Title

A Guide for Improved Resource Adequacy Assessments in Evolving Power Systems: Institutional and Technical Dimensions

Authors
Carvallo, Juan Pablo
Zhang, Nan
Leibowicz, Benjamin D
et al.

Publication Date
2024-01-21

Electricity Markets & Policy
Energy Analysis & Environmental Impacts Division
Lawrence Berkeley National Laboratory

A Guide for Improved Resource Adequacy Assessments in Evolving Power Systems

Institutional and technical dimensions

Juan Pablo Carvallo, Nan Zhang[1], Benjamin D. Leibowicz[1], Thomas Carr[2], Sunhee Baik, and Peter H. Larsen

[1]University of Texas, Austin
[2]Western Interstate Energy Board (WIEB)

June 2023

This work was supported by the U.S. Department of Energy's Grid Deployment Office under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231.

A Guide for Improved Resource Adequacy Assessments in Evolving Power Systems

Prepared for the
Grid Deployment Office
Transmission Planning Division
U.S. Department of Energy

Authors

Juan Pablo Carvallo
Nan Zhang
Benjamin D. Leibowicz
Thomas Carr
Sunhee Baik
Peter H. Larsen

Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS 90R4000
Berkeley CA 94720-8136

June 2023

The work described in this study was funded by the U.S. Department of Energy's Grid Deployment Office under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231.

Acknowledgements

The authors would like to thank Wesley Cole and Gord Stephen (NREL), Derek Stenclik (Telos Energy), Lorenzo Kristov (Consultant), Christopher Plante (MISO), David Parsons and Grace Relf (Hawaii PUC), Shahab Rastegar, Fei Weng, Peter Wong, and Steven Judd (ISO-NE), John Fazio (NWPCC), Phil Pettingill (CAISO), and Steve Johnson (WA UTC) for sharing of time for interviews that produced valuable contributions to this work. We are grateful to the members of the Technical Advisory Committee: Phil Pettingill (CAISO), Ryan Roy (NWPP), John Fazio (NWPCC), Chris Haley (SPP), Stuart Hansen (MISO), Tom Falin (PJM), John Buechler (EIPC), Joseph Bowring (Monitoring Analytics), Thomas Coleman (NERC), Gord Stephen (NREL), and Derek Stenclik (Telos Energy). We thank the insightful reviews and comments from Stephanie Lenhart (Boise State University), Bruce Biewald (Synapse Energy Economics), John Fazio (NWPCC), Gord Stephen (NREL), Ryan Roy and Sarah Edmonds (Western Power Pool), Stuart Hansen (MISO), Joseph Bowring (Monitor Analytics), Derek Stenclik (Telos Energy), Jan Porvaznik (Mainspring Energy), Paul Spitsen (DOE-EERE), and Hamody Hindi (DOE-GDO).

The work described in this study was funded by the U.S. Department of Energy's Grid Deployment Office under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231.

Table of Contents

Table of Figures

List of Tables

Executive Summary

Major weather events have caused widespread interruptions across vast regions in the U.S. In many cases, the power system did not have enough resources – including reserves – to meet demand due to a mix of higher than expected generation and transmission outages or deviations in short-term load or variable energy resources (VRE) generation forecasts. The ability of an electric power system to meet demands for electricity using its supply-side and demand-side resources is known as **resource adequacy (RA)** (NERC, 2011). The energy transition to highly decarbonized and electrified power systems with large penetration of VRE is requiring a revision of resource adequacy assessments to ensure the grid remains reliable, and potentially new types of assessment to ensure its resilience.

This paper identifies and evaluates issues in traditional resource adequacy assessment practices, and how adjusting these practices may depend on existing institutional arrangements for planning and procurement. The paper concludes by proposing a technical-institutional roadmap that would allow regulators in vertically-integrated jurisdictions and system planners and operators in restructured jurisdictions to revise resource adequacy practices across a range of components. More specifically, this paper provides answers to the following questions:

1. Who is the intended audience for this paper?
2. What are the key components of resource adequacy?
3. What are the emerging challenges with traditional RA assessments?
4. Should resilience be part of resource adequacy assessments?
5. What are key modeling practices that may improve RA assessments?
6. How may these technical changes in RA assessments affect other processes?
7. What are best practices for RA in planning processes?

Who is the intended audience for this paper?

This work caters to two distinct audiences:

- Regulators, policy-makers, and market designers will learn how the evolving power grid is prompting a need to review fundamental aspects of resource adequacy. The summary of recent developments in Section 3 should be accessible to understand the basic technical challenges, and Section 5 should provide insights on how the required changes will interact with existing planning processes developed in IRP and by ISO/RTO. These stakeholders may also be interested in latest developments in the treatment of resilience in planning processes, described in Section 2.
- System planners, researchers, analysts, and other stakeholders with a technical leaning may benefit from the organizational framework and discussion of adequacy provided in this Sections 1 and 2. The technical analysis in Section 4 and Appendix A should be relevant for these

audiences to learn about the benefits in complexity and accuracy of more detailed RA assessment models. These stakeholders may also benefit from reviewing Section 5, where the integration of RA assessment outcomes into planning and procurement practices raises new challenges.

What are the key components of RA?

Resource adequacy is a property of a power system, but the term is also used to refer to the process of tracking, assessing, and achieving adequacy. We propose decomposing adequacy in five key components or activities given the lack of an existing organizational framework (see Figure ES-1).

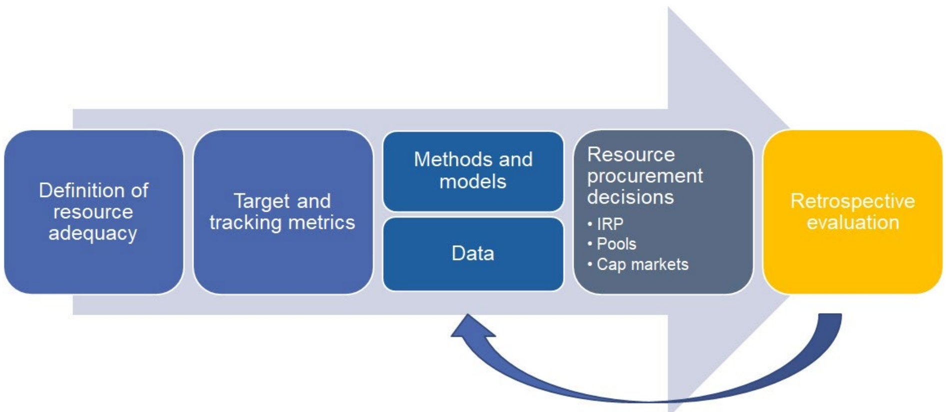

Figure ES-1 Resource adequacy framework developed in this paper

The first three components are technical in nature. How adequacy is defined; the metrics used to track adequacy and set targets; and the data, methods, and models employed have been the focus of recent research. The way adequacy assessments are translated into procurement decisions – and the potential role that this translation has in actual RA performance – has received less attention. This paper explores some aspects of this assessment-to-procurement process. Finally, the proposed RA framework offers a fifth new component focused on a retrospective evaluation of the adequacy performance of procured resources to inform their capacity accreditation, underlying data needs, and modeling approaches. RA processes in regulated jurisdictions and organized markets would benefit from this "anchoring" of the modeling process to the reality of resource performance under extreme weather events, cyberattacks, and other threats, as well as regular operational challenges.

What are the emerging challenges with traditional RA assessments?

Table ES-1 shows a sample of emerging challenges in traditional RA assessments resulting from (i) the evolution of the power systems towards decarbonization and (ii) climate change-induced extreme weather events.

Table ES-1 Emerging challenges with traditional resource adequacy assessments

Traditionally, RA assessments …	Emerging challenge …
Resources are predominantly **dispatchable**	Resources are becoming predominantly **non-dispatchable** (variable renewable resources, VRE)
The present state of most dispatchable resources **does not depend significantly on the past** and does not require chronological simulation	The present state of storage and load flexibility **resources depends on past states** and requires chronological simulation
Describe the system's high-risk conditions during the **peak demand** hour or a few top hours reasonably well	The system's high-risk conditions **may not occur during peak demand periods**, but during other hours in the year
Characterize stress conditions using **historical data**	Increases in extreme weather events, VRE adoption, and impending electrification of end uses makes **historical data less relevant** and creates challenges related to how to characterize possible reliability and resilience events
Assume that **outage events are uncorrelated** with each other and occur randomly	Evidence shows **high correlations of failures** within a class of power system assets and across infrastructure systems (e.g. natural gas and electricity)

We compile a critical review of current RA assessment practices (Section 3) based on (1) interviews with RA practitioners and (2) a review of recent technical literature. We find that:

- RA may need to **expand beyond capacity adequacy** to ensure energy adequacy – relevant for energy-limited resources such as storage – as well as ancillary service adequacy (e.g. enough ramping-up and ramping-down capability in the system). There is general agreement to include energy adequacy jointly with capacity adequacy, but it is not clear whether other system needs' assessments should be performed within the RA assessment or as separate processes.
- All studies and interviewees agreed that **basing RA assessments on the peak hour of the year or season, or on a few select top load hours, is insufficient** as peak demand may no longer predict the times when the power system is most stressed. Chronological hourly simulations are the current best practice.
- **Traditional metrics such as the Loss of Load Expectation (LOLE), Loss of Load Hours (LOLH), Loss of Load Events (LOLEv), and Loss of Load Probability (LOLP) are criticized** due to their "expected-value" nature, their focus on a single characteristic of a shortfall, and the coarse spatial resolution in their typical applications.
- An important shortcoming of current resource adequacy practices is that the **metrics and models used do not reflect economic criteria in system operation and loss of load**. Observers agree, however, that introducing economic criteria to determine adequacy levels introduces significant challenges related to valuing the loss of load for different customers, seasons, and end uses.

- There is a need to **improve representation of weather dependencies and weather data**, attending to a number of shortcomings of current practices that may hinder appropriate RA assessments under high wind and solar futures and climate change.

Should resilience be part of resource adequacy assessments?

A generally accepted definition of resilience is "the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions. Resilience includes the ability to withstand and recover from deliberate attacks, accidents, or naturally occurring threats or incidents" (EOP, 2013) Planning processes in power systems are beginning to grapple with the **need to develop metrics and methods to assess power system resilience**. We establish that from an organizational framework perspective, **resource adequacy assessments could feasibly be expanded to include resilience.** Alternatively, **a resilience-specific assessment could be developed using the same basic framework used by RA assessments**, provided that commonly-accepted resilience metrics become available. In either case, resilience assessments should complement existing resource adequacy assessments to ensure holistic power system reliability and resilience.

We review planning and RA reports for several private and public entities that plan generation and/or transmission infrastructure in the continental U.S. to look for existing practices involving resilience assessments (Section 2). We find **no systematic treatment of the costs of extreme weather and other hazards, the benefits of resilience, and resilience metrics in planning analyses** and no systematic treatment of resilience metrics, methods, and outcomes for resource adequacy purposes.

What are key modeling practices that may improve RA assessments?

This paper analyzes the effects of RA modeling choices on estimated RA outcomes to provide high-level and conceptual insights to regulators and planners on what power system operational details are important to include in a model-based RA assessment (Section 4 and Appendix A). We find that **the use of multiple years of weather and VRE performance data, the enforcement of transmission constraints, and the modeling of short-duration storage dispatch have a high impact on the accuracy of RA assessments** (Table ES-2). We develop a simplified representation of the economic operation of the grid and compare whether an RA assessment using this representation is more accurate than when utilizing non-economic dispatch assumptions in traditional RA assessments. We find that non-economic dispatch schemes that ignore economic objectives for power system operation can lead to fairly accurate RA assessments when enhanced with detailed operational strategies. This means that **current approaches that do not represent economic dispatch of power systems may still be sufficient to represent resources for RA assessments providing they implement sufficient operational data to characterize system performance**.

Table ES-2 Impacts of operational details on RA assessments

Operational or simulation characteristic	Impacts on RA assessment accuracy	Level of effort to represent in models
Multi-year data	High	Medium
Transmission limits	High	Medium
Storage dispatch	High	Medium (short duration) High (long duration)
Non-economic thermal dispatch	Medium	Low
Operational cost	Medium	High
Short-term forecast error	Low	High

In addition, we find that new **RA metrics that capture event-specific shortfall characteristics should be used as supplements to traditional metrics** to better capture the impacts of different modeling assumptions on RA outcomes, as well as to better describe the ability of the system to prevent specific high-impact shortfalls. More generally, **categorizing shortfalls based on their duration, season, and magnitude may be a promising area of RA assessment improvement**.

How may these technical changes in RA assessments affect other processes?

RA assessments are usually embedded within broader planning processes developed by utilities and system operators. In turn, planning processes are part of a set of regulatory and/or market designs that support procurement practices to ensure that the power system remains affordable, reliable, resilient, and sustainable – typically competing priorities. The mutual dependencies that arise from these relationships prompt the question: how would technical changes in resource adequacy assessments – including metrics, data, models, and methods – impact planning processes and broader institutional contexts.

It is important to know that the technical changes in resource adequacy that are suggested by recent work (i) may require upstream changes in planning and procurement practices for their successful implementation and (ii) may create opportunities to enhance planning processes due to availability of high resolution weather, load, and generator performance data, in addition to higher resolution representation of the transmission system. Section 5 presents changes required and opportunities for improvement of RA assessments and planning practices.

What are best practices for RA in planning processes?

We examine integrated resource planning (IRP) reports as well as Independent System Operator and Regional Transmission Organization (ISO/RTO) RA assessments. We use this information to propose a roadmap of evolving industry standards for resource adequacy assessments in resource planning and transmission planning (Table ES-3). The roadmap is based on three benchmarks applicable to key components of RA assessments:

- The first benchmark identifies the **bare minimum** of essential steps that entities need to implement in a reliability assessment.
- The second proposed benchmark corresponds to current **best practices** in the industry.
- The third proposed benchmark adopts a forward-looking perspective to identify RA assessment **frontier practices** that few, if any, entities are currently implementing.

The roadmap describes the potential challenges in implementing best or frontier practices. For example, the use of multiple adequacy metrics would require new methods to select portfolios that meet all, or some, of the targets set for each metric. Similarly, the use of forward-looking downscaled weather data that reflects climate change scenarios would require a parallel development of models that allow load and renewable resources to respond to these weather profiles, rather than relying on historical values. This more holistic representation of loads and resources would have repercussions on integrated resource and transmission planning processes for vertically-integrated and ISO/RTO jurisdictions, respectively.

Table ES-3 Roadmap to incorporate best practices for RA assessment into planning processes

Components of RA Framework	Planning Element	Benchmarks Minimum Practice	 Current Best Practice	 Frontier Practice
Definition of RA	Temporal resolution for RA	Meet load in a fraction of the top peak net load hours of the year	Meet load on an 8,760 hour basis	Sub-hourly analysis to meet load and ramping requirements
RA metrics and targets	RA metrics and targets	Single metric (e.g. planning reserve margin) driven by a maximum LOLP (not by the 1-in-10 rule of thumb)	Develop and explore multiple metrics produced by stochastic models that track shortfall magnitudes, frequencies, and durations	Use multiple metrics that track magnitudes, frequencies, and durations; consider full probability distributions of metrics and economic metrics
Data	Weather data	A few years of historical weather data with daily maximums/minimums	Several decades of historical weather data with variables at an hourly temporal resolution	Combine historical data with climate model data for forward-looking hourly weather forecasts
Data	Load forecasting for resource adequacy	Rely on several years of historical load data	Develop econometric or engineering-based load models that explicitly capture the dependence of load on weather	Pair weather-sensitive load models with forward-looking climate change-based weather patterns
Data	VRE characterization	Historical wind/solar performance for several years	Forward-looking wind/solar data for new sites, informed by historical empirical profiles	Climate change-induced wind/solar profiles based on downscaled climate model output
Models	Transmission and market transactions	Basic modeling of firm capacity and available exchanges	Regional simulation to accurately account for the availability of imported resources and market depth uncertainty; locational reliability analysis.	Enhanced modeling of transmission line derates; strengthen integration between generation and transmission expansion
Models	RA modeling and integration with planning process	Basic chronological Monte Carlo LOLP analysis; simplified storage representation	Iterative LOLP-CEM approach; model chronological storage operations	Stochastic CEM that internally assesses and ensures RA; include unit commitment and operational details
Procurement	Capacity credit	ELCC for renewables	ELCC for all resources, analyzed from individual and portfolio perspectives	Energy adequacy analysis; portfolio-based ELCC accounting for interactive effects

1. Introduction

1.1 Motivation

Recent extreme weather events in California and Texas have stressed the local electric grid such that system operators had to take emergency measures to avert widespread power disruptions. The factors that lead to these outcomes are numerous and complex. However, there is a recognition of the need to improve the planning processes employed in these systems to identify resource shortcomings and procure enough capacity. This need is heightened with the impending transition across the globe to power systems that are much more dependent on weather, and with changes in the climate that make historical performance less relevant and extreme weather more uncertain.

Resource adequacy (RA) refers to the ability of an electric power system to meet demands for electricity using its supply-side and demand-side resources (NERC, 2011). RA assessments are a cornerstone of planning processes that ensure that the power system meets prescribed levels of reliability.

Supply-side resources have traditionally been dispatchable and with predictable hourly capacity factors. In this context, the system would be resource adequate if enough capacity was available to meet projected peak demand plus a reserve margin to hedge against inaccuracies in projected demand, unexpected outages, and other operational deviations. The planning reserve margin is set at a level that at least roughly reflects an underlying basic standard for reliability equivalent to a maximum number of hours per year of power supply interruption.

This antiquated definition of resource adequacy may not be applicable to power systems with high penetration of variable renewable energy (VRE) for at least three reasons. First, peak demand is a poor descriptor of the times in which the system is under highest stress – either to meet load or more generally to rapidly move from one state of the system to a different state. Second, energy production becomes as uncertain as capacity to meet peak demand as monthly and annual energy production can vary substantially over time. Finally, ancillary services requirements, especially ramping capacity, are much larger in a high VRE system compared to a traditional power system. Furthermore, the current definitions of resource adequacy do not reflect any preferences or standards relative to system resilience, as they only reflect preferences for system reliability.

