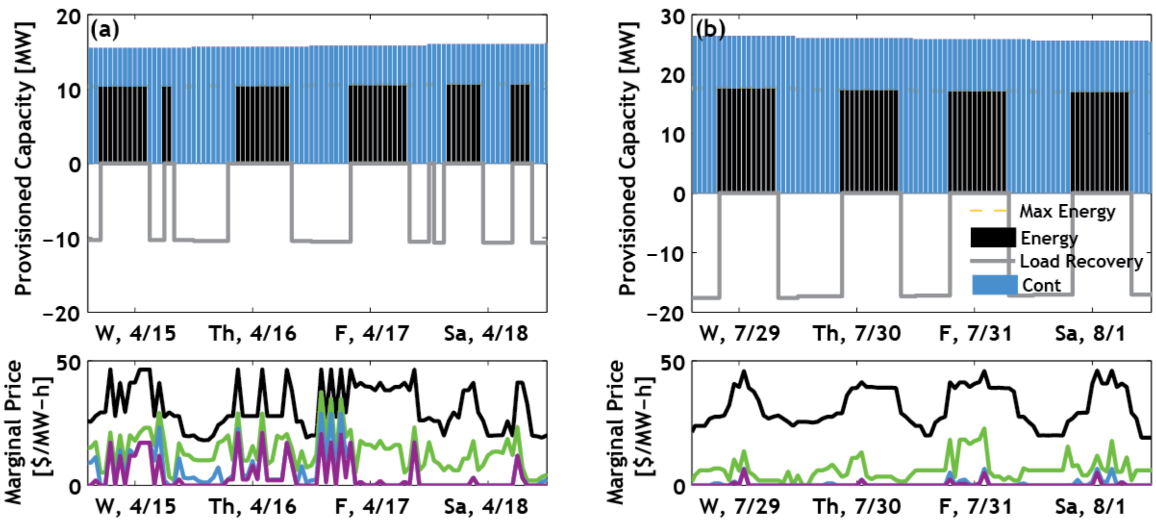

Figure 29. Agricultural pumping demand response in PSCo: (upper) Hourly capacity provision for (a) spring and (b) summer and (lower) hourly marginal price of energy and operating reserves

Municipal Pumping

Municipal pumping capacity follows a seasonal pattern as well as a daily pattern that can be observed in Figure 30a and Figure 30b. Energy is the only service that municipal pumping can provide, which is subject to full charging within 24 hours of energy shed. Figure 30b shows that there is a tendency to shift energy away from the early evening hours when energy prices are typically highest to the hours when prices are typically lower in the late evening and early morning.

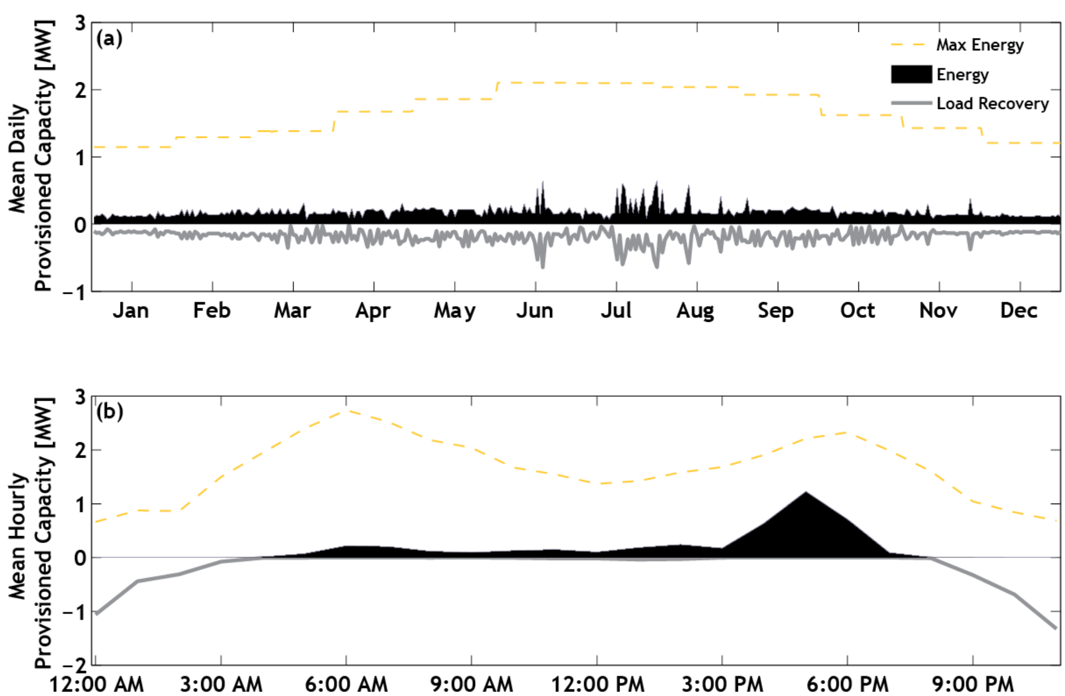

Figure 30. Municipal pumping demand response in PSCo: Mean (a) daily and (b) hourly provision of capacity

The energy shifting capability of municipal pumping is also constrained to 2 hours of energy shedding in a day and one start per day. Both of these constraints are "soft," meaning they can be broken with the payment of a fee. Utilizing the energy shifting more than 2 hours results in a \$50 violation penalty per hour, and a call to use the pumps more than once per day results in a \$100 violation penalty per call. There is also a \$5 cost for initiating DR from municipal pumping (equivalent to a start cost for a generator).

