COMMONWEALTH OF KENTUCKY
BEFORE THE PUBLIC SERVICE COMMISSION

In the Matter of

ELECTRONIC APPLICATION OF KENTUCKY UTILITIES COMPANY FOR AN ADJUSTMENT OF ITS ELECTRIC RATES AND APPROVAL OF CERTAIN REGULATORY AND ACCOUNTING TREATMENTS	)))))	CASE NO. 2025-00113

In the Matter of

ELECTRONIC APPLICATION OF LOUISVILLE GAS AND ELECTRIC COMPANY FOR AN ADJUSTMENT OF ITS ELECTRIC AND GAS RATES AND APPROVAL OF CERTAIN REGULATORY AND ACCOUNTING TREATMENTS	))))))	CASE NO. 2025-00114

RESPONSE OF JOINT INTERVENORS KENTUCKIANS FOR THE COMMONWEALTH, KENTUCKY SOLAR ENERGY SOCIETY, MOUNTAIN ASSOCIATION, AND METROPOLITAN HOUSING COALITION TO COMMISSION STAFF'S POST-HEARING REQUEST FOR INFORMATION [DATED NOVEMBER 12, 2025]

Dated: November 25, 2025

JOINT INTERVENORS KENTUCKIANS FOR THE COMMONWEALTH, KENTUCKY SOLAR ENERGY SOCIETY, MOUNTAIN ASSOCIATION, AND METROPOLITAN HOUSING COALITION

RESPONSE TO COMMISSION STAFF'S POST-HEARING REQUEST FOR INFORMATION
Dated November 12, 2025

Case No. 2025-00113

Case No. 2025-00114

Question No. 2.1

Q-2.1 Refer to the Hearing Testimony of James Fine. Provide copies of the New York Times Op-Ed and Environmental Economics study referenced in the testimony.

A-2.1 RESPONSE:

The Op-Ed Mr. Fine referred to is:

> Tyler Norris, Opinion, Guest Essay, *A Simple Fix to America's Soaring Electricity Prices*, New York Times, (Nov. 4, 2025), *available at* https://www.nytimes.com/2025/11/04/opinion/electricity-power-bills-ai-data-centers.html

The Op-Ed is not attached, as a New York Times subscription is required for viewing and reproduction is prohibited. However, it was written by one of the co-authors of a study referenced therein, attached as Attachment 1:

> Tyler H. Norris, Tim Profeta, Dalia Patino-Echeverri, and Adam Cowie-Haskell, *Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in US Power Systems*,
> Nicolas Institute, Duke University, (Feb. 2025), *available at* https://nicholasinstitute.duke.edu/publications/rethinking-load-growth

The Environmental Economics Study Mr. Fine referred to is attached as Attachment 2:

Adrian Au, Grant Freudenthaler, Ben Elsey, Hugh Somerset, Edita Danielyan and Zachary Ming, *The Capacity Accreditation of Demand Response in SPP*, Energy Economics & Environment, (Oct. 2025), *available at* https://www.ethree.com/spp-demand-response/

VERIFICATION

The undersigned, James David Fine, being first duly sworn, deposes and says that he has personal knowledge of the matters set forth in the foregoing testimony and that the information contained therein is true and correct to the best of his information, knowledge, and belief, after reasonable inquiry.

__James David Fine_James David Fine_________

State of Texas

County of Williamson

Subscribed and sworn to before me by James David Fine this 25th day of November, 2025. by James David Fine.

____
Notary Public

My commission expires: 02/18/2028

NOTARY PUBLIC ★ STATE OF TEXAS ★

Christopher Kaositadima Isibor

ID NUMBER
132363262
COMMISSION EXPIRES
February 18, 2028

Electronically signed and notarized online using the Proof platform.

Attachment 1

Duke NICHOLAS INSTITUTE *for* ENERGY, ENVIRONMENT & SUSTAINABILITY

Rethinking Load Growth

Assessing the Potential for Integration of Large Flexible Loads in US Power Systems

Tyler H. Norris, Tim Profeta, Dalia Patino-Echeverri, and Adam Cowie-Haskell

Authors and Affiliations

Tyler H. Norris, Nicholas School of the Environment, Duke University
Tim Profeta, Sanford School of Public Policy and Nicholas Institute for Energy, Environment & Sustainability, Duke University
Dalia Patino-Echeverri, Nicholas School of the Environment, Duke University
Adam Cowie-Haskell, Nicholas School of the Environment, Duke University

Acknowledgments

The authors would like to thank Jessalyn Chuang and Wendy Wen for their research assistance.

Citation

Norris, T. H., T. Profeta, D. Patino-Echeverri, and A. Cowie-Haskell. 2025. *Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in US Power Systems*. NI R 25-01. Durham, NC: Nicholas Institute for Energy, Environment & Sustainability, Duke University.
https://nicholasinstitute.duke.edu/publications/rethinking-load-growth

Cover image courtesy Gerville via iStock

Nicholas Institute for Energy, Environment & Sustainability

The Nicholas Institute for Energy, Environment & Sustainability at Duke University accelerates solutions to critical energy and environmental challenges, advancing a more just, resilient, and sustainable world. The Nicholas Institute conducts and supports actionable research and undertakes sustained engagement with policymakers, businesses, and communities—in addition to delivering transformative educational experiences to empower future leaders. The Nicholas Institute's work is aligned with the Duke Climate Commitment, which unites the university's education, research, operations, and external engagement missions to address climate challenges.

Contact

Nicholas Institute | Duke University | P.O. Box 90467 | Durham, NC 27708
1201 Pennsylvania Avenue NW | Suite 500 | Washington, DC 20004
919.613.1305 | nicholasinstitute@duke.edu

CONTENTS

INTRODUCTION

A New Era of Electricity Demand

Rapid US load growth—driven by unprecedented electricity demand from data centers, industrial manufacturing, and electrification of transportation and heating—is colliding with barriers to timely resource expansion. Protracted interconnection queues, supply chain constraints, and extended permitting processes, among other obstacles, are limiting the development of new power generation and transmission infrastructure. Against this backdrop, there is increasing urgency to identify strategies that accommodate rising demand without compromising reliability, affordability, or progress on decarbonization.

Aggregated US winter peak load is forecasted to grow by 21.5% over the next decade, rising from approximately 694 GW in 2024 to 843 GW by 2034, according to the *2024 Long-Term Reliability Assessment* of the North American Electric Reliability Corporation. This represents a 10-year compound annual growth rate (CAGR) of 2.0%, higher than any period since the 1980s (NERC 2024). Meanwhile, the Federal Energy Regulatory Commission's (FERC) latest five-year outlook forecasts 128 GW in peak load growth as early as 2029—a CAGR of 3.0% (FERC 2024b).

The primary catalyst for these updated forecasts is the surge in electricity demand from large commercial customers. While significant uncertainty remains, particularly following the release of DeepSeek, data centers are expected to account for the single largest growth segment, adding as much as 65 GW through 2029 and up to 44% of US electricity load growth through 2028 (Wilson et al. 2024; Rouch et al. 2024). Artificial intelligence (AI) workloads are projected to represent 50% to 70% of data center demand by 2030—up from less than 3% at the start of this decade—with generative AI driving 40% to 60% of this growth (Srivathsan et al. 2024; Lee et al. 2025).

Analysts have drawn parallels to the 1950s through the 1970s, when the United States achieved comparable electric power sector growth rates (Wilson et al. 2024). Yet these comparisons arguably understate the nature of today's challenge in the face of stricter permitting obstacles, higher population density, less land availability, skilled labor shortages, persistent supply chain bottlenecks, and demand for decarbonization and greater power reliability. While historical growth rates offer a useful benchmark, the sheer volume of required new generation, transmission, and distribution capacity forecasted for the United States within a condensed timeframe appears unprecedented.

The immensity of the challenge underscores the importance of deploying every available tool, especially those that can more swiftly, affordably, and sustainably integrate large loads. The time-sensitivity for solutions is amplified by the market pressure for many of these loads to interconnect as quickly as possible. In recent months, the US Secretary of Energy Advisory Board (SEAB) and the Electrical Power Research Institute (EPRI) have highlighted a key solution: load flexibility (SEAB 2024, Walton 2024a). The promise is that the unique profile of AI data centers can facilitate more flexible operations, supported by ongoing advancements in distributed energy resources (DERs).

Flexibility, in this context, refers to the ability of end-use customers to temporarily reduce their electricity consumption from the grid during periods of system stress by using on-site generators, shifting workload to other facilities, or reducing operations.[1] When system planners can reliably anticipate the availability of this load flexibility, the immediate pressure to expand generation capacity and transmission infrastructure can potentially be alleviated, mitigating or deferring costly expenditures. By facilitating near-term load growth without prematurely committing to large-scale capacity expansion, this approach offers a hedge against mounting uncertainty in the US data center market in light of the release of DeepSeek and related developments (Kearney and Hampton 2025).

Summary of Analysis and Findings

To support evaluation of potential solutions, this study presents an analysis of the existing US electrical power system's ability to accommodate new flexible loads. The analysis, which encompasses 22 of the largest balancing authorities serving 95% of the country's peak load, provides a first-order estimate of the potential for accommodating such loads with minimal capacity expansion or impact on demand-supply balance.[2]

Specifically, we estimate the gigawatts of new load that could be added in each balancing authority (BA) before total load exceeds what system planners are prepared to serve, provided the new load can be temporarily curtailed as needed. This serves as a proxy for the system's ability to integrate new load, which we term *curtailment-enabled headroom*.

Key results include (see Figure 1):

- 76 GW of new load—equivalent to 10% of the nation's current aggregate peak demand—could be integrated with an average annual load curtailment rate of 0.25% (i.e., if new loads can be curtailed for 0.25% of their maximum uptime)
- 98 GW of new load could be integrated at an average annual load curtailment rate of 0.5%, and 126 GW at a rate of 1.0%
- The number of hours during which curtailment of new loads would be necessary per year, on average, is comparable to those of existing US demand response programs
- The average duration of load curtailment (i.e., the length of time the new load is curtailed during curtailment events) would be relatively short, at 1.7 hours when average annual load curtailment is limited to 0.25%, 2.1 hours at a 0.5% limit, and 2.5 hours at a 1.0% limit
- Nearly 90% of hours during which load curtailment is required retain at least half of the new load (i.e., less than 50% curtailment of the new load is required)
- The five balancing authorities with the largest potential load integration at 0.5% annual curtailment are PJM at 18 GW, MISO at 15 GW, ERCOT at 10 GW, SPP at 10 GW, and Southern Company at 8 GW[3]

1 Note that while *curtailment* and *flexibility* are used interchangeably in this paper, *flexibility* can refer to a broader range of capabilities and services, such as the provision of down-reserves and other ancillary services.

2 For further discussion on the nuances regarding generation versus transmission capacity, see the section on limitations.

3 A complete list of abbreviations and their definitions can be found at the end of the report.

Overall, these results suggest the US power system's existing headroom, resulting from intentional planning decisions to maintain sizable reserves during infrequent peak demand events, is sufficient to accommodate significant constant new loads, provided such loads can be safely scaled back during some hours of the year. In addition, they underscore the potential for leveraging flexible load as a complement to supply-side investments, enabling growth while mitigating the need for large expenditures on new capacity.

We further demonstrate that a system's potential to serve new electricity demand without capacity expansion is determined primarily by the system's load factor (i.e., a measure of the level of use of system capacity) and grows in proportion to the flexibility of such load (i.e., what percentage of its maximal potential annual consumption can be curtailed). For this reason, in this paper we assess the technical potential for a system to serve new load under different curtailment limit scenarios (i.e., varying curtailment tolerance levels for new loads).

The analysis does not consider the technical constraints of power plants that impose intertemporal constraints on their operations (e.g., minimum downtime, minimum uptime, startup time, ramping capability, etc.) and does not account for transmission constraints. However, it ensures that the estimate of load accommodation capacity is such that total demand does not exceed the peak demand already anticipated for each season by system planners, and it discounts existing installed reserve margins capable of accommodating load that exceeds historical peaks. It also assumes that new load is constant throughout all hours.

This analysis should not be interpreted to suggest the United States can fully meet its near- and medium-term electricity demands without building new peaking capacity or expanding the grid. Rather, it highlights that flexible load strategies can help tap existing headroom to more quickly integrate new loads, reduce the cost of capacity expansion, and enable greater focus on the highest-value investments in the electric power system.

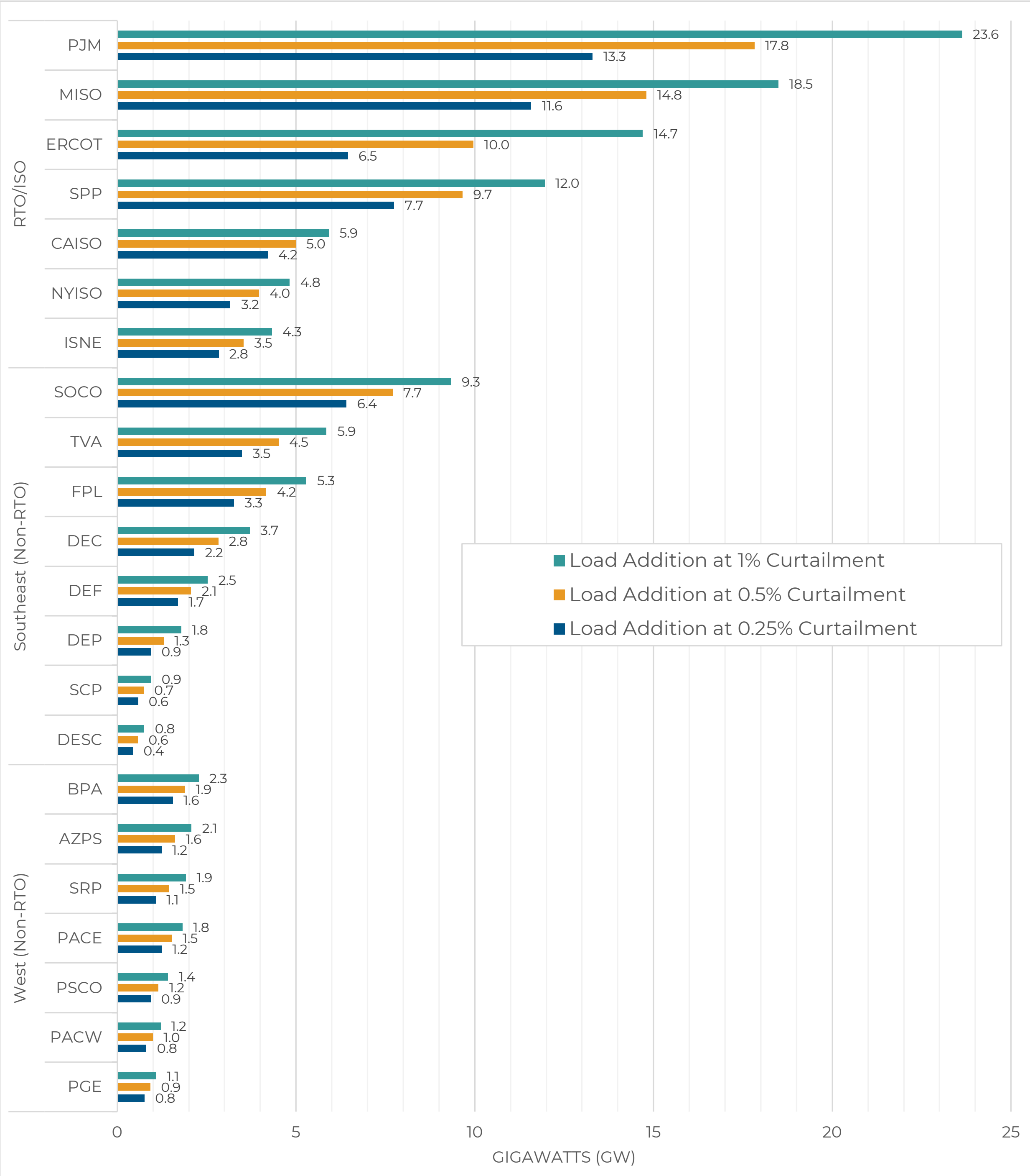
This paper proceeds as follows: the following section provides background on the opportunities and challenges to integrating large new data centers onto the grid. It explores how load flexibility can accelerate interconnection, reduce ratepayer costs through higher system utilization, and expand the role of demand response, particularly for AI-specialized data centers. We then detail the methods and results for estimating curtailment-enabled headroom, highlighting key trends and variations in system headroom and its correlation with load factors across regions. The paper concludes with a brief overview of key findings, limitations, and near-term implications.

BACKGROUND

Load Flexibility Can Accelerate Grid Interconnection

The growing demand for grid access by new large loads has significantly increased interconnection wait times, with some utilities reporting delays up to 7 to 10 years (Li et al. 2024; Saul 2024; WECC 2024). These wait times are exacerbated by increasingly severe transmission equipment supply chain constraints. In June 2024, the President's National Infrastructure Advisory Council highlighted that transformer order lead times had ballooned to two to five years—up from less than one year in 2020—while costs surged by 80% (NIAC 2024). Circuit breakers have seen similar delays: last year, the Western Area Power Administration

Figure 1. System Headroom Enabled by Load Curtailment of New Load by Balancing Authority, GW



Note: *System headroom* refers to the amount of GW by which a BA's load can be augmented every hour in the absence of capacity expansion so that, provided a certain volume of curtailment of the new load, the total demand does not exceed the supply provisioned by system planners to withstand the expected highest peak. The headroom calculation assumes the new load is constant and hence increases the total load by the same GW hour-by-hour.

reported lead times of up to four and a half years for lower voltage classes and five and a half years for higher voltage classes, alongside a 140% price hike over two years (Rohrer 2024). Wood Mackenzie reported in May 2024 that lead times for high-voltage circuit breakers reached 151 weeks in late 2023, marking a 130% year-over-year increase (Boucher 2024).

Large load interconnection delays have recently led to growing interest among data centers in colocating with existing generation facilities. At a FERC technical conference on the subject in late 2024 (FERC 2024c), several participants highlighted the potential benefits of colocation for expedited interconnection,[4] a view echoed in recent grey literature (Schatzki et al. 2024). Colocation, however, represents only a portion of load interconnections and is not viewed as a long-term, system-wide solution.

Load flexibility similarly offers a practical solution to accelerating the interconnection of large demand loads (SIP 2024, Jabeck 2023). The most time-intensive and costly infrastructure upgrades required for new interconnections are often associated with expanding the transmission system to deliver electricity during the most stressed grid conditions (Gorman et al. 2024). If a new load is assumed to require firm interconnection service and operate at 100% of its maximum electricity draw at all times, including during system-wide peaks, it is far more likely to trigger the need for significant upgrades, such as new transformers, transmission line reconductoring, circuit breakers, or other substation equipment.

To the extent a new load can temporarily reduce (i.e., curtail) its electricity consumption from the grid during these peak stress periods, however, it may be able to connect while deferring—or even avoiding—the need for certain upgrades (ERCOT 2023b). A recent study on Virginia's data center electricity load growth noted, "Flexibility in load is generally expected to offset the need for capacity additions in a system, which could help mitigate the pressure of rapid resource and transmission expansion" (K. Patel et al. 2024). The extent and frequency of required curtailment would depend on the specific nature of the upgrades; in some cases, curtailment may only be necessary if a contingency event occurs, such as an unplanned transmission line or generator outage. For loads that pay for firm interconnection service, any period requiring occasional curtailment would be temporary, ending once necessary network upgrades are completed.[5] Such "partially firm," flexible service was also highlighted by participants in FERC's 2024 technical conference on colocation.[6]

Traditionally, such arrangements have been known as *interruptible* electric service. More recently, some utilities have pursued *flexible* load interconnection options. In March 2022, for example, ERCOT implemented an interim interconnection process for large loads seeking to connect in two years or less, proposing to allow loads seeking to qualify as controllable load resources (CLRs) "to be studied as flexible and potentially interconnect more MWs" (ERCOT 2023b). More recently, ERCOT stated that "the optimal solution for grid reliability is for

4 For example, the Clean Energy Buyers Association (2024) noted, "Flexibility of co-located demand is a key asset that can enable rapid, reliable interconnection."

5 Such an arrangement is analogous to provisional interconnection service available to large generators, as defined in Section 5.9.2 of FERC's Pro Forma *Large Generator Interconnection Agreement* (LGIA).

6 MISO's market monitor representative stated, "instead of being a network firm customer, could [large flexible loads] be a non-firm, or partial non-firm [customer], and that could come with certain configuration requirements that make them truly non-firm, or partially non-firm. But, all those things are the things that could enable some loads to get on the system quicker" (FERC 2024c).

more loads to participate in economic dispatch as CLRs" (Springer 2024). Similarly, Pacific Gas and Electric (PG&E) recently introduced a Flex Connect program to allow certain loads faster access to the grid (Allsup 2024).

These options resemble interconnection services available to large generators that forgo capacity compensation, and potentially higher curtailment risk, in exchange for expedited lower-cost grid access (Norris 2023). FERC codified this approach with Energy Resource Interconnection Service (ERIS) in Order 2003 and revisited the concept during a 2024 technical workshop to explore potential improvements (Norris 2024). Some market participants have since proposed modifying ERIS to facilitate the colocation of new generators with large loads (Intersect Power 2024).

Ratepayers Benefit from Higher System Utilization

The US electric power system is characterized by a relatively low utilization rate, often referred to as the *load factor*. The load factor is the ratio of average demand to peak demand over a given period and provides a measure of the utilization of system capacity (Cerna et al. 2023). A system with a high load factor operates closer to its peak system load for more hours throughout the year, while a system with a low load factor generally experiences demand spikes that are higher than its typical demand levels (Cerna et al. 2022). This discrepancy means that, for much of the year, a significant portion of a system's available generation and transmission infrastructure is underutilized (Cochran et al. 2015).

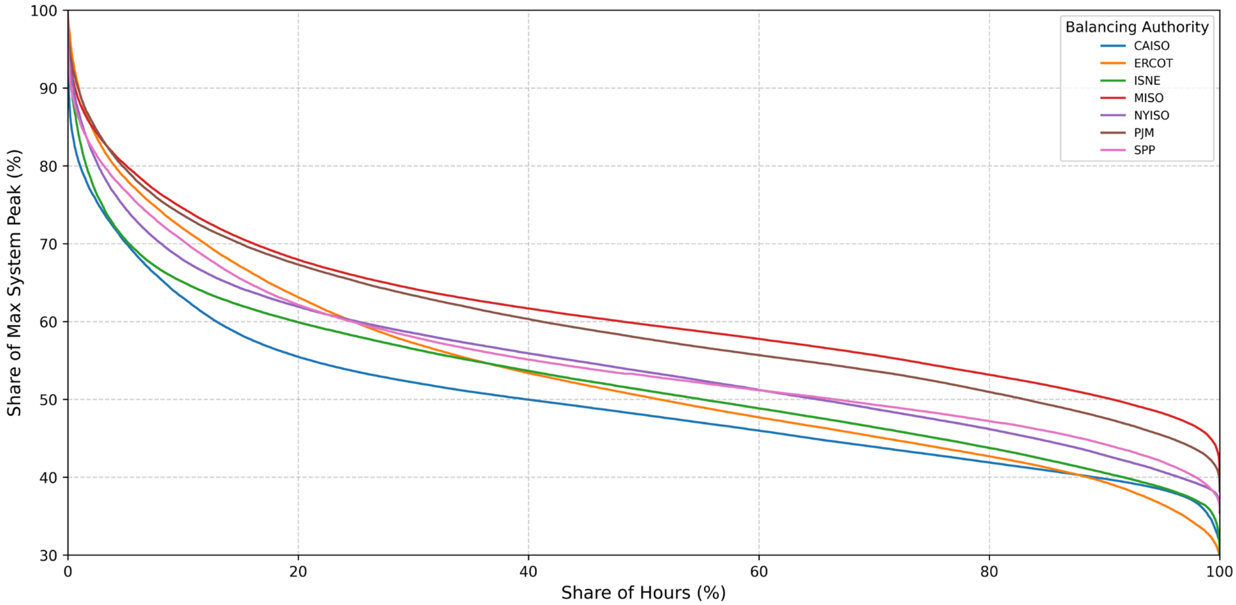
The power system is designed to handle the highest demand peaks, which in some cases may occur less than once per year, on average, due to extreme weather events. As a result, the bulk of the year sees demand levels well below that peak, leaving substantial headroom in installed capacity. Seasonal shifts add another layer of complexity: some balancing authorities may show higher load factors in summer, yet experience significantly lower utilization in winter, and vice versa.

The *load duration curve* (LDC) illustrates system utilization by ranking demand from highest to lowest over a given period. It provides a visual representation of how often certain demand levels occur, highlighting the frequency and magnitude of peak demand relative to average load. A steep LDC suggests high demand variability, with peaks significantly exceeding typical loads, while a flatter LDC indicates more consistent usage. Figure 2 presents LDCs for each US RTO/ISO based on hourly load between 2016 and 2024, standardized as a percentage of each system's maximum peak demand to allow cross-market comparisons.

A system utilization rate below 100% is expected for most large-scale infrastructure designed to withstand occasional surges in demand. Nevertheless, when the gap between average demand and peak demand is consistently large, it implies that substantial portions of the electric power system—generation assets, transmission infrastructure, and distribution networks—remain idle for much of the year (Riu et al. 2024). These assets are expensive to build and maintain, and ratepayers ultimately bear the cost.

Once the infrastructure is in place, however, there is a strong economic incentive to increase usage and spread these fixed costs over more kilowatt-hours of delivered electricity. An important consideration is therefore the potential for additional load to be added without significant new investment, provided the additional load does not raise the system's overall

Figure 2. Load Duration Curve for US RTO/ISOs, 2016–2024



This figure is adapted from the analysis section of this paper, which contains additional detail on the data and method.

peak demand and thereby trigger system expansion.[7] When new loads are flexible enough to avoid a high coincident load factor, thereby mitigating contribution to the highest-demand hours, they fit within the existing grid's headroom.[8] By strategically timing or curtailing demand, these flexible loads can minimize their impact on peak periods. In doing so, they help existing customers by improving the overall utilization rate—thereby lowering the per-unit cost of electricity—and reduce the likelihood that expensive new peaking plants or network expansions may be needed.

In contrast, inflexible new loads that increase the system's absolute peak demand can drive substantial additional needs for generation and transmission capacity. Even a modest rise in peak demand may trigger capital investments in peaking plants, fuel supply infrastructure, and reliability enhancements. These cost implications have contributed to increasingly contentious disputes in which regulators or ratepayer advocates seek to create mechanisms to pass the costs of serving large loads directly to those loads and otherwise ensure data centers do not shift costs via longer contract commitments, billing minimums, and upfront investment (Howland 2024a; Riu et al. 2024). Some examples include:

- The **Georgia Public Service Commission (GPSC)**, citing "staggering" large load growth and the need to protect ratepayers from the costs of serving those customers, recently implemented changes to customer contract provisions if peak draw exceeds 100 MW, mandating a GPSC review and allowing the utility to seek longer contracts and minimum billing for cost recovery (GPSC 2025). This follows GPSC's approval

7 See the discussion on limitations and further analysis in the following section for additional nuance.

8 Demand charges are often based on coincident consumption (e.g., ERCOT's Four Coincident Peak charge uses the load's coincident consumption at the system's expected seasonal peak to determine an averaged demand charge that may account for >30% of a user's annual bill).

of 1.4 GW of gas capacity proposed by Georgia Power in response to load growth "approximately 17 times greater than previously forecasted" through 2030/2031, a forecast it revised upward in late 2024 (GPC 2023, 2024).