1.2 Analytical framework

One of the contributions of this paper is the development of a comprehensive analytical framework that encompasses technical aspects of RA in tandem with institutional aspects (Figure 1.1). Most of the work on resource adequacy to-date has been focused on technical aspects related to properly capturing and simulating the power system's stress periods that can lead to demand shortfalls. In reality, these technical assessments are just one component of a process that leads to acquisition and performance of capacity resources, and which is governed by specific institutional and regulatory arrangements. The framework presented in this document distinguishes five components of RA that characterize the outcome of the resource adequacy process from planning to procurement. This framework serves four purposes including the:

- *Integration* of technical and institutional decisions that affect the assessment and procurement of resource adequacy in a power system to support regulators, policy makers, and market designers to understand their dependencies.
- *Identification* of key decision points that establish dependencies across the different components of the resource adequacy framework. These dependencies enable decision makers to understand how the outcome of the resource adequacy assessment and procurement depends on definitions made outside the scope of a specific component.
- *Establishment* of a retrospective evaluation of resource performance that helps planners refine the data, methods, and models employed in resource adequacy assessments, as well as identify institutional designs whose performance may be departing from the resource adequacy planning process.
- *Assistance* to different stakeholders that coordinate and collaborate based on the roles they fulfill for each of the five components of the resource adequacy framework.

A key objective of the resource adequacy framework is to understand how the emerging body of work devoted to reexamining the technical methods used to assess resource adequacy may affect or conflict with existing institutional designs for restructured and non-restructured jurisdictions.

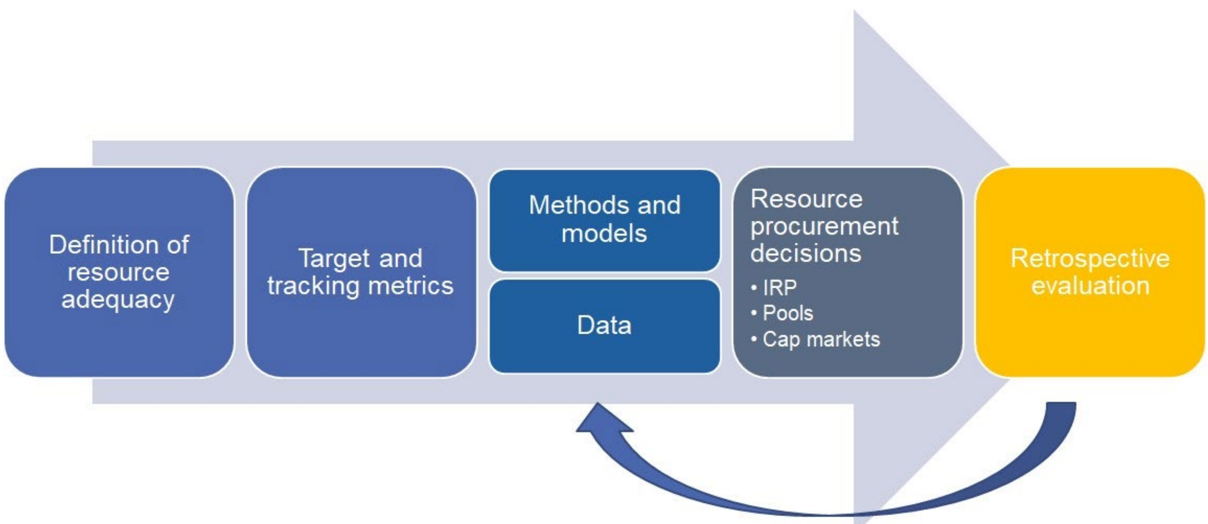

Figure 1.1 Resource adequacy framework developed in this paper

The first component of the framework relates to establishing an actionable definition of what resource adequacy is. NERC defines resource adequacy as “the ability of supply-side and demand-side resources to meet the aggregate electrical demand (including losses).” The historical implementation of this definition has focused almost exclusively on the peak hour, or a small set of high demand or net demand hours. This component explores how the implementation of RA may need to evolve, including the potential expansion to cover the resilience of the power system.

The second component describes the targets and tracking metrics used to set resource adequacy goals that should be met through procurement and to track the adequacy status of a given system, respectively. The choice of metrics for both purposes shapes the outcome of the resource adequacy process, both technically and institutionally.

The third component refers to the methods, models, and data required to perform RA assessments. Methods and models have received substantial attention in the last few years prompted by extreme weather events that have disrupted power systems around the country. This component describes the minimum and ideal modeling as well as the methodological approaches that produce rigorous assessments of a power system's resource adequacy. In addition, data needs required to run these models appropriately have not been explored in depth and this component explicitly discusses how data availability may constrain modeling performance and results.

The fourth component is largely institutional and refers to the mechanisms employed in different jurisdictions to procure the capacity resources deemed necessary by RA assessments. In non-restructured jurisdictions where utilities are vertically-integrated, the mechanism is regulatory approval of capital expenditures on resources supported by integrated resource plans (IRP). In jurisdictions that operate under mandatory capacity markets (ISO-NE, NYISO, and PJM) the mechanism is forward-looking auctions for capacity whose design is supported by RA assessments developed by the ISO/RTO. Finally, some jurisdictions have regional pools or facilitate bilateral agreements for load serving entities to meet RA obligations. The three mechanisms may be affected by the technical changes in other components of the RA framework. The identification of these changes is one of the core contributions of this paper.

The fifth component reflects the need for retrospective evaluations that would allow regulators, planners, and system operators to assess how the translation of technical resource adequacy objectives into regulatory obligations and capacity market design performs. There are several technical and non-technical reasons why actual resource performance may diverge from the outcomes of the RA assessment developed in components 2 and 3. This component therefore acts as a control mechanism to ensure that RA assessments account for actual system performance. In contrast to the first four components, retrospective evaluations of resource adequacy assessments are not currently performed in planning processes.

1.3 Objectives and organizational structure

This paper achieves a number of objectives including:

- A summary of recent literature that has studied the technical changes that resource adequacy assessment needs to incorporate to accurately and rigorously describe the reliability status of near-future decarbonized power systems with high dependence on VRE.
- An evaluation of the potential integration of resilience metrics, methods, and processes into resource adequacy as a way of effectively integrating resilience into power system planning.

- An assessment of how resource adequacy assessments would be improved by capturing the availability of resources during actual dispatch conditions that resources are subject to within economically-dispatched power systems.
- An analysis of the trade-offs between complexity and usability in resource adequacy assessment methods. Employing more technically sophisticated methods could result in more accurate resource adequacy assessments but make the assessment process more difficult to implement and less transparent to stakeholders.
- Bridging an existing gap by situating the technical changes within the broader institutional context by focusing on resource planning, power pools, and organized markets. Readers are then able to identify regulatory and policy changes that may be required to adapt current planning and procurement processes in light of advancements in resource adequacy assessments.

This paper is organized based on the five components of the resource adequacy framework. Section 2 delves into component 1 – the definition of RA – with a focus on how resilience may be considered part of its scope. Section 3 addresses components 2 and 3 by summarizing the existing work on RA and by offering new perspectives on metrics and modeling approaches that contribute to this work. Section 4 reports a deep technical analysis that employs a specially-adapted production cost model to assess RA and determine the relevance of using economic dispatch to identify resource availability. This section also explores the impact on RA assessment accuracy of several other variables of interest. Finally, Section 5 puts the technical developments in resource adequacy into the broader context of institutional environments that translate assessments into procurement, indicating how these institutional processes may need to evolve.

This work caters to two distinct audiences.

- Regulators, policy makers, and market designers will learn how the evolving power grid is prompting a need to review fundamental aspects of resource adequacy. The summary of recent developments in Section 3 should be accessible to understand the basic technical challenges, and Section 5 should provide insights on how the required changes will interact with existing planning processes developed in IRP and by ISO/RTO. These stakeholders may also be interested in the latest developments in the treatment of resilience in planning processes, described in Section 2.
- System planners, researchers, analysts, and other stakeholders with a technical leaning may benefit from the organizational framework and discussion of adequacy provided in this Section and Section 2. The technical analysis in Section 4 and Appendix A should be relevant for these audiences to learn about the benefits in complexity and accuracy of more detailed RA assessment models. These stakeholders may also benefit from reviewing Section 5, where the integration of RA assessment outcomes into planning and procurement practices raises new challenges.

2. The definition of resource adequacy

This section discusses the concept of resource adequacy and its implementation in power system planning processes, with a focus on how resource adequacy should consider power system resilience.

2.1 Defining adequacy

Resource adequacy is a concept that describes the ability to meet demand with sufficient supply- and demand-side resources (NARUC, 2021). NERC's definition of adequacy is more elaborate and precise, and clarifies that resource adequacy is one component of reliability (the other being operating reliability). As reported in NERC's latest Long Term Reliability Assessment, adequacy is "The ability of the electricity system to supply the aggregate electric power and energy requirements of the electricity consumers at all times, taking into account scheduled and expected unscheduled outages of system components" (NERC, 2021). An even more precise definition is provided by the National Renewable Energy Laboratory (NREL) in the documentation of its Probabilistic Resource Adequacy Suite (PRAS) model, which states that "An electrical power system is considered resource adequate if it has procured sufficient resources (including supply, transmission, and responsive demand) such that it runs a sufficiently low risk of invoking emergency measures (such as involuntary load shedding) due to resource unavailability or deliverability constraints" (Stephen, 2021).

The definition in PRAS refines the typical adequacy concept in four important ways. First, it does not assume that load is served at all times, but acknowledges that there are shortfalls that may occur with a tolerable risk. Second, it mentions that this risk does not mean a shortfall will indeed occur, but that emergency measures may need to be used to mitigate or eliminate a potential shortfall. Third, it specifies reasons why these emergency measures may need to be implemented, including availability of resources and deliverability of power to loads. Finally, its use of the term "procurement" reflects the need for adequacy assessments to be translated into mechanisms to address deficiencies by procuring generation, transmission, storage, and/or demand side resources. These refinements are important because they are more precise and specific, and also because they inform how the assessment should be performed and what its ultimate objective should be.

Novel implementations of adequacy assessments have recognized that resource adequacy is a system property that varies with spatial and temporal scales. For example, recent changes in the European Resource Adequacy Assessment method require assessing adequacy for current and projected demand levels at the Union, member state, and individual bidding zone levels (ACER, 2020). Similarly, the California Independent System Operator (CAISO) has system, local, and flexible resource adequacy requirements. Most regional resource adequacy assessments developed by NERC and the Northwest Power and Conservation Council (NWPCC) have an explicit temporal scale looking three to five years into the future, accounting for different levels of committed resources and policy targets, and reporting the evolution of adequacy over that period.

Emerging reasons to continue refining resource adequacy definitions may include tracking flexibility and ancillary services needs in addition to capacity needs within the same adequacy process. Resource adequacy is focused on capacity to meet demand, but the inclusion of a large amount of energy-constrained storage may require additional energy adequacy or reserve margin metrics to be set and tracked. While there is agreement on these evolving needs of a power system, it is an open question on whether resource adequacy should be the process to internalize them or whether new planning processes will be needed to ensure that capacity, energy, flexibility, and ancillary services – among others – are in adequate supply in the power system.

2.2 Treatment of resilience

Resilience is an emergent concept in power systems with no commonly-accepted metrics, methods, and processes to reflect resilience in planning processes (NARUC, 2019). A generally accepted definition of resilience is “the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions. Resilience includes the ability to withstand and recover from deliberate attacks, accidents, or naturally occurring threats or incidents” (EOP, 2013). The National Academies of Sciences finds that "Resilience is not the same as reliability. While minimizing the likelihood of large-area, long-duration outages is important, a resilient system is one that acknowledges that such outages can occur, prepares to deal with them, minimizes their impact when they occur, is able to restore service quickly, and draws lessons from the experience to improve performance in the future” (National Academies of Sciences, 2017, p. 10).

The treatment of variability in several aspects of the power system that are relevant to resource adequacy assessments make it possible for them to internalize at least certain aspects of resilience. An argument against including resilience within resource adequacy is that the latter process focuses on *risks* rather than *uncertainties*. A technical definition by Knight (1921) stipulates that risks are situations in which the outcome is not known, but the probabilities governing the outcome can be known. In contrast, uncertainty is characterized by unknown outcomes and underlying probabilities of these outcomes of occurring. This is why, in practice, uncertainty is managed through contingency and emergency procedures that operators can leverage on a short-term basis once the uncertain event or variable and its impact are known (e.g. the response to an earthquake).

Planning processes in power systems are beginning to grapple with the need to expand the risk management processes that have shaped resource adequacy assessments to incorporate resilience. This is largely due to the effects that climate change is having on weather patterns that affect supply and demand conditions in a power system. We reviewed planning documents for several private and public entities that plan generation and/or transmission infrastructure in the continental U.S. to understand whether and how resilience is incorporated in planning processes (see list of entities in Appendix 8.2). We find that utilities performing resource planning acknowledge the importance of resilience, but we find no systematic treatment of the costs of extreme weather and non-weather events, resilience benefits, and resilience metrics in their analyses. Some entities would indicate that a certain resource has a resilience benefit, but made no effort to quantify this benefit or to put it in the context of a resilience-based planning approach. Exceptions to this qualitative treatment were Puget Sound Energy and the California IOUs, who analyze the resilience benefits of a subset of resources and the entire portfolio, respectively. Some of these entities designed special “extreme weather” scenarios for load that intended to capture 90th percentile and beyond conditions[1], although entities that perform stochastic assessments implicitly assess the tails of load probability distributions in their adequacy analysis.

Regional planning entities and reliability organizations do not perform resource planning; they undertake transmission planning or no planning at all. However, all of these organizations develop resource adequacy assessments as part of their planning or reliability assessment processes. As with utilities, none of these entities has developed a systematic treatment of resilience metrics, methods, and outcomes. However, some of these organizations briefly discuss resilience issues. MISO and ISO-New England raise emerging issues that could change the future grid operation and technologies that are expected to enhance grid resiliency (e.g., introducing black-start capabilities or inverter-based resources, the introduction of VERs, HVDC and FACTS devices), but none of them describe how these issues would be considered in their future system planning or operations. NWPCC considers resilience as a type of benefit of increased energy efficiency and recommends that their regional technical forum investigate methods for quantifying the value of flexibility and resiliency for energy efficiency measures for their future studies. The Western Electricity Coordinating Council (WECC) discusses the increasing trend in extreme events and policy changes to the demand variability distribution and suggests the possibility of including extreme weather events (e.g., heat waves) in its future analysis. Finally, NERC's 2021 Long Term Reliability Assessment does not explicitly model resilience, but includes an analysis of the potential impacts of extreme weather events on each electric reliability organization in addition to their analysis of the reserve margins, energy risks, and integration of variable energy resources. The fact that in NERC's report the term "resilience" appears nine times, compared to 424 mentions of "reliability" and related terms, indicates that there is much work to do to effectively integrate resilience into power system analysis.

Our review of planning analyses and adequacy assessments from existing organizations reveals a lack of systematic treatment of resilience. Incorporating resilience into power system planning may benefit from drawing parallels with reliability assessments performed through resource adequacy analyses to help entities develop methods that extend from analyses they know how to perform. In Table 2.1, we identify four steps for reliability and resilience assessments. Three of these steps are common across both reliability and resilience. These steps include:

1. Selection of variables or events of interest. In the case of resource adequacy, these are common and well known, such as load, generation and transmission infrastructure outages, and variable renewable resource output. In the case of resilience, a stakeholder process supported by historical analysis may be used to select specific events of interest
2. Assessment of the impacts of variables or events on the power system. Resource adequacy performs simplified and repeated modeling of the power system operation to identify instances in which demand may exceed supply. Resilience assessments may use the same modeling

[1] Planning processes typically focus on 50th percentile (median) load conditions for investment purposes.

framework but focus on a limited set of rare, high-impact events that fall well outside normal operating conditions and are difficult to assign probabilities to.

3. Using outcomes of the assessment to inform decision-making. Resource adequacy assessments link risk of loss of load with levels of firm capacity, which connects investment decisions with risk. Resilience assessments may complement this decision-making process, potentially leveraging decision-making under uncertainty as a tool to make resilience assessments actionable (Stanton and Roelich, 2021). A stakeholder process may be important to assess the benefits of resilience interventions and the tradeoff with their cost.

The component related to characterizing the likelihood of occurrence of variables or events is not required for resilience assessments because their probability distribution is not known. However, the table shows that key elements from current resource adequacy assessments could be redeployed in a complementary process to internalize resilience in power system planning and decision-making.

Table 2.1 Analogous treatment of reliability and resilience in power systems

Steps	**Reliability**	**Resilience**
Select variables or events of interest	Typically load levels, infrastructure outages, renewable energy production, all chronologically and weather correlated	Earthquakes, large wildfires, cyberattacks; types of weather events that do not have historical record
Characterize likelihood of occurrence	Specify probability distributions for variables of interest	Not applicable/feasible
Assess impact of variables/events on power system	Comprehensive probabilistic assessment through chronological Monte Carlo simulation[2]	Scenario-based assessment, subjecting the power system to instances of the events of interest
Decision-making	Risk of loss of load tightly linked to firm capacity; investment decisions follow from quantitative preference for risk	Use decision-making under deep uncertainty approaches; investment decisions emerge from comparing several outcomes[3]

[2] *Chronological Monte Carlo* describes a modeling approach that simulates the availability of a resource on each time step in the simulation (e.g. every hour) by assuming a certain probability for its outage, or a certain probability for its output (in the case of VRE). More importantly, the state of a unit on a given point in time may depend on its state in a previous point. This method is also referred to as *sequential Monte Carlo*.