Figure 31 shows examples of operation for the (a) spring and (b) summer, as well as the marginal prices for energy and operating reserves for those periods. Energy is typically being shifted away from the late evening to the middle of the night, as observed by the average day in Figure 30b. It is also clear that some constraints, such as the maximum operating hours of 2 per day, are being violated in July but not in April.

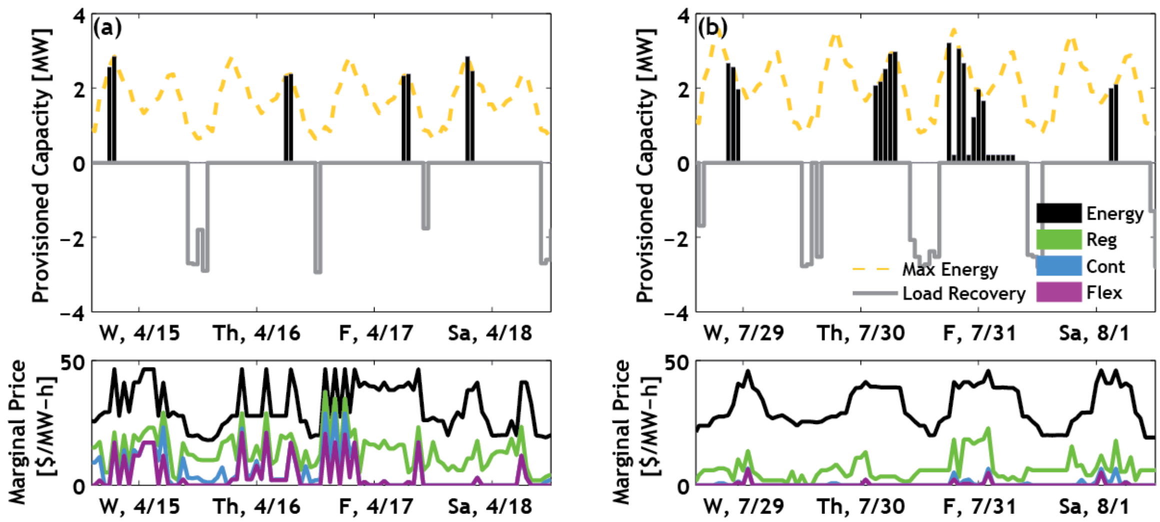

Figure 31. Municipal pumping demand response in PSCo: (upper) Hourly capacity provision for (a) spring and (b) summer and (lower) hourly marginal price of energy and operating reserves

Refrigerated Warehouse

Refrigerated warehouses have a seasonal and daily pattern (Figure 31) similar to the other "energy only" DR resources. These also tend to shift energy away from the peaking hours in late afternoon/early evening to the middle of the night. Similar to other pump-like resources, the daily runtime is restricted to 4 hours. There is no option to violate this constraint, which can be observed in the limited hours of operation in Figure 33. As expected, the energy is being shifted to the lower price times of day in both the (a) spring and (b) summer.

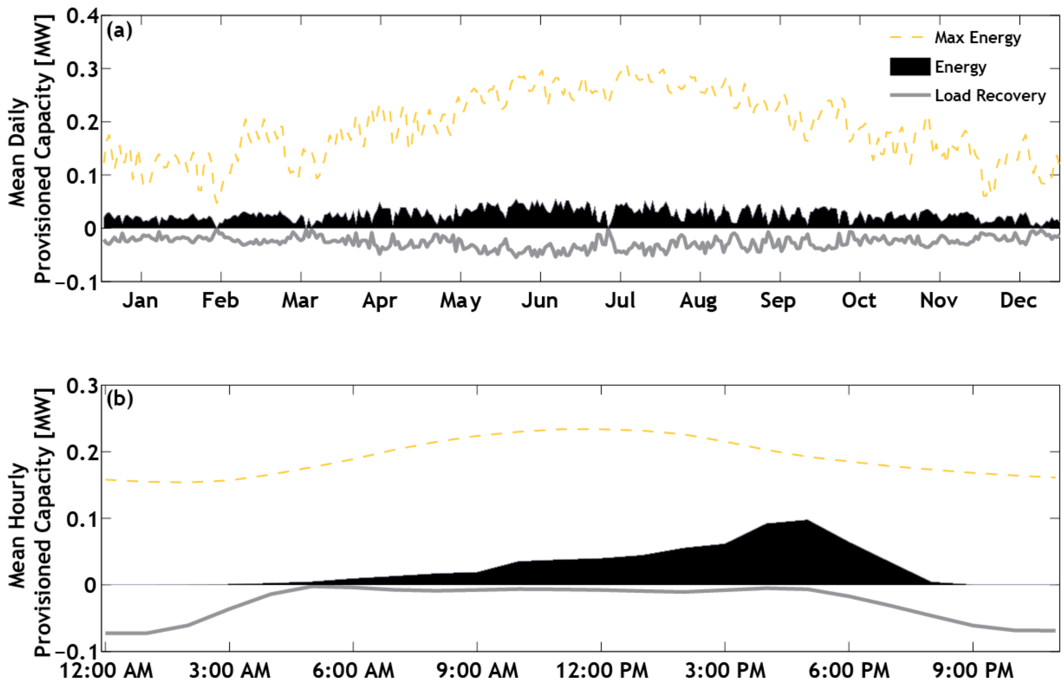

Figure 32. Refrigerated warehouse demand response in PSCo: Mean (a) daily and (b) hourly provision of capacity