- **Ohio**, where American Electric Power issued a moratorium on data center service requests, followed by a settlement agreement with the Public Service Commission staff and consumer advocates that calls for longer contract terms, load ramping schedules, a minimum demand charge, and collateral for service from data centers exceeding 25 MW (Ohio Power Company 2024).
- **Indiana**, where 4.4 GW of interconnection requests from a "handful" of data centers represents a 157% increase in peak load for Indiana Michigan Power over the next six years. Stakeholders there have proposed "firewalling" the associated cost of service from the rest of the rate base, wherein the utility would procure a separate energy, capacity, and ancillary resource portfolio for large loads and recover that portfolio's costs from only the qualifying large loads (Inskeep 2024).
- **Illinois**, where Commonwealth Edison reported that large loads have paid 8.2% of their interconnection costs while the remaining 91.8% is socialized across general customers (ComEd 2024).

These examples underscore the significance of exploring how flexible loads can mitigate peak increases, optimize the utilization of existing infrastructure, and reduce the urgency for costly and time-consuming capacity expansions.

Demand Response and Data Centers

Demand response refers to changes in electricity usage by end-use customers to provide grid services in response to economic signals, reliability events, or other conditions. Originally developed to reduce peak loads (also called *peak shaving*), demand response programs have evolved to encompass a variety of grid services, including balancing services, ancillary services, targeted deferral of grid upgrades, and even variable renewable integration (Hurley et al. 2013; Ruggles et al. 2021). Demand response is often referred to as a form of *demand-side management* or *demand flexibility* (Nethercutt 2023).

Demand response is the largest and most established form of virtual power plant (Downing et al. 2023), with 33 GW of registered capacity in wholesale RTO/ISO programs and 31 GW in retail programs as of 2023 (FERC 2024a).[9] As a share of peak demand, participation in RTO/ISO programs ranges from a high of 10.1% in MISO to a low of 1.4% in SPP. A majority of enrolled capacity in demand response programs are industrial or commercial customers, representing nearly 70% of registered capacity in retail (EIA 2024).

Following a decade of expansion, growth in demand response program participation stalled in the mid-2010s partially because of depressed capacity prices, forecasted over-capacity, and increasingly restrictive wholesale market participation rules (Hledik et al. 2019). However, the resurgence of load growth and increasing capacity prices, coupled with ongoing advancements in DERs and grid information and communication technologies (ICT) appears likely to reverse this trend.

9 RTO/ISO and retail data may overlap.

Studies of national demand response potential have identified a range of potential scenarios (Becker et al. 2024), ranging as high as 200 GW by 2030 in a 2019 study, comprising 20% of the then-forecasted system peak and yielding $15 billion in annual benefits primarily via avoided generation and transmission and distribution (T&D) capacity (Hledik et al. 2019). Notably, this research was conducted before recent load growth forecasts.

The Participation Gap: Data Centers and Demand Response

For nearly two decades, computational loads—and data centers in particular—have been identified as a promising area for demand response. Early studies explored these capabilities, such as a two-phase Lawrence Berkeley National Laboratory study drawing on six years of research, which concluded in 2010 that "data centers, on the basis of their operational characteristics and energy use, have significant potential for demand response" (Ghatikar et al. 2010) and in 2012 that "[certain] data centers can participate in demand response programs with no impact to operations or service-level agreements" (Ghatikar et al. 2012). The 2012 study provided one of the earliest demonstrations of computational load responsiveness, finding that 10% load shed can typically occur within 6 to 15 minutes.

Despite this promise, data centers have historically exhibited low participation rates in demand response programs as a result of operational priorities and economic incentives (Basmadjian 2019; Clausen et al. 2019; Wierman et al. 2014). Data centers are designed to provide reliable, uninterrupted service and generally operate under service-level agreements (SLAs) that mandate specific performance benchmarks, including uptime, latency, and overall quality of service. Deviation from these standards can result in financial penalties and reputational harm, creating a high-stakes environment where operators are averse to operational changes that introduce uncertainty or risk (Basmadjian et al. 2018).

Compounding this challenge is the increasing prevalence of large-scale colocated data centers, which represent a significant share of the data center market (Shehabi et al. 2024). These facilities house multiple tenants, each with varying operational requirements. Coordinating demand response participation in such environments introduces layers of administrative and logistical complexity, as operators must mediate cost- and reward-sharing agreements among tenants. Further, while data centers possess significant technical capabilities, tapping these capabilities for demand response requires sophisticated planning and expertise, which some operators may not have needed to date (Silva et al. 2024).

Economic considerations have further compounded this reluctance. Implementing a demand response program requires investments in advanced energy management systems, staff training, and integration with utility platforms for which costs can be material, particularly for smaller or midsized facilities. At the same time, financial incentives provided by most demand response programs have historically been modest and insufficient to offset the expenses and opportunity costs associated with curtailed operations. For operators focused on maintaining high utilization rates and controlling costs, the economic proposition of demand response participation may be unattractive.

Existing demand response program designs may inadvertently discourage participation. Many programs were originally created with traditional industrial consumers in mind, with different incentives and operational specifications. Price-based programs may require high price variability to elicit meaningful responses, while direct control programs without sufficient guardrails may introduce unacceptable risks related to uptime and performance. The complexity of active participation in demand response markets, including bidding processes and navigating market mechanisms, adds another layer of difficulty. Without streamlined participation structures, tailored incentives, and metrics that reflect the scale and responsiveness of data centers, many existing demand response programs may be ill-suited to the operational realities of modern data centers.

Table 1. Key Data Center Terms

Term	Definition
AI workload	A broad category encompassing computational tasks related to machine learning, natural language processing, generative AI, deep learning, and other AI-driven applications.
AI-specialized data center	Typically developed by hyperscalers, this type of facility is optimized for AI workloads and relies heavily on high-performance graphics processing units (GPUs) and advanced central processing units (CPUs) to handle intensive computing demands.
Computational load	A category of electrical demand primarily driven by computing and data processing activities, ranging from general-purpose computing to specialized AI model training, cryptographic processing, and high-performance computing (HPC).
Conventional data center	A facility that could range from a small enterprise-run server room to a large-scale cloud data center that handles diverse non-AI workloads, including file sharing, transaction processing, and application hosting. These facilities are predominantly powered by CPUs.
Conventional workload	A diverse array of computing tasks typically handled by CPUs, including file sharing, transaction processing, application hosting, and similar operations.
Cryptomine	A dedicated server farm optimized for high-throughput operations on blockchain networks, typically focused on validating and generating cryptocurrency.
Hyperscalers/hyperscale data centers	Large, well-capitalized cloud service providers that build hyperscale data centers to achieve scalability and high performance at multihundred megawatt scale or larger (Howland 2024b, Miller 2024).
Inferencing	The ongoing application of an AI model, where users prompt the model to provide responses or outputs. According to EPRI, inferencing represents 60% of an AI model's annual energy consumption (Aljbour and Wilson 2024).
Model training	The process of developing and training AI models by processing vast amounts of data. Model training accounts for 30–40% of annual AI power consumption and can take weeks or months to complete (Aljbour and Wilson 2024).

Rethinking Data Centers with AI-Driven Flexibility

Limited documentation of commercial data center participation in demand response has reinforced a perception that these facilities' demands are inherently inflexible loads. A variety of recent developments in computational load profiles, operational capabilities, and broader market conditions, however, suggest that a new phase of opportunity and necessity is emerging.

In a July 2024 memo on data center electricity demand, the SEAB recommended the Department of Energy prioritize initiatives to characterize and advance data center load flexibility, including the development of a "flexibility taxonomy and framework that explores the financial incentives and policy changes needed to drive flexible operation" (SEAB 2024). Building on these recommendations, EPRI announced a multi-year Data Center Flexible Load Initiative (DCFlex) in October 2024 with an objective "to spark change through hands-on and experiential demonstrations that showcase the full potential of data center operational flexibility and facility asset utilization," in partnership with multiple tech companies, electric utilities, and independent system operators (Walton 2024a).[10]

The central hypothesis is that the evolving computational load profiles of AI-specialized data centers facilitate operational capabilities that are more amendable to load flexibility. Unlike the many real-time processing demands typical of conventional data center workloads, such as cloud services and enterprise applications, the training of neural networks that power large language models and other machine learning algorithms is deferrable. This flexibility in timing, often referred to as *temporal flexibility*, allows for the strategic scheduling of training as well as other delay-tolerant tasks, both AI and non-AI alike. These delay-tolerant tasks are also referred to as *batch processing* and are typically not user-prompted (AWS 2025).

This temporal flexibility complements the developing interest in *spatial flexibility*, the ability to dynamically distribute workloads across one or multiple data centers in different geographic locations, optimizing resource utilization and operational efficiency. As stated by EPRI in a May 2024 report, "optimizing data center computation and geographic location to respond to electricity supply conditions, electricity carbon intensity, and other factors in addition to minimizing latency enables data centers to actively adjust their electricity consumption … some could achieve significant cost savings—as much as 15%—by optimizing computation to capitalize on lower electric rates during off-peak hours, reducing strain on the grid during high-demand periods" (EPRI 2024). For instance, having already developed a temporal workload shifting system, Google is seeking to implement spatial flexibility as well (Radovanović 2020).

In addition to temporal and spatial flexibility, other temporary load reduction methods may also enable data center flexibility. One approach is dynamic voltage and frequency scaling, which reduces server power consumption by lowering voltage or frequency at the expense of processing speed (Moons et al. 2017; Basmadjian 2019; Basmadjian and de Meer 2018). Another is server optimization, which consolidates workloads onto fewer servers while idling or shutting down underutilized ones, thereby reducing energy waste (Basmajian 2019; Chaurasia et al. 2021). These load reduction methods are driven by advances in virtual workload management, made possible by the "virtualization" of hardware (Pantazoglou et al. 2016).

10 Pointing to EPRI's new DCFlex Initiative, Michael Liebreich noted in a recent essay, "For instance, when they see how much it costs to work 24/7 at full power, perhaps data-center owners will see a benefit to providing some demand response capacity… " (Liebreich 2024).

Finally, temperature flexibility leverages the fact that cooling systems account for 30% to 40% of data center energy consumption (EPRI 2024). For instance, operators can increase cooling during midday when solar energy is abundant and reduce cooling during peak evening demand.[11] While these methods may be perceived as uneconomic due to potential impacts on performance, hardware lifespan, or SLAs, they are not intended for continuous use. Instead, they are best suited for deployment during critical hours when grid demand reduction is most valuable.

Beyond peak shaving, data centers also hold potential to participate in ancillary services, particularly those requiring rapid response, such as frequency regulation. Studies have described how data centers can dynamically adjust workloads to provide real-time support to the grid, effectively acting as "virtual spinning reserves" that help stabilize grid frequency and integrate intermittent renewable resources (McClurg et al. 2016; Al Kez et al. 2021; Wang et al. 2019). This capability extends beyond traditional demand response by providing near-instantaneous balancing resources (Zhang et al. 2022).

Three overarching market trends create further opportunities for load flexibility now than in the past. The first is constrained supply-side market conditions that raise costs and lead times for the interconnecting large inflexible loads, when speed to market is paramount for AI developers. The second is advancements in on-site generation and storage technologies that have lowered costs and expanded the availability of cleaner and more commercially viable behind-the-meter solutions, increasing their appeal to data center operators (Baumann et al. 2020). The third is the growing concentration of computational load in colocated or hyper-scale data centers—accounting for roughly 80% of the market in 2023—which is lending scale and specialization to more sophisticated data center operators. These operators, seek-ing speed to market, may be more likely to adopt flexibility in return for faster interconnec-tion (Shehabi et al. 2024; Basmadjian et al. 2018). The overarching trends underpinning this thesis are summarized in Table 2.

An important consideration for future data center load profiles is the balance between AI-specialized data centers focused on model development and those oriented toward inferencing. If fewer AI models are developed, a larger proportion of computing resources will shift toward inferencing tasks, which is delay-intolerant and variable (Riu et al. 2024). According to EPRI, training an AI model accounts for 30% of its annual footprint, compared to 60% for inferencing the same model (EPRI 2024).

In the absence of regulatory guidance, most advancements in data center flexibility to date are being driven by voluntary private-sector initiatives. Some hyperscalers and data center developers are taking steps to mitigate grid constraints by prioritizing near-term solutions for load flexibility. For example, one such company, Verrus, has established its business model around the premise that flexible data center operations offer an effective solution for growth needs (SIP 2024). Table 3 highlights additional initiatives related to facilitating or demonstrating data center flexibility.

11 Cooling demand for servers is inherently dependent on server workloads. Therefore, reducing workloads saves on cooling needs as well.

Table 2. Trends Enabling Data Center Load Flexibility

Category	Legacy	Future
Computational load profile	• Conventional servers with CPU-dominated workloads (Shehabi et al. 2024) • Real-time, delay-intolerant, and unscheduled processing (e.g., cloud services, enterprise apps) • Low latency critical	• AI-specialized servers with GPU or tensor processing unit (TPU)-favored workloads (Shehabi et al. 2024) • Greater portion of delay-tolerant and scheduled machine learning workloads (model training, non-interactive services) • Higher share of model training affords greater demand predictability • Highly parallelized workloads (Shehabi et al. 2024)
Operational capabilities	• Minimal temporal load shifting • Minimal spatial load migration • High proximity to end users for latency-sensitive tasks • Reliance on Tier 2 diesel generators for backup • Limited utilization of on-site power resulting from pollution concerns and regulatory restrictions (Cary 2023)	• More robust and intelligent temporal workload shifting (Radovanović et al. 2022) • Advanced spatial load migration and multi–data center training (D. Patel et al. 2024) • Flexibility in location for model training • Backup power diversified (storage, renewables, natural gas, cleaner diesel) • Cleaner on-site power enables greater utilization
Market conditions	• Minimal electric load growth • High availability of T&D network headroom • Standard interconnection timelines and queue volumes • Low supply chain bottlenecks for T&D equipment • Low capacity prices and forecasted overcapacity • High cost of clean on-site power options • Small-scale "server room" model	• High electric load growth • Low availability of T&D network headroom • Long interconnection timelines and overloaded queues • High supply chain bottlenecks for T&D equipment • High capacity prices and forecasted undercapacity (Walton 2024b) • Lower cost of clean on-site power options (Baranko et al. 2024) • Data center operations concentrating in large-scale facilities and operators

Table 3. Implementations of Computational Load Flexibility

Category	Examples
Operational flexibility	• Google deployed a "carbon-aware" temporal workload–shifting algorithm and is now seeking to develop geographic distribution capabilities (Radovanović 2020). • Google data centers have participated in demand response by reducing non-urgent compute tasks during grid stress events in Oregon, Nebraska, the US Southeast, Europe, and Taiwan (Mehra and Hasegawa 2023). • Enel X has supported demand response participation by data centers in North America, Ireland, Australia, South Korea, and Japan, including use of on-site batteries and generators to enable islanding within minutes (Enel X 2024). • Startup companies like Emerald AI are developing software to enable large-scale demand response from data centers through recent advances in computational resource management to precisely deliver grid services while preserving acceptable quality of service for compute users
On-site power	• Enchanted Rock, an energy solutions provider that supported Microsoft in building a renewable natural gas plant for a data center in San Jose, CA, created a behind-the-meter solution called Bridge-to-Grid, which seeks to provide intermediate power until primary service can be switched to the utility. At that point, the on-site power transitions to flexible backup power (Enchanted Rock 2024, 2025).
Market design and utility programs	• ERCOT established the Large Flexible Load Task Force and began to require the registration of large, interruptible loads seeking to interconnect with ERCOT for better visibility into their energy demand over the next five years (Hodge 2024). • ERCOT's demand response program shows promise for data center flexibility, with 750+ MW of data mining load registered as CLRs, which are dispatched by ERCOT within preset conditions (ERCOT 2023a). • PG&E debuted Flex Connect, a pilot that provides quicker interconnection service to large loads in return for flexibility at the margin when the system is constrained (Allsup 2024, St. John 2024).
Cryptomining	• A company generated more revenue from its demand response participation in ERCOT than from Bitcoin mining in one month, at times accommodating a 95% load reduction during peak demands (Riot Platforms 2023).

ANALYSIS OF CURTAILMENT-ENABLED HEADROOM

In this section we describe the method for estimating the gigawatts of new load that could be added to existing US power system load before the total exceeds what system planers are prepared to serve, provided that load curtailment is applied as needed. This serves as a proxy for the system's ability to integrate new load, which we term *curtailment-enabled headroom*.[12] We first investigated the aggregate and seasonal load factor for each of the 22 investigated balancing authorities, which measures a system's average utilization rate. Second, we computed the curtailment-enabled headroom for different assumptions of ac-

12 SEAB proposed a similar term, *available flex capacity*, in its July 2024 report *Recommendations on Powering Artificial Intelligence and Data Center Infrastructure*.

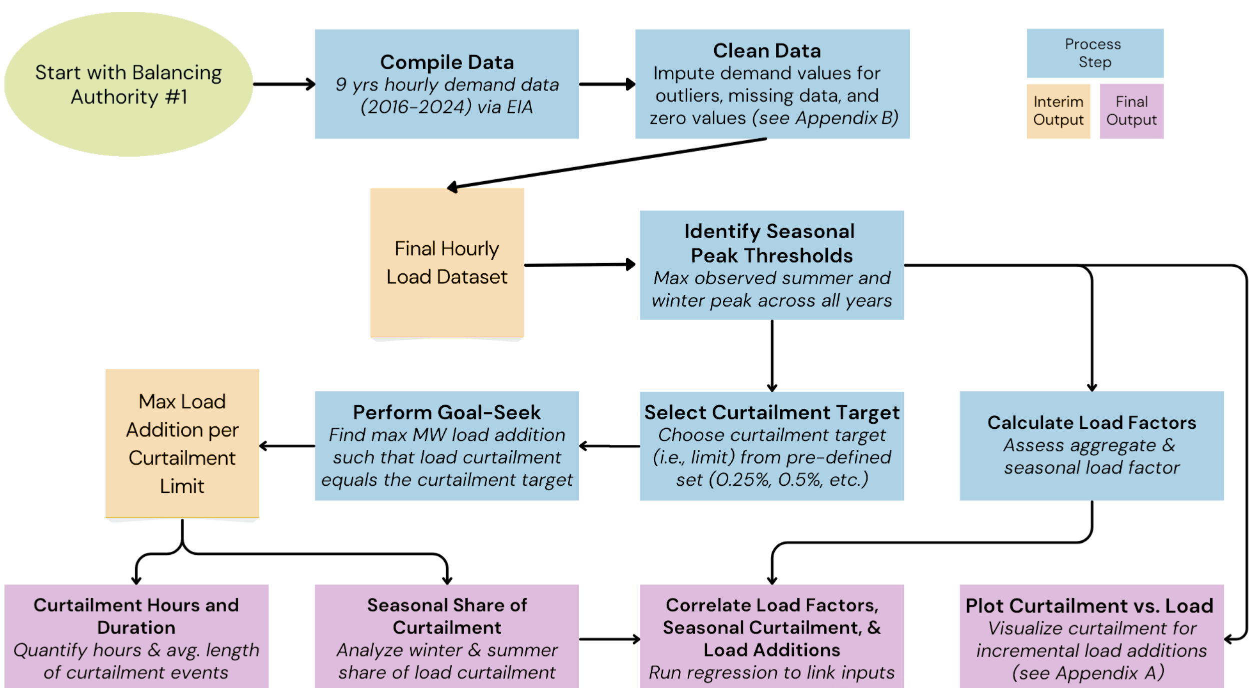
ceptable new load curtailment rates. In this context, *curtailment* refers to instances where the new load temporarily reduces its electricity draw—such as by using on-site generation resources, shifting load temporally or spatially, or otherwise reducing operations—to ensure system demand does not exceed historical peak thresholds. Third, we quantified the magnitude, duration, and seasonal concentration of the load curtailment for each balancing authority. Finally, we examined the correlation between load factor, seasonal curtailment, and max potential load additions. This process is summarized in Figure 3.

Data and Method

Data

We considered nine years of hourly load data aggregated for each of the 22 balancing authorities, encompassing seven RTO/ISOs,[13] eight non-RTO Southeastern BAs,[14] and seven non-RTO Western BAs.[15] Together, these balancing authorities represent 744 of the approximate 777 GW of summer peak load (95%) across the continental United States. The dataset, sourced from the EIA Hourly Load Monitor (EIA-930), contains one demand value per hour

Figure 3. Steps for Calculating Headroom and Related Metrics



13 CAISO, ERCOT, ISO-NE, MISO, NYISO, PJM, and SPP.

14 DEC; DEP; DEF; DESC; FPL; Santee Cooper, SCP; Southern Company (SOCO); and TVA. Note the different BA codes used by EIA: DUK for DEC, CPLE for DEP, SCEG for DESC, FPC for DEF, and SC for SCP. Also note that Southern Company includes Georgia Power, Alabama Power, and Mississippi Power. A complete list of abbreviations and their definitions can be found at the end of the paper.

15 AZPS, BPA, PACE, PACW, PGE, PSCO, and SRP. Note that EIA uses the code BPAT for BPA. A complete list of abbreviations and their definitions can be found at the end of the paper.

and spans January 1, 2016, through December 31, 2024.[16] Data from 2015 were excluded because of incomplete reporting.[17] The dataset was cleaned to identify and impute values for samples with missing or outlier demand values (see details in Appendix B).

Determining Load Additions for Curtailment Limits

An analysis was conducted to determine the maximum load addition for each balancing authority that can be integrated while staying within predefined curtailment limits applied to the new load. The load curtailment limits (0.25%, 0.5%, 1.0%, and 5.0%) were selected within the range of maximum curtailment caps for existing interruptible demand response programs.[18] The analysis focused on finding the load addition volume in megawatts that results in an average annual load curtailment rate per balancing authority that matches the specified limit. To achieve this, a goal-seek technique was used to solve for the load addition that satisfies this condition,[19] for which the mathematical expression is presented in Appendix C. The calculation assumed the new load is constant and hence increases the total system load by the same gigawatt volume hour-by-hour. To complement this analysis and visualize the relationship between load addition volume and curtailment, curtailment rates were also calculated across small incremental load additions (i.e., 0.25% of the BA's peak load).

Load Curtailment Definition and Calculation

Load curtailment is defined as the megawatt-hour reduction of load required to prevent the augmented system demand (existing load + new load) from exceeding the maximum seasonal system peak threshold (e.g., see Figure 4). Curtailment was calculated hourly as the difference between the augmented demand and the seasonal peak threshold. These hourly curtailments in megawatt-hours were aggregated for all hours in a year to determine the total annual curtailment. The curtailment rate for each load increment was defined as the total annual curtailed megawatt-hours divided by the new load's maximum potential annual consumption, assuming continuous operation at full capacity.

Peak Thresholds and Seasonal Differentiation

Balancing authorities develop resource expansion plans to support different peak loads in winter and summer. To account for variation, we defined seasonal peak thresholds for each balancing authority. Specifically, we identified the maximum summer peak and the maximum winter peak observed from 2016 to 2024 for each balancing authority.[20] These thresholds serve as the upper limits for system demand during their respective seasons, and all

16 Additional detail on EIA's hourly load data collection is available at https://www.eia.gov/electricity/gridmonitor/about.

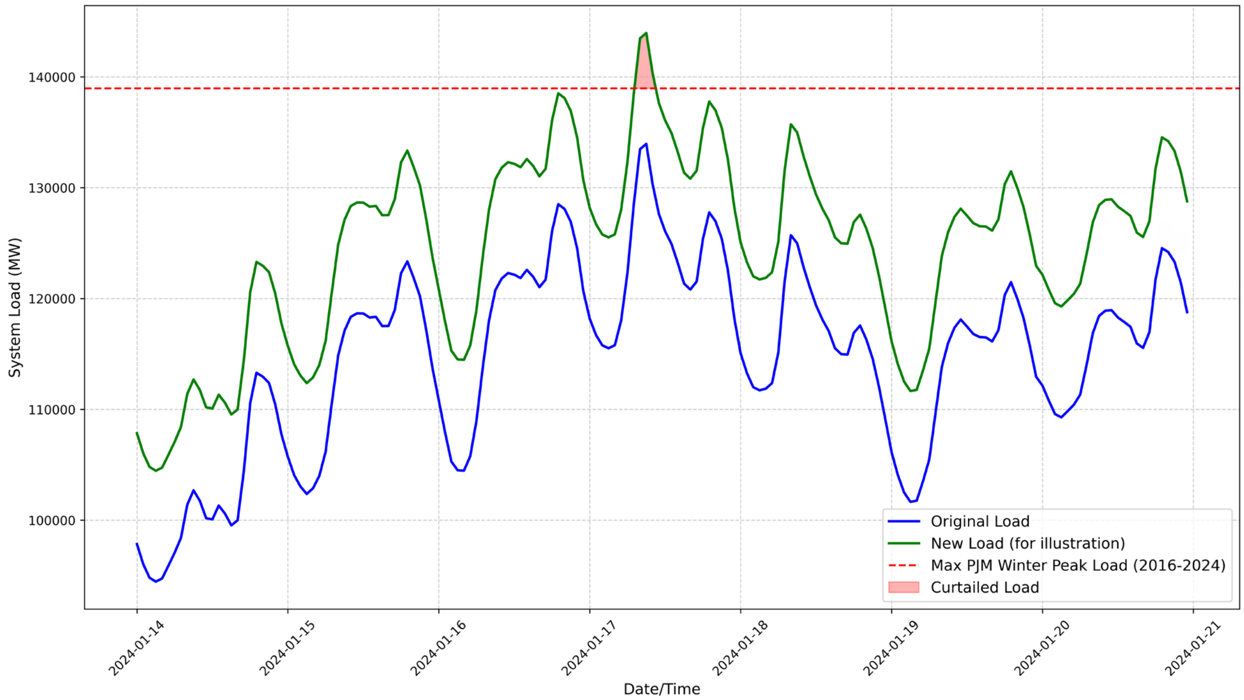
17 Fewer than half of the year's load hours were available, making the data unsuitable for inclusion.

18 For example, PG&E's and Southern California Edison's Base Interruptible Programs limit annual interruption for registered customers to a maximum of 180 hours (2.0% of all annual hours) or 10 events per month.

19 The goal-seek approach was implemented using Python's scipy.optimize.root_scalar function from the SciPy library. This tool is designed for solving one-dimensional root-finding problems, where the goal is to determine the input value that satisfies a specified equation within a defined range.

20 To identify the max seasonal peak load, summer was defined as June–August, while winter encompassed December–February. In a few cases, the BA's seasonal peak occurred within one month of these periods (AZPS winter, FPL winter, CAISO summer, CAISO winter), which were used as their max seasonal peak. To account for potential (albeit less likely) curtailment in shoulder months, the applicable summer peak was applied to April–May and September–October and the winter peak to November and March.

Figure 4. Illustrative Load Flexibility in PJM



megawatt-hours that exceeded these thresholds was counted as curtailed energy. This seasonal differentiation captures the distinct demand characteristics of regions dominated by cooling loads (summer peaks) versus heating loads (winter peaks).