[3] A good reference for deep uncertainty is Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M., Popper, S.W. (Eds.), 2019. Decision Making under Deep Uncertainty: From Theory to Practice. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-05252-2

3. Diagnosing current RA assessment practices: metrics, methods, models, and data

As the definition of resource adequacy continues to be refined and evolve, the technical components of resource adequacy assessments including the metrics, methods, models, and data are also evolving. Existing technical work on RA assessments has diagnosed and identified the limitations of current assessments to capture the high-stress states of evolving power systems that will substantially rely on variable renewable energy, storage, and demand-side resources. Table 3.1 reports assumptions that have traditionally been used in the methods and data employed in resource adequacy assessments (left column) and how these assumptions are or may become invalid as power systems are decarbonized and climate change drives changes in weather patterns (right column). This section expands on these emerging challenges and the practices that have been identified in the literature to address them. Section 4 delves into one specific aspect that has been relatively understudied: how would a more accurate representation of the *operation* of the power system reflect the actual availability of resources to contribute to system adequacy?

Table 3.1 Emerging challenges with traditional resource adequacy assessments

Traditionally, RA assessments …	Emerging challenge …
Resources are predominantly **dispatchable** and available to the operator based on their planned outage rates	Resources are becoming predominantly **non-dispatchable** (variable renewable resources, VRE) and their availability and contribution to RA are stochastic
The present state of most dispatchable resources **does not depend significantly on the past** and does not require chronological simulation	The present state of storage **resources depends on past states** and requires chronological simulation
Describe the system's high-risk conditions during the **peak demand** hour or a few top hours reasonably well	The system's high-risk conditions **may not occur during peak demand periods**, but during other hours in the year
Characterize stress conditions using **historical data**	Increases in extreme weather events and VRE adoption makes **historical data less relevant** and creates challenges related to how to characterize possible reliability and resilience events
Assume that **outage events are uncorrelated** with each other and occur randomly	Evidence shows **high correlations of failures** within a class of power system assets and across infrastructure systems (e.g. natural gas and electricity)

Part of the motivation to examine whether operational details should be included in resource adequacy assessment relates to a trend to make the assessments significantly more complex. As noted in previous studies on integrated resource planning, planning processes tend to become increasingly more complex over time, with both regulators and regulated entities adding layers of analysis, expanded scopes, and more comprehensive scenarios, among others (e.g. Carvallo et al., 2019, 2018). There has been comparatively less work on how to make these processes *simpler* (e.g. by removing analyses or requirements that may have become obsolete or not proven to be useful) and little critical evaluation of the advantages and disadvantages of assessing RA using ever more complex technical methods. An obvious tradeoff is that as studies become more complex, it becomes increasingly more challenging for regulators, policy makers, customers, and other non-technical stakeholders to understand the data, methods, and outcomes of these processes. More complex studies are computationally intensive, which increases the financial and time requirements to run them. This study does not delve into the detailed tradeoffs between added complexity and increased RA assessment accuracy, but does identify when these tradeoffs may be present and examine which methodological choices meaningfully influence estimated RA metrics.

A small but recent body of literature has sought to critically review existing resource adequacy practices, diagnose potential issues in evolving power systems, and propose both technical and (to some extent) institutional solutions. We focus on studies developed by practitioners or consultants working closely with planning entities and regulators – in lieu of academic literature – to capture proposals that are grounded in the reality of U.S. power systems. Papers cited in this section include Cigre (2018), ESIG (2021), Fazio and Hua (2019), Frew (2018), Gramlich (2021), Lannoye and Tuohy (2020), Mauch et al. (2022), and NERC (2018). We complement this review of literature with interviews conducted with researchers, consultants, regulators, and ISO/RTO staff acknowledged at the beginning of this report. We organize the presentation of topics and practices according to the RA framework introduced in Section 1.

3.1 RA definition

In general, studies to-date have not criticized the definition of resource adequacy, although as reported in Section 2 the definition has been refined to be more specific about its intent. A few observers have commented that the actual implementation of adequacy has historically focused exclusively on capacity, but that this focus may need to evolve (ESIG, 2021; Frew, 2018). The main suggestion is that adequacy may not only need to focus on having supply- or demand-side resources to meet demand in a given hour, but that the system may also need enough ancillary services and flexibility to maintain reliability as well as energy to supply over predetermined periods of time. An open question is whether resource adequacy should be the process to assess whether sufficient ancillary services and flexibility exist in the system, or whether these needs should be assessed separately in a parallel process.

An interesting proposal remarks that resource adequacy focuses on having enough resources to meet prescribed demand levels with a certain risk – this is, that resources exist to meet demand (Cigre, 2018). The proposal suggests that the evolution in demand flexibility and responsiveness prompts a complementary and symmetric definition: that there is enough flexible demand to absorb power injections into the system at any point. The latter complementary definition is particularly important when explicitly considering distributed generation and storage and suggests the need for a new paradigm that integrates bulk power system adequacy with distribution system capacity adequacy to absorb these injections.

With the evolving integration of resource and distribution system planning, the adequacy of the bulk power system on its own may become less meaningful when relatively large portions of customer needs may be met locally. This will be particularly complex for ISOs and RTOs that have historically had no visibility into the distribution system planning and procurement processes. Exploring how a joint distribution-bulk power system adequacy framework could technically and institutionally work – especially considering the challenges of assessing distributed resource deliverability through distribution networks – is an emerging area of research.

3.2 Target and tracking metrics

Resource adequacy assessments have coalesced around a handful of technical metrics. The Loss of Load Expectation (LOLE) measures the number of event-periods, where an event-period is a specified time period (day, month, and year) in which one or more shortfall events occur.. Complementary metrics such as the Loss of Load Hours (LOLH) and Loss of Load Events (LOLEv) count the expected number of shortfall hours per year and the expected number of shortfall events per year, respectively. The LOLE is typically used as a target setting metric and has historically taken a value of 1 event-day in 10 years, commonly (and incorrectly) interpreted as 2.4 hours per year[4] (Stephen et al., 2022). These frequency (LOLE, LOLEv) and duration (LOLH) metrics are complemented by a cumulative energy lost metric, the Expected Unserved Energy (EUE) that reflects the cumulative unserved energy due to the shortfall events on a year.

As we will review in Section 4 and Appendix A, some implementations of RA assessments rely on the Loss of Load Probability (LOLP) that can be calculated from repeated simulations that draw values from probability distributions of key variables, also known as a Monte Carlo analysis[5]. Planners that employ LOLP have generally used a 5% annual LOLP, which is the likelihood that a future year will experience one or more shortfalls of any duration or magnitude.

The LOLP is an aggregate metric that summarizes the results of a complex and expansive analysis process that tests the current power system thousands of times for load shortfalls. This process is not amenable to optimal capacity expansion modeling used to identify investments that maintain adequacy. Most planners then roughly translate the LOLP into a planning reserve margin (PRM) that can be programmed into a capacity expansion model to constrain it not to meet peak demand, but to meet peak demand plus a margin. This margin provides a "cushion" for the system to remain reliable under uncertainty and allows the capacity expansion model to select the least-cost combination of resources that satisfies the PRM constraint.

[4] As indicated in Stephen et al. (2022) p.1 *"LOLE is not a measure of expected total shortfall duration, […] a 2.4 hours per year LOLE target implies a less reliable system than a 1 day in 10 years (0.1 days per year) LOLE target, and […] exact conversions between hourly and daily LOLE targets are not generally possible."*

[5] Power systems that are not energy-constrained can be assessed more efficiently using convolution methods instead of Monte Carlo analysis (Stephen, 2021).

The LOLE, LOLH, LOLEv, and LOLP metrics have been criticized in a number of ways due to their "expected-value" nature, their focus on a single type of shortfall, and the spatial resolution in their typical applications (see Table 3.2)

Table 3.2 Criticisms of common resource adequacy metrics

Criticism	Comment
The metrics identify a single property of a shortfall or group of shortfalls, either their frequency, duration, or magnitude.	Observers have highlighted the need to move towards compound metrics or multiple metrics that provide a broader overview of the types of adequacy challenges that systems exhibit, and that may better inform the types of resources needed to cure these deficiencies. There are outstanding questions on how a planning process would internalize multiple metrics, especially what standards should be adopted as targets, how to make capacity expansion models respond to more than one metric, and what and how these metrics should be used for tracking adequacy.
These metrics characterize shortfalls on average or in aggregate, but do not provide any information on the distribution of frequencies, durations, or magnitudes and hence the characteristics of individual shortfalls.	Some of these limitations would be addressed with the use of multiple metrics, but there is still valuable information in the distribution of shortfall events that the commonly-used metrics do not currently capture. Recommendations of moving in the direction of using these distributions do not identify, however, the exact benefits of using these distributions nor the mechanisms through which these distributions would be applied in planning processes.
No metric is currently able to capture the timing of a shortfall event	Timing is relevant because customer needs change over a day and across seasons, and the economic impacts of shortfalls will most likely be different across these different temporal scales.
Metrics are established for entire systems, typically a balancing area or the footprint of an ISO or RTO.	In many cases, the size, transmission constraints, and technology heterogeneity of these footprints may warrant the development of multiple subsystem metrics that respond to specific local needs in terms of duration, frequency, magnitude, or timing of events. Even if a single set of targets is used for the entire footprint, tracking adequacy at smaller spatial scales would allow for better investment decision to address deficiencies.

Finally, an important shortcoming of current resource adequacy practices is that the metrics and models used to assess adequacy do not reflect economic criteria. The typical metrics introduced earlier establish technical criteria based on the characteristics of the shortfalls, but not their economic impact. This means that there is no explicit agreement on how much the power system as a whole is willing to pay to achieve certain levels of reliability; the cost is embedded in the choice of adequacy targets and actual outcomes. Observers agree, at the same time, that introducing economic criteria to determine adequacy levels introduces significant challenges. For example, the value of lost load (VOLL), which would be used to determine how much load is economically efficient to not meet, is very hard to calculate as it differs by customers (or at least across customer segments), across geographies, and over time. Providing each customer or groups of customers with customized reliability service may be feasible with advanced metering infrastructure. However, work will be needed to translate customer-level reliability choices into system-level adequacy outcomes and to ensure that no new equity issues arise with customer-level reliability choices.

3.3 Methods, models, and data

Perhaps unsurprisingly, most of the diagnoses, criticisms, and recommendations for resource adequacy assessments relate to the methods, models, and data components.

All studies and interviewees agreed that basing RA assessments on the peak hour of the year or season, or on a few select top load hours, is insufficient. The reasoning is that peak demand may no longer predict the times when the power system is most stressed, as outages, load, and VRE production all depend on weather patterns. Using net demand was deemed an alternative in the long term, although it has the critical disadvantage of entangling the probability distributions for load and VRE that will actually differ. One recommended practice is to run Monte Carlo simulations over every hour of the year, in which the model compares a stochastic realization of weather-dependent load to stochastic realizations of VRE, hydropower production, generation and transmission outages, and other variables to determine whether a shortfall would occur. This process would then be run over several hundred or thousands of “test” years to encompass a wide range of possible stochastic realizations. These Monte Carlo runs can be non-sequential (each hour is considered independent of other hours) or chronological (in any given hour, the state of the system is inherited from the previous hour) (Stephen, 2021).

The use of chronological Monte Carlo methods is recommended to capture the operation of energy-limited resources such as battery storage and reservoir hydropower, as well as time-dependent resources such as demand response and other load flexibility measures. In all these cases, however, RA assessment tools operate these resources to minimize shortfalls or unserved energy, which is not the way resources are operated in power systems. Indeed, unit commitment and economic dispatch are two operational aspects of bulk power systems that greatly inform the actual availability and state of a resource in a given hour. In response to this need, this paper devotes Section 4 and Appendix A to investigating the implications of using economic dispatch logic in resource adequacy assessments.

Many current RA assessments simplify the transmission system, generally assuming a copper-plate[6] model with no transmission constraints. Deliverability studies are typically conducted by ISOs and RTOs

[6] A copper-plate model is a power system approximation that assumes all resources connected to a single point, effectively removing the transmission system from the analysis.

to determine whether resources that are considered by a load serving entity to meet its resource needs are indeed available at the expected levels and timing. However, these studies are not used for planning purposes and hence are not an integral part of resource adequacy assessments. Ignoring transmission constraints during a resource adequacy assessment has several consequences: (1) it precludes an appropriate representation of transmission outage failures, especially in areas where the grid is most vulnerable, (2) does not allow the assessment of deliverability of resources within a footprint, potentially undercounting shortfalls, and (3) does not allow accounting for the reliability benefits of neighboring systems. Expanding transmission system representations in RA assessments would make simulations computationally more complex, and may also require representations of neighboring systems that expand the scope of the entire exercise.

Some observers are becoming critical of the practice of subtracting distributed resources from customer load to represent the net load that resources should meet for adequacy purposes, especially in areas of the country with higher penetration of distributed resources. On the one hand, they argue that the distributed resources have a distinct risk profile that differs from that of load, and more importantly in the case of rooftop PV, is highly correlated with utility-scale PV. On the other hand, confounding distributed resources with load does not allow identifying the flexibility that each resource can provide to the system. An idea has been proposed to explicitly separate customer demand into flexible and inflexible components in an effort to improve the treatment of load within RA assessments.

Many observers emphasize the need to improve representation of weather dependencies and weather data, attending to a number of shortcomings of current practices that may hinder appropriate RA assessments under high wind and solar futures and climate change. First, most models represent generator and transmission outages using Markov chains with failure rates that are independent of each other and of other variables in the system. The recent experience of heat waves in the Pacific Northwest in 2019 and winter storm Uri in Texas in 2021 demonstrated the importance of replacing these uncorrelated failures with weather-dependent, and hence correlated, ones. Second, the use of weather-dependent load models is not extended; making load and generation depend on the same weather patterns would better capture their operational correlation and the actual needs of the system. Finally, it has been recognized that using historical weather patterns may not appropriately reflect how future weather patterns will evolve under climate change. There will be a need for highly spatially- and temporally-resolved weather datasets that can be used in forward-looking adequacy assessments.

4. Analyzing the use of power dispatch models to improve RA assessments – Non-technical summary

Due to the increasing penetration of VRE and storage in modern power systems, RA assessments are becoming sensitive to the combination of operational uncertainties and a series of dispatch decisions, including the scheduling of generators and the charging and discharging of storage units. Therefore, in order to accurately assess the RA of a power system, it may be necessary to simulate the system's chronological operations using a dispatch model. In this section, we develop an RA assessment framework based on an existing production cost model, *Prescient*, to simulate a hypothetical system's power dispatch and analyze its impacts on RA assessments. Most significantly, we investigate how important it is to represent each of the following aspects of power system operations in a more sophisticated – but more computationally burdensome – manner in order to accurately estimate the system's RA performance:

- Multi-year data
- Transmission limits
- Storage dispatch
- Non-economic thermal dispatch
- Operational cost
- Short-term forecast error

This section provides a high-level summary of our methodology and findings for non-technical audiences. For more details, the full analysis is described in Appendix A.

4.1 Summary of the methodology

Prescient is a software toolkit developed by Sandia National Laboratories to simulate electric grid operations (Watson et al., 2020). It simulates the power system's economic dispatch by developing a day-ahead and real-time simulation cycle framework that combines commitment optimization and dispatch optimization. The day-ahead simulation is implemented as a single unit commitment optimization that determines the daily system operations with respect to the forecasted load and renewable generation. After that, the real-time simulation consists of a series of chronological hourly economic dispatch problems that re-optimize the daily system operations subjected to actual load, actual renewable generation, and the commitment status determined by the day-ahead unit commitment solution. Each hourly simulation uses the system state after the previous hour as its initial condition to simulate the inter-temporal dispatch constraints.

Given that Prescient is inherently a production cost model, we make several modifications to translate it into an RA assessment dispatch model. These modifications make the model less computationally burdensome so that it can be solved many times within a large-scale Monte Carlo simulation. First, we replace the unit commitment problem with the linearized merit order model in the day-ahead simulation. Second, we represent power flows between zones via the transmission network using a simplified transportation model. Third, we model storage operations by embedding storage decisions into the day-ahead dispatch problem to determine state of charge targets that the system will attempt to follow during real-time operations.

We apply our RA assessment model to a case study based on the IEEE Reliability Test System - Grid Modernization Laboratory Consortium (RTS-GMLC) (Barrows et al., 2020). The IEEE RTS region covers desert areas of Southern California, Nevada, and Arizona. It is intended to serve as a platform for analyzing power system operation strategies and issues, with given power system topology, load obligations, and generation resources. We make some modifications to the raw RTS data (described in more detail in Appendix A.2) in order to produce a tractable case study that leads to an interesting RA assessment. The zones and simplified transmission network in our case study are depicted in Figure 4.1.