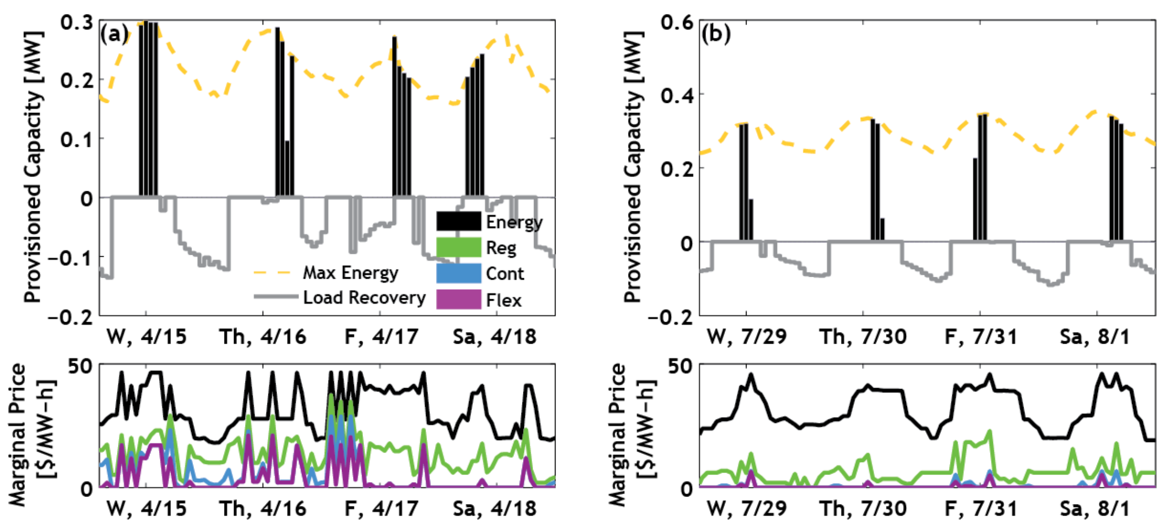

Figure 33. Refrigerated warehouse demand response in PSCo: (upper) Hourly capacity provision for (a) spring and (b) summer and (lower) hourly marginal price of energy and operating reserves

Commercial Ventilation

Commercial ventilation and outdoor lighting are DR resources that are not able to offer energy as a service; however, their capacity is co-optimized between the three operating reserves. Each DR resource has a constrained capacity available for each ancillary service response in each time interval. The sum of the capacity allocated to all operating reserves, by each DR resource in each time interval, cannot exceed the largest capacity available for any one ancillary service during that time interval. Thus, DR capacity for one ancillary service is provisioned exclusive of any other.

Commercial ventilation has very similar ancillary service capabilities to commercial lighting in magnitude and pattern (Figure 34), although the daily profile for commercial ventilation does not decrease as much overnight (Figure 35). As a result, commercial ventilation provides a nearly constant 1 MW of regulation reserves throughout the year. The majority of the remaining capacity is provisioned for contingency reserves.

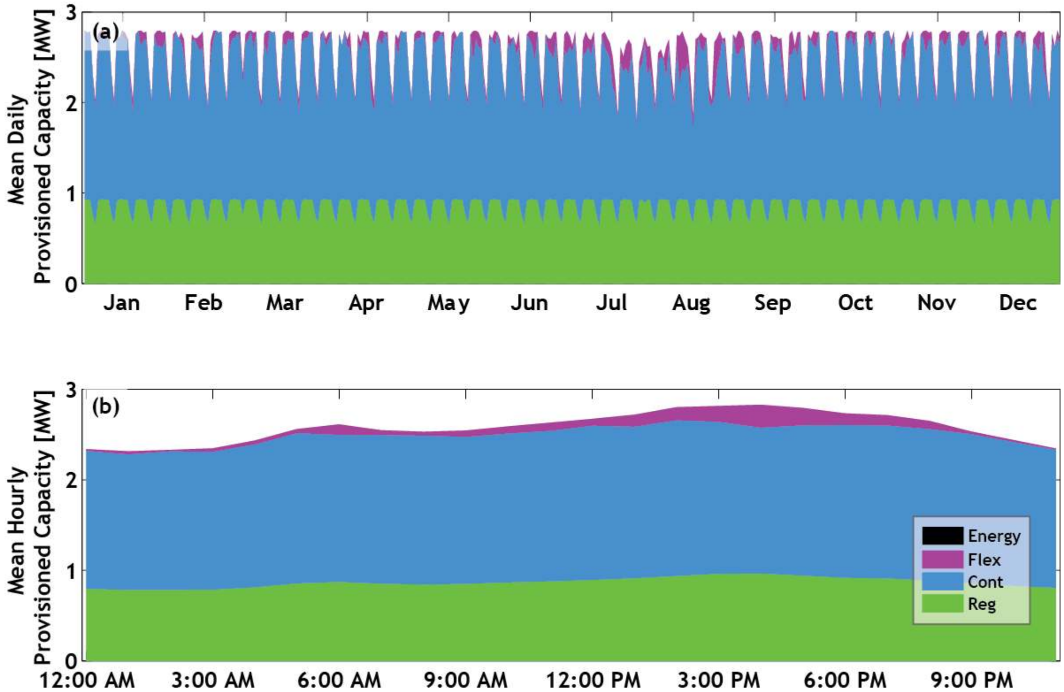

Figure 34. Commercial ventilation demand response in PSCo: Mean (a) daily and (b) hourly provision of capacity