Year-by-Year Curtailment Analysis

Curtailment was analyzed independently for each year from 2016 to 2024. This year-by-year approach captures temporal variability in demand patterns, including the effects of extreme weather events and economic conditions. For each year, curtailment volumes were calculated across all load addition increments, resulting in a list of annual curtailment rates corresponding to each load increment. To synthesize results across years, we calculated the average curtailment rate for each load addition increment by averaging annual curtailment rates over the nine years. This averaging process smooths out year-specific anomalies and provides an estimate of the typical system response to additional load. This analysis was also used to calculate the average number of hours of curtailment for each curtailment limit and the seasonal allocation of curtailed generation.[21] We also assessed the magnitude of load curtailment required during these hours as a share of the new load's maximum potential draw to calculate the number of hours when 90%, 75%, and 50% or more of the load would still be available.

21 Consistent with the curtailment analysis, summer was defined as June–August and winter as December–February. For BAs located on the Pacific coast (BPA, CAISO, PGE, PACE, PACW), November was counted as winter given the region's unique seasonal load profile.

Figure 5. Load Factor by Balancing Authority and Season, 2016–2024

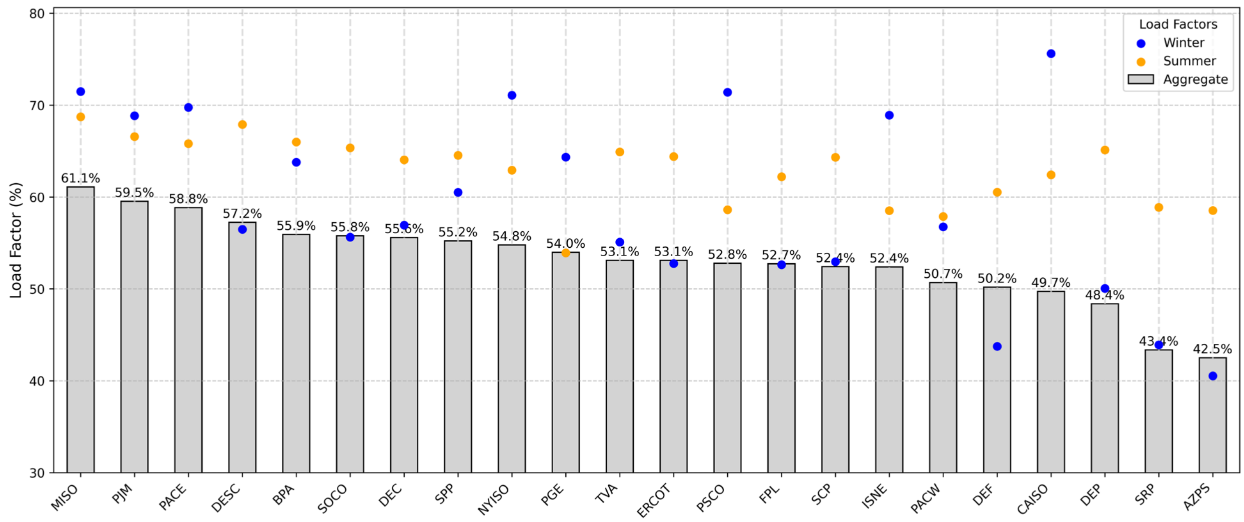
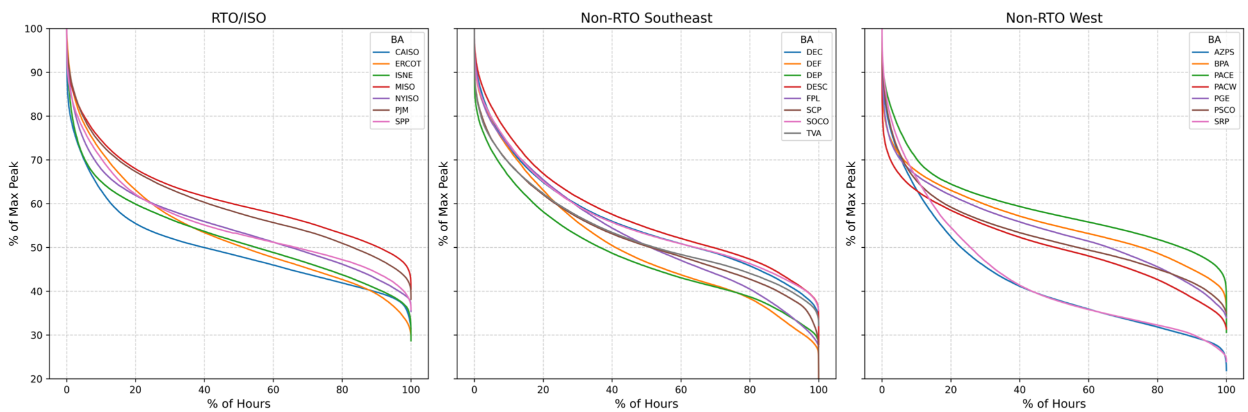


Figure 6. Load Duration Curves by Balancing Authority, 2016–2024



Results

Load Factor

In examining data for 22 balancing authorities, we found that aggregate load factors ranged between 43% to 61% (Figures 5 and 6), with an average and median value of 53%. The BAs with the lowest aggregate load factors were those in the desert southwest, Arizona Public Service Company (AZPS) and Salt River Project Agricultural Improvement and Power District (SRP). In terms of seasonal load factor, defined here as the average seasonal load as a share of seasonal maximum load (i.e., not as a share of the maximum all-time system load), winter load factors were notably lower than summer. The average and median winter load factor was 59% and 57% respectively, compared to 63% and 64% for summer. A majority of the balancing authorities had higher summer load factors (14) than winter (8).

Headroom Volume

Results show that the headroom across the 22 analyzed balancing authorities is between 76 to 215 GW, depending on the applicable load curtailment limit. This means that 76 to 215 GW of load could be added to the US power system and yet the total cumulative load would remain below the historical peak load, except for a limited number of hours per year

Figure 7. Headroom Enabled by Load Curtailment Thresholds, GW

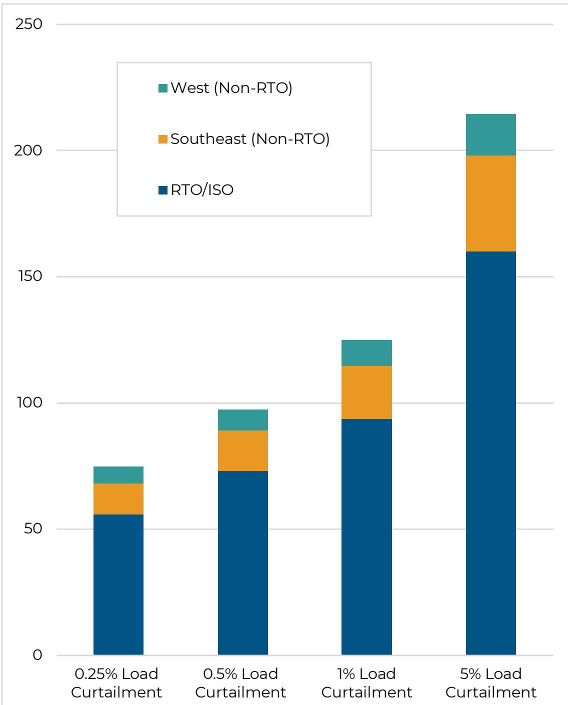


Figure 8. Headroom Enabled by 0.5% Load Curtailment by Balancing Authority, GW

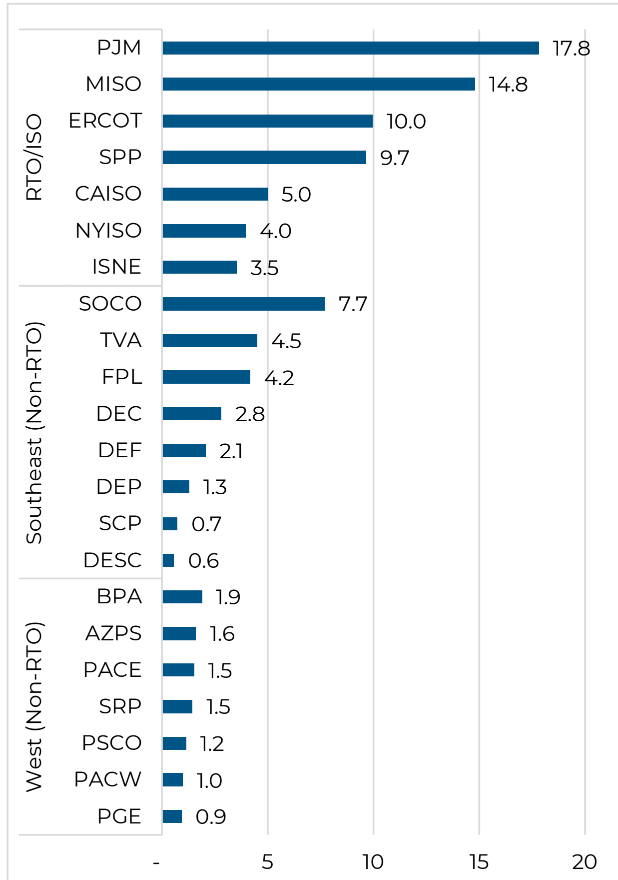
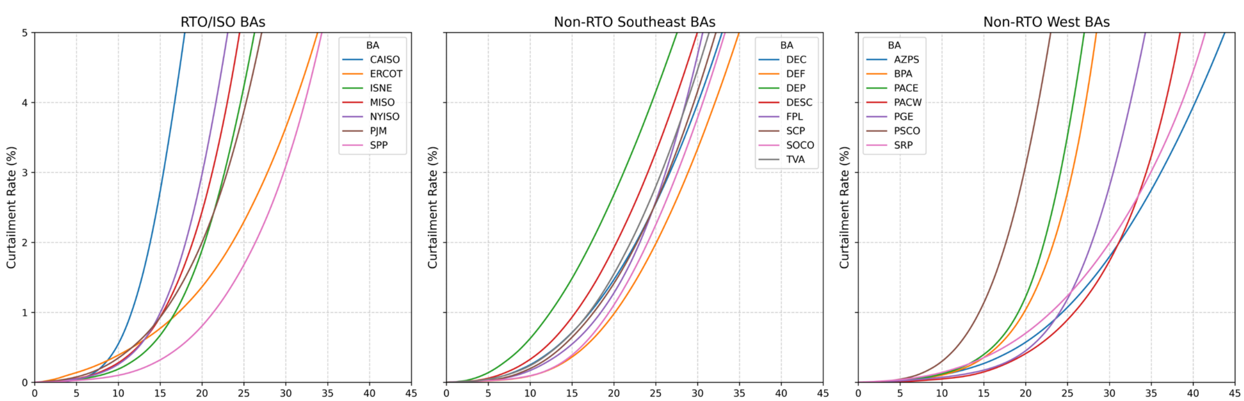


Figure 9. Load Curtailment Rate Due to Load Addition, % of System Peak



when the new load would be unserved. Specifically, 76 GW of headroom is available at an expected load curtailment rate of 0.25% (i.e., if 0.25% of the maximum potential annual energy consumption of the new load is curtailed during the highest load hours, or 1,643 out of 657,000 GWh). This headroom increases to 98 GW at 0.5% curtailment, 126 GW at 1.0% curtailment, and 215 GW at 5.0% curtailment (Figure 7). Headroom varies by balancing authority (Figure 8), including as a share of system peak (Figure 9). The five balancing authorities with the highest potential volume at 0.5% annual curtailment are PJM at 18 GW, MISO at 15 GW, ERCOT at 10 GW, SPP at 10 GW, and Southern Company at 8 GW. Detailed plots for each balancing authority, including results for each year, can be found in Appendix A.

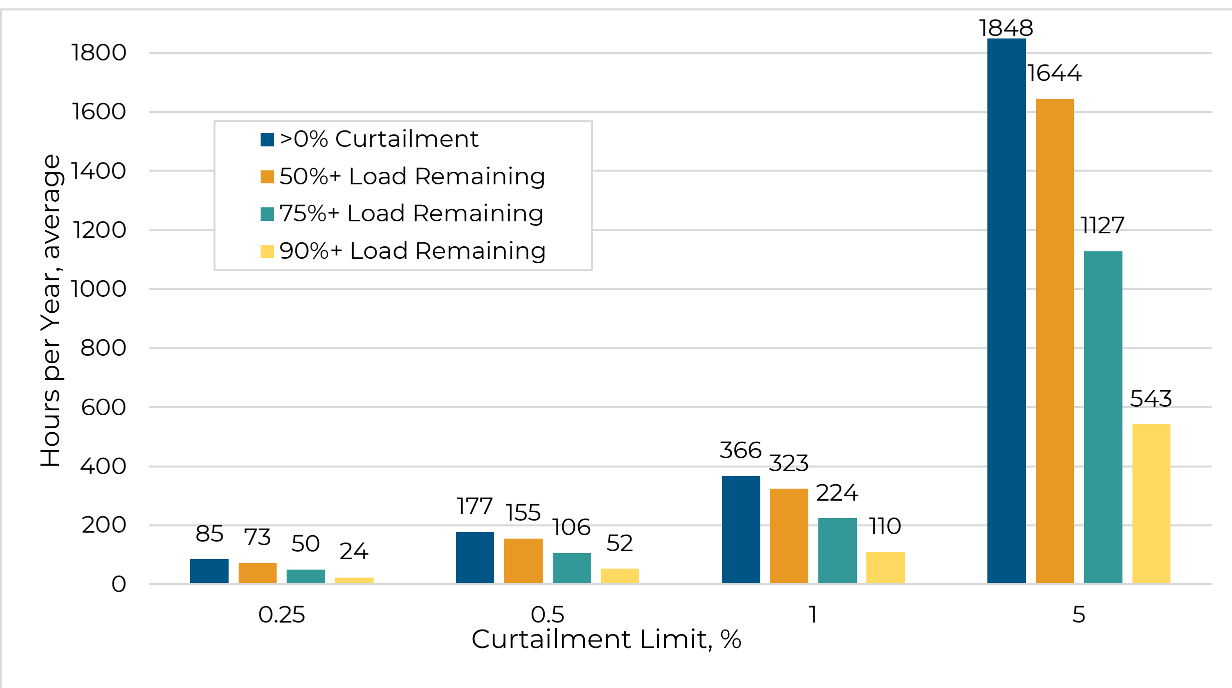
Curtailment Hours

A large majority of curtailment hours retain most of the new load. Most hours during which load reduction is required entail a curtailment rate below 50% of the new load. Across all 22 BAs, the average required load curtailment times are 85 hours under the 0.25% curtailment rate (~1% of the hours in a year), 177 hours under the 0.5% curtailment rate, 366 hours under the 1.0% curtailment rate, and 1,848 hours under the 5.0% curtailment rate (i.e., ~21% of the hours). On average, 88% of these hours retain at least 50% of the new load (i.e., less than 50% curtailment of the load is required), 60% of the hours retain at least 75% of the load, and 29% retain at least 90% of the load (see Figure 10).

Curtailment Duration

The analysis calculated the average hourly duration of curtailment events (i.e., the length of time the new load is curtailed during curtailment events). All hours in which any curtailment occurred were included, regardless of magnitude. The results for each balancing authority and curtailment limit are presented in Figure 11. The average duration across BAs was 1.7 hours for the 0.25% limit, 2.1 hours for the 0.5% limit, 2.5 hours for the 1.0% limit, and 4.5 hours for the 5.0% limit.

Figure 10. Hours of Curtailment by Load Curtailment Limit



Seasonal Concentration of Curtailment

The analysis reveals significant variation in the seasonal concentration of curtailment hours across balancing authorities. The winter-summer split ranged from 92% to 1% for CAISO (California Independent System Operator), where curtailment is heavily winter-concentrated, to 0.2% to 92% for AZPS,[22] which exhibited a heavily summer-concentrated curtailment profile (Figure 12a).[23]

Figure 11. Average Curtailment Duration by Balancing Authority and Curtailment Limit, Hours

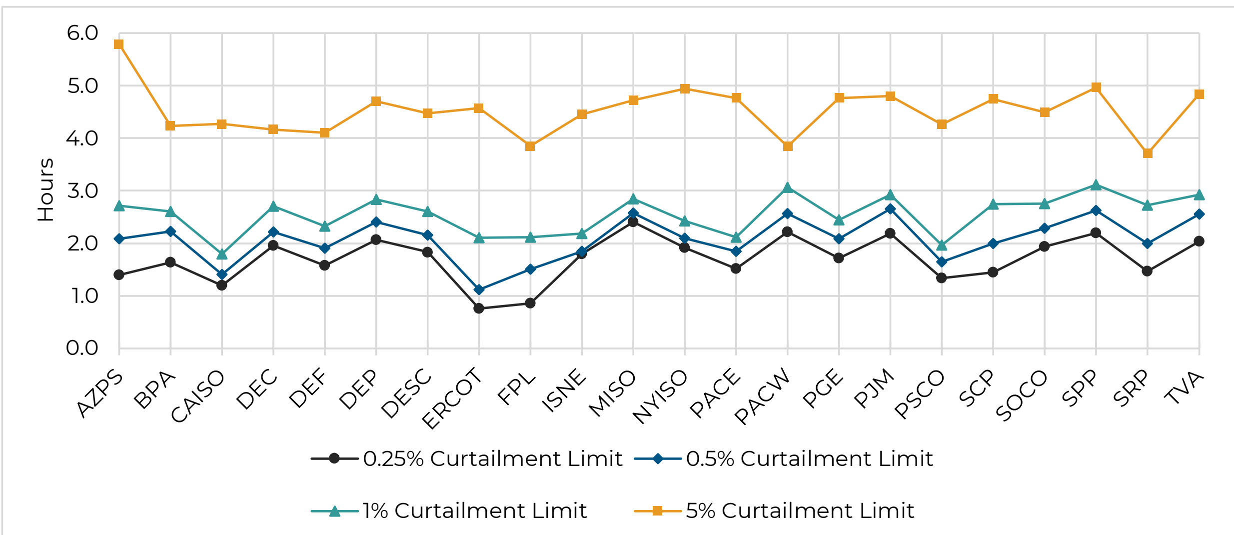
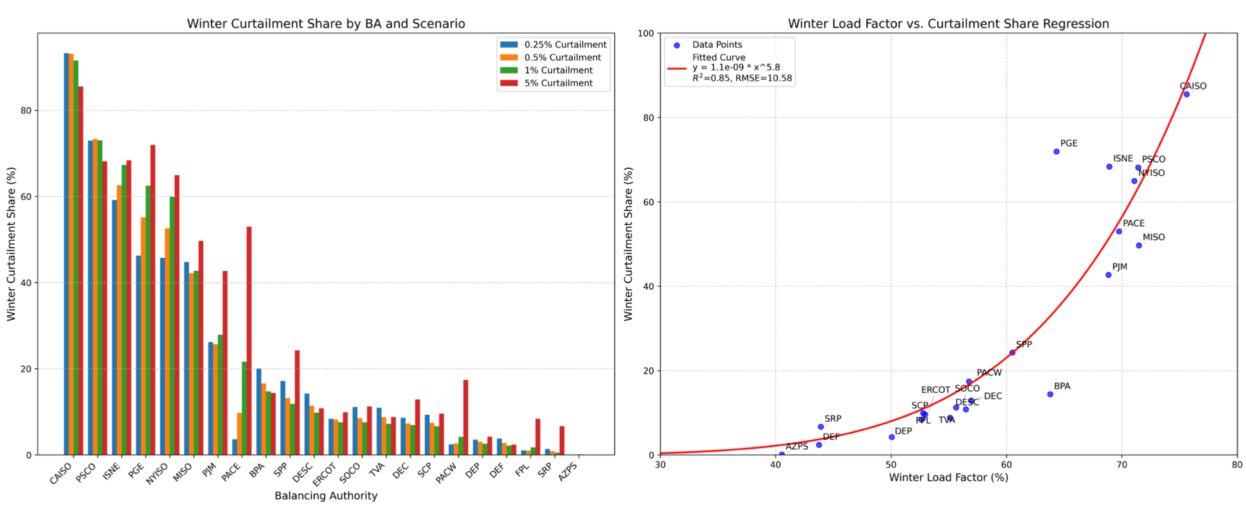


Figure 12. Seasonal Curtailment Analysis



(a) (b)

22 Note the remainder of the curtailment occurred in these BAs in shoulder months (i.e., not summer, not winter).

23 These values correspond to the seasonal curtailment concentration for the 1% curtailment limit.

A key observation is the strong correlation between the winter load factor (system utilization during winter months) and the seasonal allocation of curtailment hours (Figure 12b). BAs with lower winter load factors—indicating reduced system utilization during winter—tend to have greater capacity to accommodate additional load in winter while experiencing a dispro-portionately higher share of curtailment during summer months. This trend is particularly pronounced in balancing authorities located in the Sun Belt region, resulting in a lower win-ter concentration of curtailment hours.

While most BAs exhibited relatively stable seasonal curtailment shares across increasing load addition thresholds, some demonstrated notable shifts in seasonal allocation as load additions increased (e.g., PACW, FPL, NYISO, ISO-NE, PACE, PGE). These shifts highlight the dynamic interplay between system demand patterns and the incremental addition of new load.

Figure 12a illustrates this variability, showcasing the relationship between winter load factor and winter curtailment share across curtailment scenarios.[24]

Discussion

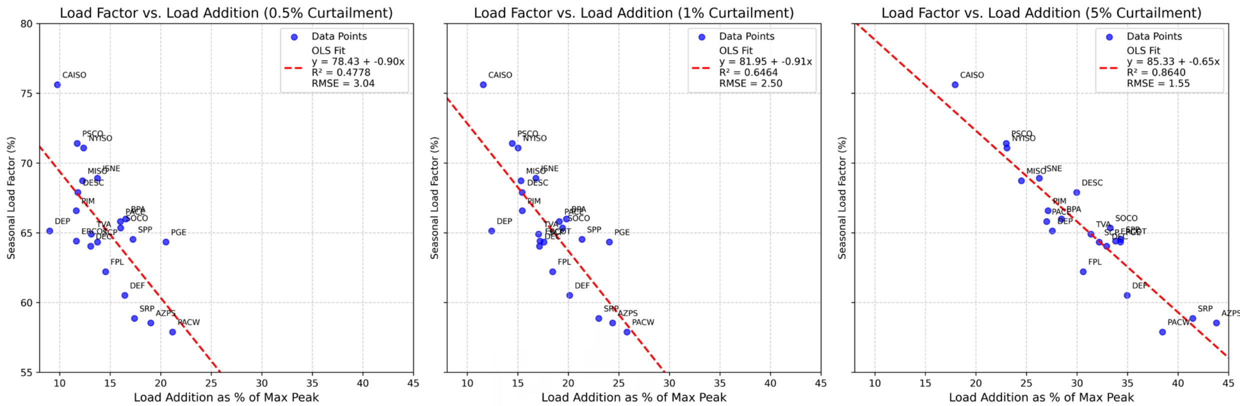
The results highlight that the significant headroom in US power systems—stemming from their by-design low load factors—could be tapped to enable the integration of substantial load additions with relatively low rates of load curtailment. They also underscore substantial variation in flexibility across balancing authorities, driven by differences in seasonal and aggregate load patterns. This variation suggests that seasonal load factors may be strongly linked to how much additional load a balancing authority can integrate without requiring high curtailment rates.

To explore this relationship, we analyzed system load factors in relation to the additional load that each balancing authority could accommodate while limiting the load curtailment rate to 0.5%, 1.0%, and 5.0% (i.e., the load curtailment limit). To allow for meaningful comparison across BAs, the additional load was standardized as a percentage of the BA's historical peak load. To account for whether a balancing authority's curtailment was concentrated in the summer or winter, the seasonal load factor was selected corresponding to the season with the highest share of curtailment.

The analysis revealed that BAs with higher seasonal load factors tended to have less headroom for the load curtailment limits examined (Figure 13). In simpler terms, systems with higher utilization during their busiest season had less power generation capacity planned to be available that could serve new load without hitting curtailment limits. For example, CAISO, with a seasonal load factor of 76%, could accommodate less additional load compared to PacifiCorp West (PACW) and AZPS, which exhibited lower seasonal load factors and supported larger load additions as a share of peak system load. This relationship grew in statistical significance as the load curtailment limit increased, yielding an R^2 value of 0.48 and an RMSE of 3.04 at the 0.5% curtailment limit, and an R^2 value of 0.86 and an RMSE of 1.55 at the 5% curtailment limit (i.e., 86% of the variation in load addition capacity across balancing authorities can be explained by differences in load factor at a curtailment limit of 5.0%).

24 Note in Figure 12b that a high-degree polynomial function captures the nonlinear growth in the area under the load curve as curtailed load exceeds a fixed peak threshold. This fit generally aligns with expectations, demonstrating that higher-degree terms are necessary to capture the relationship between load factor and curtailed load.

Figure 13. Load Factor Versus Max Load Addition as Share of Peak Load



These findings emphasize the importance of load factor as a predictor of curtailment-enabled headroom. BAs with more uneven peak seasonal demand—characterized by relatively low system utilization in winter or summer—tend to have greater capacity to integrate new loads with limited curtailment. Conversely, systems with more consistent demand across the winter and summer face tighter limits, as their capacity to absorb additional load is already constrained by elevated baseline usage.

Limitations

This analysis provides a first-order assessment of power generation capacity available for serving new curtailable loads, and hence is an exploration of the market potential for large-scale demand response. The primary focus of the analysis is to ensure that total demand, subject to curtailment limits for new load, stays below the system peak for which system planners have prepared. Other considerations important for planning—such as ensuring adequate transmission capacity, ramping capability, and ramp-feasible reserves, among others—are beyond the scope of this study and therefore the results cannot be taken as an accurate estimate of the load that can be added to the system. Additionally, the analysis assumes the new loads do not change current demand patterns but rather shift the existing demand curves upward, and a more precise assessment of the potential for integration of new loads would require detailed characterization of the temporal patterns of the load. There is significant variation in how system operators forecast and plan for system peaks, accounting for potential demand response, and as a result there will be differences in the methods used to estimate potential to accommodate new load. Despite these limitations, the results presented here signal a vast potential that, even if overstated, warrants further research.

On the other hand, some aspects of this study may have contributed to an underestimation of available headroom. First, the analysis assumes that each BA's maximum servable load in the winter and summer is equivalent to the BA's highest realized seasonal peak demand based on the available historical data. However, the available generation capacity in each balancing authority should materially exceed this volume when accounting for the installed reserve margin. In other words, system operators have already planned their systems to accommodate load volume that exceeds their highest realized peak. Second, the analysis removed outlier demand values in some BAs to avoid using unreasonably high maximum peak thresholds, which would understate the curtailment rates. However, if some of the removed outliers properly represent a level of system load that the system is prepared to serve reliably, this analysis may have understated the curtailment-enabled headroom. Third, the analysis assumed all new load is constant and hence increases the total system load by the same gigawatt hour-by-hour, which would tend to overstate the absolute level of required gigawatt hour curtailment for a load that is not constant.

Future Analysis

Enhancing this analysis to more accurately assess the capacity to integrate large curtailable load would require addressing the following considerations:

Network Constraints

This analysis does not account for network constraints, which would require a power flow simulation to evaluate the ability of the transmission system to accommodate additional load under various conditions. As such, the results should not be interpreted as an indication that the identified load volumes could be interconnected and served without any expansions in network capacity. While the existing systems are planned to reliably serve their peak loads, this planning is based on the current load topology and the spatial distribution of generation and demand across the transmission network. A large new load could avoid exceeding aggregate peak system demand by employing flexibility, yet still cause localized grid overloads as a result of insufficient transmission capacity in specific areas. Such overloads could necessitate network upgrades, including the expansion of transmission lines, substations, or other grid infrastructure. Alternatively, in the absence of network upgrades, localized congestion could be addressed through the addition of nearby generation capacity, potentially limiting the flexibility and economic benefits of the new load. These factors underscore the importance of incorporating network-level analyses to fully understand the operational implications of large flexible load additions.