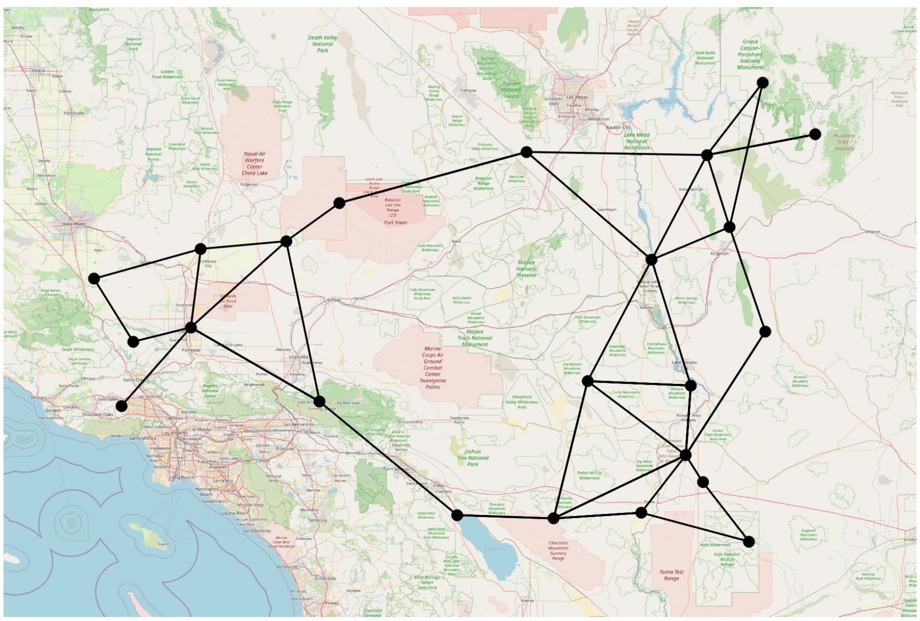

Figure 4.1 Zones and transmission network

To investigate how the model representations of six aspects of power system operations affect estimates of the system's RA metrics, we run numerous scenarios (described in more detail in Appendix A.3) that differ in the following respects:

- Multi-year data: We simulate power system operations over 500 randomly generated years of operating condition data that include time series for loads, wind outputs, solar outputs, and stochastic thermal generator outages. We examine how sensitive the system's RA performance is to the particular data year.
- Transmission limits: We run scenarios with and without transmission limits that constrain power flows between zones.
- Storage dispatch: In some scenarios storage is operated during the real-time stage to meet state of charge targets that were either exogenously defined or optimized in the day-ahead

dispatch problem. In other scenarios, storage is treated as either a cost-free resource (charge when there is excess renewable generation, discharge to ramp down thermal generators) or as reserves (discharge only when load exceeds available generation capacity, otherwise charge as much as possible).

- Non-economic thermal dispatch: In scenarios where thermal generators are not economically dispatched, we consider prioritizing them based on their reliability or at random.
- Operational cost: We compare scenarios with non-economic dispatch schemes to scenarios in which units are dispatched based on their marginal production costs to minimize total operational cost.
- Short-term forecast error: We compare scenarios with perfect day-ahead forecasts to scenarios where real-time operating conditions can differ from what was expected.

Based on the results of each scenario, we compute RA metrics including LOLE, LOLP, LOLEV, and EUE. Comparing the estimated values of these metrics across scenarios provides insight into how sensitive or robust the system's estimated RA performance is to the choices that modelers make about how to represent power system operations. In addition to the aforementioned RA metrics, we also examine the full distribution of loss of load events in each scenario including the times when they arise during the year, their durations, and their variability across many simulated years. This helps highlight whether new RA metrics are needed as supplements to the traditional metrics in order to capture additional important characteristics of shortfalls. Lastly, we calculate the total production cost in each scenario to assess whether scenarios with superior RA performance tend to involve considerably higher production costs.

4.2 Summary of findings

In this analysis, we have created a technical framework for probabilistic RA assessment and used it to study how key choices about how to model power system operations affect the values that are obtained for RA metrics. As summarized in Table 4.1, the results helped us distinguish operational details that are critical to include in any accurate RA assessment from details that are computationally burdensome but do not significantly affect the evaluation of RA. Our detailed scenario results that informed the conclusions summarized in Table 4.1 are provided in Appendix A.4.

Table 4.1 Impacts of operational details on RA assessments

Operational or simulation characteristic	Impacts on RA assessment accuracy	Level of effort to represent in models
Multi-year data	High	Medium
Transmission limits	High	Medium
Storage dispatch	High	Medium (short duration) High (long duration)
Non-economic thermal dispatch	Medium	Low
Operational cost	Medium	High
Short-term forecast error	Low	High

Several high-level findings emerged from the case study explored using our technical RA framework.

- First, non-economic dispatch schemes that ignore economic objectives can lead to fairly accurate RA assessments when coordinated with detailed operational strategies.
- Second, representing the detailed chronological operations of thermal generators and storage units is essential for accurate RA assessments, but simplified dispatch models can be used as screening tools (e.g., to identify critical hours) due to their low computational complexity.
- Multi-year data is important to include to capture inter-annual variations in system conditions.
- Neglecting to incorporate transmission limits into RA assessment could lead to substantial underestimation of traditional "expected value" RA metrics. However, experiments with alternate metrics that focus on individual event characteristics show that neglecting transmission limits will not mask the most critical shortfall events.
- New RA metrics that capture event-specific shortfall characteristics should be used as supplements to traditional metrics to better capture the impacts of different modeling assumptions on RA outcomes, as well as better describe the ability of the system to prevent specific high-impact shortfalls.

This study has been motivated by the consensus that current RA assessments are inadequate for modern power systems, and with the goal of prioritizing ways to improve them. We believe that our findings provide several useful insights to stakeholders to support decision-making in RA assessments. Future work can consider incorporating the operations of energy-constrained resources (e.g., hydroelectric resources) and long-duration storage to expand the capabilities of our technical framework. In addition, thermal generator failures that may lead to low-probability, but high-impact shortfalls can be included in the economic dispatch model, which can potentially bring resilience analysis into RA modeling. Finally, our framework can also be deployed alongside capacity expansion models that enable planners to optimize their least-cost resource portfolios while maintaining robust RA levels. The developed power system operation model may have a promising capability to support decision-making in planning processes, such as the capacity accreditation of variable renewable resources or the location-sizing problem for battery storage investments.

5. Impact of changes in resource adequacy assessments on planning processes

This report has established that traditional resource adequacy assumptions and practices need to evolve as electric power systems shift to higher levels of variable renewable energy resources and energy-limited storage resources while attempting to remain reliable and resilient against climate change. RA assessments are usually embedded within broader planning processes developed by utilities and system operators (Figure 5.1). In turn, planning processes are part of a set of regulatory and/or market designs that support procurement practices to ensure that the power system remains affordable, reliable, resilient, and sustainable. The mutual dependencies that arise from these relationships prompt the question: how would technical changes in resource adequacy assessments – including metrics, data, models, and methods – ripple through and require changes in planning processes and broader institutional contexts.

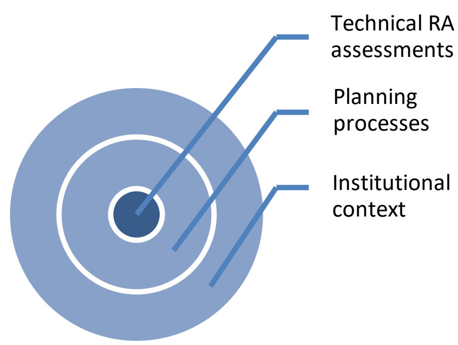

Figure 5.1 Process relationship between resource adequacy assessments, planning, and broader institutional context

This section explores how changing RA standards and practices may affect the planning practices that are used to define investments in power systems, with a focus on integrated resource planning (IRP) developed by regulated utilities and transmission planning developed by RTOs and ISOs. Resource adequacy is an integral part of the IRP process (Carvallo et al., 2021) as well as a key responsibility for ISOs and RTOs.

- IRP: resource planners develop IRP to identify the least-cost resource portfolio to meet future loads. If utility IRPs select preferred portfolios that do not meet basic reliability standards and expectations, then IRP as a process would be fundamentally flawed as a tool for procurement and capital investments. In contrast, if utilities exceed reliability standards, they may not be procuring a least-cost system, breaching a core tenet of IRP. The tension between reliability and affordability makes RA and portfolio optimization two key components of IRP.
- ISO/RTO: these entities are mandated to diagnose, track, and ensure that the bulk power system they operate meets certain reliability objectives. One of the main components of these mandates are the transmission planning processes that these entities develop, but they also conduct resource adequacy assessments that are needed to ensure reliability. Many ISOs and

RTOs have implemented capacity market constructs to translate their RA assessments and planning process outcomes into procurement of firm resources for adequacy.

The challenge for resource and transmission planners (and in some cases, their regulators) is how to update the RA assessment that feeds into the planning process given the current state of knowledge with an industry in transition over its RA criteria. Many of the existing RA metrics and processes are well known and understood by the stakeholders in IRP and transmission planning processes. Moving to new and more complex metrics and modeling tools may be challenging to most of the participants in these processes given its ripple effects across many components of the planning process. To address this issue, this section proposes a roadmap of evolving industry standards for resource adequacy assessments in resource and transmission planning (Table 5.1). Planners, state regulators, and other stakeholders would match their entity with a benchmark for each resource adequacy component and identify potential improvements to their RA assessment practices.

The section begins with summaries of a reviews of how six different investor owned utilities (IOUs) are performing resource adequacy analyses within their most recent IRPs, complemented by a summary of four ISO and regional planning entities that are developing RA assessments to support their transmission planning processes.

5.1 Summary of review of LSE Approaches to RA in IRPs

This subsection reports on the RA assessment approaches implicit in the IRPs filed by six load-serving entities (LSEs) located in the Pacific, Mountain West, and Southeast regions of the U.S. We selected entities that do not operate under an ISO/RTO, such that they fully depend on their resource adequacy assessments for firm capacity resource selection. The collected data is summarized following our framework, reporting the metrics and targets employed and the general modeling approach – data, methods, and models (see Table 5.1). The detailed reviews of each entity are reported in Appendix C.1.

Table 5.1 Overview of utility RA metrics, targets, and models for six IRPs

Entity / IRP Year	Metrics and targets	Models and modeling approach
PacifiCorp (PAC)/ 2021 IRP	Planning portfolios evaluated for a risk-adjusted present value of revenue requirement (PVRR). Adopted the capacity factor approximation method (CF Method) to derive the capacity contributions of resources	PAC used 3 components of the Plexos model: (1) *Capacity expansion model* to evaluate resource portfolios (Long-Term planning model); (2) a *stochastic model* that performs analysis for periods of unserved load (Medium-Term model); and (3) a *production cost model* with hourly granularity (the Medium Term model). Modeling process begins with the capacity expansion model to create planning portfolios. The stochastic model and production cost model identifies periods of unserved load and iterates on best reliable portfolios.
Puget Sound Energy (PSE) 2021 IRP	Adopted 5% loss of load probability (LOLP) metric consistent with the NWPCC. PSE derives multiple metrics	PSE used three modeling tools: (1) PSE's Resource Adequacy Model (RAM); (2) the Genesys regional model developed by the NWPCC and BPA; and (3) the Wholesale Purchase Curtailment Model (WPCM) developed by PSE. PSE starts with the Genesys

	including: the Loss of Load Probability (LOLP), Expected Unserved Energy (EUE), Loss of Load Hours (LOLH), Loss of Load Expectation (LOLE), and Loss of Load Events (LOLEv).	model to develop a probabilistic regional resource adequacy assessment. PSE uses its RAM model to assess the physical supply risks over time, guide peak load planning standards, and quantify the peak capacity contributions of renewable and energy-limited resources. PSE uses WPCM is used to identify opportunities for relying on regional transactions.
Portland General Electric (PGE) 2019 IRP	Target Loss of Load Expectation (LOLE) of no more than 2.4 hours per year, or one day in 10 years	PGE used a suite of models for reliability assessment including: (1) Aurora production cost model by Energy Exemplar; (2) Renewable Energy Capacity Planning model (RECAP) developed by Energy and Environmental Economics; (3) Lucas which is a revenue requirement model that estimates the levelized fixed cost of existing and new resources; (4) the ROM production cost model which contains a more granular data on PGE resource performance and multi-stage optimal unit commitment and used to assess resource flexibility value; and (5) ROSE-E which is an optimal capacity expansion model that produces optimal resource portfolios for specific objectives. RECAP calculates the loss of load probability (LOLP) for each month, day type, and hour for a test year, and the capacity needed to achieve a reliability target and the marginal capacity contributions for additional resources. The Aurora production cost model is used to simulate market prices across the WECC region for the relevant time period. Resource portfolios are refined with the use of the other models to ensure least cost production and attainment of policy constraints.
Tucson Electric Power (TEP) 2020 IRP	System flexibility to meet ramping capability with targeted higher levels of renewable generation	TEP performed a flexibility study in its 2020 IRP. The flexibility study developed six scenarios with a range of renewable energy penetration levels from 28%, 35%, and 50% and varying mixes of solar and wind. The study performed Monte Carlo statistical analysis with 250 iterations of net load over one year to determine the 99th percentile values of peak net load for 3-hour net load ramps, 10-minute net load ramps, and curtailment levels of renewable generation. An additional analysis used the Aurora production cost model to simulate loads, renewable generation, and dispatch on a 10-minute interval for one year. The model output was analyzed for flexibility shortfalls across the different portfolios.
Public Service Company of Colorado - Xcel Energy (PSCo) ERP 2021	LOLE of 0.1 standard, with corresponding reserve margin 17.4%.	PSCo used the EnCompass model to develop initial resource portfolios with various reliability requirements. The modeling output was reviewed and prompted adjustments that would be incorporated into another round of modeling. This iterative process effectively enforces numerous reliability requirements such as reserve requirements, flex reserve requirements, and ELCC requirements. PSCo's resource plan included portfolios designed to meet carbon reduction target of 80% by 2030 from

		2005 levels.
Duke Energy Carolinas (DEC) IRP 2021	LOLE of 0.1 standard, with corresponding reserve margin 16.75 % for "combined case"	DEC retained Astrape Consulting to perform the resource adequacy analysis for its 2020 IRP using the SERVM model. DEC agreed to adopt the one day in 10 year LOLE standard. Astrape performed probabilistic modeling to derive a planning reserve margin under several different assumptions and cases. Astrape and DEC agreed to a 17.0 % target planning reserve margin for IRP purposes. DEC initially released an IRP that proposed six portfolios that would lower carbon emissions over the 15-year time horizon of the IRP. The North Carolina Public Utility Commission ordered DEC to modify the IRP with different modeling assumptions and nine supplemental portfolios. DEC's modified IRP was released in December 2020.

The case studies of Load Serving Entities (LSEs) summarized in Table 5.1 show that a majority of the LSEs examined here have adopted reliability metrics and targets with a primary focus on the LOLE, with a couple of cases using the LOLP. Many of these LSEs build upon this metric to derive a corresponding equivalent planning reserve margin metric. Puget Sound Energy has taken the additional step to develop and report multiple reliability metrics including Expected Unserved Energy (EUE), Loss of Load Hours (LOLH), Loss of Load Expectation (LOLE), and Loss of Load Events (LOLEv).

These LSEs have adopted complex modeling tools to ensure future portfolios meet more resource adequacy targets. The LSEs typically have a suite of models that generally include: (1) a capacity expansion model that identifies the least cost portfolio of new resource additions; (2) probabilistic modeling tools that generate a large number of runs for stochastic analysis to better plan for risk; and (3) production cost models that model every hour in a year to ensure sequential operability and system interdependencies in the power system. These models are used sequentially and, in some cases, iteratively, to derive a preferred least cost portfolio that meets reliability criteria.

Unique regional variations have prompted some variation among the modeling needs of utilities. In the Pacific Northwest, Puget Sound Energy and Portland General utilize granular modeling tools to inform their practice of using market transactions in a region with abundant hydro resources. In the southwest, Tucson Electric Power investigated the ramping capability of its system to meet large ramping capability anticipated in a region with excellent solar resources.

5.2 Survey of RTO and ISO approaches to RA in transmission planning

This subsection reports a summary of RA practices for three ISOs: MISO, ISO-NE, and PJM, which were chosen due to their deployment of capacity markets as a firm capacity procurement mechanism. In addition, we survey the Northwest Power and Conservation Council, a non-ISO entity that nonetheless develops a sophisticated resource adequacy assessment of the Western Interconnection with a special focus on hydropower. The collected data is summarized following our framework, reporting the metrics and targets employed and the general modeling approach – data, methods, and models (see Table 5.2). The detailed reviews of each entity are reported in Appendix C.2.

Table 5.2 Overview of RTO, ISO and other regional approaches to RA in transmission planning

Entity/Year	Metrics and targets	Models and modeling approach
Midcontinent Independent System Operator (MISO) Regional Resource Assessment 2022	Adopted LOLE target for the MISO region: 0.1 days/year.	MISO's RA analysis begins with a Regional Resource Assessment (RRA) that projects the region's generation portfolio over a 20-year time horizon to meet member plans and state policy objectives. Used a capacity expansion model (Electric Generation Expansion Analysis System) to find the least-cost resource mix that meets renewable and carbon targets under various resource forecasts, examines the flexibility of the regional system, and ensures the build out meets the established LOLE requirement. MISO used Plexos to review the potential impacts of changes in the planning reserve margin requirement and capacity contributions of wind, solar, solar plus storage, and battery and energy storage systems. The analysis requires a set of inputs, including load forecasts and renewable energy production profiles created for different historical years under the pre-defined electrification and weather scenarios (MISO Futures).
ISO New England (ISO-NE) Regional System Plan 2021	Average of 0.1 days/year LOLE	ISO-NE produces a Regional System Plan once every 3 years. This analysis produces a comprehensive ten-year peak and energy forecast and determines whether the region has enough capacity to satisfy the resource adequacy requirements. If the system has shortfalls, it identifies generation and transmission plans to address any identified needs. The forecast of peak and annual electricity consumption was projected using gross energy modeling and gross demand modeling tools (both are regression models constructed using weather, economic, and time variables) with forecast inputs related to the economy and weather. The system-wide and local-area capacity needs were calculated using the General Electric Multi-Area Reliability Simulation Program (GE MARS), which conducts a chronological Monte Carlo simulation to incorporate uncertainties in loads and resources under a wide range of existing and future system conditions
PJM RA Studies	Average of 1 day every 10 years (1 day every 25 years for locational assessments); 2022 PRM of 14.7%	PJM conducts resource adequacy planning studies to determine the capacity resources needed to serve forecasted loads while satisfying reliability criteria. The planning study consists of three components. First, the Generator Availability Data System (GADS), provides information on the availability of generating units over the past five years and identifies the parameters for generation forecasting. Second, load forecasting over the next 15 years at the monthly level generated using multiple regression analysis of the hourly metered

		load data. Third, the installed reserve margin (IRM) analysis, which determines whether PJM has enough generation capacity to serve forecasted peak and IRM while satisfying the reliability criteria. PJM used the Probability Reliability Index Study Model (PRISM) to calculate PJM's reserve requirement while incorporating uncertainties from generation performance, load forecasting, and connections to its adjacent regions. The final reserve margin value calculating in the planning study serves as an input to the Reliability Pricing Model (RPM), which ensures long-term resource adequacy while benefitting from competitive pricing. RPM determines the plans to procure the region's unforced capacity obligation through market mechanisms.
Northwest Power and Conservation Council (NWPCC) 2021 Power Plan	Annual LOLP of 5% or less as a standard to determine resource adequacy	The resource adequacy assessment process consists of three steps. First, a range of electricity demand forecasts were developed based on the predetermined sets of key economic assumptions and weather scenarios. council's demand forecasting system, Energy 2020, was used to develop the load scenarios under various economic and temperature conditions and then aggregated up to the regional level The council uses three models to inform decisions: (1) Aurora, which estimates the electricity prices and projects the buildout of the western grid; (2)the Genesys model that dispatches generation resources while incorporating uncertainties in generation resources and market dynamics and assesses the adequacy of regional power supply; and (3) the Regional Portfolio Model that determines regional capacity expansion decisions.