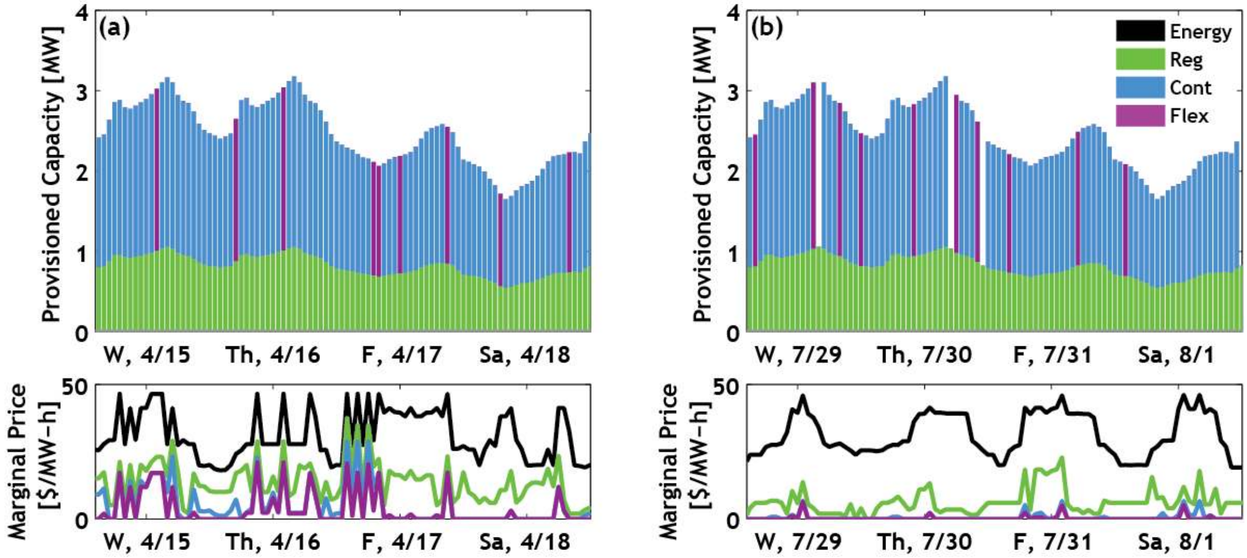

Figure 35. Commercial ventilation demand response in PSCo: (upper) Hourly capacity provision for (a) spring and (b) summer and (lower) hourly marginal price of energy and operating reserves

Outdoor Lighting

Outdoor lighting has a strong daily cycle, with almost no availability during daylight hours and seasonal variation due to length of day variations. Outdoor lighting is available to serve both regulation and flexibility, which are identical in magnitude. Figure 36 and Figure 37 (upper plots) show that outdoor lighting provides regulation reserves over flexibility, which is expected from the consistently higher marginal price of regulation (see Figure 37b).

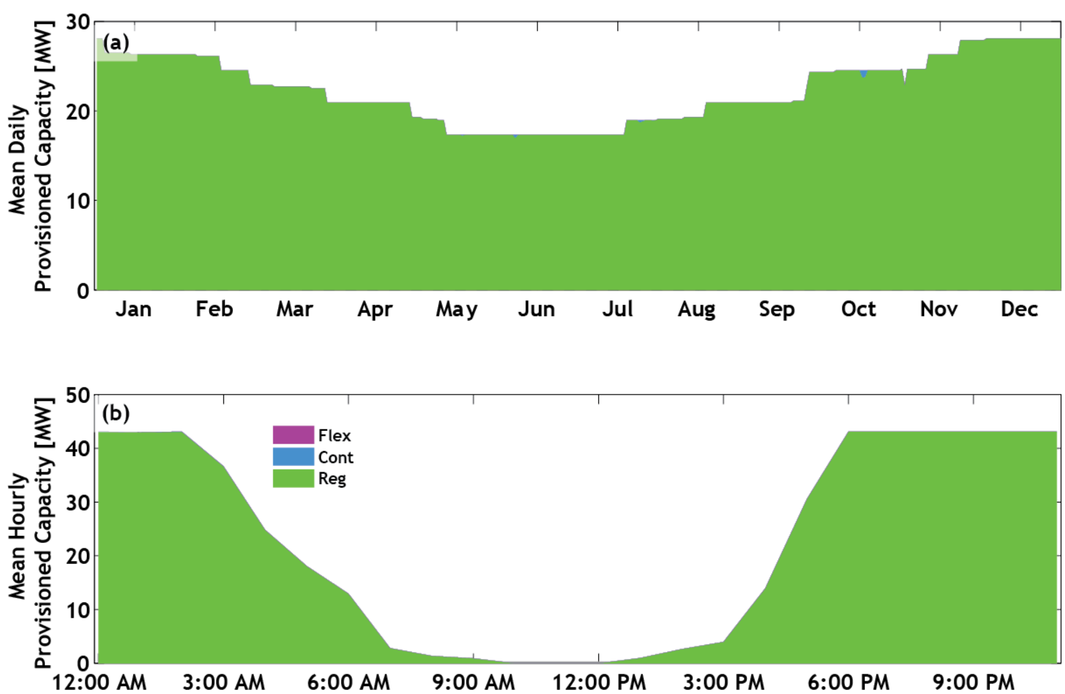

Figure 36. Outdoor lighting demand response in PSCo: Mean (a) daily and (b) hourly provision of capacity