Intertemporal Constraints

This analysis does not account for intertemporal constraints related to load and generator operations. For load operations, response times affect system operations and management of operational reserves. Faster response times from flexible loads could alleviate system stress more effectively during peak demand periods, potentially reducing the reliance on reserve capacity. Conversely, slower response times may require additional reserves to bridge the gap between the onset of system imbalances and the load's eventual response. Moreover, the rapid ramp-down of large flexible loads could lead to localized stability or voltage issues, particularly in regions with weaker grid infrastructure. These effects may necessitate more localized network analyses to evaluate stability risks and operational impacts. On the generation side, intertemporal constraints such as ramping limits, minimum up and down times, and startup times can affect the system's ability to integrate fast-response demand. For instance, ramping constraints may restrict how quickly generators can adjust output to align with the curtailment of flexible loads, while minimum uptime and downtime requirements can limit generator flexibility.

Loss of Load Expectation

Peak load is a widely used proxy for resource adequacy and offers a reasonable indicative metric for high-level planning analyses. However, a more granular assessment would incorporate periods with the highest loss of load expectation (LOLE), which represent the times when the system is most likely to experience supply shortfalls. Historically, LOLE periods have aligned closely with peak load periods, making peak load a convenient and broadly applicable metric. However, in markets with increasing renewable energy penetration, LOLE periods are beginning to shift away from traditional peak load periods. This shift is driven by the variability and timing of renewable generation, particularly solar and wind, which can alter the temporal distribution of system stress. As a result, analyses focused solely on peak load may understate or misrepresent the operational challenges associated with integrating large new loads into these evolving systems.

CONCLUSION

This study highlights extensive potential for leveraging large load flexibility to address the challenges posed by rapid load growth in the US power system. By estimating the curtailment-enabled headroom across balancing authorities, the analysis demonstrates that existing system capacity—intentionally designed to accommodate the extreme swings of peak demand—could accommodate significant new load additions with relatively modest curtailment, as measured by the average number, magnitude, and duration of curtailment hours.

The findings further emphasize the relationship between load factors and headroom availability. Balancing authorities with lower seasonal load factors exhibit greater capacity to integrate flexible loads, highlighting the importance of regional load patterns in determining system-level opportunities. These results suggest that load flexibility can play a significant role in improving system utilization, mitigating the need for costly infrastructure expansion and complementing supply-side investments to support load growth and decarbonization objectives.

This analysis provides a first-order assessment of market potential, with estimates that can be refined through further evaluation. In particular, network constraints, intertemporal operational dynamics, and shifts in loss-of-load expectation periods represent opportunities for future analyses that can offer a deeper understanding of the practical and operational implications of integrating large flexible loads.

In conclusion, the integration of flexible loads offers a promising, near-term strategy for addressing structural transformations in the US electric power system. By utilizing existing system headroom, regulators and market participants can expedite the accommodation of new loads, optimize resource utilization, and support the broader goals of reliability, affordability, and sustainability.

REFERENCES

Al Kez, D., A. M. Foley, F. W. Ahmed, M. O'Malley, and S. M. Muyeen. 2021. "Potential of Data Centers for Fast Frequency Response Services in Synchronously Isolated Power Systems." *Renewable and Sustainable Energy Reviews* 151(November): 111547. https://doi.org/10.1016/j.rser.2021.111547.

Aljbour, J., and T. Wilson. 2024. *Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption*. Palo Alto, CA: Electric Power Research Institute. https://www.wpr.org/wp-content/uploads/2024/06/3002028905_Powering-Intelligence_-Analyzing-Artificial-Intelligence-and-Data-Center-Energy-Consumption.pdf.

Allsup, M. 2024 "PG&E Is Laying the Groundwork for Flexible Data Center Interconnection." *Latitude Media*, November 15. www.latitudemedia.com/news/pg-e-is-laying-the-groundwork-for-flexible-data-center-interconnection/.

AWS. 2025. "What Is Batch Processing?" *Amazon Web Services*. aws.amazon.com/what-is/batch-processing/.

Baranko, K., D. Campbell, Z. Hausfather, J. McWalter, N. Ransohoff. 2024. "Fast, Scalable, Clean, and Cheap Enough: How Off-Grid Solar Microgrids Can Power the AI Race." *OffgridAI*, December. https://www.offgridai.us/.

Basmadjian, R. 2019. "Flexibility-Based Energy and Demand Management in Data Centers: A Case Study for Cloud Computing." *Energies* 12(17): 3301. https://doi.org/10.3390/en12173301.

Basmadjian, R., J. F. Botero, G. Giuliani, X. Hesselbach, S. Klingert, and H. De Meer. 2018. "Making Data Centers Fit for Demand Response: Introducing GreenSDA and GreenSLA Contracts." *IEEE Transactions on Smart Grid* 9(4): 3453–64. https://doi.org/10.1109/TSG.2016.2632526.

Basmadjian, R., and H. de Meer. 2018. "Modelling and Analysing Conservative Governor of DVFS-enabled Processors." In *Proceedings of the Ninth International Conference on Future Energy Systems (e-Energy '18)*. New York: Association for Computing Machinery. 519–25. https://doi.org/10.1145/3208903.3213778.

Baumann, C. 2020. *How Microgrids for Data Centers Increase Resilience, Optimize Costs, and Improve Sustainability*. Rueil-Malmaison, France: Schneider Electric. https://www.se.com/us/en/download/document/Microgrids_for_Data_Centers/.

Becker, J., K. Brehm, J. Cohen, T. Fitch, and L. Shwisberg. 2024. *Power Shift: How Virtual Power Plants Unlock Cleaner, More Affordable Electricity Systems*. RMI. https://rmi.org/wp-content/uploads/dlm_uploads/2024/09/power_shift_report.pdf.

Boucher, B. 2024. "The Challenge of Growing Electricity Demand in the US and the Shortage of Critical Electrical Equipment." *Wood Mackenzie*, May 2. https://www.woodmac.com/news/opinion/the-challenge-of-growing-electricity-demand-in-the-us-and-the-shortage-of-critical-electrical-equipment/.

Cary, P. 2023. "Virginia Environmental Regulators Drop Plan to Allow Data Centers to Rely on Diesel Generators." *Prince William Times*, April 12. www.princewilliamtimes.com/news/virginia-environmental-regulators-drop-plan-to-allow-data-centers-to-rely-on-diesel-generators/article_b337df48-d96a-11ed-8861-4b1de9b9963f.html.

Cerna, F., E. Naderi, M. Marzband, J. Contreras, J. Coelho, and M. Fantesia. 2022. "Load Factor Improvement of the Electricity Grid Considering Distributed Resources Operation and Regulation of Peak Load." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4293004.

Cerna, F. V., J. K. Coêlho, M. P. Fantesia, E. Naderi, M. Marzband, and J. Contreras. 2023. "Load Factor Improvement of the Electricity Grid Considering Distributed Energy Resources Operation and Regulation of Peak Load." *Sustainable Cities and Society* 98(November): 104802. https://doi.org/10.1016/j.scs.2023.104802.

Chaurasia, N., M. Kumar, R. Chaudhry, and O. P. Verma. 2021. "Comprehensive Survey on Energy-Aware Server Consolidation Techniques in Cloud Computing." *The Journal of Supercomputing* 77(10): 11682–737. https://doi.org/10.1007/s11227-021-03760-1.

Clausen, A., G. Koenig, S. Klingert, G. Ghatikar, P. M. Schwartz, and N. Bates. 2019. "An Analysis of Contracts and Relationships between Supercomputing Centers and Electricity Service Providers." In *Workshop Proceedings of the 48th International Conference on Parallel Processing*, 1–8. Kyoto: ACM. https://doi.org/10.1145/3339186.3339209.

Clean Energy Buyers Association. 2024. *Post-Technical Conference Comments of the Clean Energy Buyers Association*. FERC Docket No. AD24-11-000. Washington, DC: Federal Energy Regulatory Commission. https://elibrary.ferc.gov/eLibrary/filelist?accession_number=20241209-5198.

Cochran, J., P. Denholm, B. Speer, and M. Miller. 2015. *Grid Integration and the Carrying Capacity of the U.S. Grid to Incorporate Variable Renewable Energy*. Golden, CO: National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy15osti/62607.pdf.

ComEd. 2024. "Commonwealth Edison Company's Response to Constellation Energy Generation LLC's ('Constellation') Data Request Constellation-ComEd 5.01 RGP—Constellation-ComEd 5.04 RGP." *ICC Docket* Nos. 22-0486 / 23-0055 (Consol.) Refiled Grid Plan. Received May 30. https://www.icc.illinois.gov/docket/P2023-0055/documents/352947/files/617782.pdf.

Downing, J. N. Johnson, M. McNicholas, D. Nemtzow, R. Oueid, J. Paladino, and E. Bellis Wolfe. 2023. *Pathways to Commercial Liftoff: Virtual Power Plants*. Washington, DC: US Department of Energy. https://liftoff.energy.gov/wp-content/uploads/2023/10/LIFTOFF_DOE_VVP_10062023_v4.pdf.

EIA. 2024. "Electric Power Annual 2023." *US Energy Information Agency*. https://www.eia.gov/electricity/annual/.

Enchanted Rock. 2024. "Microsoft Is a Multinational Technology Company Producing Software, Electronics, PC's and Related Services like Data Centers." Case Study. December 12. https://enchantedrock.com/microsoft-is-a-multinational-technology-company-producing-software-electronics-pcs-and-related-services-like-data-centers/.

Enchanted Rock. 2025. "Bridge-To-Grid." enchantedrock.com/bridge-to-grid/.

Enel X. 2024. "How Data Centers Support the Power Grid with Ancillary Services?" July 10. www.enelx.com/tw/en/resources/how-data-centers-support-grids.

EPRI. 2024. *Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption*. Palo Alto, CA: Electric Power Research Institute. https://www.epri.com/research/products/000000003002028905.

ERCOT. 2023a. *Ancillary Services*. Austin: Electric Reliability Council of Texas. https://www.ercot.com/files/docs/2023/06/06/Ancillary-Services-Handout-0524.pdf

ERCOT. 2023b. *Large Loads—Impact on Grid Reliability and Overview of Revision Request Package*. Presented at the NPRR1191 and Related Revision Requests Workshop, August 16. https://www.ercot.com/files/docs/2023/11/08/PUBLIC-Overview-of-Large-Load-Revision-Requests-for-8-16-23-Workshop.pptx.

FERC. 2024a. *2024 Assessment of Demand Response and Advanced Metering*. Washington, DC: Federal Energy Regulatory Commission. https://www.ferc.gov/news-events/news/ferc-staff-issues-2024-assessment-demand-response-and-advanced-metering.

FERC. 2024b. *Form No. 714—Annual Electric Balancing Authority Area and Planning Area Report*. Washington, DC: Federal Energy Regulatory Commission. https://www.ferc.gov/industries-data/electric/general-information/electric-industry-forms/form-no-714-annual-electric/data

FERC. 2024c. *Transcript of Technical Conference on Large Loads Co-Located at Generating Facilities, November.* Washington, DC: Federal Energy Regulatory Commission. https://www.ferc.gov/media/transcript-technical-conference-regarding-large-loads-co-located-generating-facilities.

FERC. 2024d. Pro Forma *Large Generator Interconnection Agreement (LGIA).* 18 C.F.R. § 35.28, Appendix C, § 5.9.2. Washington, DC: Federal Energy Regulatory Commission. https://www.ferc.gov/sites/default/files/2020-04/LGIA-agreement.pdf.

GPC. 2023. *2023 Integrated Resource Plan Update*. Atlanta: Georgia Power Company. https://www.georgiapower.com/content/dam/georgia-power/pdfs/company-pdfs/2023-irp-update-main-document.pdf.

GPC. 2024. *Large Load Economic Development Report for Q3 2024 PD.* Atlanta: Georgia Power Company. https://psc.ga.gov/search/facts-document/?documentId=220461.

GPSC. 2025. "PSC Approves Rule to Allow New Power Usage Terms for Data Centers," January 23. Atlanta: Georgia Public Service Commission. https://psc.ga.gov/site/assets/files/8617/media_advisory_data_centers_rule_1-23-2025.pdf.

Ghatikar, G., M. A. Piette, S. Fujita, A. T. McKane, J. Q Han, A. Radspieler, K. C. Mares, and D. Shroyer. 2010. *Demand Response and Open Automated Demand Response Opportunities for Data Centers*. Berkeley, CA: Lawrence Berkeley National Laboratory. https://eta.lbl.gov/publications/demand-response-and-open-automated.

Ghatikar, G., V. Ganti, N. Matson, and M. A. Piette. 2012. *Demand Response Opportunities and Enabling Technologies for Data Centers: Findings From Field Studies*. Berkeley, CA: Lawrence Berkeley National Laboratory. https://doi.org/10.2172/1174175.

Gorman, W., J. Mulvaney Kemp, J. Rand, J. Seel, R. Wiser, N. Manderlink, F. Kahrl, K. Porter, and W. Cotton. 2024. "Grid Connection Barriers to Renewable Energy Deployment in the United States." *Joule* (December): 101791. https://doi.org/10.1016/j.joule.2024.11.008.

Hledik, R., A. Faruqui, T. Lee, and J. Higham. 2019. *The National Potential for Load Flexibility: Value and Market Potential Through 2030*. Boston: The Brattle Group. https://www.brattle.com/wp-content/uploads/2021/05/16639_national_potential_for_load_flexibility_-_final.pdf.

Hodge, T. 2024. "Data Centers and Cryptocurrency Mining in Texas Drive Strong Power Demand Growth." *Today in Energy*, October 3. www.eia.gov/todayinenergy/detail.php?id=63344.

Howland, E. 2024a. "FERC Rejects Basin Electric's Cryptocurrency Mining Rate Proposal." *Utility Dive*, August 21. https://www.utilitydive.com/news/ferc-basin-electricscryptocurrency-bitcoin-mining-rate-proposal/724811/.

Howland, E. 2024b. "FERC Rejects Interconnection Pact for Talen-Amazon Data Center Deal at Nuclear Plant." *Utility Dive*, November 4. https://www.utilitydive.com/news/ferc-interconnection-isa-talen-amazon-data-center-susquehanna-exelon/731841/.

Hurley, D., P. Peterson, and M. Whited. 2013. *Demand Response as a Power System Resource: Program Designs, Performance, and Lessons Learned in the United States*. Cambridge, MA: Synapse Energy Economics. https://www.synapse-energy.com/sites/default/files/SynapseReport.2013-03.RAP_.US-Demand-Response.12-080.pdf.

Inskeep, B. 2024. *Testimony on Behalf of Citizens Action Coalition of Indiana, Inc.* "In the Matter of the Verified Petition of Indiana Michigan Power Co. for Approval of Modifications to Its Industrial Power Tariff." Before the Indiana Regulatory Commission. October 15, 2024. https://iurc.portal.in.gov/docketed-case-details/?id=b8cd5780-0546-ef11-8409-001dd803817e

Intersect Power. 2024. *Post-Technical Conference Comments of the Intersect Power LLC*. FERC Docket No. AD24-11-000. Washington, DC: Federal Energy Regulatory Commission. https://elibrary.ferc.gov/eLibrary/filelist?accession_number=20241209-5237.

Jabeck, B. 2023. "Flexible Capacity: The Secret Weapon for Securing Interconnection." *Data Center Frontier*, July 28. https://www.datacenterfrontier.com/sponsored/article/33008776/flexible-capacity-the-secret-weapon-for-securing-interconnection.

Kearney, L., and L. Hampton. 2025. "US Power Stocks Plummet as DeepSeek Raises Data Center Demand Doubts." *Reuters*, January 27. https://www.reuters.com/business/energy/us-power-stocks-plummet-deepseek-raises-data-center-demand-doubts-2025-01-27/.

Lee, V., P. Seshadri, C. O'Niell, A. Choudhary, B. Holstege, and S. A Deutscher. 2025. *Breaking Barriers to Data Center Growth*. Boston: Boston Consulting Group. https://www.bcg.com/publications/2025/breaking-barriers-data-center-growth.

Li, F., L. Braly, and C. Post. "Current Power Trends and Implications for the Data Center Industry." *FTI Consulting*, July. https://www.fticonsulting.com/insights/articles/current-power-trends-implications-data-center-industry.

Liebreich, M. 2024. "Generative AI—The Power and the Glory." *BloombergNEF*, December 24. https://about.bnef.com/blog/liebreich-generative-ai-the-power-and-the-glory/.

McClurg, J., R. Mudumbai, and J. Hall. 2016. "Fast Demand Response with Datacenter Loads." In *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 1–5. Minneapolis: IEEE. https://doi.org/10.1109/ISGT.2016.7781219.

Mehra, V., and R. Hasegawa. 2023. "Supporting Power Grids with Demand Response at Google Data Centers." *Google Cloud Blog*, October 3. cloud.google.com/blog/products/infrastructure/using-demand-response-to-reduce-data-center-power-consumption.

Miller, R. 2024. "The Gigawatt Data Center Campus Is Coming." *Data Center Frontier*, April 29. https://www.datacenterfrontier.com/hyperscale/article/55021675/the-gigawatt-data-center-campus-is-coming.

Moons, B., W. Uytterhoeven, W. Dehaene, and M. Verhelst. 2017. "DVAFS: Trading Computational Accuracy for Energy Through Dynamic-Voltage-Accuracy-Frequency-Scaling." *Design, Automation & Test in Europe Conference & Exhibition*. 488-93. https://doi.org/10.23919/DATE.2017.7927038.

NERC. 2024 *Long-Term Reliability Assessment*. Atlanta: North American Electric Reliability Corporation. https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_Long%20Term%20Reliability%20Assessment_2024.pdf.

Nethercutt, E. J. 2023. *Demand Flexibility within a Performance-Based Regulatory Framework*. Washington, DC: National Association of Regulatory Utility Commissioners. https://pubs.naruc.org/pub/2A466862-1866-DAAC-99FB-E054E1C9AB13

NIAC. 2024. *Addressing the Critical Shortage of Power Transformers to Ensure Reliability of the U.S. Grid*. The National Infrastructure Advisory Council. Washington, DC: The President's National Infrastructure Advisory Council. https://www.cisa.gov/sites/default/files/2024-09/NIAC_Addressing%20the%20Critical%20Shortage%20of%20Power%20Transformers%20to%20Ensure%20Reliability%20of%20the%20U.S.%20Grid_Report_06112024_508c_pdf_0.pdf.

Norris, T. 2023. *Beyond FERC Order 2023: Considerations on Deep Interconnection Reform*. NI PB 23-04. Durham, NC: Nicholas Institute for Energy, Environment & Sustainability, Duke University. https://hdl.handle.net/10161/31260.

Norris, T. 2024. *Pre-Workshop Comments for FERC Staff-led Workshop on Innovations and Efficiencies in Generator Interconnection*. Docket No. AD24-9-000. Washington, DC: Federal Energy Regulatory Commission. https://nicholasinstitute.duke.edu/publications/comments-ferc-workshop-innovations-efficiencies-generator-interconnection.

Ohio Power Company. 2024. *Joint Stipulation and Recommendation before the Public Service Commission of Ohio*. "In the Matter of the Application of Ohio Power Company for New Tariffs Related To Data Centers and Mobile Data Centers," Case No. 24-508-EL-ATA, October 23. https://dis.puc.state.oh.us/ViewImage.aspx?CMID=A1001001A24J23B55758I01206.

Pantazoglou, M., G. Tzortzakis, and A. Delis. 2016. "Decentralized and Energy-Efficient Workload Management in Enterprise Clouds." *IEEE Transactions on Cloud Computing* 4(2): 196–209. https://doi.org/10.1109/TCC.2015.2464817.

Patel, D., D. Nishball, J. Eliahou Ontiveros. "Multi-Datacenter Training: OpenAI's Ambitious Plan To Beat Google's Infrastructure." *SemiAnalysis*, September 4. https://semianalysis.com/2024/09/04/multi-datacenter-training-openais/.

Patel, K., K. Steinberger, A. DeBenedictis, M. Wu, J. Blair, P. Picciano, and P. Oporto, et al. 2024. *Virginia Data Center Study: Electric Infrastructure and Customer Rate Impact*. San Francisco: Energy and Environmental Economics, Inc. https://jlarc.virginia.gov/pdfs/presentations/JLARC%20Virginia%20Data%20Center%20Study_FINAL_12-09-2024.pdf.

Radovanović, A. 2020. "Our Data Centers Now Work Harder When the Sun Shines and Wind Blows." *The Keyword*, April 22. https://blog.google/inside-google/infrastructure/data-centers-work-harder-sun-shines-wind-blows/.

Radovanović, A., R. Koningstein, I. Schneider, B. Chen, A. Duarte, B. Roy, D. Xiao, et al. 2023. "Carbon-Aware Computing for Datacenters." *IEEE Transactions on Power Systems* 38(2): 1270–80. https://doi.org/10.1109/TPWRS.2022.3173250.

Riot Platforms. 2023. "Riot Announces August 2023 Production and Operations Updates." September 6. www.riotplatforms.com/riot-announces-august-2023-production-and-operations-updates/.

Riu, I., D. Smiley, S. Bessasparis, and K. Patel. 2024. *Load Growth Is Here to Stay, but Are Data Centers? Strategically Managing the Challenges and Opportunities of Load Growth*. San Francisco: Energy and Environmental Economics, Inc. https://www.ethree.com/data-center-load-growth/.

Rohrer, J. 2024. "Supply Chains Impact Power Transmission Systems." *Closed Circuit*, April 23. https://www.wapa.gov/supply-chains/.

Rouch, M. A. Denman, P. Hanbury, P. Renno, and E. Gray. 2024. *Utilities Must Reinvent Themselves to Harness the AI-Driven Data Center Boom*. Boston: Bain & Company. https://www.bain.com/insights/utilities-must-reinvent-themselves-to-harness-the-ai-driven-data-center-boom/.

Ruggles, T. H., J. A. Dowling, N. S. Lewis, and K. Caldeira. 2021. "Opportunities for Flexible Electricity Loads Such as Hydrogen Production from Curtailed Generation." *Advances in Applied Energy* 3(August): 100051. https://doi.org/10.1016/j.adapen.2021.100051.

Saul, J. 2024. "Data Centers Face Seven-Year Wait for Dominion Power Hookups." *Bloomberg*, August 29. https://www.bloomberg.com/news/articles/2024-08-29/data-centers-face-seven-year-wait-for-power-hookups-in-virginia.

Schatzki, T., J. Cavicchi, and M. Accordino. 2024. *Co-Located Load: Market, Economic, and Ratemaking Implications*. Analysis Group. https://www.analysisgroup.com/globalassets/insights/publishing/2024_co_located_load_market_economic_and_ratemaking_implications.pdf.

SEAB. 2024. *Recommendations on Powering Artificial Intelligence and Data Center Infrastructure*. Washington, DC: US Secretary of Energy Advisory Board. https://www.energy.gov/sites/default/files/2024-08/Powering%20AI%20and%20Data%20Center%20Infrastructure%20Recommendations%20July%202024.pdf.

Shehabi, A., S. J. Smith, A. Hubbard, A. Newkirk, N. Lei, M. A. B. Siddik, and B. Holecek, et al. 2024. *2024 United States Data Center Energy Usage Report*. Berkeley, CA: Lawrence Berkeley National Laboratory.

Silva, C. A., R. Vilaça, A. Pereira, and R. J. Bessa. 2024. “A Review on the Decarbonization of High-Performance Computing Centers.” *Renewable and Sustainable Energy Reviews* 189(January): 114019. https://doi.org/10.1016/j.rser.2023.114019.

SIP. 2024. “Data Center Flexibility: A Call to Action Improving the Grid with a New Approach to Data Center Development.” *Sidewalk Infrastructure Partners*, March. https://www.datacenterflexibility.com/.

Springer, A. 2024. *Large Loads in ERCOT—Observations and Risks to Reliability*. Presented at the NERC Large Load Task Force, October 8. https://www.nerc.com/comm/RSTC/LLTF/LLTF_Kickoff_Presentations.pdf.

Srivathsan, B., M. Sorel, P. Sachdeva., H. Batra, R. Sharma, R. Gupta, and S. Choudhary. 2024. “AI Power: Expanding Data Center Capacity to Meet Growing Demand.” *McKinsey & Company*, October 29. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-power-expanding-data-center-capacity-to-meet-growing-demand.

St. John, J. 2024 “A New Way to Fix Grid Bottlenecks for EV Charging: Flexible Connection.” *Canary Media*, December 10. https://www.canarymedia.com/articles/transmission/a-new-way-to-fix-grid-bottlenecks-for-ev-charging-flexible-connection.

Walton, R. 2024a. “EPRI Launches Data Center Flexibility Initiative with Utilities, Google, Meta, NVIDIA.” *Utility Dive*, October 30. https://www.utilitydive.com/news/epri-launches-data-center-flexibility-initiative-with-NVIDIA-google-meta/731490/.

Walton, R. 2024b. “‘Explosive’” Demand Growth Puts More than Half of North America at Risk of Blackouts: NERC.” *Utility Dive*, December 18. www.utilitydive.com/news/explosive-demand-growth-blackouts-NERC-LTRA-reliability/735866/.

Wang, W., A. Abdolrashidi, N. Yu, and D. Wong. 2019. “Frequency Regulation Service Provision in Data Center with Computational Flexibility.” *Applied Energy* 251(October): 113304. https://doi.org/10.1016/j.apenergy.2019.05.107.

WECC. 2024. “State of the Interconnection.” *Western Electricity Coordinating Council*, September. https://feature.wecc.org/soti/topic-sections/load/index.html.

Wierman, A., Z. Liu, I. Liu, and H. Mohsenian-Rad. 2014. “Opportunities and Challenges for Data Center Demand Response.” In *International Green Computing Conference*, 1–10. Dallas, TX: IEEE. https://doi.org/10.1109/IGCC.2014.7039172.

Wilson, J. D., Z. Zimmerman, and R. Gramlich. 2024. *Strategic Industries Surging: Driving US Power Demand*. Bethesda, MD: GridStrategies. https://gridstrategiesllc.com/wp-content/uploads/National-Load-Growth-Report-2024.pdf.

Zhang, Y., D. C. Wilson, I. C. Paschalidis and A. K. Coskun. 2022. “HPC Data Center Participation in Demand Response: An Adaptive Policy With QoS Assurance.” *IEEE Transactions on Sustainable Computing* 7(1): 157–71. http://doi.org/10.1109/TSUSC.2021.3077254.