The ISO/RTO and other regional approaches to RA are technically similar in many respects to LSE RA analysis, alas with some important differences. ISO/RTO plan for a region-wide footprint that is typically larger than individual LSEs. LSE planning typically involves procurement of generation resources. By contrast, ISO/RTO do not procure generation resources but do build the transmission system and manage market operations across their respective footprint. ISO/RTO emphasize transmission aspects of resource adequacy – especially deliverability – significantly more than LSEs conducting IRP.

The metrics and RA targets adopted by ISO/RTO are similar to the LSE metrics and targets. RTOs and ISOs perform RA modeling to assess the region's future demand for electricity, future resource additions, and the corresponding impact on the transmission system. The modeling approach employed is similar to that of LSE, with a similar mix of users of commercially available tools and proprietary developments. An expected difference is that ISO/RTO do need to translate their RA modeling outcome into a capacity reserve margin to be procured via capacity markets or bilateral arrangements; IRPs translate their RA needs into capacity that becomes part of the preferred portfolio.

5.3 Roadmap for resource adequacy assessments in planning processes

This section proposes a roadmap of evolving industry standards for resource adequacy assessments in resource planning and transmission planning. The roadmap is based on three benchmarks applicable to key components of RA assessments:

- The first benchmark identifies the basic **minimum** of essential steps that entities need to implement in a reliability assessment. Failing to meet this minimum standard implies that the steps taken constitute an insufficient approach to performing reliability assessments given the emerging challenges in decarbonized power systems.
- The second proposed benchmark corresponds to **current best practices** in the industry. This standard is what planners should be implementing or striving to implement to ensure that RA assessment practices keep up with trends in electricity systems and incorporate the most accurate data, methods, tools, and results reporting available.
- The third proposed benchmark adopts a forward-looking perspective to identify RA assessment practices that no or few entities are currently implementing, but that they may need to pursue in the future as electricity systems continue to evolve. We refer to these as **frontier practices**.

Table 5.3 below summarizes the key findings of our proposed roadmap. State regulators and planners can match their existing practices to one of three benchmarks for each resource adequacy component and consider potential improvements to the RA assessments that are embedded in their planning processes. Most topics in the table apply to both IRP and transmission planning processes, although we identify the few cases in which the topic applies more specifically to IRP.

Table 5.3 Roadmap to incorporate best practices for RA assessment into planning processes

Components of RA Framework	Planning Element	Benchmarks Minimum Practice	Current Best Practice	Frontier Practice
Definition of RA	Temporal resolution for RA	Meet load in a fraction of the top peak net load hours of the year	Meet load on a chronological hourly basis	Sub-hourly analysis to meet load and ramping requirements
RA metrics and targets	RA metrics and targets	Single metric (e.g. planning reserve margin) driven by a maximum LOLP (not by the 1-in-10 rule of thumb)	Develop and explore multiple metrics produced by stochastic models that track shortfall magnitudes, frequencies, and durations	Use multiple metrics that track magnitudes, frequencies, and durations; consider full probability distributions of metrics and economic metrics
Data	Weather data	A few years of historical weather data with daily maximums/minimums	Several decades of historical weather data with variables at an hourly temporal resolution	Combine historical data with climate model data for forward-looking hourly weather forecasts
Data	Load forecasting for resource adequacy	Rely on several years of historical load data	Develop econometric or engineering-based load models that explicitly capture the dependence of load on weather	Pair weather-sensitive load models with forward-looking climate change-based weather patterns
Data	VRE characterization	Historical wind/solar performance for several years	Forward-looking wind/solar data for new sites, informed by historical empirical profiles	Climate change-induced wind/solar profiles based on downscaled climate model output
Models	Transmission and market transactions	Basic modeling of firm capacity and available exchanges	Regional simulation to accurately account for the availability of imported resources and market depth uncertainty; locational reliability analysis.	Enhanced modeling of transmission line derates; strengthen integration between generation and transmission expansion
Models	RA modeling and integration with planning process	Basic chronological Monte Carlo LOLP analysis; simplified storage representation	Iterative LOLP-CEM approach; model chronological storage operations	Stochastic CEM that internally assesses and ensures RA; include unit commitment and operational details
Procurement	Capacity credit	ELCC for renewables	ELCC for all resources, analyzed from individual perspectives	Energy adequacy analysis; portfolio-based ELCC accounting for interactive effects

5.3.1 Temporal resolution for RA

As a foundational premise, resource adequacy is the ability of an electric LSE to meet its loads with a reasonable tolerance for risk by using its available resources. The practical interpretation of "meeting load" in resource adequacy has evolved over time. Early approaches focused on measuring the ability of the system to meet a single peak load. The peak load periods typically occurred in either the hot summer months or the cold winter periods, with a few dual-peaking utilities developing RA analyses for both seasons. Electric utility resources in this era were predominantly dispatchable thermal resources and hydropower. If the utility had sufficient resources to meet the seasonal peak plus a reasonable reserve margin, then it was assumed to be capable of meeting loads throughout the rest of the year. IRPs in the early 2000s are examples of this single peak load approach to resource adequacy.

With the shift towards higher levels of renewable solar and wind, and storage resources, electric utilities can no longer assume that the ability to satisfy the peak load implies RA during all other times of the year. There is a fundamental limitation to any RA assessment that omits any hours over the year. At a **minimum**, a utility should identify a subset of "stress" hours for the system and include those hours in its RA assessment. The typical implementation of this approach continues to focus on peak load, but covering a fraction of the top load or net load hours. Entities that focus their assessments on a subset of hours may also want to develop methods to identify critical hours of the year that are not necessarily the top load hours. For example, including summer evening hours in the assessment would capture periods with rapidly declining solar PV generation and the RA challenges that they present.

The minimum standard will probably yield reasonable results with relatively low penetration of VRE, but it will be insufficient when the system's critical hours start depending on other variables that are related to load. A preferred approach that we consider a **best practice** for the industry requires that electric utilities plan to utilize their full portfolios of renewable, storage, thermal, hydro, and demand-side resources to meet loads for every hour over the year. A full 8,760-hour analysis helps capture the complex interdependency of the diurnal cycle, seasonal weather patterns, and other factors that influence resource adequacy. All the utilities in our sample survey evaluate resource adequacy over the 8,760 hours in a year. Furthermore, the technical advisory committee unanimously agreed that there is no technical reason for an RA assessment to ignore any hour of the year.

We envision that a potential **frontier** practice, motivated by very high penetration of renewable generation and storage resources, would be to shift from hourly modeling to sub-hourly modeling such as 15-minute or 5-minute intervals. This is partly due to demonstrated swings in renewable production due to wind lulls and clouds covering PV arrays, whose effects are exacerbated when these resources account for a large fraction of the overall mix. The movement to sub-hourly modeling could also be motivated by electricity markets that establish sub-hourly trading or operational commitments – including ancillary services markets – and the need to reflect these commitments when representing the state of the system for resource adequacy purposes. Tucson Electric Power has been running production cost modeling with sub-hourly intervals to better understand the system flexibility and the need to cope with high ramping conditions with increased levels of renewables.

The foundational principle of resource adequacy is not changing. Planners need to evaluate whether there are sufficient resources to meet loads with a reasonable tolerance for risk. What is changing are the metrics, methodologies, and tools used to perform RA assessment. This section proceeds to examine the emerging targets, timeframes, data, and modeling tools that are evolving given the changing nature of the power system.

5.3.2 RA metrics and targets

Resource adequacy analysis examines the risk of shortfalls in a given power system. Utilities can generally reduce the risk of a shortfall by adding more resources, which begs the question: for consumers and regulators, what is the acceptable level of risk of shortfalls?

Historically, electric utilities adopted the business practice that a power system should be built to have a loss of load event with a frequency of no more than one day every 10 years. This standard has been translated into a loss of load expectation (LOLE) of 0.1 day per year, although this translation does not express the same level of risk (Stephen et al., 2022). The capacity needs resulting from meeting this standard were typically converted into a planning reserve margin that was used as an input to a capacity expansion model. In many cases, utility regulators in some states have adopted regulations mandating a specified reserve margin for utilities under their jurisdiction. In jurisdictions that have historically set planning reserve margins administratively or by statute, these benchmarks could be re-examined with the benefit of better information on consumer preferences and probabilistic analysis.

As a **minimum** industry practice, utilities should adopt planning reserve margins that are grounded upon and linked to a loss of load probability analysis, rather than administratively set margins that do not reflect the current state of the power system. Probabilistic modeling can establish a relationship between a loss of load probability and a utility's planning reserve margin. For example, a utility could have a portfolio of resources that provides a 15% planning reserve margin that also equates to a 0.1 LOLE based upon the probabilistic modeling. Participants in an IRP planning process may prefer to rely on the planning reserve margin instead of the loss of load probability, or the reverse.

There is a growing recognition that the impact of a given reliability event can vary significantly based on the magnitude of the outage across a system, the duration of the outage, the frequency at which such events occur, and the timing of the outage. LSEs could pursue survey techniques to better understand the impacts of outages on customers depending on their magnitudes, durations, frequencies, and timing. A **best practice** is then to maintain a single-metric-based RA standard – for example, a frequency-based metric such as the 1-day-in-10-years rule – to inform investments, but to track and report the system's resource adequacy using multiple metrics. Seattle City Light and Tacoma Power have both calculated and reported multiple metrics after surveying their customers for tolerance ranges for outages.

We foresee that the **frontier** standard will be for LSEs to adopt multiple reliability metrics that track the magnitudes, frequencies, and durations of potential outages and *integrate* them to make adequacy investment decisions. Expanding the number and types of RA metrics would necessitate several key adjustments to the planning process. First, suitable standards would need to be defined for RA metrics that have not been used in decision-making yet, such as the LOLH and EUE. Second, there is no precedent for how a capacity expansion model would incorporate multiple metrics for investment decisions. One possibility would be to define planning reserve margins based on multiple standards (e.g., a LOLP, a LOLH, and an EUE standard) and then use the most stringent margin as an input to the investment decision process. In turn, this would require that ELCC calculations to determine portfolio contributions to adequacy also be developed for each of the standards. The exercise would then choose a portfolio that meets all standards, although it is possible that one or more of them will be exceeded in the process of meeting the others. Further refinements to this process may allow planners to fine-tune additions of specific resources that may affect one standard more than another may.

Two additional frontier developments in RA metrics would involve replacing simple metrics with entire probability distributions of outcomes and incorporating economically-driven metrics in addition to current technical reliability standards. The use of entire distributions for frequencies and durations of events offers similar challenges to those for the integration of multiple metrics, namely the need to establish a standard and a process to decide investments based on how they affect the shape of the distribution. As indicated in Section 2, economic criteria to determine optimal adequacy levels and inform assessments are desirable, but technically challenging. However, the adoption of advanced metering infrastructure and mechanisms for direct control could support "disconnection" bidding programs that would produce economic signals for the value of lost load of participating customers. Customers may be able to bid into these disconnection schemes on an event basis, revealing their willingness to be paid for losing load and the overall economic value of reliability for the power system.

5.3.3 Weather data

Weather is a key data input to resource adequacy analyses. Seasonal high and low air temperatures affect peak loads in the electricity system and load-serving entities have traditionally calculated weather-adjusted peak load to isolate the variability of weather from the long-term load trend. In resource adequacy assessments, temperature extremes can reduce the efficiencies of thermal generators and affect the availability of natural gas due to higher space heating use and, in extreme cases, freezing of pipelines. Seasonal rain and snow levels drive the amount of water captured by reservoirs and determine the availability of hydro generation for a year. Air temperature affects the efficiency and operational capacity of transmission lines. Finally, weather can directly affect solar and wind generation levels and their spatiotemporal patterns.

The most important characteristic of weather is that it can simultaneously affect loads, transmission, distribution, and generation. Section 3 highlighted the relevance of capturing the correlated or joint probability outcomes of all these power system variables that depend on the same weather patterns. Subsection 5.3.7 provides more details about the modeling approaches needed to capture these dependencies, and subsections 5.3.4 and 5.3.5 report on how load and renewable generation models can be designed to capture weather's influence. Underlying these models are the data needs to represent weather in RA assessment in an appropriate manner.

We find that, at a **minimum**, planners should use multiple years of historical weather data to inform probability distributions used in their resource adequacy assessments. Planners typically employ p50 weather to represent median load conditions for portfolio analysis, but they should ensure that their RA assessment includes at least p95 weather conditions or beyond. Basing adequacy analysis on a single year or few years of weather data is insufficient to capture enough variability. At a minimum, the temporal resolution of the weather data should reflect daily minimums or maximums, depending on the season.

A **best practice** is to employ several years to decades of historical weather data to capture a wide range of possible realizations of weather. Temporal resolution of this data should be at the hourly level to match the resolution used in RA assessments that model chronological power system operations. The NWPCC, for example, uses 80 years of historical hydrological data to characterize inflows into its basins – capturing the less frequent but critical drought years – and 88 years of historical temperatures to capture the effect on load. There is no set guideline for what period length is needed to reflect extreme weather conditions that may induce very low generation or very high load levels. The best practice has tended to collect as much weather data as possible and ideally for different weather stations when a planning entity's footprint is large.

Climate change will cause future weather to deviate from historical patterns, and possibly exacerbate the frequencies and intensities of extreme weather events. These extreme weather events are the types of events that a resource adequacy assessment should strive to capture. A **frontier practice** may include forward-looking forecasts of weather patterns that are responsive to an array of possible climate change realizations. These models use large amounts of historical weather data and couple it with the long-term trends implied in downscaled climate change projections to forecast how weather patterns may change over 10-30 year periods. The historical data should be carefully weighted to avoid reproducing decades-old weather patterns that will shift due to climate change. A model recently developed by Sandia National Laboratories – the Multiscenario Extreme Weather Simulator (MEWS[7]) – can produce these types of long-term weather pattern forecasts.

Expansive weather data coupled with appropriate load and generation models (see 5.3.4 and 5.3.5) may benefit the overall IRP process beyond its resource adequacy applications. Characterization of renewable resource production is very important to resource selection in IRP portfolios. A large range of sites and hourly production profiles would allow a more accurate valuation of VRE resources for least-cost planning purposes.

5.3.4 Load forecasting for resource adequacy

A foundational component of any RA analysis is the development of a load forecast. The load forecast determines the expected quantity of electricity demanded by the consumers of an LSE for its planning period, typically between five to 40 years into the future. Load forecasts are typically disaggregated by customer class (e.g., residential, commercial, and industrial) and account for seasonal and daily patterns that influence the demand for electricity differently among customer classes. The aggregate load forecast drives the need for resources to meet future loads.

As indicated in Section 5.3.3, all resource adequacy variables depend in one way or another on weather. A weather-sensitive (or weather-responsive) load model produces estimates of hourly

[7] See https://www.osti.gov/servlets/purl/1885888

consumption as a function of several underlying variables, including weather. These types of models are able to translate changes in temperature, humidity, and other meteorological variables into changes in load based on physical and behavioral responses. When weather-sensitive load models are not available, a **minimum** practice is to rely on several years of historical load data that implicitly capture seasonal and annual temperature sensitivity, as well as long-term trends.

The current **best practice** is to develop econometric or engineering-based load models that explicitly capture the sensitivity of load to weather. For example, these models would represent weather-sensitive end uses such as space heating, air conditioning, and water heating, in addition to non-weather-sensitive load. In a future with deep electrification of end uses, these load models will be relevant to represent electric vehicle and heat pump load accurately, especially given the latter's temperature and humidity sensitivity.

Pairing these weather-sensitive models with forward-looking climate change-based weather patterns would combine best practices in weather data (see Section 5.3.3) with best practices in load models for resource adequacy, resulting in a **frontier practice** for RA load forecasting. This frontier practice is consistent with the application of "deep uncertainty" principles for risk management under climate change. These principles suggest developing multiple load models based on several downscaled versions of Intergovernmental Panel on Climate Change (IPCC) climate models. In this context, RA assessments may evolve to include different load models that reflect weather sensitivity and weather probability distributions, with multiple chronologically correlated Monte Carlo analyses. The WIEB-Stanford project surveyed numerous planning entities to identify whether they had adopted deep uncertainty approaches to load forecasting that included use of climate model forecasts and consideration of multiple scenarios. The researchers found that very few planning entities in the West use this type of uncertainty analysis (Hofgard and Savage, 2022).

In addition to weather-sensitive load models, load forecasts may need to improve their representations of distributed energy resources (DER) including rooftop PV and battery storage. Currently, most planning entities treat demand-side resources as negative load (thus lowering net load) for resource adequacy purposes. Their RA assessments continue to reflect these resources in net load by subtracting a deterministic 8,760 realization from their hourly load forecast. A best practice for handling DER for RA assessment recognizes that the probability distribution of generation from these resources can be very different from that of load, especially when it is influenced by weather patterns. As DER adoption grows, it will be necessary to move away from netting DER from the baseline load forecast and instead treat DER generation as a separate variable with its own characterization for resource adequacy purposes. This distinction will allow assessments to explicitly capture the capacity contributions of DER, leading to a more accurate analysis of system needs.