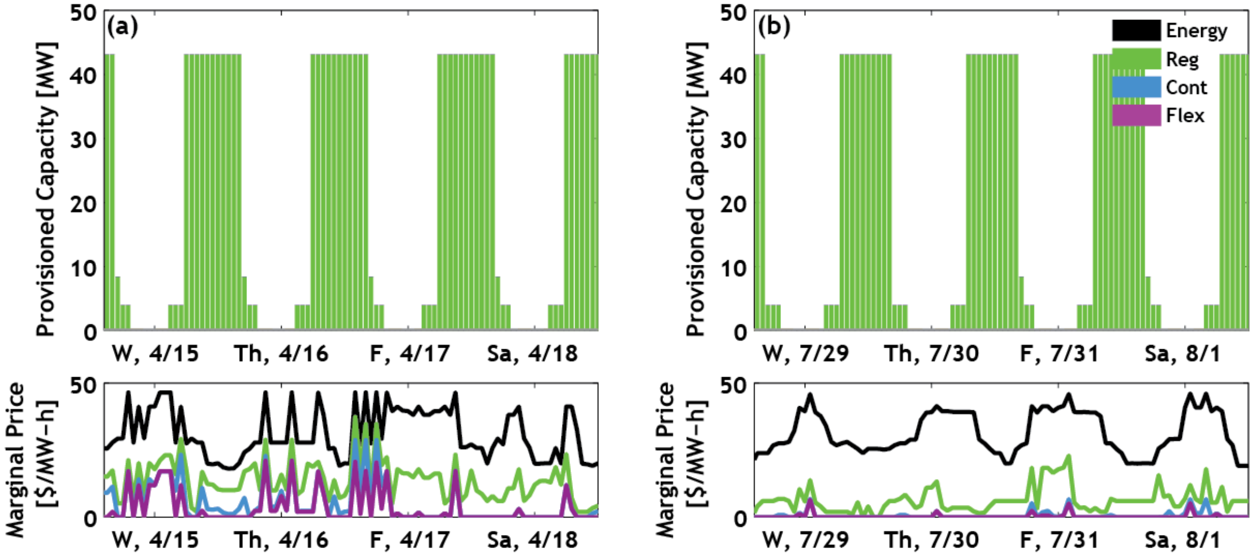

Figure 37. Outdoor lighting demand response in PSCo: (upper) Hourly capacity provision for (a) spring and (b) summer and (lower) hourly marginal price of energy and operating reserves

Seasonal Daily Average Energy and Reserve Availability and Provision

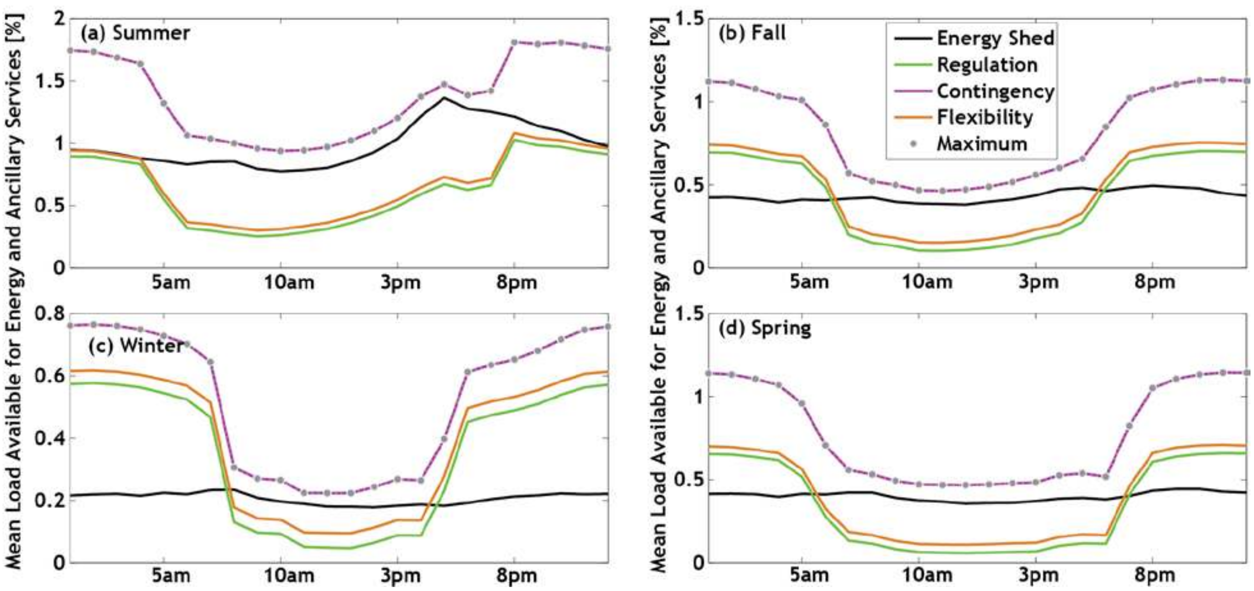

Figure 38. Seasonal daily average capacity available from demand response for each service