ABBREVIATIONS

AI	Artificial intelligence
AZPS	Arizona Public Service Company
BA	balancing authority
BPA	Bonneville Power Administration
CAGR	compound annual growth rate
CAISO	California Independent System Operator
CLRs	controllable load resources
CPUs	central processing units
DEC	Duke Energy Carolinas
DEF	Duke Energy Florida
DEP	Duke Energy Progress East
DERs	distributed energy resources
DESC	Dominion Energy South Carolina
EIA	Energy Information Administration
EPRI	Electrical Power Research Institute
ERCOT	Electric Reliability Council of Texas
ERIS	Energy Resource Interconnection Service
FERC	Federal Energy Regulatory Commission’s
FPL	Florida Power & Light
GPUs	graphics processing units
ICT	information, and communication technology
ISO-NE	ISO New England
LGIA	Large Generator Interconnection Agreement
LOLE	loss of load expectation
MISO	Midcontinent Independent System Operator
NYISO	New York Independent System Operator
PACE	PacifiCorp East
PACW	PacifiCorp West
PG&E	Pacific Gas and Electric
PGE	Portland General Electric Company
PJM	PJM Interconnection
PSCO	Public Service Company of Colorado
RMSE	Root mean square error
RTO/ISO	Regional transmission organization/independent system operator
SCP	Santee Cooper, South Carolina Public Service Authority
SEAB	Secretary of Energy Advisory Board
SLAs	service-level agreements
SOCO	Southern Company
SPP	Southwest Power Pool
SRP	Salt River Project Agricultural Improvement and Power District
TPU	tensor processing unit
TVA	Tennessee Valley Authority

APPENDIX A: CURTAILMENT-ENABLED HEADROOM PER BALANCING AUTHORITY

Figure A.1. Curtailment Rate Versus Load Addition by RTO/ISO, MW

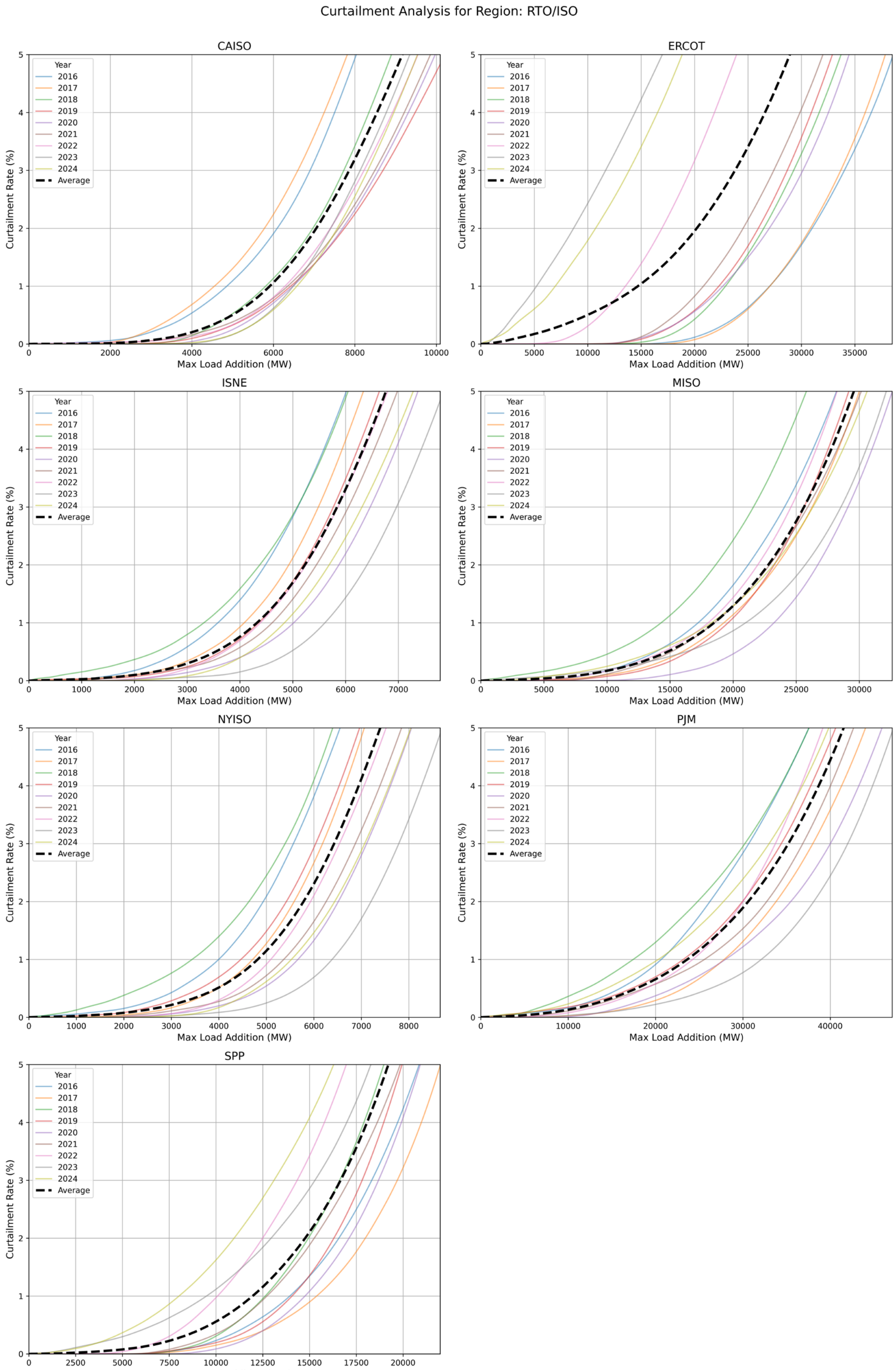


Figure A.2. Curtailment Rate Versus Load Addition by Non-RTO Southeastern Balancing Authority, MW

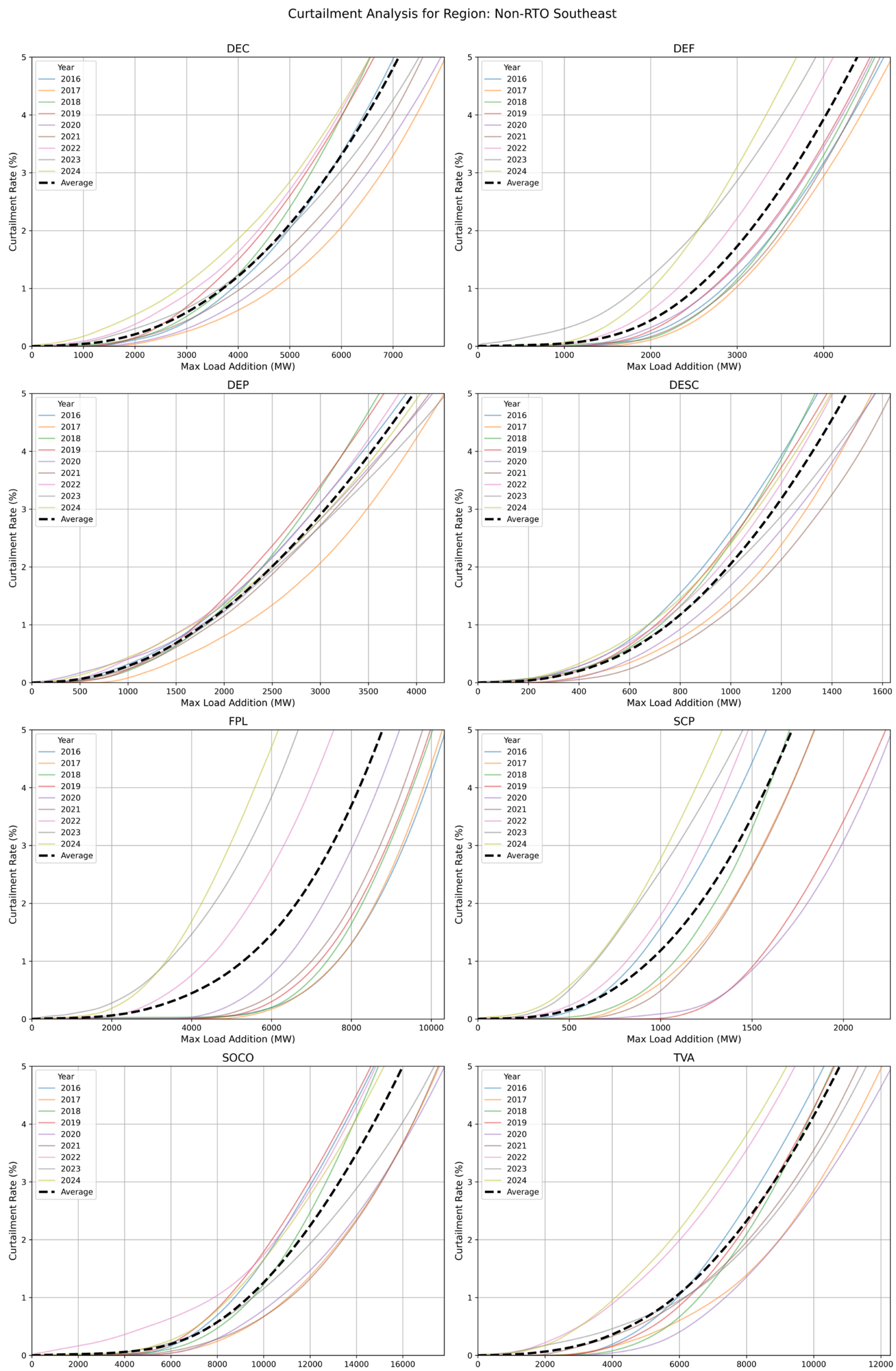
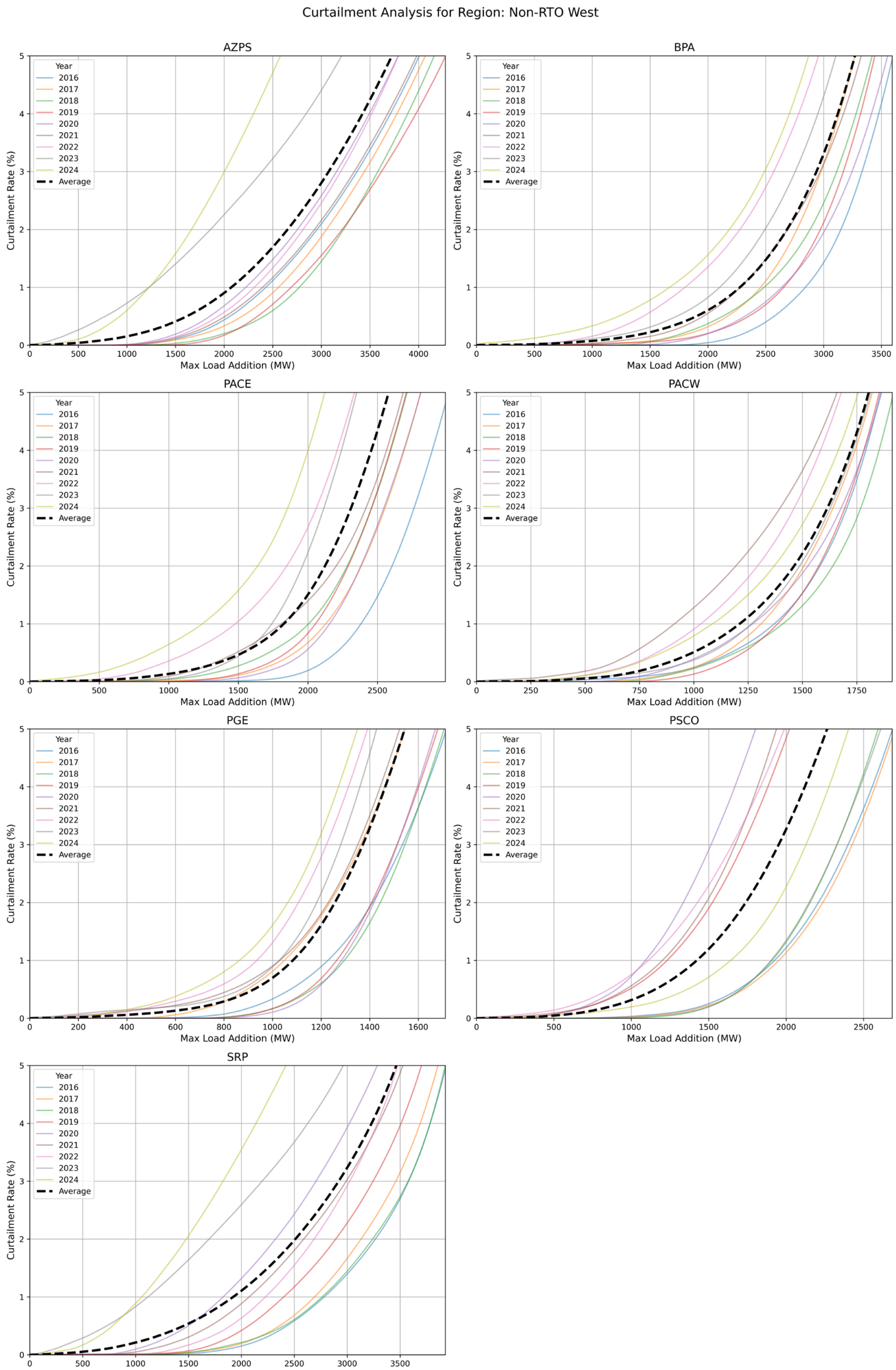


Figure A.3. Curtailment Rate Versus Load Addition by Non-RTO Western Balancing Authority, MW



APPENDIX B: DATA CLEANING SUMMARY

The data cleaning process attempted to improve the accuracy of nine years of hourly load data across the 22 balancing authorities, including the following steps:

1. Data normalization
    - **Dates:** Date-time formats were verified to be uniform.
    - **Demand data:** Where the balancing authority had an "Adjusted demand" value for a given hour, this value was used, otherwise its "Demand" value was used. The final selected values were saved as "Demand" and a log was kept.
    - **BA labels:** Labels were mapped to align with widely used acronyms, including:
        - CPLE → DEP
        - DUK → DEC
        - SC → SCP
        - SWPP → SPP
        - SCEG → DESC
        - FPC → DEF
        - CISO → CAISO
        - BPAT → BPA
        - NYIS → NYISO
        - ERCO → ERCOT
2. Identifying and handling outliers
    - **Missing and zero values:** Filled using linear interpolation between adjacent data points to maintain temporal consistency.
    - **Low outliers:** Demand values below a predefined cutoff threshold (such as 0 or extremely low values inconsistent with historical data) were flagged. Imputation for flagged low outliers involved identifying the closest non-outlier value within the same balancing authority and time period and replacing the flagged value.
    - **Spikes:** Sudden demand spikes that deviated significantly from historical patterns were flagged. Corrections were applied based on nearby, consistent data.
    - **Erroneous peaks:** Specific known instances of demand peaks that are outliers (e.g., caused by reporting errors) are explicitly corrected or replaced with average values from adjacent time periods.
3. **Data validation**:
    - Seasonal and annual peak loads, load factors, and other summary statistics were computed and inspected to ensure no unexpected results. Max peaks were compared to forecasted peaks collected by FERC to ensure none were out of range.
    - Logs summarizing corrections, including the number of spikes or outliers addressed for each balancing authority, were saved as additional documentation.

APPENDIX C: CURTAILMENT GOAL-SEEK FUNCTION

Mathematically, the function can be expressed as

$$\frac{1}{N}\sum_{y=1}^{N}\left(\frac{Curtailment_y(L)}{L\ \cdot\ 8{,}760}\cdot 100\right) = \ CurtailLimit$$

where

L	=	load addition in MW (constant load addition for all hours)
N	=	total number of years in the analysis (2016–2024)
$Curtailment_y(L)$	=	curtailed MWh for year y at load addition L
$L \cdot 8{,}760$	=	maximum potential energy consumption of the new load operating continuously at full capacity
$CurtailLimit$	=	predefined curtailment limit (e.g., 0.25%, 0.5%, 1.0%, or 5.0%).

For each hour t in year y, the curtailment is defined as

$$Curtailment_t(L) = \max(0, Demand_t + L - Threshold_t)$$

where

L	=	load addition being evaluated in MW
$Demand_t$	=	system demand at hour t in MW
$Threshold_t$	=	seasonal peak threshold applicable for hour t in MW (i.e., the maximum winter or summer peak across all years)

These hourly curtailments are aggregated to find the total annual curtailment

$$Curtailment_y(L)\ =\ \sum_{t\,\in\,T_y} Curtailment_t(L)$$

where

T_y	=	all hours in year y.

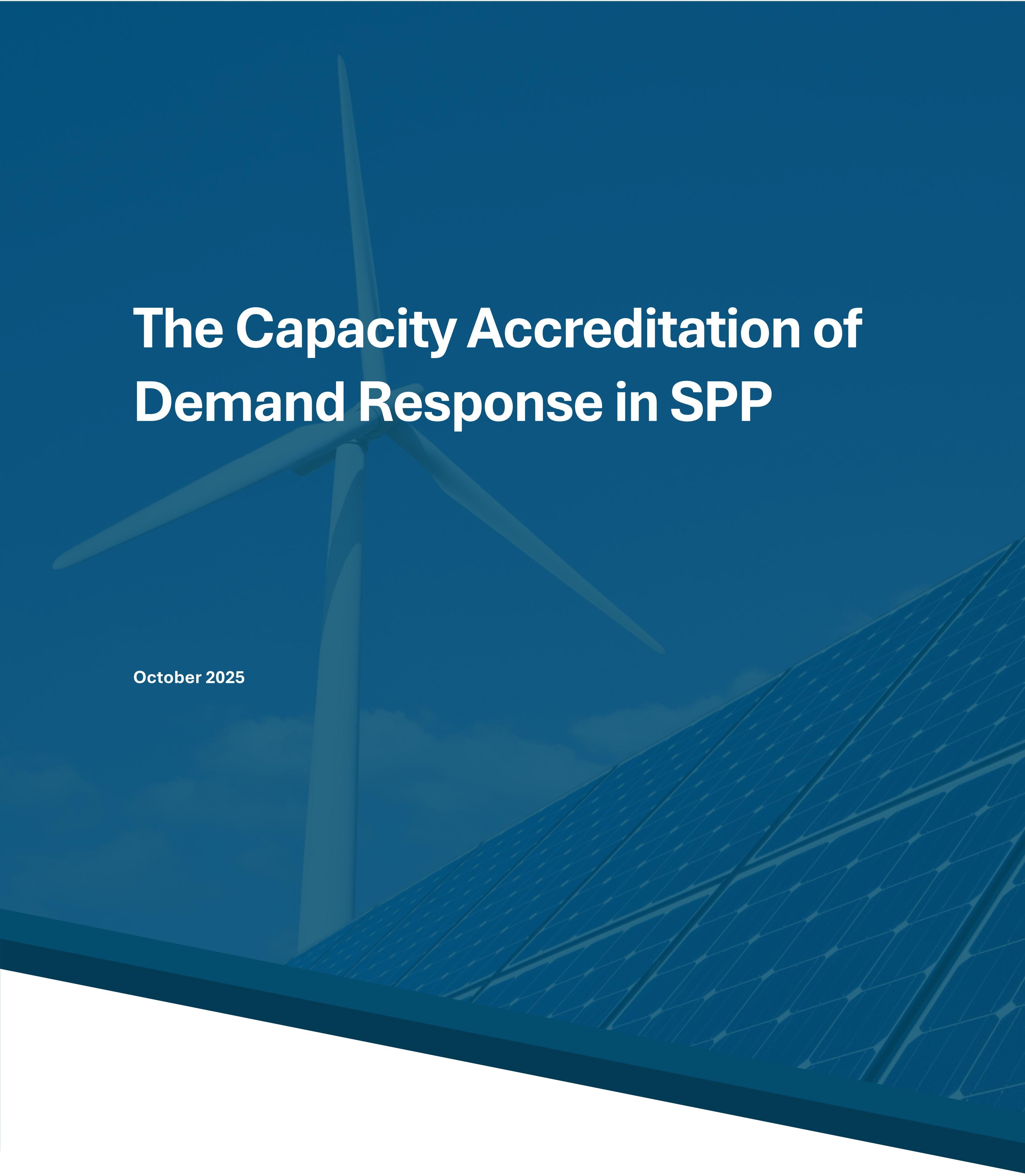
Replacing $Curtailment_y(L)$ in the original formula, the integrated formula becomes

$$\frac{1}{N}\sum_{y=1}^{N}\left(\frac{\sum_{t\,\in\,T_y}\max(0, Demand_t +\ L\ -\ Threshold_t)}{L\ \cdot\ 8{,}760} * 100\right) = \ CurtailLimit$$

Attachment 2

The Capacity Accreditation of Demand Response in SPP

October 2025



Energy+Environmental Economics

Authors & Acknowledgements

Project Team

Energy and Environmental Economics, Inc. (E3) is a leading economic consultancy focused on the clean energy transition. For over 30 years, E3's analysis has been utilized by the utilities, regulators, developers, and advocates that are writing the script for the clean energy transition in leading-edge jurisdictions such as California, New York, Hawaii and elsewhere. E3 has offices in San Francisco, Boston, New York, Denver, and Calgary.

E3 Study Team: Adrian Au; Grant Freudenthaler; Ben Elsey; Hugh Somerset; Edita Danielyan; Zachary Ming.

Acknowledgments

This paper was partially funded by Google LLC (Google), who reviewed the analysis prior to publication. Modeling and analysis for this study was conducted in 2025. Inputs and assumptions reflect the best available data from public sources at the time. E3 is an independent economic consulting firm, and all paper conclusions and analyses are our own.

Table of Contents

Executive Summary

The Southwest Power Pool (SPP) is undergoing significant changes, driven by a transformation in generation mix, scheduled fossil retirements, extreme weather events like Winter Storm Uri in 2021, and growing electricity demand from data centers and other large loads. These dynamics have revealed increasing resource adequacy (RA) risks that occur outside of traditional Summer peak periods. In response, SPP has updated its planning reserve margin (PRM) requirements, prompting Load Responsible Entities (LREs) to procure new resources to meet the new Winter resource adequacy requirement in addition to their existing Summer obligations. Like many other power markets, the SPP also faces challenges in interconnection timelines and developers face supply chain constraints for new development. These issues continue to pose risk to meeting the system's near-term reliability needs.

New loads coming online is part of SPP's growing resource adequacy need but is also a potential source of system flexibility. As with most US electricity markets today, data centers and other large loads represent the largest component of near-term load growth. In the longer term, drivers of load growth may also include electrification of transportation, buildings, and industrial processes. To help integrate this unprecedented demand, the industry is exploring whether data centers and other large loads can use load flexibility to defer some of the incremental resource adequacy need.[1]

This paper looks to further the industry's understanding of the capacity accreditation of demand response (DR) within an established capacity market framework for Independent System Operators (ISOs)/Regional Transmission Operators (RTOs). Specifically, this study focuses on calculating DR capacity accreditation under an **Effective Load Carrying Capability (ELCC)** framework within the SPP capacity market design and aims to answer the following key questions:

+ What is the ELCC based capacity accreditation value of DR in SPP?
+ Which DR parameters (duration, number of hours available, and total available curtailable load) have the largest impact on its capacity accreditation in SPP?
+ How often and at what times should DR be expected to be called to achieve high capacity accreditation?

This paper evaluates DR under different scenarios using RECAP,[2] E3's loss-of-load-probability (LOLP) model that has been used across North America to answer similar questions. We simulated the SPP market within RECAP to determine critical hours within the Summer and Winter seasons, as the first step toward determining the capacity accreditation of DR. SPP's critical hours occur both in the Summer and the Winter periods, during the abnormally hot or cold periods with high loads and low resource availability. In the Summer, heat waves lead to higher-than-normal cooling demands in the afternoon, but electricity demand falls as the region cools at night. This results in Summer critical

[1] For example see some recent articles on data center DR: How Data Centers Can Set the Stage for Larger Loads to Come - RMI, Existing US grid can handle 'significant' new flexible load: report | Utility Dive, Internet data centers participating in demand response: A comprehensive review - ScienceDirect

[2] E3's RECAP Model: RECAP - E3, and documentation applicable to this study: RECAP-Documentation.pdf

hours during the afternoon lasting 2-6 hours, with critical hours occurring in several consecutive days. On the other hand, cold snaps often last longer than 6 hours. Heating demand may even persist across multiple days, leading to multi-hour or multi-day periods of consecutive critical hours in Winter. The ability to meet demand during these critical hours across both seasons impacts all resources' capacity accreditation, including that of DR.

This study uses SPP's definition of ELCC as the metric to quantify DR capacity accreditation. SPP currently uses ELCC to accredit variable resources (wind and solar) and battery storage. Increasingly, markets and planners use the ELCC as a method to calculate the capacity accreditation of intermittent and energy-limited resources, to reflect their performance during system critical hours. ELCC captures the interactions of increasing penetrations of the same resource and the interactions between other resources within the portfolio, which makes it useful to measure the changing accreditation of DR in this study.[3]

Quantifying DR's Capacity Accreditation in SPP

To evaluate the capacity accreditation of DR in SPP, we quantify DR's ELCC across a suite of scenarios. These scenarios are designed to provide insight into the impact that key parameters have on DR's capacity accreditation within the SPP system. The scenarios test duration, calls, and penetration of total available curtailable load for the Summer and Winter season across two test years, 2025 and 2030.[4]

Table 1: Eight Study Scenarios

Duration	Seasonal Call Limit	Penetration	Scenario Description
4 hours	40 hours/season	2 GW	Approximates the current level of DR adoption and duration
4 hours	40 hours/season	4 GW	Tests moderate DR growth and entry of new large loads
4 hours	40 hours/season	6 GW	Bookend scenario of high DR entry
4 hours	No Seasonal Call Limit	6 GW	Designed to understand the impact of adding more call hours with high saturation
10 hours	100 hours/season	2 GW	Approximates current levels of DR adoption and tests the value of longer duration
10 hours	100 hours/season	4 GW	Tests moderate DR growth and the value of longer duration
10 hours	100 hours/season	6 GW	Bookend scenario of high DR entry and duration
10 hours	No Seasonal Call Limit	6 GW	Designed to estimate the impact of adding more call hours under high DR saturation

[3] For more detail on ELCC and its application to markets, see E3's 2025 paper *Resource Adequacy for the Energy Transition: A Critical hours Reliability Framework and its Applications in Planning and Markets* and 2020 paper *Capacity and Reliability Planning in the Era of Decarbonization.*

[4] E3 defines total available curtailable load as the total MW available to be curtailed by the market operator

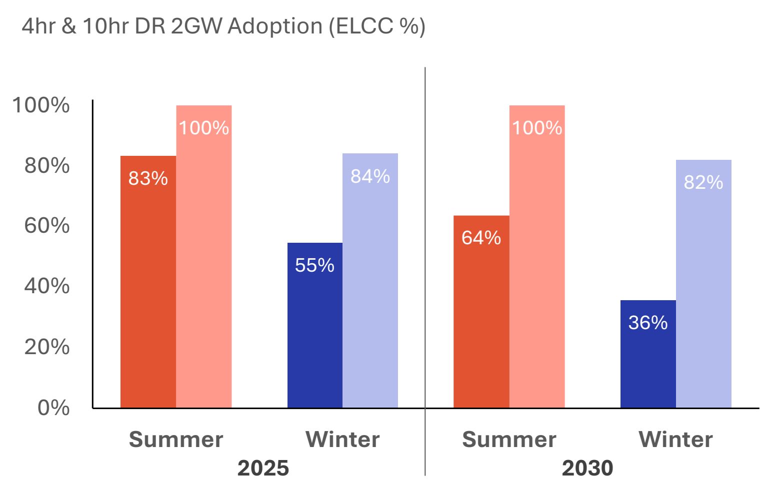
The results of the scenarios from Table 1 revealed the following insights**:**

- **Call Duration:** Extending event lengths from four to ten hours increases ELCC materially, with the largest increase in capacity accreditation in the Winter.
- **Call Limits:** Increasing call limits increases ELCCs only under high penetrations of DR.
- **DR Penetration:** First 2 GW of DR yields high capacity accreditation, but with increasing DR penetration, saturation effects emerge.
- **Critical Hours:** Across simulations in this study, SPP Summer exhibits critical hour periods lasting 3-5 hours whereas Winter critical hour periods often exceed 10 hours. DR has accordingly higher ELCCs in Summer than Winter.[5]
- **Portfolio Composition:** DR ELCCs are positively influenced by increasing renewables penetration due to diversity benefits but negatively impacted by competition from other energy limited resources (ELRs) like battery storage.