5.3.5 VRE characterization

Resource planners need accurate information on the expected hourly generation of their fleet of wind and solar resources when developing preferred portfolios and assessing resource adequacy. Wind and solar resource capacity contributions depend on their spatial locations as well as the production profiles of existing assets, and their interactions become more important as renewable penetration increases.

Given these conditions, it is insufficient to characterize VRE based on a single year and single location for RA purposes. Proper VRE representation must (i) be responsive to changing weather patterns, (ii) represent a wide range of sites within a planning entity's footprint, (iii) include many years of weather-driven VRE production, and (iv) reflect technological changes in converting wind and solar radiation into electricity. The first component was already addressed in Section 5.3.4; this subsection expands on practices for the remaining three components.

A **minimum** industry practice would be for planners to use many years of actual historical renewable generation to perform sophisticated probabilistic modeling of solar and wind generators for every hour of a year. Existing units should be represented by their specific historical hourly generation patterns; new resources may reuse some of these patterns for simplicity. However, the use of historical data entails limitations. Historical wind and solar production do not reflect technological change should the units deployed in these sites be upgraded to the latest available technology. In addition, historical renewable production data typically encompass a relatively short period from a probabilistic perspective given that wind and solar resources have been developed in earnest only in the last 15 to 20 years. Meteorological analyses by the National Oceanic and Atmospheric Administration (NOAA) typically use 90 to 100 years of weather data to properly characterize meteorological variables and capture decadal events and trends.

A **best practice** trend that improves upon the use of historical data is to develop simulated or synthetic profiles for wind and solar generation that combine historical performance with technology representation. These simulated data would reflect relevant technological changes in PV (e.g., more efficient cells and tracking mechanisms) and wind turbines (e.g., greater height and longer blade diameter) that affect their production profiles.

Historical weather data may become less useful for predicting future weather conditions and renewable generation. VRE production models must then go beyond reflecting changes in technology but also

Collecting, curating, and managing weather datasets

A relevant factor for incorporating detailed weather data and weather-dependent VRE generation is the complexity of producing, accessing, and managing this data. Resource planners face a tremendous challenge to obtain relevant, granular weather data to develop the best forecasts of renewable generation for planning and RA purposes. Confidentiality concerns over data by developers tends to limit the broader use of this important information. One potential strategy to improve the weather data available to the industry is to expand the data collection and analysis activities of existing governmental entities such as the National Weather Service and the national laboratories that currently collect data related to renewable energy generation. NREL currently develops data on the performance of both solar and wind generators based on historical weather patterns. This work could be expanded and refined further with access to actual generator performance and very granular geographic locations.

A second strategy for collecting and providing access to granular data on renewable generation is to shift the responsibility for data collection and analysis from the local planning entity to a regional planning entity such as an RTO, an ERO Enterprise regional entity, or a regional planning entity. The regional entity could collect the required weather data and perform the appropriate analyses that could then be made available to member LSEs. LSE planners could use this information for developing individual IRPs and resource adequacy assessments. This strategy would lower the cost of data collection and analysis by spreading the costs across more entities and lowering unit costs. The challenge to implement this approach would be to overcome proprietary data concerns to ensure better data for the collective of participating LSEs in the region.

reflect underlying changes in weather patterns that affect the resources harnessed by these technologies. A **frontier** area for resource adequacy is to capture how climate change will affect renewable generation levels and profiles. The development of forward-looking solar and wind generation profiles requires the development of highly granular, downscaled weather pattern generators consistent with one or more IPCC-supported climate realizations. This resource potential would be paired with credible estimates for technological improvement in harnessing wind and solar radiation that may change the generation profiles and capacity credits of these resources.

5.3.6 Transmission and market transactions

Individual utilities that want to add resources to their portfolios to meet future loads have the option to build and own generation resources or rely on market transactions for delivery of power for a specified period. Some entities prefer the conservative approach and own all or most of their generation resources. Other utilities are comfortable with the risk associated with market transactions as part of their resource portfolio. RTOs/ISOs manage by definition interconnected systems that require explicit representations of transmission systems. A necessary condition to rely on market transactions for both utilities and regional organizations is the availability of transmission to deliver power from the seller to the buyer.

In the Western Interconnection, there are no organized markets except for the California Independent System Operator (CAISO) and the Alberta Electric System Operator (AESO). Many utilities in the West rely on bilateral market transactions as a reliable resource to meet their resource adequacy requirements. In the Eastern Interconnection, by contrast, most power is transacted in organized markets. Even though bilateral transactions do exist, they are facilitated, coordinated, and made visible by the ISO. Regardless of their differences, both types of arrangements – vertically integrated utilities and organized markets – need to characterize and model market and bilateral transactions in their RA assessments. In this context, it is important to note that LSE planning and IRPs have traditionally omitted transmission investments from their RA assessments, assuming that all resources are deliverable within their service territories. This assumption may no longer hold as renewable resource development is driving the need for transmission expansion, and hence including transmission development in RA assessments will be relevant. Most RA assessments do not explicitly model transmission, but our research shows that ignoring transmission constraints leads to overestimation of system reliability.

A **minimum** industry practice would be for planners to perform basic modeling of firm capacity and potential exchanges available in bilateral or regional markets. One recent example of analyzing transmission as a resource is the 2021 IRP by Idaho Power. The resource adequacy assessment was developed for scenarios with and without considering the proposed Boardman to Hemingway (B2H) transmission project as well as with sensitivities related to the Gateway West transmission project. The analysis demonstrated that deploying the B2H transmission line reduced by five times the amount of firm capacity needed to maintain a 1-in-10 LOLE through the analysis period.

A current **best practice** to analyze the availability and value of transmission and market transactions is to perform regional simulations that include modeling transmission ties outside the planning entity's footprint. These simulations should model transmission lines as a resource with basic flow constraints and outage models, and additionally simulate the outages and availability of resources outside the footprint that are providing firm capacity to the planning entity. This practice applies to both RTOs/ISOs as well as LSEs, although the latter are especially more sensitive to proper modeling of neighboring resource availability given the much smaller size of their footprints compared to regional organizations. An example of this practice in the Western U.S. are the regional simulations performed by Puget Sound Energy (PSE). PSE uses these simulations to forecast and anticipate precise exchanges in the region and the availability of using short-term market transactions in their portfolio. PSE uses this information to enter transactions that were designed to identify opportunities to lower costs and improve reliability.

It is worth noting that it may not be efficient for a single load-serving entity to conduct a thorough regional analysis on its own. In general, it will be more efficient for a regional entity to perform transmission and market transaction analyses in the way that RTOs and ISOs do. This centralized role also ensures that any other entity in the regional footprint has access to the same high-quality analysis and that the assumptions used by entities involved in a market transaction are shared and consistent. In the case of PSE, a new entity that will run such a regional resource adequacy analysis is the Western Power Pool (WPP). The WPP established a contractual agreement among its members to pool resources and share resources if other members become unexpectedly short of resources to meet loads. The institutional formation of a regional resource adequacy program serves to improve information and accounting of resources, especially related to the transmission needs and market transactions involved in resource adequacy assessments.

A **frontier** practice may involve an enhanced representation of failure rates by modeling transmission lines with weather-dependent stochastic failures – rather than simplified or even heterogeneous Markov chains – to more realistically simulate line derating. For entities that develop integrated resource plans, a frontier practice should look to integrate more tightly capacity expansion decisions for generation and transmission, especially considering some of the modeling practices described in the next subsection.

5.3.7 RA modeling and integration with planning process

RA assessment models integrate all of the components and approaches that we have described in previous subsections. In practice, the requirements that we have outlined for input data condition the structure of the model. There are, however, two additional design components for the model itself that are relatively independent of the input data: (i) the fidelity with which the model represents the system operation and (ii) the integration between the RA model and resource planning portfolio selection.

In the past, RA would be analyzed through a convolution method that simply estimated the probability of a load shortfall by multiplying the individual and independent probabilities of component failures. Convolution methods like this are insufficient, as has been established in Section 3. The **minimum** standard for current resource adequacy models is based on the chronological modeling of system operations, integrating most of the “minimum” standards for temporal resolution, RA metrics, weather data, load and VRE models, and market transactions. This minimum standard employs a chronological Monte Carlo approach with simplified representation of the power system’s operational details. In most cases, line power flows are represented using simple transportation models, short-duration storage operation is not represented but rather assumed to be available with a certain capacity, and no other operational details are modeled in the assessment. From a planning perspective, RA deficiencies of the resource portfolio being analyzed will be corrected by simply assuming that natural gas peakers with "perfect capacity" can be deployed, with little additional analysis.

A **best practice** approach improves on both the modeling – especially the representation of the system's operational status – as well as the integration of RA assessment with the broader planning process. State-of-the-art models use a chronological Monte Carlo approach that includes a representation of system dispatch to reflect the status of energy-limited dispatchable resources such as energy storage and reservoir hydropower. The integration between the planning models and the RA assessment is much more rigorous. A best practice approach – like that developed by PSE – is to calculate the planning reserve margin required to meet a certain LOLP target and then use the ELCC for each type of resource to determine how they contribute to meeting this margin. The resulting portfolio is then analyzed through a probabilistic assessment, iterating back to the capacity expansion model to ensure that the portfolio remains least-cost.

A **frontier practice** should aim to improve the fidelity of the hourly – and possibly sub-hourly – representation of the state of the system to identify potential shortfalls and stress conditions. The actual response of power systems to contingencies depends on not only the capacities of available resources, but on the probabilities of finding them in certain states. This probability depends on several external factors that have been reviewed in this paper, in addition to economically driven operational decisions as outlined in Sections 3 and 4. Since RA assessments do not reflect the economics of the system, they currently cannot capture unit commitment decisions for generation units, transmission lines, and storage, and they cannot reflect the accuracy of day-ahead and hour-ahead forecasts that determine the potential use of operational reserves. Recent developments by the NWPCC aim to incorporate these features in RA assessments.

On the integration of resource adequacy assessments and resource planning, a frontier practice may be the development of a modeling tool that merges the functionality a traditional capacity expansion model with a probabilistic model that employs thousands of simulations to estimate the probability of meeting reliability targets. To date, we are not aware of a single commercial model that could perform both the capacity expansion function and the probabilistic reliability check[8]. This is due to the large computational burden of solving a stochastic capacity expansion model with all the scenario

[8] For an exploration on this topic in the academic literature, see Stephen, G., Kirschen, D., 2022. Enhanced Representations of Thermal Generator Outage Risk in Capacity Expansion Models, in: 2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). https://doi.org/10.1109/PMAPS53380.2022.9810560

combinations and granularity required by an IRP process, in addition to the complexity of formulating such a model.

Verification of RA model output
Resource adequacy assessments and resource planning decisions are fundamentally modeling exercises. As such, the outcomes of these processes may not necessarily reflect the reality of the system they are emulating. Carvallo et al. (2019) demonstrated that the outcomes of integrated resource plans do not necessarily match the procurement choices made in reality by the planning entities. Similarly, there is no process to validate that the RA performance of a system as determined through stochastic analysis reflects the performance of the actual power system in its real-time operation with a predetermined accuracy. The potential issue is that an RA assessment without an anchor to reality may render modeling results that do not correspond to actual system performance. It is possible that the actual system LOLP or LOLE differs from that estimated via modeling; that the representation of load and demand- and supply-side resources does not correspond to the way that resources behave in reality; or that the operational reserve margins and contingency measures are larger or smaller than needed to ensure certain reliability levels. We recommend that planning entities develop a method to track contingencies, “close calls,” use of reserves, or other events that reflect that the system was under stress. This data would be used to confirm what characterizes these events in terms of timing, location, correlation with other system variables (including weather), and so on, such that it provides support for the assumptions made in the RA assessment. Planners could also study several decades of power system performance to identify generation and transmission outages and other operational conditions that caused actual interruptions, load shedding events, or the need to deploy emergency measures.

5.3.8 Capacity credit

Capacity accreditation is not inherent to resource adequacy assessments, especially when these are performed following the best practices outlined in this section. We recommend avoiding the representation of the capacity contribution of a resource with a single, deterministic value, instead representing resources with probability distributions that depend on several variables. However, calculating the capacity credit for resources is very important for several planning and operational processes, including compensation for firm capacity contributions and demonstration of sufficient capacity to meet planning reserve margins, among others. As explained earlier, capacity accreditation is a critical process to connect RA outcomes with resource planning decisions in IRP given that these models are independent.

In the early days of renewable generation investment, most entities estimated the capacity contributions of wind and solar generators for RA purposes based on an average capacity factor for the peak hour of a year. For example, the hourly dependable capacity approach establishes the capacity contributions of renewable resources based on a very limited statistical treatment of the resources’ performance. As has been established above, this practice does not recognize synergies between load and renewable resource profiles, nor does it recognize the reduction in marginal capacity credit with increased adoption of renewable resources, especially solar (Mills and Wiser, 2015).

A **minimum** standard for determining capacity credit for renewable generation and storage resources is the Effective Load Carrying Capability (ELCC) metric. The ELCC metric can measure and compare the marginal contributions of renewable resources with the capacity contributions of thermal power plants and other types of resources. The ELCC measures the amount of load that can be added to a system with the addition of a resource while keeping the reliability of the system unchanged (ESIG, 2021). Use of the ELCC metric is very important to assess the declining marginal contributions of adding more of the same type of resource to a utility's portfolio, as well as potential synergistic impacts between resources such as solar and battery storage, and demand-side resources. Moreover, the ELCC can be useful for identifying how regional diversification of wind generation profiles can be leveraged to reduce variation in total wind generation across the whole system.

The **best practice** is to use the ELCC metric beyond just renewable generation and apply it to other resources as well. ELCC can be applied to fossil plants to account for temperature derates and seasonal hydropower; to transmission to account for weather derates; and to demand response resources based on their temporal profiles and probabilities of load reductions. In the future, with the growing use of probabilistic modeling of power systems, the ELCC will likely serve as the standard for comparisons across supply- and demand-side resources, including renewables, storage, thermal, and hydro resources. The broader use of ELCC for all resources would replace the historic practice of comparing resources based on the notion of the "perfect capacity" benchmark.

In the short-term, a **frontier practice** that is based on the ELCC should calculate ELCCs for portfolios of resources instead of individual resources, in order to capture their interactive effects. Past the short-term, as power systems evolve to incorporate larger amounts of VRE, storage, and flexible demand, estimating the ELCC of each individual resource will become increasingly difficult – not only computationally, but also conceptually. These trends make the interactions among resources and chronological system operations more important for determining system-level RA. At some point, RA will be a system-level property to such an extent that the resource-level ELCC, however estimated, may cease to be useful for understanding the contribution of a resource to improving the power system's RA. For instance, even if a sophisticated ELCC analysis suggests that a new gas power plant and a new battery storage unit would have the same ELCC, it is doubtful that these two resources are perfectly equivalent in terms of their implications for system-level RA. It could thus be problematic to treat them identically for purposes such as capacity accreditation and remuneration in a capacity market or constructing portfolios that meet desired RA standards in an IRP process. Addressing the gap between the system-level nature of RA and the need to perform capacity accreditation will require new innovation to develop promising long-term frontier practices. In a post-ELCC world, developers will still need a reasonable method to estimate the capacity payment revenue for their units and these payments will ideally be stable and predictable. This is an emerging area of research as the power system transitions into higher levels of renewable penetration.

Energy adequacy

A separate challenge relates to energy adequacy, or ensuring that there is enough energy to meet demand over a given (generally long) time period and accounting for stochastic fluctuations in fuel availability and VRE production. Energy adequacy for energy-constrained resources will become more relevant as penetration of these resources increases, and possibly due to more frequent and severe droughts that affect hydro energy availability. However, there are no widely accepted practices to evaluate the energy adequacy of systems that follow the same stochastic approach as in resource adequacy.

6. Summary of findings and additional research needs

This paper identifies and evaluates issues in traditional resource adequacy assessment practices, and how adjusting these practices may impact and depend on existing institutional arrangements for planning and procurement. The paper proposes a technical-institutional roadmap that would allow regulators in vertically-integrated jurisdictions and system planners and operators in restructured jurisdictions to revise resource adequacy practices across a range of components.

We compile a critical review of current RA assessment practices based on (1) interviews with RA practitioners and (2) a review of recent technical literature. We find that:

- RA may need to expand beyond capacity adequacy to ensure energy adequacy – relevant for energy-limited resources such as storage – as well as ancillary service adequacy (e.g. enough ramping-up and ramping-down capability in the system). There is general agreement to include energy adequacy jointly with capacity adequacy, but it is not clear whether other system needs' assessments should be performed within the RA assessment or as separate processes.
- All studies and interviewees agreed that basing RA assessments on the peak hour of the year or season, or on a few select top load hours, is insufficient as peak demand may no longer predict the times when the power system is most stressed. Chronological hourly simulations are the current best practice.
- Traditional metrics such as the Loss of Load Expectation (LOLE), Loss of Load Hours (LOLH), Loss of Load Events (LOLEv), and Loss of Load Probability (LOLP) are criticized due to their "expected-value" nature, their focus on a single characteristic of a shortfall, and the coarse spatial resolution in their typical applications.
- An important shortcoming of current resource adequacy practices is that the metrics and models used do not reflect economic criteria in system operation and loss of load. Observers agree, however, that introducing economic criteria to determine adequacy levels introduces significant challenges related to valuing the loss of load for different customers, seasons, and end uses.
- There is a need to improve representation of weather dependencies and weather data, attending to a number of shortcomings of current practices that may hinder appropriate RA assessments under high wind and solar futures and climate change.