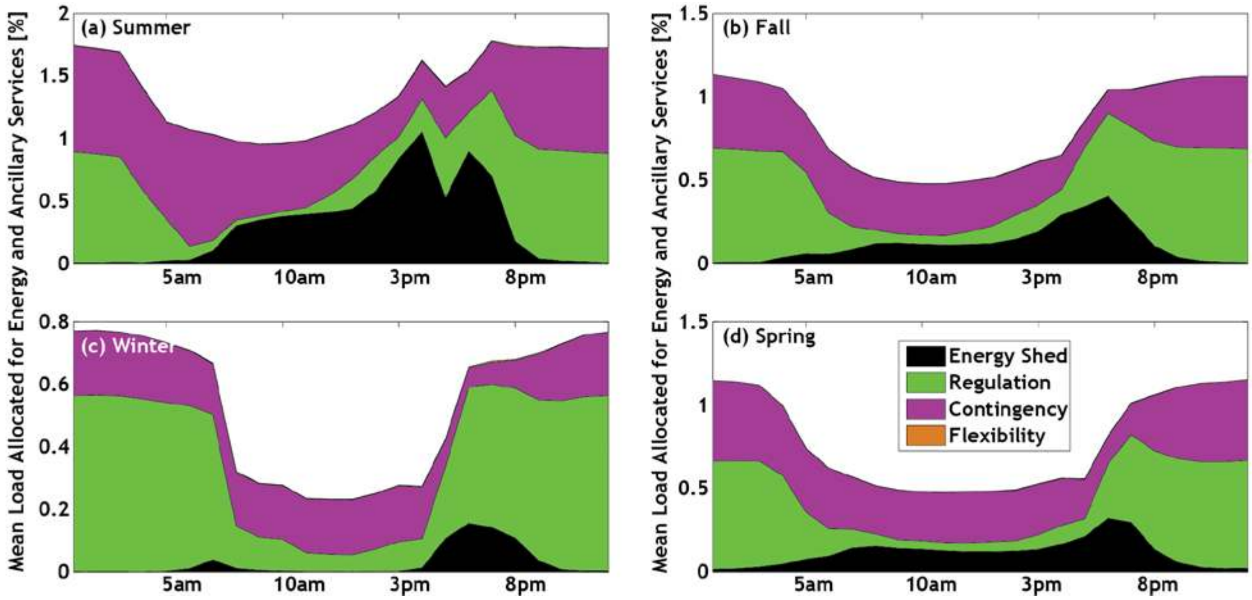

Figure 39. Mean daily allocation of demand response capacity for energy and operating reserves, by season

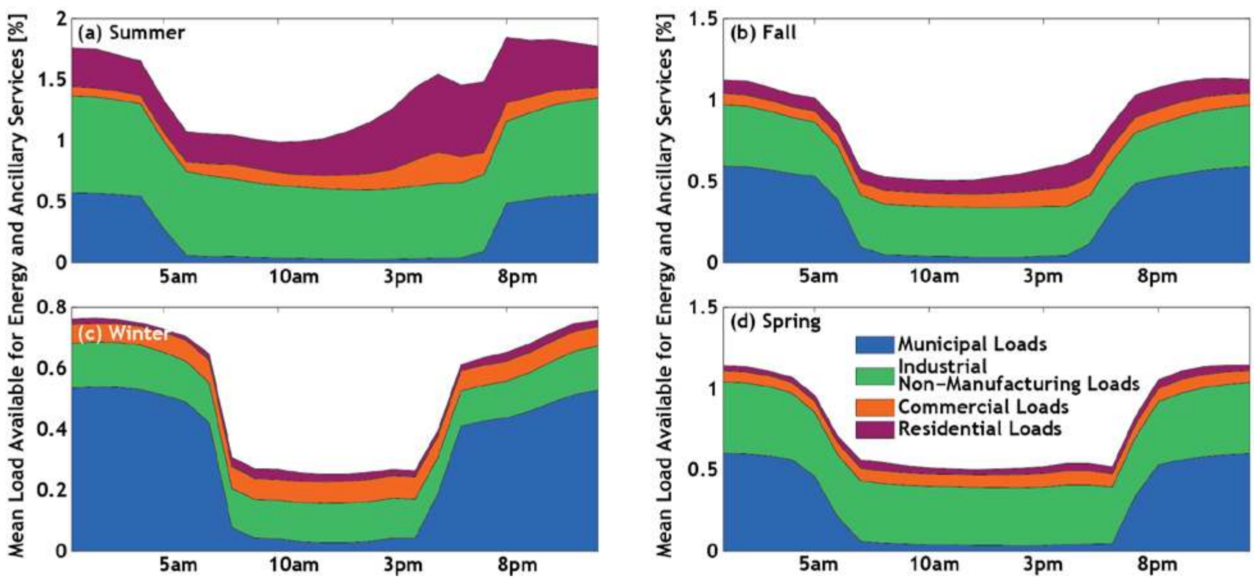

Figure 40. Mean daily allocation of demand response capacity for energy and operating reserves, by season