Call Duration

Four-hour DR has a Summer ELCC of 83% and 64% in 2025 and 2030 respectively, and a Winter ELCC of 55% and 36% in 2025 and 2030 respectively. Ten-hour DR achieves higher ELCCs than four-hour, 100% ELCC in the Summer for 2025 and 2030, and between 82% and 84% ELCC in the Winter from 2025-2030. These ELCC findings highlight the impact that duration has on capacity accreditation in the SPP. They also highlight the seasonal impact of accreditation as both four- and ten-hour DR perform well during concentrated Summer afternoon risk windows, while ten-hour DR provides much more reliability benefits during multi-day cold snaps that drive extended loss-of-load conditions in the Winter.

Figure 1: SPP DR ELCC Results with Seasonal Call Limits at 2GW Adoption

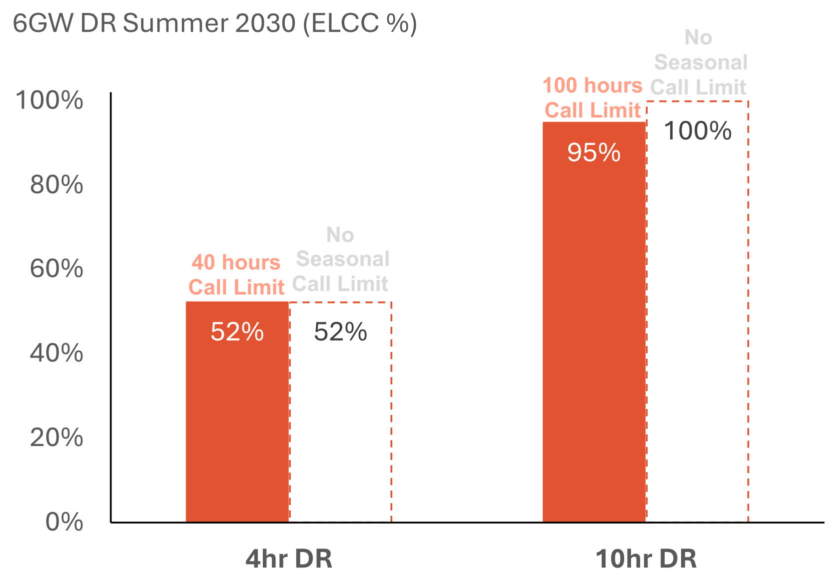


[5] SPP Summer season is June 1 – September 30, Winter season is December 1 – March 31. Shoulder months do not have resource adequacy requirements in SPP.

Call Limits

Under high penetrations of DR, the ELCC results highlight saturation effects, and the impact of increasing call hours under high DR adoption. In 2030 with ten-hour duration, at 6GW of penetration, increasing the call limit increased capacity accreditation by 5%.

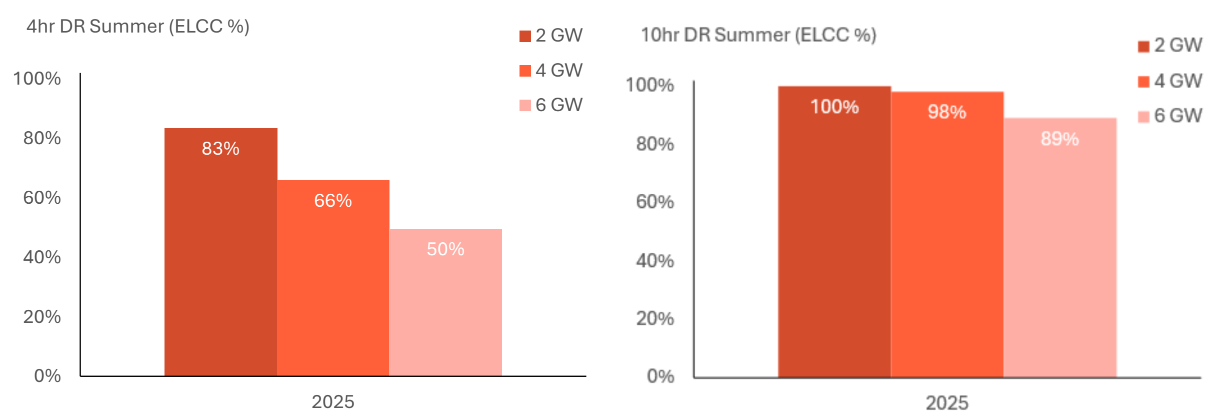
Figure 2: Impact of Seasonal Call Limits Under High DR Penetration



DR Saturation

With respect to the impact of saturation, four-hour DR saw stronger saturation effects than ten-hour. In 2025 four-hour ELCCs decline from 83% to 50% as penetration moves from 2GW-6GW. Ten-hour is more resilient to saturation effects and only declines by 11%.

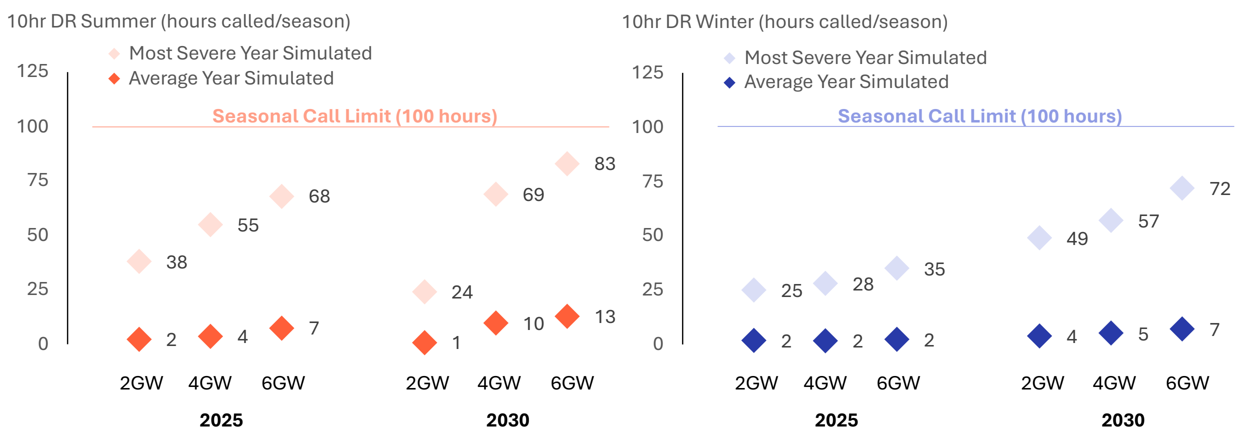
Figure 3: Saturation Effects of DR ELCCs



Critical Hours Required

To achieve these ELCC results, DR requires only limited hours of response. Across the weather years and forced outage simulations in this study, an ***average*** of 10 hours of DR is required across the year to achieve the ELCCs described above. Further, under the most severe Winter event observed in the LOLP model, 72 hours are needed across the season to achieve an 82% ELCC for ten-hour DR. During the hottest Summer observed in the LOLP model, 83 hours are needed across the season to achieve 100% ELCC again on ten-hour duration DR. Figure 4 below highlights the max and average number of hours required across both study years for each duration tested for ten-hour DR.

Figure 4: Total Hours DR is Called by Season (Ten-Hour DR with 100 Hour Seasonal Call Limit, Average vs. Most Severe Simulated Year)



Portfolio Composition

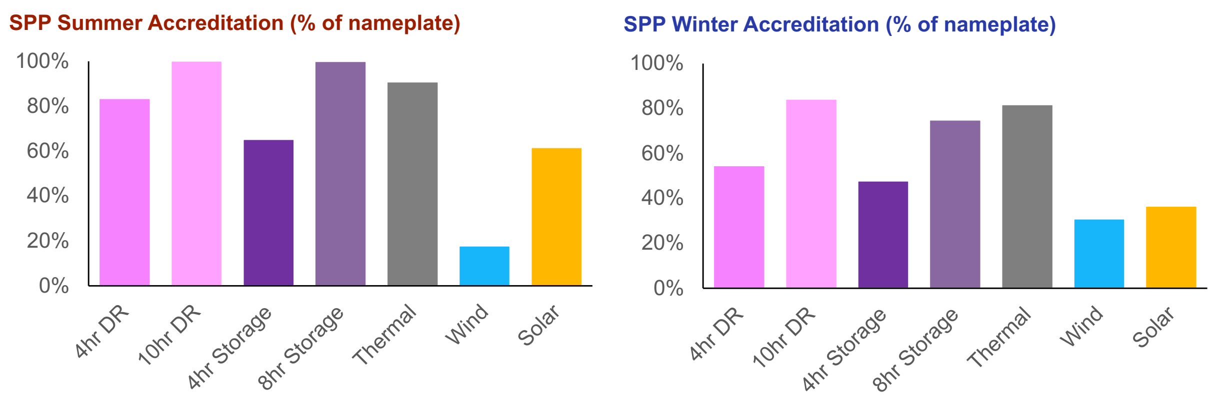
Similar to other ELRs like battery storage, the value of DR also depends on the portfolio mix, specifically if the presence of variable resources shortens or lengthens the critical hours. Moving from 2025-2030 our forecast of the SPP portfolio sees several changes, which is a key driver of DR ELCCs. From 2025-2030, E3's forecast includes 13GW of wind additions, 2GW of solar additions, 5GW of four-hour battery storage additions, and 4GW of gas additions. The forecast also includes 3GW of coal retirements. DR ELCCs decline from 2025-2030, indicating a net negative impact from the saturation of ELRs due to increased battery storage, which outweigh the diversity benefits of increased wind and solar.[6]

[6] For more information in diversity benefits and saturation effects see these again see these E3 papers: [Resource Adequacy for the Energy Transition: A Critical hours Reliability Framework and its Applications in Planning and Markets]() and [Capacity and Reliability Planning in the Era of Decarbonization.]()

DR Has Comparable ELCCs to Other Supply Resources

Both four-hour and ten-hour DR achieve capacity accreditation on par with traditional supply-side resources. DR shares similar dispatchable characteristics to storage and even gas peakers, giving operators one additional resource to balance load during critical hours. Figure 5 shows the achieved DR ELCCs measured in this study and the SPP capacity accreditation of multiple resource classes. When appropriately accredited, DR under these parameters can give system planners and operators another resource to meet near- and long-term resource adequacy needs. Given the current delays in development of supply-side resources, this makes DR a potentially faster solution for facilitating load growth and mitigating near-term resource adequacy risks.

Figure 5: Comparison of 2025 Capacity Accreditation Values by Resource*



* DR accreditations are E3's 2025 RECAP ELCC results (2GW adoption, 40hrs/yr for four-hour DR, 100hrs/yr for ten-hour DR). Storage, Wind, and Solar accreditations are Tier 1 ELCCs from SPP's 2024 ELCC study. Thermal accreditation is conventional fleet average ACAP from SPP's 2025 ACAP informational posting.

Conclusion

When appropriately accredited, DR can be just as valuable as any other resource in meeting resource adequacy needs. In the SPP, four-hour DR achieves a 55-83% ELCC and ten-hour DR achieves an 84-100% ELCC across seasons in 2025. This requires an average of ten hours of dispatch per year, and no more than 83 hours across the most severe Winter and Summer observed. Removing the seasonal call limit to DR does not increase its ELCC unless under very high DR adoption. In contrast, extending the duration capability of each call does increase ELCCs. As DR penetration increases, so does the length of critical hours, making shorter duration DR less effective. In the near-term, ten-hour DR ELCC stays above 80% even when testing the high adoption scenario of 6GW of DR. With appropriate accreditation and market design, SPP can encourage load flexibility and provide dependable capacity even with limits on how often the DR can be called. DR supports a balanced path for SPP: maintaining near-term reliability while long-term transmission and clean generation investments advance.

For DR to be fully integrated into SPP's capacity framework:

- Accreditation should be dependent on the program parameters, tied to availability and delivery during critical periods, aligning DR with storage and renewables.
- Market products must evolve with market needs. This means calculating an appropriate capacity accreditation for different classes or parameters of DR. For example, quantifying the number of well-timed events required to achieve high ELCC, rather than unlimited seasonal hours.
- DR can provide additional resources for LREs navigating rapid demand growth, enabling new loads before utility-scale resources can come online.

Introduction

SPP, like many other U.S. markets, is experiencing a period of rapid and unprecedented change. This change is driven by increases in electricity demand, further adoption of renewables and energy storage, ongoing thermal generation retirements, shifting weather patterns, and transmission system constraints. To navigate these challenges and uncertainties, SPP has launched a range of initiatives aimed at strengthening reliability, optimizing resource utilization, and expanding market access for both load and generation. Current initiatives include:

+ Large Load Integration
    - High-Impact Large Load (HILL) integration (RR696). New processes to interconnect and operate very large, fast-growing loads (e.g., data centers), [7] including the Conditional High-Impact Large Load (CHILL) pathway that enables faster interconnection for interruptible large loads.
+ SPP's Changes to RA
    - FERC-accepted Base PRM increases from 15% to 16% for Summer 2026, and a new Winter PRM of 36% for 2026/2027, followed by a planned 2029 increase to 17% (Summer) and 38% (Winter).[8] This corresponds to an ACAP PRM of 7% for Summer 2026 and 15% in Winter 2026/2027.
+ Performance-Based DR Accreditation
    - SPP proposed shifting its DR framework from the current enrollment-based accreditation system to a performance-based accreditation model.[9]
+ Western Expansion/Markets+
    - FERC-approved Markets+ tariff and funded Phase 2 implementation for SPP's Western day-ahead/real-time market; broader RTO expansion to increase reliability and efficiency of dispatch in SPP.[10]
+ Transmission System Planning and Investment
    - Historic $7.7B 2024 ITP portfolio approval including large bulk system investments, short-term reliability projects, and coordination born out of Holistic Integrated Tariff Team (HITT) recommendations.[11]

---

[7] High Impact Large Load (HILL) Integration - Southwest Power Pool
[8] SPP board approves new planning reserve margins to protect against high Winter, Summer use - Southwest Power Pool
[9] SPP Documents & Filings - Southwest Power Pool
[10] SUF Monthly Update
[11] 2024-itp-assessment-report-v10.pdf

With growing interest from data centers and other large loads in SPP, alongside the introduction of the CHILL/HILL process and evolving RA requirements, this paper examines the role DR can play as the market undergoes significant transformation.

Today, SPP uses a load modifier regime in its RA framework that includes about 1.6 GW of DR from interruptible/curtailable and non-dispatchable programs, typically supplied by utilities and cooperatives with existing DR/interruptible rate customers. SPP recently proposed a performance-based accreditation for DR, which means only DR resources that can meet reliability/response performance metrics will be counted toward RA. Further, the CHILL pathway is designed to bring in large loads (like data centers and industrial/advanced manufacturing) that are willing to accept potential load-shedding or curtailment during system stress to connect more quickly. This combination of performance-based DR in combination with CHILL providers could broaden the pool of DR providers beyond traditional interruptible rate customers to large commercial/industrial loads and possibly aggregators with strong response performance.

Building on these recent developments in DR in SPP, in this white paper, we assess the reliability contribution of DR in the SPP market. Using E3's SPP outlook as a baseline portfolio, we estimate the ELCC of DR to quantify its capacity accreditation. Our analysis evaluates multiple DR configurations, including variations in response duration and adoption levels. These different configurations illuminate what parameters are most important for strong DR contributions to RA.

The structure of this paper is as follows:

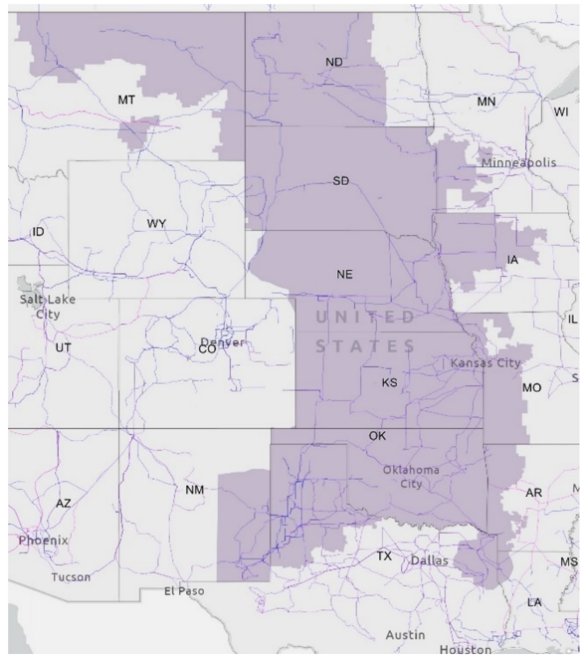
- **Section 1:** Background on the SPP market and the current state of DR
- **Section 2:** Capacity accreditation of DR, with ELCC results across DR configurations and scenarios
- **Section 3:** Comparison of DR ELCC and other resources
- **Section 4:** Conclusions

SPP Background

SPP is a regional transmission organization (RTO) that coordinates the reliable operation of the bulk electric system and wholesale electricity markets. SPP's footprint in the Eastern Interconnection spans 14 states across the central United States, from the Canadian border to Texas and Louisiana, serving over 17 million people within a 545,000 square mile area.

SPP's extensive high-voltage transmission network enables power flows across the region and with neighboring systems, supporting both reliability and market operations. SPP administers day-ahead and real-time energy markets, as well as transmission congestion management and ancillary services. SPP also is responsible for regional transmission planning, ensuring that investments in the grid support long-term reliability, economic efficiency, and integration of diverse energy resources.

Figure 6: SPP Footprint



Energy Transition in SPP

SPP has an all-time Summer peak demand of 56 GW occurring in 2023 and saw an annual 2024 energy consumption of approximately 278 TWh.

The region is distinguished by its industry-leading wind resources, with a current peak wind output of 24 GW from approximately 35 GW of installed capacity. Alongside wind, SPP relies on a diverse generation mix that includes 22 GW of coal, 36 GW of natural gas-fired generation, and smaller contributions from nuclear (2 GW), oil-fired (2 GW), and hydroelectric (3 GW) resources. While solar and battery storage currently make up a small portion of the SPP resource mix, their capacity is expected to grow significantly in the coming years.

In 2024, wind generation supplied 38% of SPP's total energy, making it the single largest contributor to the mix. Natural gas followed with 29%, while coal accounted for 24%. Nuclear and hydro combined provided 7%, with the remaining 2% coming from solar, oil-fired units, and other sources.[12]

---

[12] 2024_Seasonal_state_of_the_market_report.pdf

Figure 7: SPP 2024 Installed Capacity

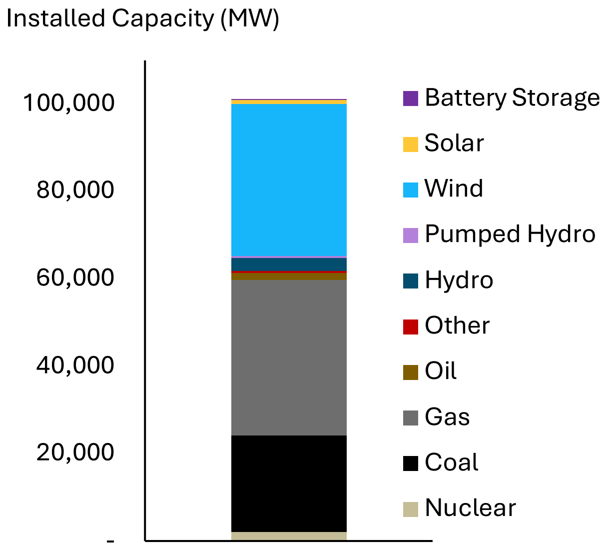
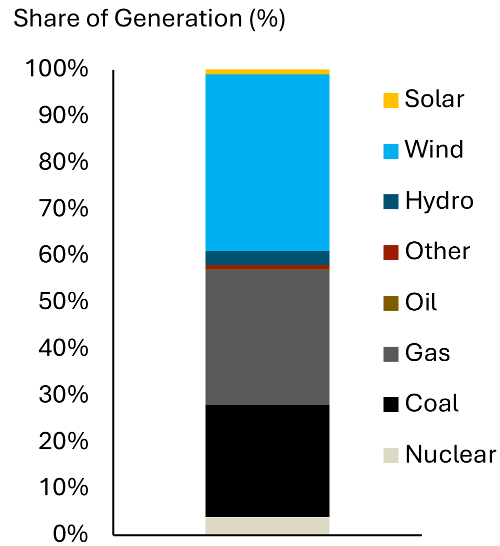
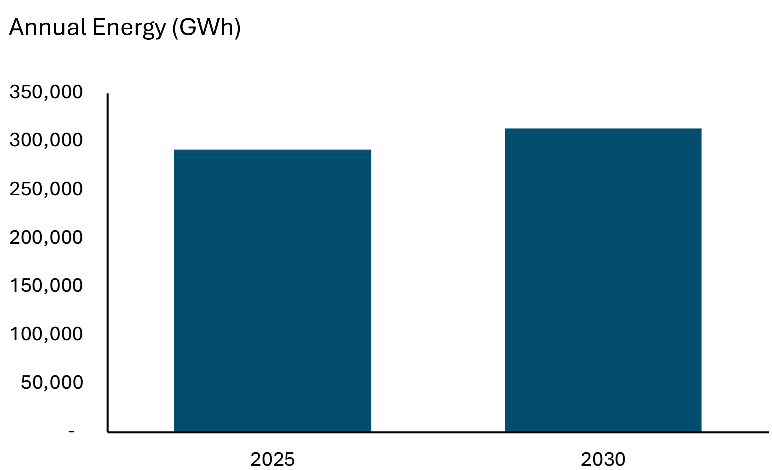


Figure 8: SPP 2024 Energy Production



To estimate the ELCC value of DR, we rely on our SPP market price forecast for its load and supply portfolio.[13] This outlook projects a 4% increase in both annual energy demand and peak load between 2025 and 2030, with peak demand reaching 61 GW by 2030.[14]

Figure 9: E3 Forecast of SPP Load Growth



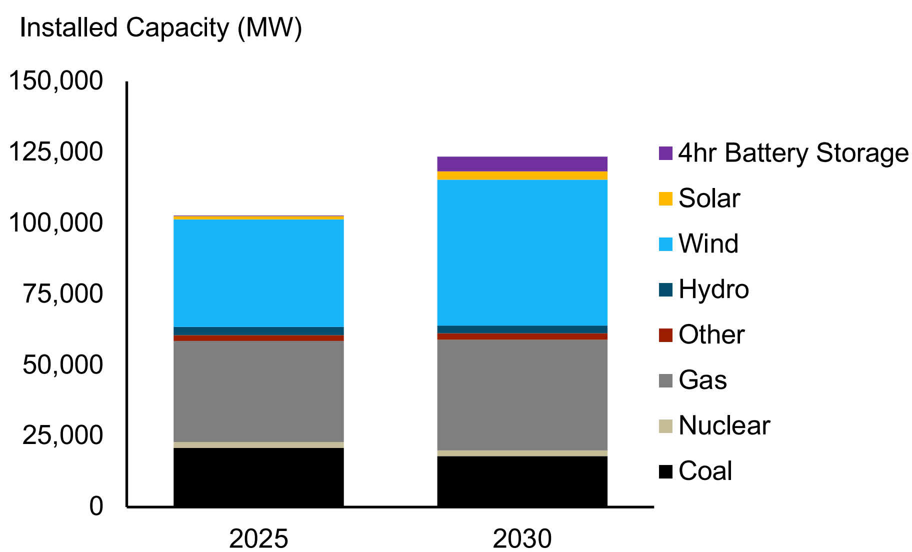
From 2025-2030, our outlook utilizes information from utility IRPs and SPP's interconnection queue to forecast near-term capacity additions, leveraging late-stage projects. We then combine the utility's long-term IRP results with our own long-term expansion model for generation 2030+ based

[13] More information on our SPP market forecast: SPP Price Forecast – 2025 Edition – Core Case | Energy + Environmental Economics

[14] E3's 2025 load outlook assumes near-term datacenter growth, stabilizing through 2030.

on our view of gas prices, load, state policy, PRM requirements, and resource costs to forecast the generation that will meet load. The figure below shows the capacity portfolio utilized in the simulations from our off-the-shelf view. The forecast sees increases in wind generation, along with gas additions to meet load growth and offset coal retirements. We also forecast energy storage adoption in 2030, with solar investment also contributing.[15]

Figure 10: E3 2025-2030 Forecast of SPP Resource Portfolio



SPP RA Market

Unlike the centralized capacity markets in PJM or ISO-NE, SPP's RA framework is a requirement within its open access transmission tariff. Each Load Responsible Entity (LRE) in SPP must demonstrate, via seasonal RA Workbooks submissions, that it holds enough accredited capacity to meet its forecasted Net Peak Demand plus an ACAP PRM requirement.

While SPP does not clear a centralized RA auction, LREs face deficiency charges if they cannot demonstrate compliance, creating an incentive to secure capacity. This makes SPP's RA system a bilateral, self-supply model where utilities and other entities balance their portfolios through a mix of owned generation, firm power purchases, and increasingly, qualifying demand response programs.

The SPP RA market has been undergoing several design changes, impacting both how LRE capacity obligations are determined and how resources are accredited. SPP's recently adopted ACAP PRM is 7% for the Summer Season and 15% for the Winter Season.[16] Currently, each LRE's "Net Peak Demand" is their seasonal Non-Coincident Peak (NCP) load reduced by qualifying DR programs, firm

[15] For more information on our resource cost forecasts and how this incorporates recent trends in tariffs and tax credits please see our RECOST model: [E3 Forecasts Higher Resource Costs Under 2025 Policy in Q3 RECOST Update - E3]

[16] Corresponds to Base PRMs of 16% in Summer and 36% in Winter: [ACAP PRM – 2025 Informational]

imports, and other load modifiers. This process treats DR as a load modifier rather than an accredited resource. The following section explains how supply-side resources are accredited in SPP, and how this framework can be applied to DR.

SPP Capacity Accreditation

SPP uses different methods to determine the accredited capacity of resources for its RA framework. For variable resources like wind and solar, and for battery energy storage, SPP determines accreditation through annual ELCC studies. For conventional resources like natural gas and coal-fired plants, SPP recently introduced a performance-based accreditation model. This methodology is based on a resource's demonstrated net generating capability and its historical performance, including its equivalent forced outage rate on demand (EFORd), which measures how often a resource is forced offline when needed.[17] Hydro resources are accredited according to firm capacity obligations and historical performance, while external imports receive accreditation only when they're backed by firm contracts.

SPP's annual ELCC studies determine seasonal capacity accreditation for wind, solar, and battery storage. The studies use probabilistic LOLP modeling to determine the resource-class ELCCs for the upcoming delivery seasons.[18] ELCC studies are repeated annually since results are dependent on SPP's resource portfolio mix and load. Wind and solar are accredited through a tiered framework: Tier 1 for resources with firm transmission service and Tier 2 for all others. The ELCC value of each resource Tier is allocated across the fleet of registered resources based on each registered resource's performance during SPP's highest 3% of net-load hours. Storage accreditation is differentiated by both duration and Tier. For example, eight-hour storage can approach 100% accreditation, while four-hour has lower values (~65% in Summer, <50% in Winter).