We review planning and RA reports for several private and public entities that plan generation and/or transmission infrastructure in the continental U.S. to look for existing practices involving resilience assessments. We find no systematic treatment of the costs of extreme weather and other hazards, the benefits of resilience, and resilience metrics in planning analyses and no systematic treatment of resilience metrics, methods, and outcomes for resource adequacy purposes.

We examine integrated resource planning (IRP) reports as well as Independent System Operator and Regional Transmission Organization (ISO/RTO) RA assessments and use this information to propose a guide of evolving industry standards for resource adequacy assessments in resource planning and transmission planning. Reporting minimum, best, and frontier practices will allow regulators and planners to benchmark their current practices across a range of relevant dimensions in adequacy analysis.

In particular, the frontier practices reported in this paper support additional research to:

- Validate the need for sub-hourly analysis for more accurate characterization of system needs.
- Develop multiple reliability metrics from resource adequacy assessments, and harmoniously integrate these metrics in capacity procurement processes such as IRP, bilateral transactions, and capacity markets.
- Develop forward-looking climate change-based weather data that can be utilized with outage, load, and generation production models explicitly developed to account for the impact of weather on these variables.
- Study the importance of considering DER separate from load, and properly characterize its capacity contributions. In particular, this line of research would support the use of virtual power plants as capacity resources in organized markets.
- Integrate resilience-based planning into resource adequacy or as a complement to traditional adequacy analysis, including specific analysis of resilience scenarios to ensure a broad assessment of power system reliability and resilience.
- Refine the characterization of transmission in adequacy assessments, in particular including weather-sensitive dynamic line ratings and strengthening the treatment of transmission as a capacity resource.
- Continue exploring computational methods to merge or integrate capacity expansion modeling with stochastic resource adequacy assessments, and determine what level of sophistication in the representation of the operation of the power system is needed for accurate resource availability assessments.
- Continue developing portfolio-based capacity accreditation methods that provide predictable and stable revenue to developers to incentivize resource deployment for reliability purposes.

This paper should be useful to regulators, policy-makers, system planners, and analysts that require information on how the evolving power grid is prompting a need to review fundamental aspects of resource adequacy. The paper offers an accessible guide to support technical improvements in metrics, data, models, and practices; R&D developments; and regulatory and market reform.

7. References

ACER, 2020. Methodology for the European resource adequacy assessment (No. Annex I), ACER Decision on the ERAA methodology. Agency for the Cooperation of Energy Regulators.

Barrows, C., Preston, E., Staid, A., Stephen, G., Watson, J.-P., Bloom, A., Ehlen, A., Ikaheimo, J., Jorgenson, J., Krishnamurthy, D., Lau, J., McBennett, B., O'Connell, M., 2020. The IEEE Reliability Test System: A Proposed 2019 Update. IEEE Trans. Power Syst. 35, 119–127. https://doi.org/10.1109/TPWRS.2019.2925557

Black, J., Rojo, V., 2019. Long-Term Load Forecast Methodology Overview. ISO-NE.

Borges, C.L.T., Dias, J.A.S., 2017. A Model to Represent Correlated Time Series in Reliability Evaluation by Non-Sequential Monte Carlo Simulation. IEEE Trans. Power Syst. 32, 1511–1519. https://doi.org/10.1109/TPWRS.2016.2585619

Carvallo, J.P., Larsen, P.H., Sanstad, A.H., Goldman, C.A., 2018. Long term load forecasting accuracy in electric utility integrated resource planning. Energy Policy 119, 410–422. https://doi.org/10.1016/j.enpol.2018.04.060

Carvallo, J.P., Sanstad, A.H., Larsen, P.H., 2019a. Exploring the relationship between planning and procurement in western U.S. electric utilities. Energy 183, 4–15. https://doi.org/10.1016/j.energy.2019.06.122

Carvallo, J.P., Sanstad, A.H., Larsen, P.H., 2019b. Exploring the relationship between planning and procurement in western U.S. electric utilities. Energy 183, 4–15. https://doi.org/10.1016/j.energy.2019.06.122

Carvallo, J.P., Zhang, N., Leibowicz, B.D., Carr, T., Galbraith, M., Larsen, P.H., 2021. Implications of a regional resource adequacy program for utility integrated resource planning. Electr. J. 34, 106960. https://doi.org/10.1016/j.tej.2021.106960

Cebulla, F., Fichter, T., 2017. Merit order or unit-commitment: How does thermal power plant modeling affect storage demand in energy system models? Renew. Energy 105, 117–132. https://doi.org/10.1016/j.renene.2016.12.043

Chen, J., Rabiti, C., 2017. Synthetic wind speed scenarios generation for probabilistic analysis of hybrid energy systems. Energy 120, 507–517. https://doi.org/10.1016/j.energy.2016.11.103

Cigre, 2018. The future of reliability–Definition of reliability in light of new developments in various devices and services which offer customers and system operators new levels of flexibility.

DEC, 2021. 2020 Integrated Resource Plan Modified. Duke Energy Carolinas.

DEC, 2020. 2020 Integrated Resource Plan. Duke Energy Carolinas.

Draxl, C., Clifton, A., Hodge, B.-M., McCaa, J., 2015. The Wind Integration National Dataset (WIND) Toolkit. Appl. Energy 151, 355–366. https://doi.org/10.1016/j.apenergy.2015.03.121

EOP, 2013. Presidential Policy Directive 21: Critical Infrastructure Security and Resilience. Executive Office of the President, Washington, D.C.

ESIG, 2021. Redefining Resource Adequacy for Modern Power Systems.

Fazio, J., Hua, D., 2019. Three probabilistic metrics for adequacy assessment of the Pacific Northwest power system. Electr. Power Syst. Res. 174, 105858. https://doi.org/10.1016/j.epsr.2019.04.036

Frew, B., 2018. Beyond Capacity Adequacy. ESIG.

Gramlich, R., 2021. Ensuring low cost reliability: resource adequacy recommendations for a cleaner grid. Grid Strategies.

Hawaii PUC, 2022. Addressing the Hawaiian Electric Companies' motion for clarification and/or partial reconsideration of decision and order No 38482, PUC Order.

Hofgard, J., Savage, E., 2022. Incorporating Temperature and Precipitation Trends in Long-Term

Planning. WIEB.
ISO-NE, 2023. Key Grid and Market Stats [WWW Document]. URL https://www.iso-ne.com/about/key-stats/ (accessed 2.9.23).
ISO-NE, 2022. 2021 Capacity, Energy, Loads, and Transmission Report. ISO-NE.
ISO-NE, 2021. Regional System Plan.
Knight, F., 1921. Risk, Uncertainty and Profit. The Riverside Press Cambridge, New York, NY.
Lannoye, E., Tuohy, A., 2020. Resource Adequacy Challenges: Issues Identified Through Recent Experience in California (No. 3002019972). Electric Power Research Institute, Palo Alto, CA.
Mauch, B., Millar, D., Dorris, G., 2022. Resource Adequacy Modeling for a High Renewable Future. National Regulatory Research Institute, Washington, DC.
Mills, A.D., Wiser, R.H., 2015. Strategies to mitigate declines in the economic value of wind and solar at high penetration in California. Appl. Energy 147, 269–278. https://doi.org/10.1016/j.apenergy.2015.03.014
MISO, 2023. MISO Fact Sheet [WWW Document]. URL https://www.misoenergy.org/about/media-center/corporate-fact-sheet/ (accessed 2.9.23).
MISO, 2022a. 2022 Regional Resource Assessment Report. Midcontinent Independent System Operator.
MISO, 2022b. 2022 RRA_Technical_Appendix. Midcontinent Independent System Operator.
NARUC, 2021. Resource adequacy - Primer for state regulators. Nationa Association of Regulatory Utility Commissioners.
NARUC, 2019. The Value of Resilience for Distributed Energy Resources: An Overview of Current Analytical Practices.
National Academies of Sciences, E., 2017. Enhancing the Resilience of the Nation's Electricity System. https://doi.org/10.17226/24836
NERC, 2021. 2021 Long Term Reliability Assessment. North American Electric Reliability Corporation.
NERC, 2018. Probabilistic Adequacy and Measures (Technical refernece report). North American Electric Reliability Corporation, Atlanta, GA.
NERC, 2011. Standard BAL-502-RFC-02.
NWPCC, 2023. GENESYS Model [WWW Document]. URL https://www.nwcouncil.org/2021powerplan_genesys-model/ (accessed 2.9.23).
NWPCC, 2021a. AURORA Model [WWW Document]. URL https://www.nwcouncil.org/2021powerplan_aurora-model/#_ftnref1 (accessed 2.9.23).
NWPCC, 2021b. Regional Portfolio Model [WWW Document]. URL https://www.nwcouncil.org/2021powerplan_regional-portfolio-model-0/ (accessed 2.9.23).
NWPCC, 2018. Pacific Northwest Power Supply Adequacy Assessment for 2023.
Pacificorp, 2021. 2021 Integrated Resource Plan.
PGE, 2019. 2019 Integrated Resource Plan. Portland General Electric Company.
PJM, 2022. PJM at a glance.
PJM, 2021a. PJM Manual 19: Load Forecasting and Analysis (No. 35).
PJM, 2021b. PJM Manual 20: PJM Resource Adequacy Analysis (No. 12).
Preston, E., Barrows, C., 2018. Evaluation Of Year 2020 IEEE RTS Generation Reliability Indices, in: 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Presented at the 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), pp. 1–5. https://doi.org/10.1109/PMAPS.2018.8440394
PSCo, 2021. 2021 Electric Resource Plan. XCel Energy - Public Service Company of Colorado.
PugetSound, 2021. 2021 Integrated Resource Plan. Puget Sound Energy.
Shen, Z., Wei, W., Wu, D., Ding, T., Mei, S., 2021. Modeling arbitrage of an energy storage unit without

binary variables. CSEE J. Power Energy Syst. 7, 156–161. https://doi.org/10.17775/CSEEJPES.2019.03340

Stanton, M.C.B., Roelich, K., 2021. Decision making under deep uncertainties: A review of the applicability of methods in practice. Technol. Forecast. Soc. Change 171, 120939. https://doi.org/10.1016/j.techfore.2021.120939

Stephen, G., 2021. Probabilistic Resource Adequacy Suite (PRAS) v0.6 Model Documentation (No. NREL/TP--5C00-79698, 1785462, MainId:35919). https://doi.org/10.2172/1785462

Stephen, G., Tindemans, S.H., Fazio, J., Dent, C., Acevedo, A.F., Bagen, B., Crawford, A., Klaube, A., Logan, D., Burke, D., 2022. Clarifying the Interpretation and Use of the LOLE Resource Adequacy Metric, in: 2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). Presented at the 2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), pp. 1–4. https://doi.org/10.1109/PMAPS53380.2022.9810615

Systematic Solutions Inc., 2021. Energy 2020 Overview.

Talbot, P.W., Rabiti, C., Alfonsi, A., Krome, C., Kunz, M.R., Epiney, A., Wang, C., Mandelli, D., 2020. Correlated synthetic time series generation for energy system simulations using Fourier and ARMA signal processing. Int. J. Energy Res. 44, 8144–8155. https://doi.org/10.1002/er.5115

TEP, 2020. 2020 Integrated Resource Plan. Tucson Electric Power.

Watson, J.P., Knueven, B., Concepcion, R., Melander, D., Short, A., Zhang, P., Woodruff, D., 2020. Prescient. Sandia National Lab. (SNL-NM), Albuquerque, NM (United States). https://doi.org/10.11578/dc.20201101.1

Wu, L., Shahidehpour, M., Li, T., 2008. Cost of Reliability Analysis Based on Stochastic Unit Commitment. Power Syst. IEEE Trans. On 23, 1364–1374. https://doi.org/10.1109/TPWRS.2008.922231

Appendix A. Analyzing the use of power dispatch models to improve RA assessments - Full description

Due to the increasing penetration of VRE and storage in modern power systems, RA assessments are becoming sensitive to the combination of operational uncertainties and a series of dispatch decisions, including the scheduling of generators and the charging and discharging of storage units. Therefore, in order to accurately assess the RA of a power system, it may be necessary to simulate the system's chronological operations using a dispatch model. In this section, we develop an RA assessment framework based on an existing production cost model, *Prescient*, to simulate a hypothetical system's power dispatch and analyze its impacts on RA assessments.

A.1 Methodology

Prescient is a software toolkit developed by Sandia National Laboratories to simulate electric grid operations (Watson et al., 2020). It simulates the power system's economic dispatch by developing a day-ahead and real-time simulation cycle framework that combines commitment optimization and dispatch optimization. The day-ahead simulation is implemented as a single unit commitment optimization that determines the daily system operations with respect to the forecasted load and renewable generation. After that, the real-time simulation consists of a series of chronological hourly economic dispatch problems that re-optimize the daily system operations subjected to actual load, actual renewable generation, and the commitment status determined by the day-ahead unit commitment solution. Each hourly simulation uses the system state after the previous hour as its initial condition to simulate the inter-temporal dispatch constraints. Given that Prescient is inherently a production cost model, we make several modifications to translate it into an RA assessment dispatch model.

A.1.1 Overall framework

The formulation of the unit commitment problem would introduce numerous binary decision variables to represent the chronological availability status of each thermal generator, which significantly increases the computational time required to solve the daily dispatch problem. Though it can be incorporated into reliability analysis (Wu et al., 2008), a unit commitment model is too computationally burdensome to be solved many times within a large-scale Monte Carlo simulation. In our framework, we simplify the commitment problem by replacing it with the linearized merit order model. Both models determine the optimal generation schedule, minimizing costs of power dispatch, subject to device and operational constraints such as production limits and ramp rates, but the merit order model can yield slightly different dispatch decisions and may underestimate the system's reliability (Cebulla and Fichter, 2017). Though we omit the original unit commitment problem, we keep the two-stage framework with day-ahead and real-time simulations, to coordinate the storage operational decisions and thermal generator failure modeling. It should be noted that while decisions are made in two stages, our framework is not a two-stage stochastic program, as its day-ahead decision-making does not explicitly consider the uncertainties that arise in the real-time simulation. Prescient has multiple formulation options available to represent the transmission system. For transmission, we apply a simplified transportation model to represent power flows between zones in order to maintain computational tractability.

A.1.2 Storage operations

An accurate representation of storage status and operations can be particularly important for RA assessment. If the operation of storage is neglected or oversimplified, such as by assuming that storage is a firm resource that can always provide maximum power, then system reliability would likely be over-estimated since the actual status of the storage level is not being identified or monitored. This would not capture energy limitations, an increasingly important constraint in RA assessments. By contrast, assuming overly rigid storage operations, such as fixing the charging/discharging behavior of storage in advance, would likely lead to underestimated system reliability due to not considering the flexibility of storage to help a system prepare for and cope with peak net load events. Storage can also be very useful for operators during contingency events and emergency operations, but these conditions are not part of the scope of this analysis.

To model storage operations, we employ the Prescient logic for the two-stage simulation framework that uses the day-ahead forecast to optimize the actual evolution of the storage level. Specifically, we model the storage operations problem of each unit as a linear program without introducing the complementary binary variables (Shen et al., 2021), and embed it into the dispatch framework. At the beginning of each day, we implement a single day's dispatch problem as the day-ahead simulation to determine the state of charge (SoC) of storage units, while the unit commitment decisions in Prescient are ignored. After that, we run a series of hourly chronological dispatch problems to simulate the actual system operations, where the SoC targets of storage units at each hour have already been determined in the day-ahead simulation. Failing to meet the predetermined SoC targets will induce a penalty cost in the objective and thus the model will follow the predetermined targets unless unexpected challenges arise during real-time operations (e.g., a large plant outage, low renewable generation). With these settings, real-time storage dispatch will be somewhat suboptimal, to the extent that real-time operating conditions diverge from the conditions that were expected when the SoC targets were computed in the day-ahead stage. The model will prefer to charge a storage unit when its real-time SoC is below the target, and try to discharge it when the real-time SoC is above the target.

Due to limits related to data availability and available computational resources, we do not incorporate the operations of long-duration storage — including reservoir hydropower and pumped storage — into our framework. However, embedding long-duration storage operations may be essential for conducting RA assessments of regions with high shares of hydro resources (e.g., U.S. Pacific Northwest), as they need to address potential energy adequacy issues and have some ability to shift energy inter-seasonally.

A.1.3 Setting up model objective functions

Power systems in the real world are dispatched economically, with the goal of minimizing operational costs subject to operational constraints. However, RA assessments have often not modeled dispatch at all, and when they have, the system is generally represented with RA as its sole objective (instead of cost minimization), so that the system will satisfy demand as long as it is technically possible to do so regardless of the operational cost. Incorporating economic dispatch into RA assessment modeling is appealing because this is how power systems are actually operated in practice and hence the state of the system in a given hour is better described by its economic dispatch decisions. However, economic dispatch is more computationally demanding than alternative dispatch assumptions, which has generally led to RA assessments neglecting economic dispatch. In our analysis, we assess RA using economic and various non-economic dispatch schemes in order to investigate whether the dispatch formulation meaningfully affects estimated system RA.

Our formulation of economic dispatch does not include constraints requiring that all loads be satisfied. Rather, a penalty term in the objective function assigns a cost to any unmet load, with the cost established by the value of lost load (VoLL) parameter. The system is thus free to decide whether to incur the costs of supplying enough electricity to satisfy the loads or to allow a shortfall and incur the VoLL-based penalty. As long as the VoLL is set high enough, then the system will satisfy all loads that it can feasibly meet in the course of operations, which is indeed the case in our scenarios (with the VoLL set at $30,000/MWh).

We complement our economic dispatch implementation with several non-economic dispatch schemes that are used by other RA models (Stephen, 2021). These non-economic dispatch models dispatch power based on the following descending priority rules in both the day-ahead and real-time simulations. Their objective functions are entirely penalty-based with penalty terms reflecting outcomes that the system would like to avoid. The non-economic dispatch models developed for this work are described as follows:

- <u>Minimize shortfalls</u>: The model will first optimize dispatch decisions in order to prevent any loss of load events. A high penalty is assigned to each MW of load that cannot be satisfied.