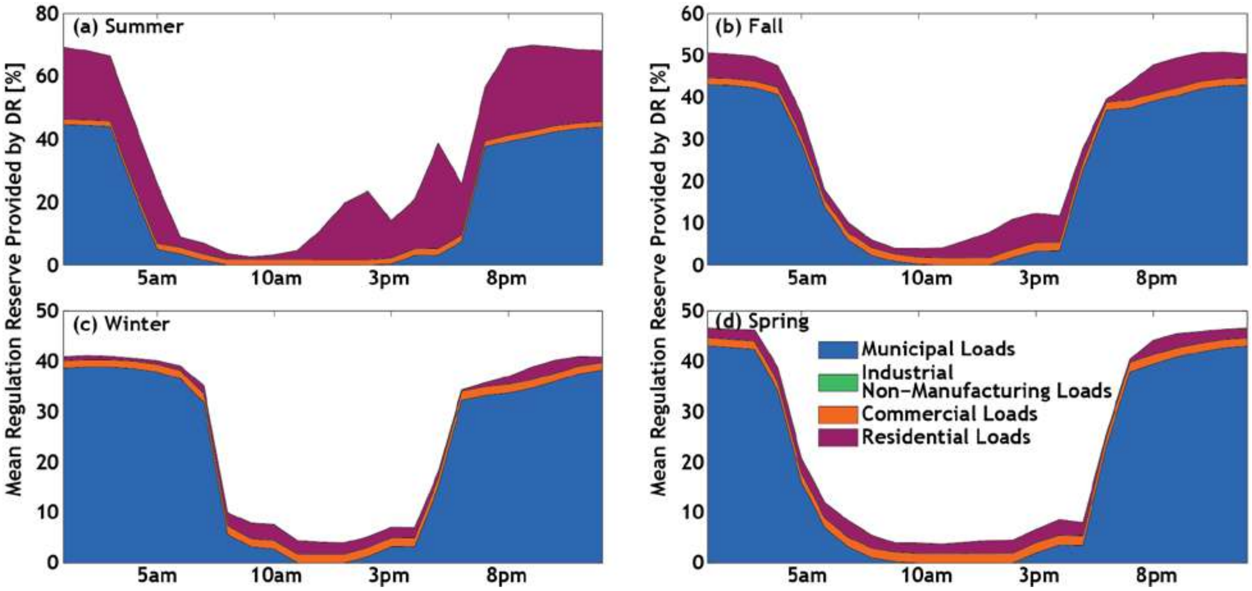

Figure 41. Average seasonal daily provision of regulation by demand response in the Colorado test system

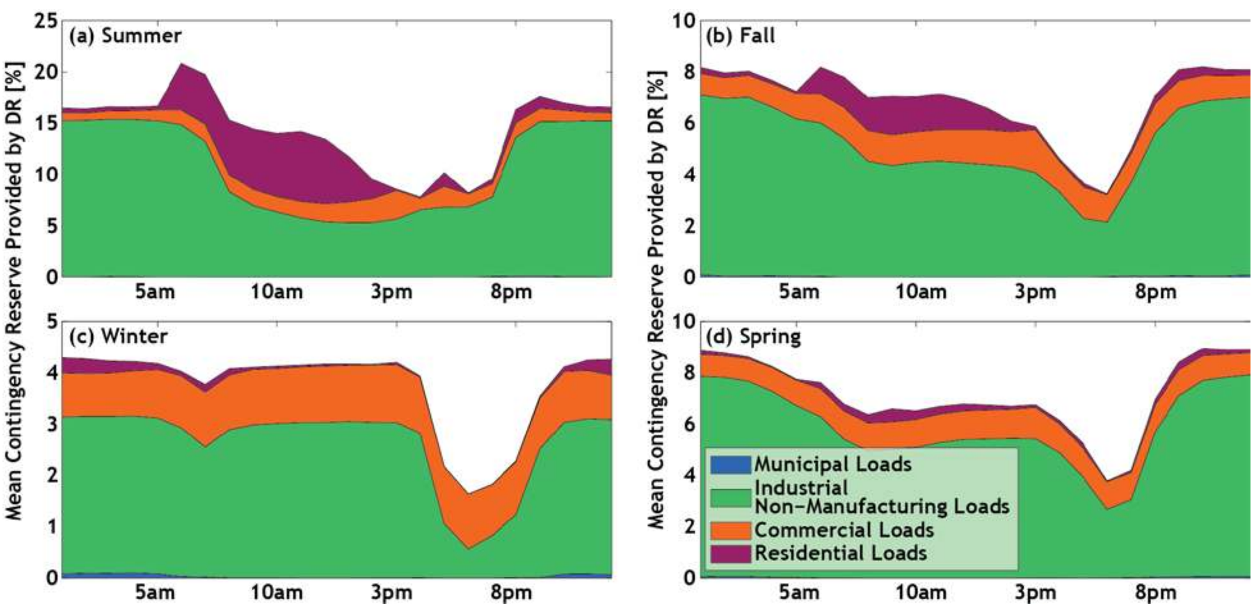

Figure 42. Average seasonal daily provision of contingency by demand response in the Colorado test system