Recent discussions within the SPP working group, as documented in Supply Adequacy Working Group (SAWG) minutes and scope documents, show active efforts to refine how DR is considered. In the July 2025 session, SPP proposed a new DR framework to establish an accreditation structure that integrates DR into RA on the same footing as generation, moving away from simply deducting DR from load forecasts. The framework creates two primary categories: Market Registered DR (MRDR), which participates fully in the Integrated Marketplace and is deployed economically like supply resources, and Reliability Registered DR (RRDR), which is deployed only during conservative operations or energy emergencies to support grid reliability. Both MRDR and RRDR require registration, seasonal capability and operational testing, and performance-based accreditation using lookback periods and event-hour measurements.[19]

These DR efforts are also being considered in alignment with new large-load and HILL and CHILL policies. Specifically, CHILL offers large-load customers a faster interconnection path, with the trade-off of potential temporary curtailments, in exchange for expedited study results that allow

---

[17] FUEL ASSURANCE AND ACAP PRM OVERVIEW
[18] SPP 2025 ELCC Study Scope
[19] SPP Documents & Filings - SAWG Meeting Materials

them to integrate and operate as quickly as possible.[20] Interconnecting as a CHILL currently appears to have the CHILL counted like DR in RA planning, but it does not auto-enroll the load as a market DR resource.

Table 2: SPP Current Capacity Accreditation

Resource Type	Accreditation Method	2024-2025 Accreditation (% of nameplate capacity)[21]
Gas, Coal, Nuclear	ACAP - Installed capacity minus equivalent forced outage rates on demand (EFORd) & adjusted equivalent forced outage factor (EFOF)[22]	91% (Summer), 82% (Winter)
Hydro	Firm capacity obligations	100% of capacity obligation
Wind	ELCC by Tier, Tier 1 (with firm TX) & Tier 2 (without)	18% (Summer), 31% (Winter)
Solar	ELCC by Tier, Tier 1 (with firm TX) & Tier 2 (without)	62% (Summer), 37% (Winter)
Battery Storage	ELCC by duration (4hr, 6hr, 8hr) and by Tier	4hr: 65% (Summer), 48% (Winter) 6hr: 94%(Summer), 75% (Winter) 8hr: 100%(Summer), 75% (Winter)

Demand Response (DR)

What is DR?

DR encompasses a range of approaches that differ in how they reduce or reshape electricity demand during periods of system stress. DR is the most traditional form, where participating loads, such as industrial facilities, commercial buildings, or aggregated residential demand, are curtailed outright during an event, lowering system load in real time (curtailed DR). Shift DR instead moves consumption from critical periods to lower-risk periods, for example by pre-cooling buildings in the afternoon before a Summer peak or shifting EV charging to overnight hours. Other forms include shape DR, which permanently modifies load profiles through smart appliances or pricing signals. Together, these DR types provide different durations, magnitudes, and response speeds, allowing them to complement supply-side resources and enhance system reliability. **This study specifically looks at the capacity accreditation of curtailed *DR* programs in SPP.**

DR in SPP

DR in SPP is currently concentrated among large industrial and commercial users. These participants often operate under interruptible tariffs or direct load control agreements with their LREs. The most common participants include industrial manufacturing facilities, pulp and paper mills, chemical processors, agricultural pumping loads, and municipal water treatment systems.

[20] [High Impact Large Load (HILL) Integration]

[21] Accreditation values shown for aggregate SPP resource class, prior to allocation of accreditation to individual SPP-registered resources.

[22] SPP thermal resources are accredited with net generating capability under Base PRM accounting. Thermal resources receive ACAP values under ACAP PRM accounting, as part of Performance Based Accreditation revisions. For ELCC resources, the same values are used under both Base PRM and ACAP PRM accounting.

These large-scale participants can reduce load anywhere from tens to hundreds of megawatts with relatively short notice, providing significant support to the grid during critical hours. Programs are often structured for four-hour durations, which aligns with SPP's resource adequacy qualification requirements. However, many of these programs can accommodate longer durations when needed during extended peak demand periods or emergency conditions. As per SPP's 2025 Summer Resource Adequacy Report, LREs collectively reported ~1.6 GW of controllable and dispatchable DR, with a projected ~57% increase over the next five years.[23]

Table 3: Types of DR in SPP

DR Type	Accreditation Method	Utilization	Status[24]
Peak Demand DR Programs	Reduction in Load-Responsible Entity (LRE) Peak Demand Forecast	Typically called during EEA Events	Active
Market Registered DR Resource	ELCC + Energy Market Offer Curves	Economic dispatch into energy market, called before reliability registration	In development
Reliability Registered DR Resource *Focus of this paper*	ELCC	Called during SPP critical periods	In development

[23] SPP 2025 Summer Season RA Report
[24] Status as of September 2025

Capacity Accreditation of Demand Response

This section examines the capacity accreditation of DR in SPP using E3's resource adequacy modeling tool, RECAP. This study uses an ELCC methodology and aligns with SPP's definition of ELCC as the metric to quantify DR capacity accreditation. ELCC is the amount of perfectly reliable capacity (a "perfect capacity resource") that can be replaced by a given resource while maintaining the same system reliability standard. SPP currently uses ELCC to accredit variable (wind and solar) and battery storage.

We begin this section by describing the characteristics of the DR examined in this paper. Next, we characterize the distinct nature of loss-of-load events in SPP, which occur under different conditions in Summer versus Winter. We then present results from our ELCC analysis across a range of DR configurations and adoption levels. Finally, we summarize the findings and highlight the role of DR in supporting reliability during a period of rapid system change.

For this study, we modeled the SPP footprint as a single-zone system over 15 Monte-Carlo draws & 66 historical weather years (1954-2019). We modeled SPP's existing (2025) and 2030 system using assumptions from our 2025e SPP Market Price Forecast.

What Determines ELCC?

Any resource's capacity accreditation, and thus ELCC, is determined by the alignment between SPP's timing of critical hours and resource performance. To evaluate the reliability contribution of DR, we test multiple configurations that capture its operational limits and system interactions: duration, call limits, and penetration level.

SPP Critical Hours

While Summer critical hours have long defined reliability planning, recent history shows that extreme Winter weather events can be just as, if not more, disruptive. In February 2021, during Winter Storm Uri, SPP declared an Energy Emergency Alert 3 (EEA 3) and initiated controlled outages for the first time in its history as gas supply froze, wind output sagged, and demand surged. More recently, in December 2022 and January 2024, severe cold snaps again strained the system, with SPP relying heavily on imports and favorable wind conditions to avoid broader outages. These events underscore that SPP's reliability risks are not confined to peak Summer afternoons but span both seasons.

Generally, loss-of-load events in SPP emerge from a combination of severe weather, leading to higher than normal load conditions, and unexpected resource unavailability, leading to a lack of energy. These conditions are SPP's **critical hours**. Critical hours can happen any time of year but are typically categorized into seasonal critical hours. The heatmaps below show the critical periods in SPP as modeled in RECAP. These heatmaps represent the critical hours once SPP meets a 1-day-in-10-year loss of load expectation (or 0.1 LOLE) standard.

During the Summer, critical hours are concentrated in July and August and are most pronounced between hours 12-17. These periods coincide with higher cooling demand, lower than expected thermal availability, and lower than expected wind generation, resulting in a 4 to 5 hour window of risk. In 2030, the Summer loss-of-load hours widen and move later into the evening with higher adoption of solar. This creates an additional 3 to 4 hour window during and after sunset, an ideal scenario for energy limited resources like storage or demand response to be effective.

In the Winter, loss of load risk typically coincides with the most severe Winter days. Compared to Summer, the peak demand during the coldest days is typically much higher and longer, leading to an extended need for heating demand. Loss of load risk is therefore prevalent from the morning ramp-up until late in the evening. Cold snaps, wind droughts, and thermal forced outages pose a strong reliability risk for the entire Winter day.

Figure 11: SPP Summer Loss of Load Patterns

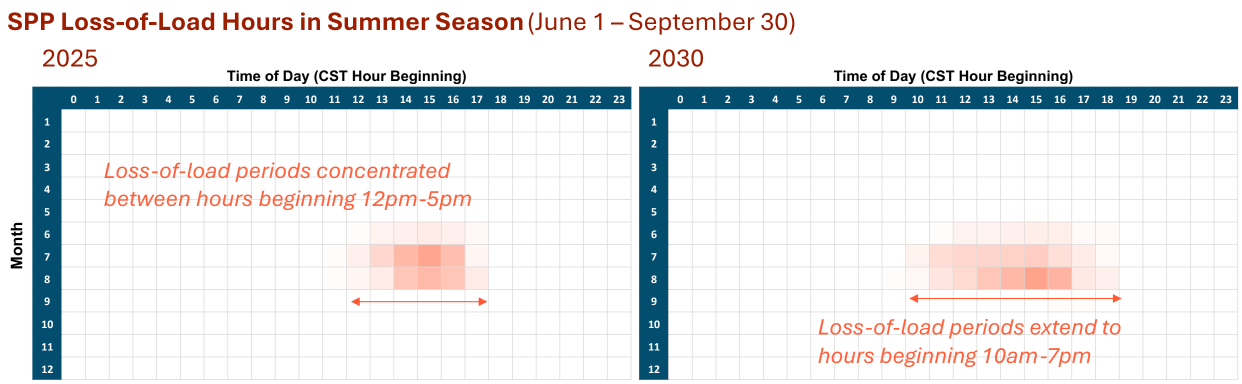
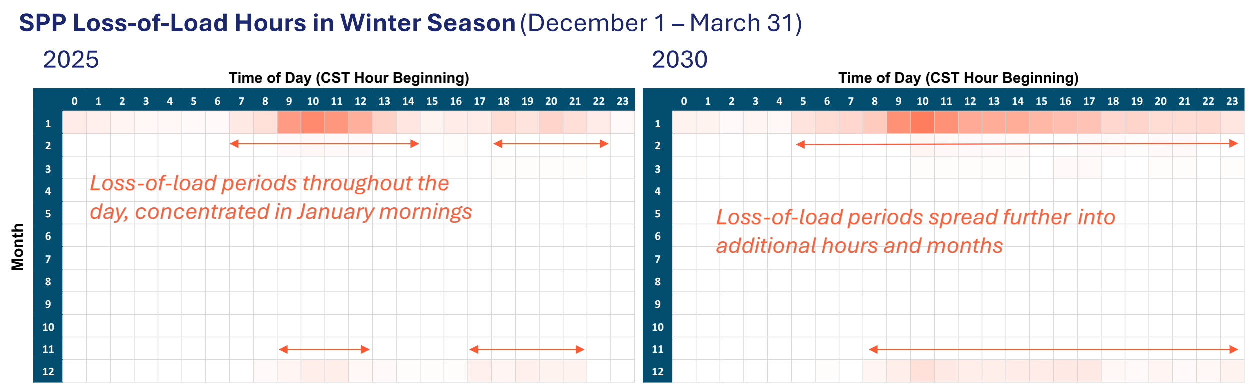


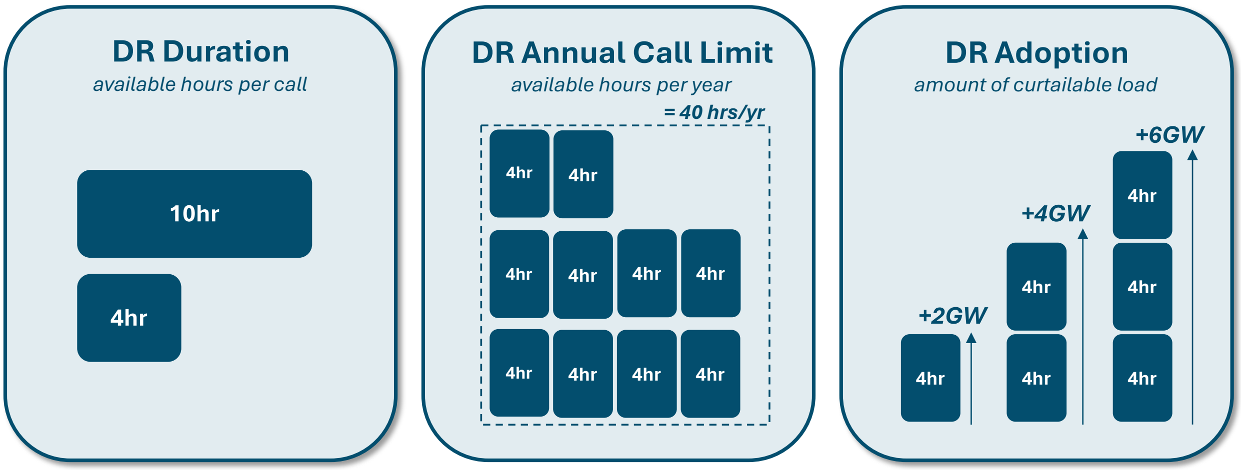
Figure 12: SPP Winter Loss of Load Patterns



DR Program Constraints and Availability During Critical Risk Periods

The capacity accreditation of DR is determined by three key parameters: DR duration, DR seasonal call limit, and DR adoption. DR duration is the maximum number of hours DR is available per call. DR calls are restricted to once per day. The DR seasonal call limit is the total number of hours per season that DR can be dispatched. Finally, DR adoption is the magnitude (measured in MW) of load that will respond to the DR call for its full duration. DR adoption may therefore be lower than the average or maximum magnitude of load associated with the DR providers. Only the fraction of load that is registered to respond in each SPP Season is counted. For example, a single SPP customer may have 500MW of average load but only register 200MW as DR in Summer and 100MW as DR in Winter. This study quantifies the ELCC of DR in SPP in 2025 and 2030 across a range of expected and bookend values for these three parameters.

Figure 13: Demand Response (DR) Program Parameters Explored in this Study



First, event duration is directionally linked to effectiveness: longer-duration DR can sustain support through extended loss-of-load periods, yielding higher ELCC values than shorter-duration programs.

Second, the number of allowable calls per season can matter: if a DR program commits to ten events but an eleventh is needed, DR provides no additional load reduction for the last event, reducing its effective contribution. Finally, like all energy-limited resources, DR experiences saturation effects as penetration increases—the incremental ELCC of each additional MW declines. Understanding these dynamics is essential for accurately quantifying DR's role in reducing loss-of-load risk and for designing programs that maximize its capacity accreditation. The following table outlines the suite of parameters tested:

Table 4: Range DR Operating Parameters & Adoption Levels Evaluated

DR Duration *Available Hours per day*	**DR Seasonal Call Limit** *Available Hours per season*	**DR Adoption Level** *Available GW*
4 hours	40 - 120	2 - 6 GW
10 hours	100 - 300	2 - 6 GW

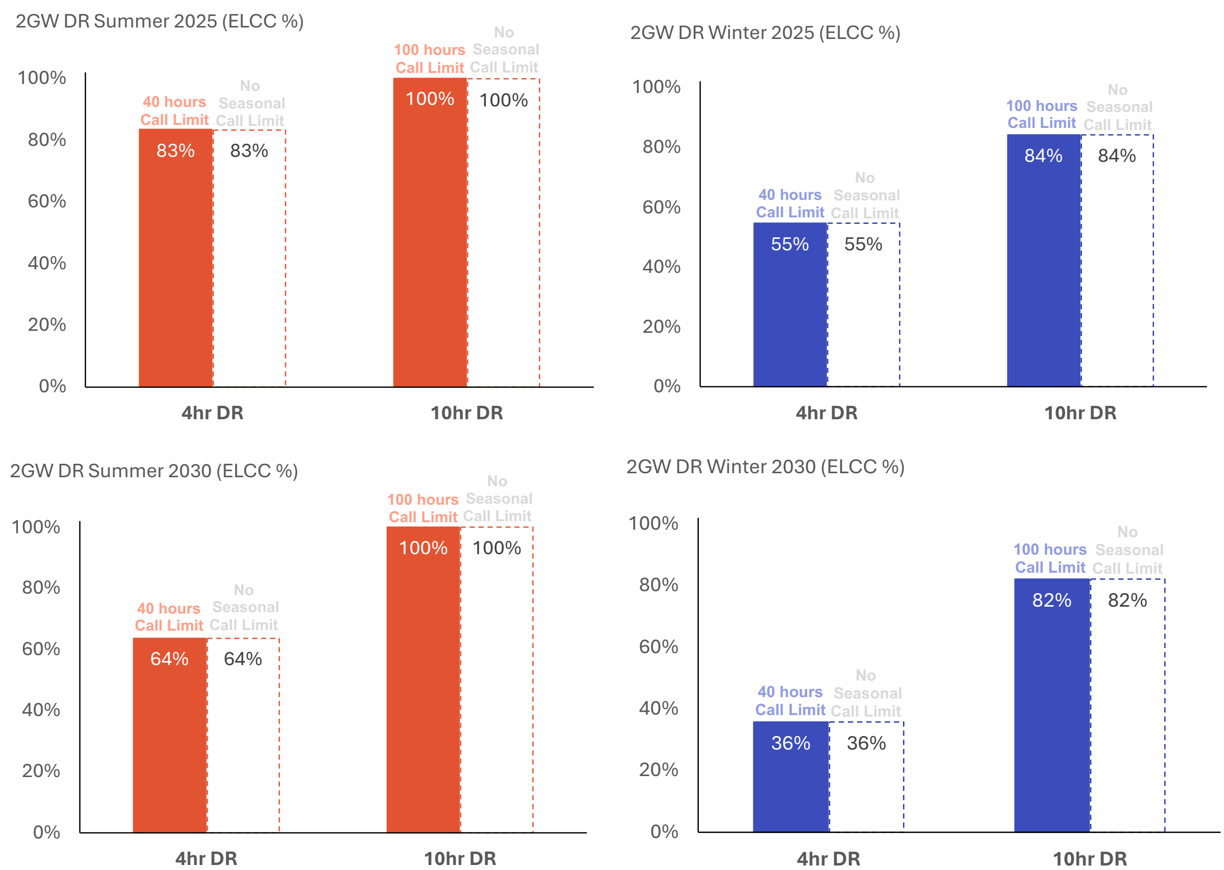
Demand Response ELCCs

In this section, we demonstrate the results of our SPP LOLP modeling for DR. Specifically, we measure the ELCC of DR across different scenarios and configurations.

DR ELCCs

In 2025, four-hour DR with a 40 hours per season call limit achieves an 83% ELCC in the Summer and a 55% ELCC in the Winter. Ten-hour DR with a 100 hours per season call limit achieves a 100% ELCC in the Summer and 82% ELCC in the Winter. DR ELCCs change between 2025 and 2030 due to changes in SPP critical hours which are in turn due to changes in SPP's resource portfolio and load. SPP critical hours, resource portfolio forecasts, and load forecasts are in above sections. In 2030, four-hour DR with a 40 hours per season call limit achieves an 64% ELCC in the Summer and a 36% ELCC in the Winter. Ten-hour DR with a 100 hours per season call limit achieves a 100% ELCC in the Summer and 82% ELCC in the Winter. Under 2 GW of DR adoption, removing the seasonal call limit had no impact on the ELCC results in 2025 and 2030. DR duration is the most impactful parameter on DR ELCCs in both seasons and years, with ten-hour DR ELCCs receiving 27% to 28% higher ELCCs than four-hour DR at 2GW of DR adoption.

Figure 14: DR ELCCs by Duration and Seasonal Call Limit, 2GW of Adoption

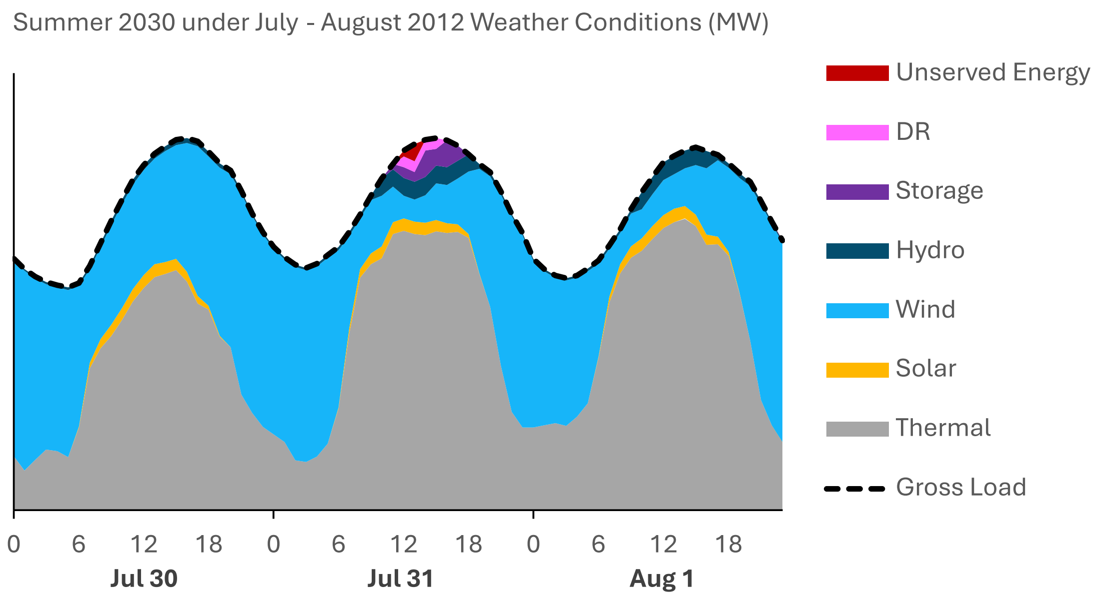


The importance of seasonal call limits was evaluated by calculating DR ELCCs with and without the limit imposed. Figure 14: DR ELCCs by Duration and Seasonal Call Limit, 2GW of Adoption

above shows no difference in ELCC when the total calls per season are unlimited at the 2GW adoption level. This is primarily due to (1) the low frequency of critical hour periods occurring in any given year and (2) DR being limited to 1 call per day. Ten-hour DR receives higher ELCCs than four-hour DR in both seasons, revealing the importance of maximum duration capabilities in a single day. The seasonal call limit did not bind in any of the scenarios with 2GW of DR adoption, resulting in DR with no seasonal call limits having the exact same ELCC as DR with the seasonal call limits applied.

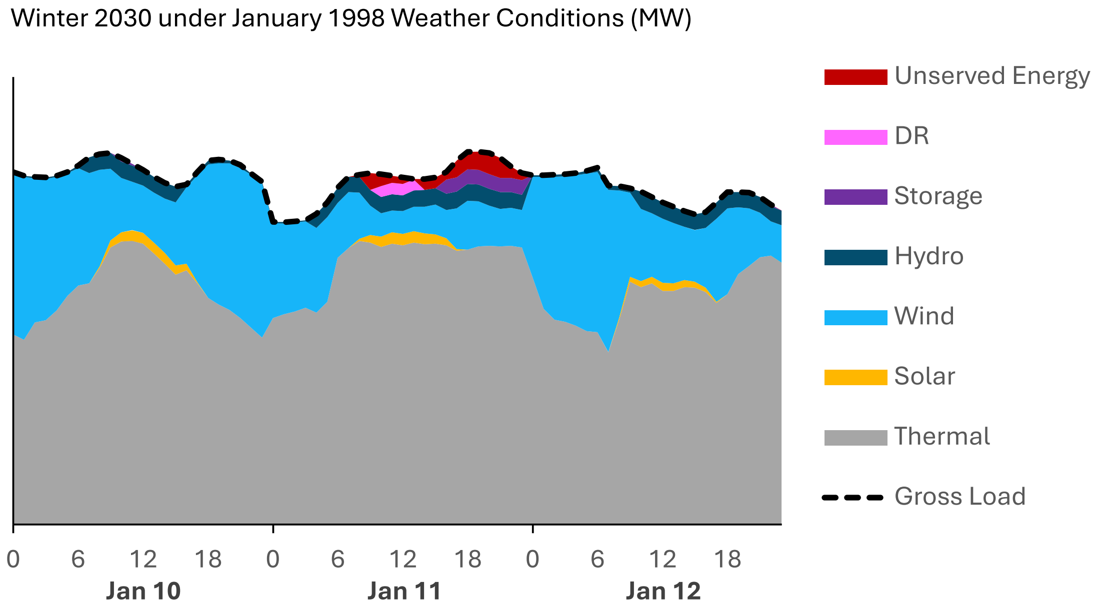
The dispatch plots below demonstrate examples of critical hours in Summer and Winter of 2030. The example Summer event shown below features two hours with unserved energy (in red) starting at midday when SPP load is high. Four-hour DR (pink) performs at full capacity for four hours, including the period with unserved energy. A four-hour battery storage (purple) discharges at less than full capacity over seven hours. The Summer of 2030 exhibits critical hour periods which occur mostly but not entirely within four consecutive hours, resulting in four-hour DR ELCC of 83% at 2GW of adoption.

Figure 15: Example of Summer Loss of Load Conditions in RECAP with Four-hour DR



In the Winter, critical hours periods are longer resulting in lower DR ELCCs. The example Winter event shown below features ten-hour with unserved energy (in red) starting at 9am when SPP load is increasing and wind generation is low. Four-hour DR (pink) perform at full capacity for four hours, but the duration limit prevents it from performing over the entire period with unserved energy. Four-hour battery storage (purple) is similarly unable to perform at full capacity over the ten hours with unserved energy. The Winter of 2030 exhibits a large fraction of critical hour periods which are significantly longer than four consecutive hours, such as the event in this example, which results in four-hour DR ELCC of 36%.

Figure 16: Example of Winter Loss of Load Conditions in RECAP with Four-hour DR

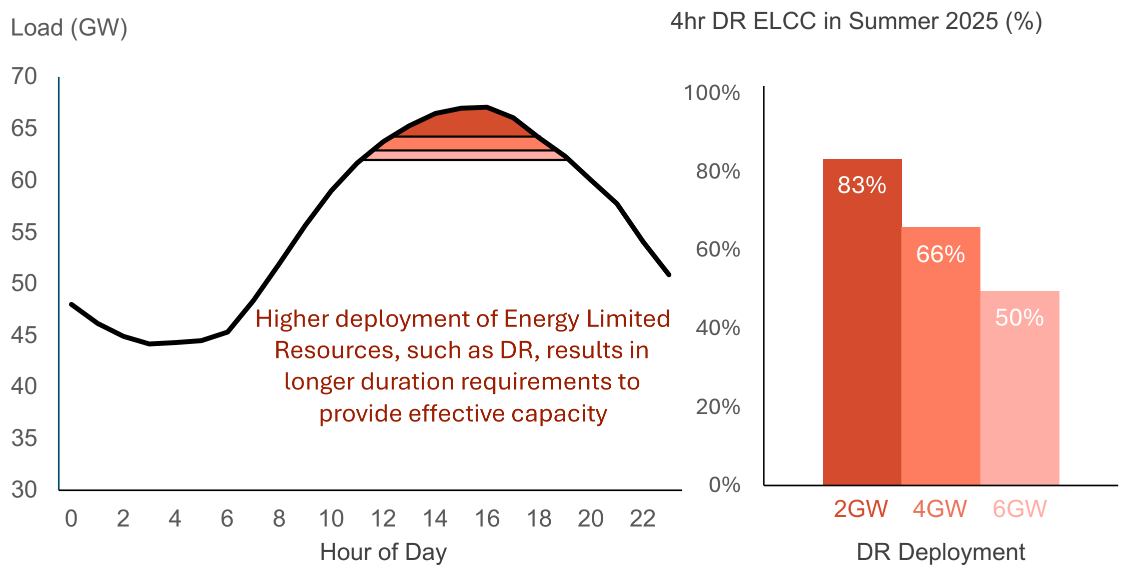


DR ELCCs Decline with Incremental DR Penetration

ELCCs of certain resources experience saturation effects where ELCCs decline as the resource type becomes a larger fraction of the system's capacity supply mix. Figure 17 illustrates the declining ELCC phenomenon with increasing penetrations of energy limited resources (storage and DR). Increased penetrations of energy limited resources result in longer periods of critical hours, this in turn reduces the value of shorter-duration energy limited resources. The saturation of energy systems with energy limited resources is a consistent finding across multiple resource adequacy studies.[25, 26]

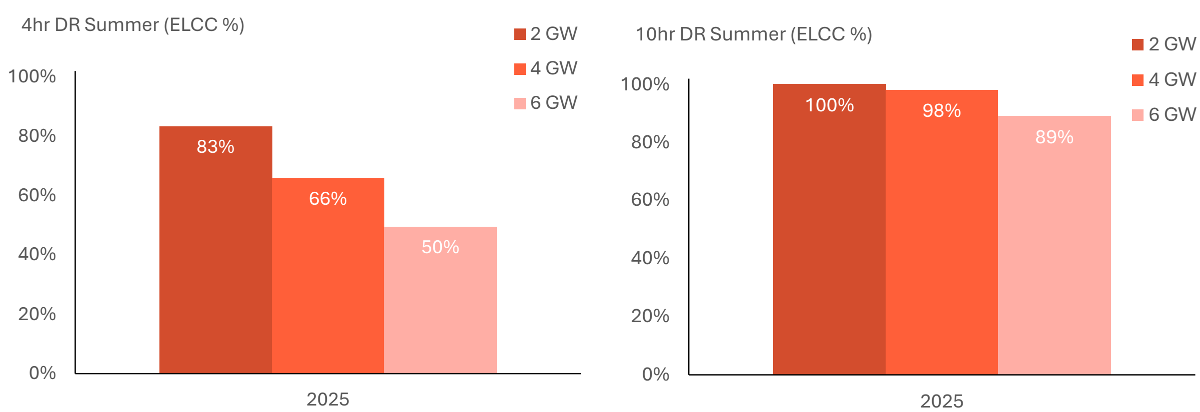
[25] Practical Application of Effective Load Carrying Capability in Resource Adequacy | E3
[26] A Critical Periods Reliability Framework and its Applications in Planning and Markets | E3

Figure 17: DR in SPP Exhibiting Saturation Effect of Energy-Limited Resources



To examine DR's sensitivity to higher penetrations, we tested additional scenarios with 4GW and 6GW adoption of DR in addition to the 2GW scenario. In 2025, four-hour DR starts at 83% ELCC but sees its value decline to 50% with 4GW incremental amounts of four-hour DR. For ten-hour DR, the saturation is less pronounced, but still present.

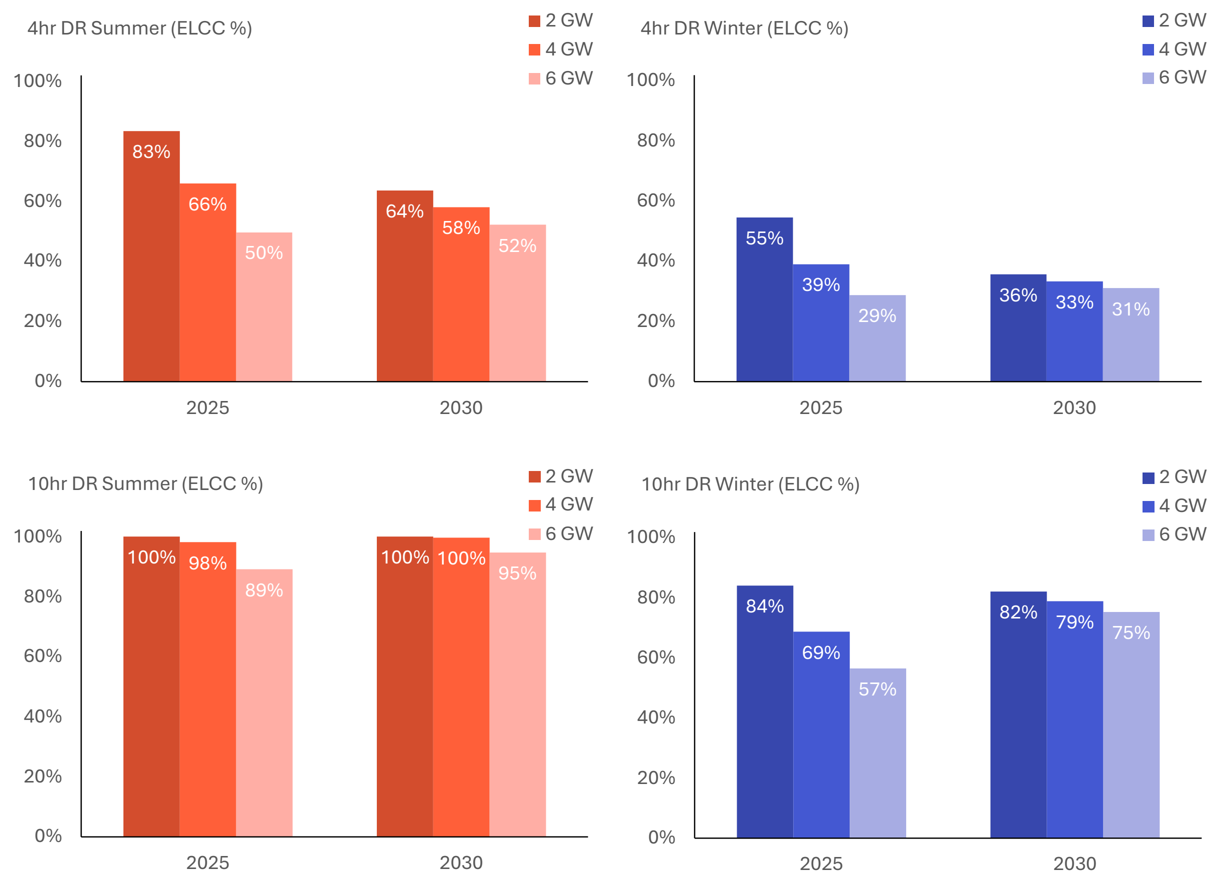
Figure 18: Four-hour and Ten-hour DR ELCCs by Adoption Level, Summer 2025



DR ELCCs in 2030: Effects on Capacity Accreditation as SPP Portfolio Mix Changes

By 2030, SPP's portfolio shifts significantly, with most incremental capacity additions coming from thermal, wind, solar, and storage resources. From 2025-2030, SPP portfolio undergoes several changes including 13GW of additional wind, 2GW of additional solar, and 5GW of additional four-hour battery storage. The rapid expansion of storage affects DR ELCCs by contributing to the saturation of energy limited resources on the SPP system, while rising wind and solar adoption provide diversity benefits that help support DR ELCCs over time. Diversity benefits between energy limited resources (DR and storage) with variable resources (solar and wind) is a phenomenon consistently observed across resource adequacy studies.[27,28] The net impact of changes in the SPP system from 2025-2030 is a decrease in four-hour and ten-hour DR Summer ELCCs, whereas the impact on Winter ELCCs is more complex and described further in this section.

Figure 19: DR ELCCs by Year and Market Adoption Level



Compared to 2025, four-hour DR ELCC declines by 2030 in both SPP's Summer and Winter seasons at 2GW and 4GW of adoption. In 2030, DR ELCCs benefit from the diversity benefit of additional variable resources (solar and wind) coming online. However, the 5GW of additional energy limited resources (four-hour battery storage) by 2030 contributes to the saturation of energy limited resources in SPP.

Notably, four-hour DR ELCCs at 6GW adoption is stable between 2025 and 2030. This indicates the diminishment of the saturation effect, which occurs when the duration of reliability events stabilizes

[27] Practical Application of Effective Load Carrying Capability in Resource Adequacy | E3
[28] [A Critical Periods Reliability Framework and its Applications in Planning and Markets | E3]

despite additional increments of energy limited resources. As seen in these results, the saturation stabilization effect typically happens at relatively low ELCC levels.

In contrast to four-hour DR, ten-hour DR sees a slight increase in ELCC between 2025 and 2030. Unlike four-hour DR which is on the same saturation curve as four-hour battery storage, ten-hour DR is less impacted by the four-hour battery storage resource since it has additional performance capabilities. Under optimal dispatch, the ten-hour and four-hour resource can be dispatched strategically to minimize or prevent loss of load in a manner which is not possible when only a single duration of energy limited resource is available to the system. The slight increase of ELCCs for ten-hour DR indicates that the diversity benefit from more renewables in 2030 outweighs the saturation effect of additional energy limited resources on the system.

However, if longer duration battery storage were to enter the system, ten-hour DR would see ELCCs lower than presented in this study. Since resources receive new ELCC accreditation values annually, based on system conditions in the forthcoming delivery period, the sensitivity of ELCCs to market portfolio changes is a key consideration. These results highlight how energy limited resources interact with one another, and how longer duration resources can benefit from ELCCs which decline more slowly over time.

Other Market Impacts on DR Operations and ELCC

While not comprehensively explored in this study, changes in the resource mix, especially renewables and storage resources, also affect DR's reliability contributions. ELCCs are shaped by saturation and diversity effects. Solar saturation occurs when additional solar provides diminishing incremental capacity value, since critical evening hours increasingly fall after sunset. Storage saturation similarly arises when large quantities of short-duration storage flatten net load curves, requiring subsequent tranches to cover longer and less frequent events. However, when ELRs like storage are paired with variable renewables, they generate meaningful diversity benefits: storage can shift surplus solar into evening peaks, while solar reduces the depth and duration of storage dispatch needed.

Table 5: Impact of Market Conditions on DR Capacity Value

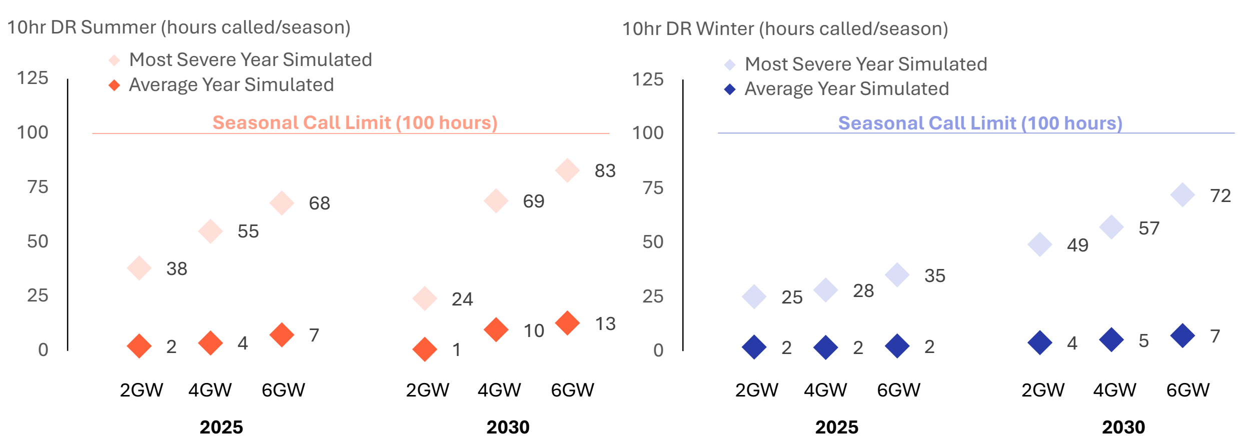
Scenario	Impact on DR ELCCs	Impact on DR Calls
Accelerated renewable adoption	Increase *Due to diversity benefit*	Increase/Minimal Change *Depends on portfolio sufficiency*
Accelerated DR adoption	Decrease *Due to ELR saturation*	Increase/Minimal Change *Depends on portfolio sufficiency*
Accelerated storage adoption	Decrease *Due to ELR saturation*	Minimal Change *Depends on portfolio sufficiency*
Accelerated load growth	Uncertain *Depends on load types*	Increase/Minimal Change *Depends on portfolio sufficiency*
Other resource mix changes	Uncertain *Depends on portfolio composition*	Uncertain *Depends on portfolio composition*

DR Duration is more important than its Seasonal Call Limit

Across the DR durations (four-hour, ten-hour), adoption levels (2GW, 4GW, 6GW) and test years (2025, 2030) examined for SPP, removing the seasonal call limit only impacts DR ELCCs in the Summer 2030 scenario with 6GW of ten-hour DR, this result is discussed further below. For all other scenarios, DR calls are not reaching the seasonal call limit, hence removing the seasonal call limit does not change DR ELCCs. In addition to calculating DR ELCCs for each scenario, we analyzed total hours DR is called by season. Since RECAP simulates each scenario over 960 simulated years, the average and highest number of hours DR is called by season could be extracted. The figure below shows the average (average year simulated) and highest (most severe year simulated) number of hours ten-hour DR is called per season, when seasonal call limits are imposed.

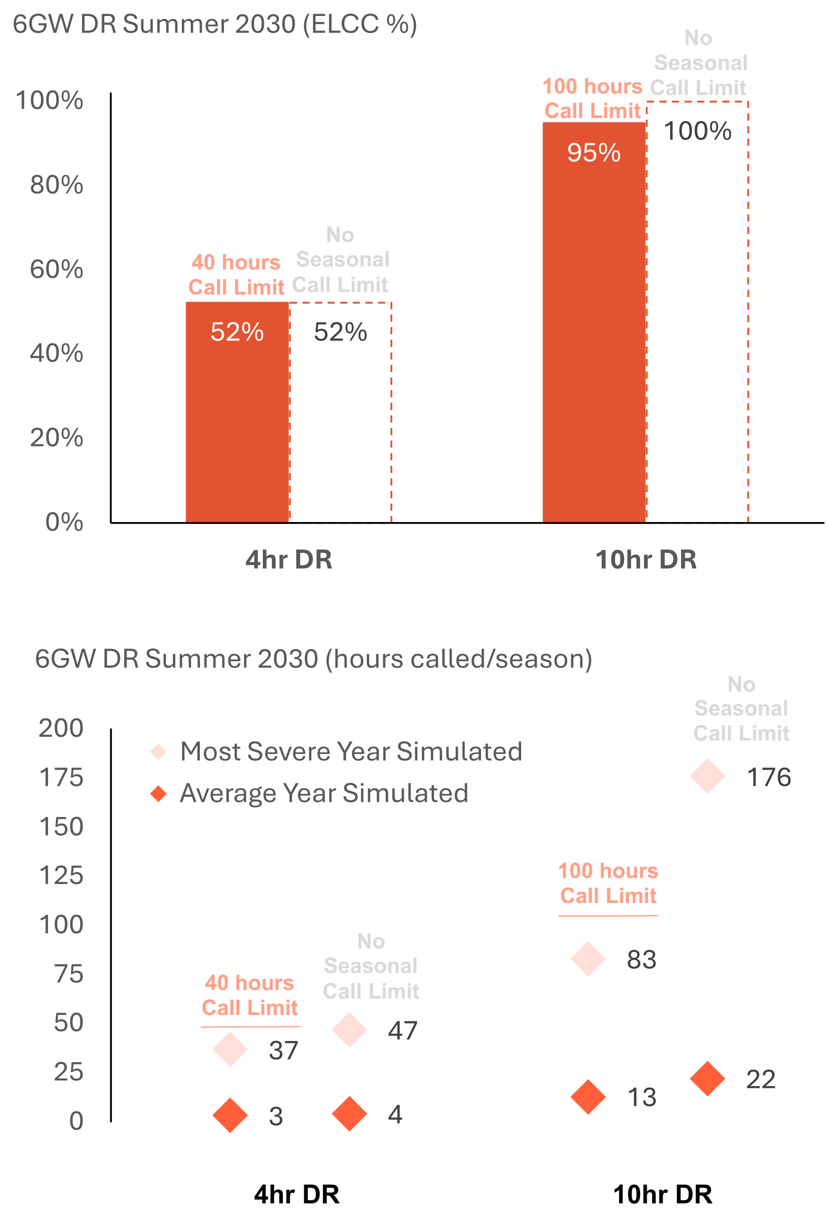
In Summer 2025-2030, ten-hour DR is called for 1 to 13 hours over the entire season of the average simulated weather year. In Winter 2025-2030, ten-hour DR is called for 2 to 7 hours over the entire season of the average simulated year. Over the entire year, ten-hour DR is called for 3 to 20 hours, significantly less than the seasonal call limit of 100 hours. Over the 960 simulations of each scenario, total DR calls reach 83 hours in the most severe Summer simulated and 73 hours in the most severe Winter simulated, both at 6GW of ten-hour DR adoption in 2030. In the average year, DR is called far less than the seasonal call limit, however, achieving the DR ELCCs presented in the above sections requires the capability of DR to be utilized up to its specified seasonal call limit.

Figure 20: Ten-hour DR Utilization by Year and Adoption Level



The seasonal call limit does impact DR ELCCs in SPP under certain system conditions, involving the high DR penetration levels. The Summer of 2030 with 6GW of ten-hour DR is the only scenario tested where removing the seasonal call limit increases ELCCs. In 2030, ten-hour DR receives a 100% ELCC when it does not have a seasonal call limit and is used for 176 hours in the most severe weather year. In contrast, ten-hour DR with the seasonal call limit receives 95% ELCC.

Figure 21: DR ELCCs in Summer 2030 by Duration and Call Limit, 6GW Adoption



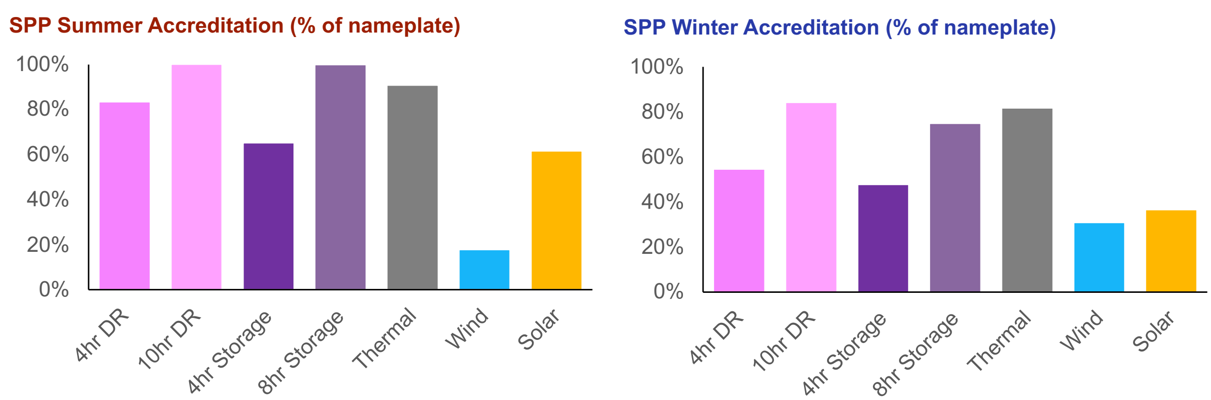
Meeting SPP's Near- and Long-Term Reliability Needs

This section seeks to explore the viability of DR as a capacity resource within the SPP market. We compare DR ELCCs to other resources in SPP to understand how DR could add to the resource adequacy market in SPP. Further, we explore the impact DR could make in the short term if load growth outpaces supply additions in the SPP footprint due to procurement and interconnection bottlenecks for supply-side resources.

DR vs. Accredited Supply Side Resources

To understand DR's potential role in SPP's resource mix, we take the ELCC results from the previous section and compare them to accreditation values of supply-side resources from SPP's most recent studies.[29] The figure below shows that for today's system, four-hour DR has ELCC values comparable to four-hour battery storage ELCCs in both Summer and Winter. Ten-hour DR has ELCC values comparable to 8hr battery storage ELCCs and thermal ACAP values in both Summer and Winter. The DR ELCCs shown are associated with the lower seasonal call limit (40 hours/season for four-hour DR & 100 hours/season for ten-hour DR) at 2GW of market adoption. As discussed in the Results section, the duration (measured as maximum daily availability) is the key performance characteristic determining ELCCs for energy limited resources (both DR and battery storage) in SPP.

Figure 22: SPP Summer and Winter Capacity Accreditation by Resource



While SPP has not performed ELCC studies out to 2030 yet, based on the results for DR, we anticipate that wind, solar, and battery storage will also face saturation headwinds, as we anticipate that wind, solar, and storage provide the most growth in the near-term based on late-stage

---

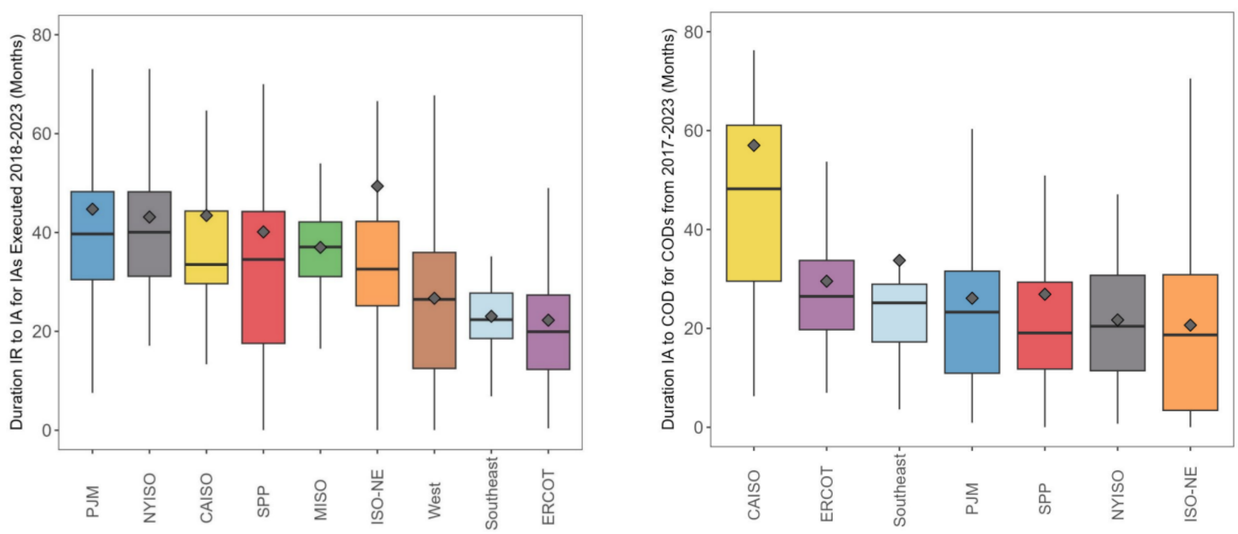
[29] DR accreditations are E3's 2025 RECAP ELCC results (2GW adoption, 40hrs/yr for 4hr DR, 100hrs/yr for ten-hour DR). Storage, Wind, and Solar accreditations are Tier 1 ELCCs from SPP's 2024 ELCC study. Thermal accreditation is conventional fleet average ACAP from SPP's 2025 ACAP informational posting.
The comparisons of accreditation across multiple data sources are not exactly "apples to apples" due to differences in study timing, given Seasonal changes in SPP resource mix, loads, and accreditation methodologies. SPP's ELCC and ACAP values are published Seasonal, with the 2025 ELCC study expected in late 2025.

interconnection projects. As such, we anticipate that the 2025 relative ranking to hold similar for 2030.

RA Risks and Opportunities in SPP

One potential benefit to the integration of DR from large loads looking to enter SPP is the speed at which DR can deploy. As Lawrence Berkeley National Laboratory demonstrates in their 2024 *Queued Up* report, SPP is facing an average of 40 months from the interconnection request coming in to signing interconnection agreements. SPP then sees an average of 30 months from interconnection agreement signature to interconnection commercial operations. That is a total of ~70 months or almost six years. This also does not consider pre-interconnection application development activities.

Figure 23: Total Interconnection Timelines for SPP Generation



Source: [LBNL Queued Up: 2024 Edition - Energy Markets & Policy]

Similarly, industry is facing delays on items like gas turbines. Recently, large developers like NextEra have announced on earnings calls that they are facing challenges getting new turbines installed by 2030.[30] This problem is systemic across industry. As developers have increased orders in response to demand growth, alongside demand from other non-power sector turbine customers, turbine manufacturers have reached production capacity.[31, 32] Even if interconnection speeds are accelerated for reliability purposes,[33] supply chain constraints will be binding unless the project is already under construction or has its major equipment secured.

As SPP sees increased near-term load growth and as the HILL/CHILL processes develop, SPP LREs may find themselves in the scenario of needing new forms of capacity resources to manage growing

[30] Utility Dive - NextEra partners with GE Vernova to build 'gigawatts' of gas generation
[31] Gas turbine manufacturers expand capacity, but order backlog could prove stubborn | Utility Dive
[32] Texas' $7.2 billion loan program for gas power plants has approved two projects in two years | News From The States
[33] PJM fast-tracks 11.8 GW, mainly gas, to bolster power supplies | Utility Dive

RA needs amidst supply-side constraints. DR from new large loads can achieve a high capacity accreditation and can be provided as soon as the load comes online.

Given the challenges SPP may face in a high load growth and constrained supply environment, DR should be considered a RA resource and used to provide valuable load reduction when the system needs it. Establishing clear and appropriate DR performance requirements, as explored in the rest of this paper, can facilitate fair and efficient adoption of DR in the SPP market. This will help SPP LREs manage difficult supply-demand balances as the market adopts large loads and continues to decarbonize.

Conclusion

This paper demonstrates that Demand Response (DR) is a valuable capacity resource in SPP, even when subject to a seasonal call limit. Our loss-of-load probability modeling of SPP in 2025-2030 showed that DR with duration and seasonal call limits has significant resource adequacy value in SPP's critical hours. Four-hour DR with a 40 hour/season call limit provides significant system value in both Summer and Winter, measured in ELCC. Longer-duration DR resources provide even greater effectiveness in both seasons, with especially pronounced benefits in Winter, when cold snaps create multi-hour or multi-day reliability challenges that shorter-duration resources cannot fully address. These findings highlight the importance of duration as a key driver of DR's ELCC.

Importantly, DR remains effective even with seasonal call limits, based on hundreds of simulations across diverse weather years and portfolio conditions. In most years, the system requires far fewer calls (often fewer than five) to address the highest-risk events. DR with no seasonal call limits was shown to have slightly higher ELCC than DR with seasonal call limits, but only in 1 of the 8 scenarios tested. This suggests that DR providers do not need to commit to unlimited dispatch to achieve high accredited capacity in SPP. Instead, carefully designed programs that balance duration, call frequency, and availability can result in DR providing dependable capacity at low system cost.

An efficient and scalable DR market design should have resource class performance characteristics which are clearly designed, supportive of SPP resource adequacy needs, and mitigate risks to DR providers. DR utilization in the average and most severe simulated weather year analyzed in this study provide guidance on the appropriate seasonal call limits that DR programs should consider.

As SPP continues to refine its accreditation and resource adequacy frameworks, integrating DR as a resource with performance-based accreditation will be essential. Recognizing DR's capacity accreditation alongside storage and other energy-limited resources ensures fair crediting and enables LREs to deploy DR as a cost-effective hedge against near-term uncertainty. In this role, DR can provide a critical bridge resource, reducing the risk of loss-of-load events while the region advances new generation and transmission development needed for long-term reliability and decarbonization.