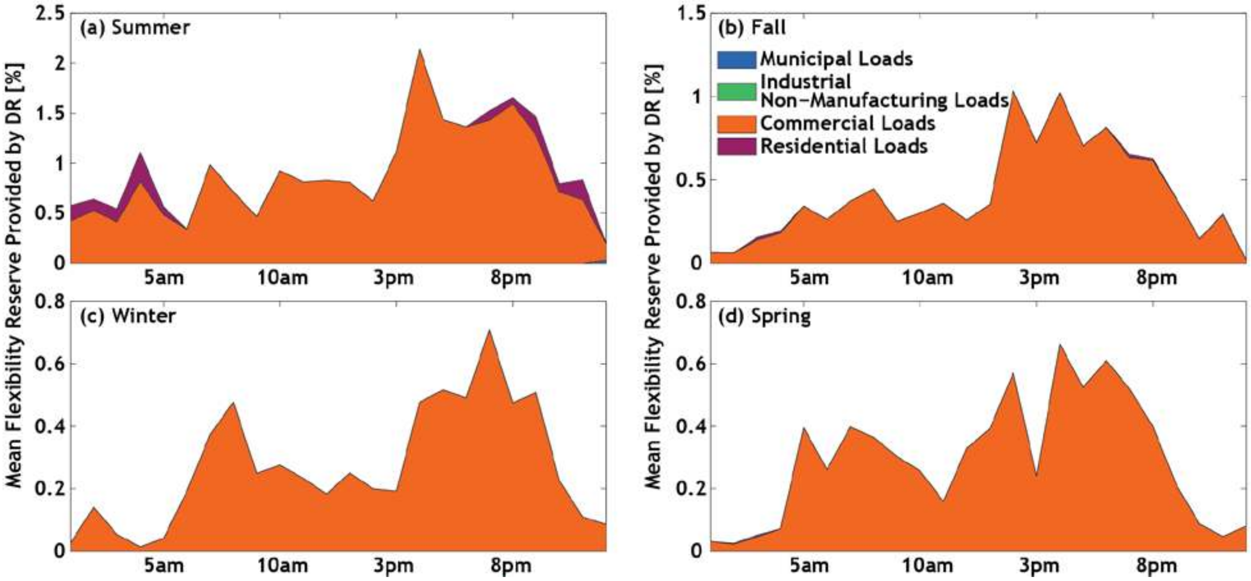

Figure 43. Average seasonal daily provision of flexibility by demand response in the Colorado test system

Marginal Price Duration Curves for Reserves

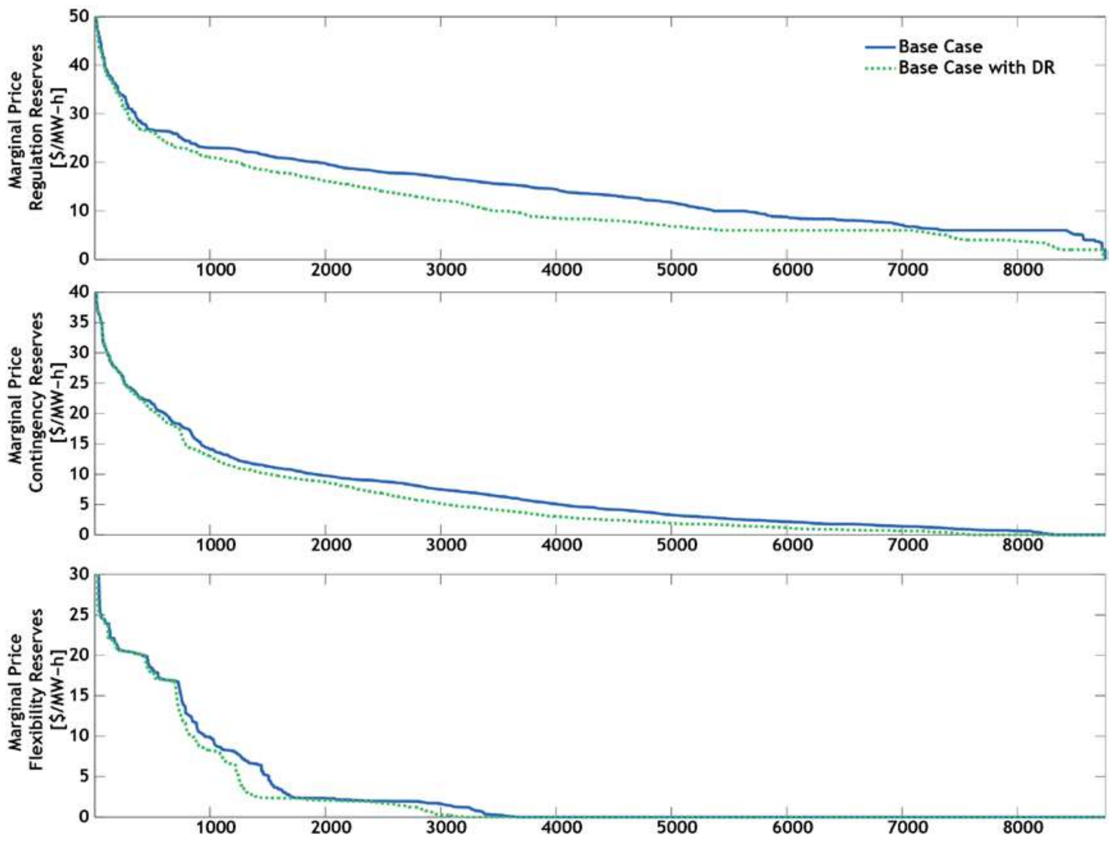

Figure 44. Price duration curves for test system with and without demand response resources providing energy and operating reserves—(top) regulation, (middle) contingency, and (bottom) flexibility