- <u>Minimize renewable curtailment</u>: Renewable generation in our framework is represented as having no variable generation costs and thus should be efficiently used. The model will try to charge storage units or ramp down thermal generators in order to prevent renewable curtailment.

- <u>Minimize storage SoC deviation</u>: In the real-time stage, each storage unit is dispatched according to its storage SoC levels that were determined in the day-ahead stage. It will try to charge when its SoC is lower than the target, and discharge when its SoC is higher than the target.

- <u>Minimize expected power reduction due to thermal generator failures</u>: In the merit order model, thermal generation is dispatched according to the marginal generation cost of each thermal generator. In this version of our framework, we dispatch thermal generation according to each generator’s reliability. In other words, the more reliable generator with a lower forced outage rate (FOR) will be dispatched first.

The four priority rules listed above are integrated into a single objective function by assigning properly scaled penalty coefficients. In some scenarios, we will experiment with assigning different weights to the four dispatch priority criteria. In addition, we will compare scenarios based on these non-economic dispatch schemes to other scenarios where the system is economically dispatched. We will describe the details in Subsection A.3.

A.2 Data preparation

We develop our case study based on the IEEE Reliability Test System - Grid Modernization Laboratory Consortium (RTS-GMLC) (Barrows et al., 2020). The IEEE RTS region covers desert areas of Southern California, Nevada, and Arizona. It is intended to serve as a platform for analyzing power system operation strategies and issues, with given power system topology, load obligations, and generation resources. We make some modifications to the raw RTS data in order to produce a tractable case study that leads to an interesting RA assessment.

A.2.1 System configuration

IEEE RTS is a test power system with 73 buses, 158 generators, and 120 lines. In order to keep our case study computationally tractable, we simplify the default system topology. We cluster the 73 buses in IEEE RTS and concentrate adjacent buses into 23 zones. We then identify the lines that link those zones and aggregate them to produce a simplified transmission network with 35 lines. The zones and the simplified transmission network are depicted in Figure A.1.

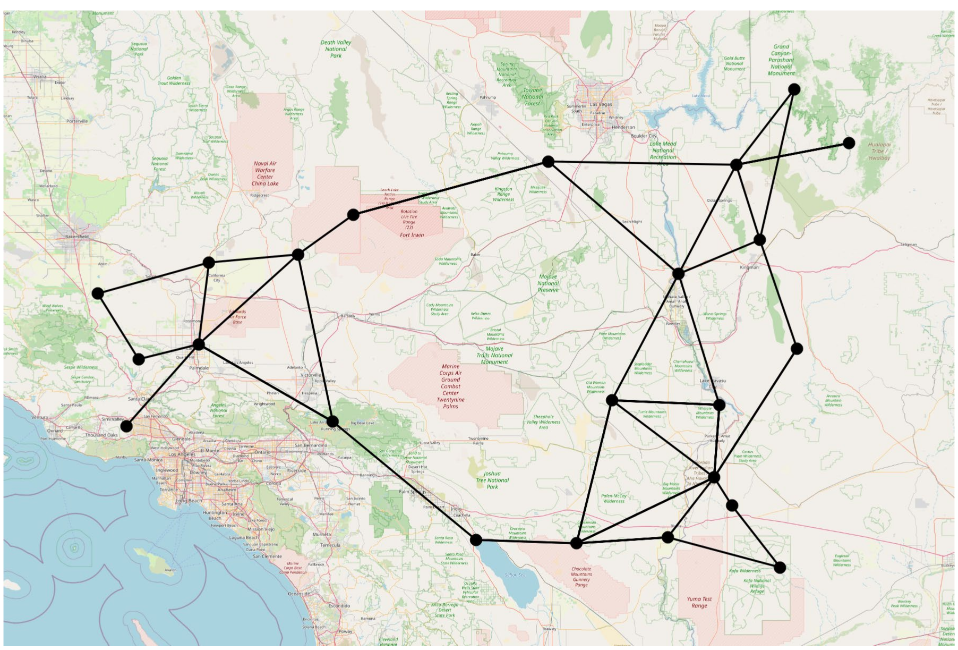

Figure A.1 Zones and transmission network

IEEE RTS represents a power system with strong RA performance. Computing RA metrics for this system would not clearly show the potential differences in RA assessment outcomes stemming from different modeling approaches because shortfall events are so limited. In order to produce a more enlightening analysis of how modeling choices affect RA assessment, we make several adjustments to the generation resources to create a test power system that is less resource adequate. Specifically, all oil plants, a nuclear plant, and two 350 MW coal plants are treated as retired, which reduces the total thermal capacity from 8.08 GW to 6.65 GW. Furthermore, we remove all hydro resources with a total nameplate capacity of 1.0 GW in the system since their operations require long-term dispatch decisions. In addition, we redesign the short-duration storage units in the system. One battery storage unit with a power capacity of 40 MW, a duration of 4 hours, and a round-trip efficiency of 85% is added in each zone, which adds a total of 920 MW to the system. Overall, the simplified system has 1.7 GW of coal plants, 5.0 GW of gas plants, and solar and wind resources with total nameplate capacities of 2.9 GW and 2.5 GW respectively.

Before proceeding, it should be noted that the simplified test power system we created does not represent any real-world power system. Resource portfolios with similar properties should not really exist in reality due to their low reliability levels. By contrast, the test power system after adjustments is more likely to represent a prototypical portfolio that a utility is considering, which features the retirements of several thermal plants but does not yet include investments in new resources to replace them. Therefore, the case study we conduct in this section does not aim to study the RA status of an existing power system. Instead, our case study is developed to elucidate the effects of RA modeling choices on estimated RA outcomes, and in doing so, to provide high-level and conceptual insights to regulators and planners on what power system operational details are important to include in a model-based RA assessment.

A.2.2 Time-series data expansion

IEEE RTS only includes a single year time-series for loads and intermittent renewable generation. This is insufficient for conducting long-term RA assessment, which requires a few decades' worth of chronologically correlated load and renewable generation data for each load zone. To address this shortcoming, we expand the one year of raw data into five years of synthetic load and renewable generation time series data in order to capture inter-annual variability.

Many methods have been proposed to capture the correlation among energy time-series (Borges and Dias, 2017) and generate synthetic energy data (Chen and Rabiti, 2017; Talbot et al., 2020). In this study, we apply traditional decomposition time-series methods to generate synthetic data for load and solar generation. A decomposition method usually deconstructs a time-series into three distinct parts: trend, seasonality, and residual. The trend represents the long-term progression of the time series, while the seasonality component defines the repeating short-term interval cycle of the series. At last, the residual contains all non-systematic components remaining for further analysis. We apply additive decomposition for the time-series as follows.

$$\boldsymbol{X_t} = \boldsymbol{T_t} + \boldsymbol{W_t} + \boldsymbol{D_t} + \boldsymbol{R_t}$$

In this formula, $\boldsymbol{X_t}$ represents the triplet of raw load time-series across three RTS regions, while $\boldsymbol{T_t}$ represents the trend, $\boldsymbol{W_t}$ represents the weekly periodicity, $\boldsymbol{D_t}$ represents the daily periodicity, and $\boldsymbol{R_t}$

represents the residual. The residual is assumed to be a multivariate stationary time-series, and we deploy the Probabilistic AutoRegressive (PAR) model to generate synthetic residuals (Patki et al., 2016).

$$\widehat{\boldsymbol{R_t}} = PAR(\boldsymbol{R_t})$$

The synthetic load time-series are then recovered.

$$\widehat{\boldsymbol{X_t}} = \boldsymbol{T_t} + \boldsymbol{W_t} + \boldsymbol{D_t} + \widehat{\boldsymbol{R_t}}$$

We apply the above procedure for every four consecutive weeks of raw load time-series. A sample of synthetic load that we generate for one region is shown in Figure A.2. The procedure for synthetic solar generation is similar to that for the load but does not account for weekly periodicity and does not capture correlation between solar radiation and load. By deploying this method, we expand the load and solar generation time-series from a single year to five years.

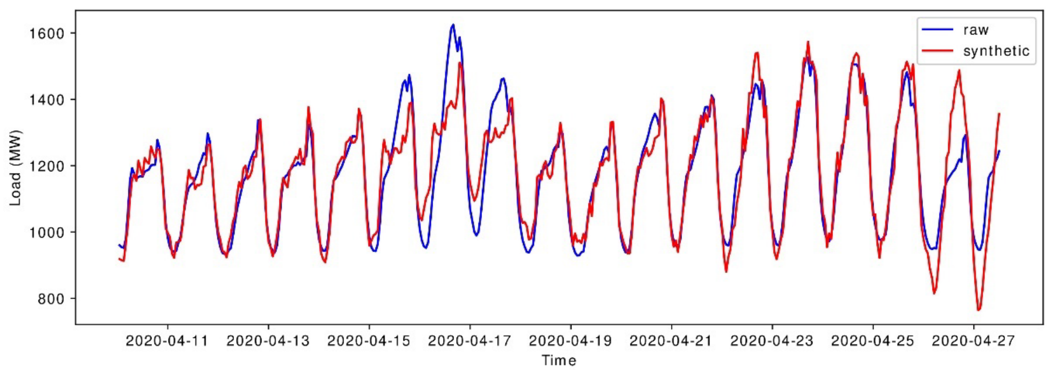

Figure A.2 Synthetic load sample

By contrast, we do not observe significant trend and seasonality in the raw wind data. Therefore, the above decomposition method is not suitable for synthetic wind data creation. Given that the drivers of wind generation are approximately stable over time and we have historical data on wind from the IEEE RTS region, we base our synthetic wind time-series closely on historical wind data. We obtain the wind capacity factors from the dataset developed by the Wind Integration National Dataset (WIND) Toolkit (Draxl et al., 2015). Specifically, we select four wind farms from the dataset to represent the four wind plants in IEEE RTS according to their geographical locations and average capacity factors. We download the power generation of those four wind farms from 2010 to 2013 and re-scale them as the multi-year wind data in our case study. A sample of wind generation that we create for one region is shown in Figure A.3. Compared to load and solar, the five-year wind time-series include more day-to-day variations due to changing weather patterns and their large impacts on wind generation. Note that our wind time-series are created independently from the load and solar profiles, and thus there is no consideration of common underlying weather drivers.

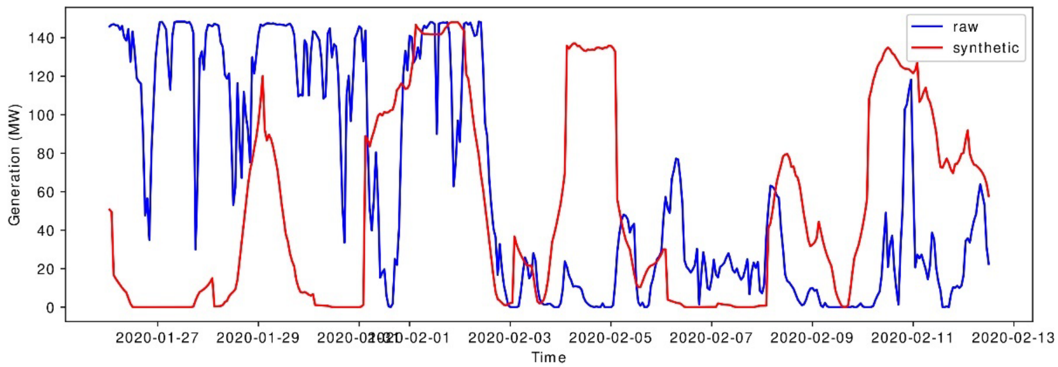

Figure A.3 Synthetic wind generation sample

We conclude the data preparation subsection by constructing a Monte Carlo simulation set that consists of 500 simulation instances. Each instance is constructed by a random draw of a single year time-series from the five years of solar, wind, and load data and a unique stochastic sequence that represents all random outage events in this year for all thermal generators.

A.3 Scenarios

We first investigate how the dispatch scheme (economic or non-economic) affects the outcomes of an RA assessment. In the economic dispatch scenario, we keep the production costs of thermal generators in the model and dispatch the system to minimize costs, reflecting real-world operating principles. The production cost only includes the variable production cost, which is a convex increasing function of the power generated by the unit. In addition, we calculate the shortfall cost by multiplying the quantity of unserved energy by the assumed VoLL of 30,000$/MWh. The system's total cost (objective value to minimize) is then calculated as the sum of the production cost and shortfall cost.

By contrast, the non-economic dispatch scenarios do not consider dollar values at all and simply dispatch units based on various prioritization rules. While real-world power systems are economically dispatched, these non-economic dispatch schemes are much less computationally burdensome to include in a probabilistic RA assessment based on a large Monte Carlo simulation. Therefore, we intend to determine whether using non-economic dispatch formulations actually results in less accurate estimates of system RA performance than modeling economic dispatch. In this study, we construct non-economic dispatch scenarios according to three strategies for thermal dispatch and four strategies for storage dispatch that are characterized as follows:

- Thermal dispatch
    - **Random-priority strategy**: thermal generators will be dispatched at random priority when available.
    - **Reliability-focused strategy**: thermal generators will be dispatched based on their reliability, i.e., a more reliable generator with a lower FOR will be dispatched first.
    - **Economic strategy**: thermal generators will be dispatched based on their convex

production costs, i.e., generators with lower marginal production costs will be dispatched with higher priority.

- Storage dispatch
  - **Passive strategy**: each storage unit will be dispatched to meet its predetermined storage SoC target at that time. The target is given as an exogenous constant.
  - **Active strategy**: each storage unit will be dispatched to meet its predetermined storage SoC target at that time. The target is determined by the optimal values derived in the day-ahead problem.
  - **Cost-free resource strategy**: storage units will only charge when there is excess renewable generation and discharge as much as possible if there is a thermal generator that is able to ramp down.
  - **Reserve strategy**: storage units will operate as reserves. They will only discharge when the load exceeds available generation capacity and charge as much as possible if there is a thermal generator that is able to ramp up.

The first set of scenarios that we formally define is designed primarily to investigate how the dispatch scheme (e.g., economic vs. non-economic dispatch) affects estimates of RA metrics (see Table A.1). The *RADp-4* scenario is based on the most detailed formulation of non-economic dispatch that we consider. Hence, *RADp-4* acts as a benchmark for the remaining scenarios and will be included in subsequent sets of scenarios constructed to explore how other power system modeling choices influence RA outcomes.

Table A.1 Scenarios designed to investigate the effects of dispatch schemes on RA metrics

Scenario	Model objective	Thermal dispatch	Storage dispatch
RADp-4	Minimize total shortfalls (MWh)	Reliability-focused	Active
EcDp	Minimize total operational costs ($)	Economic	Active
AgDp	Minimize total operational costs ($)	Economic	Cost-free resource
CsDp	Minimize total shortfalls (MWh)	Reliability-focused	Reserve

The four scenarios are formally defined as follows:

- *RADp-4*: Non-economic dispatch. In this scenario, thermal generators will be dispatched based on the reliability-focused strategy and storage units will be dispatched based on the active strategy, both of which were described above.
- *EcDp*: Economic dispatch. In this scenario, thermal generators will be dispatched based on the economic strategy and storage units will continue to be dispatched based on the active strategy.

- *AgDp*: Aggressive dispatch. In this scenario, thermal generators will be dispatched according to the economic strategy, as in *EcDp*. However, storage units will be dispatched according to the cost-free resource strategy.
- *CsDp*: Conservative dispatch. In this scenario, thermal generators will be dispatched according to the reliability-focused strategy, as in *RADp-4*. However, storage units will be dispatched based on the reserve strategy.

In scenarios *EcDp* and *AgDp*, the objective function of the model is to minimize the system's operational cost. In scenarios *RADp-4* and *CsDp*, the objective function of the model is to minimize total shortfalls, which is an RA-oriented objective. The total operational costs in these two scenarios are recalculated post-solve after the dispatch decisions have been determined in order to calculate the additional cost that the model incurs in pursuit of an RA-focused objective rather than an economic objective. Scenario *EcDp* simulates the way that real-world power systems generally operate, and scenarios *AgDp* and *CsDp* show two extreme cases that focus exclusively on system economics and system reliability, respectively.

As explained above, the *RADp-4* scenario represents a relatively sophisticated formulation of non-economic dispatch, with day-ahead and real-time decision-making stages, reliability-focused thermal plant dispatch, and active storage dispatch. With our next set of non-economic dispatch scenarios defined in Table A.2 below, we experiment with making various simplifications to the dispatch scheme to determine whether we can reduce the computational complexity of the model without sacrificing much accuracy in terms of the RA metrics we estimate. These additional non-economic dispatch scenarios include the traditional convolution method (*Conv*) and RA dispatch models that are simpler versions of *RADp-4* (*RADp-1, RADp-2, RADp-3*). Summaries of these scenarios and qualitative indications of their computational complexity are provided in Table A.2.

Table A.2 Scenarios designed to investigate whether less complex implementations of non-economic dispatch lead to a significant loss of accuracy when estimating RA metrics

Scenario	Stages	Thermal dispatch	Storage dispatch	Complexity
Conv	N/A	N/A	N/A	Easy
RADp-1	Real-time	Random-priority	N/A	Moderate
RADp-2	Real-time	Random-priority	Passive	Moderate
RADp-3	Real-time	Reliability-focused	Passive	Moderate
RADp-4	Day-ahead/Real-time	Reliability-focused	Active	Difficult

The *Conv* scenario is constructed according to the work of Preston and Barrows (2018). It is based on a simple convolution method that only considers the net load duration curve and thermal generator forced outages, with no explicit modeling of the power network or chronological system operations. Specifically, the *Conv* and *RADp-1* scenarios do not incorporate storage operations, and they treat storage units as firm resources with certain capacities. We run each of these scenarios twice as two sub-scenarios: the first sub-scenario run treats storage units as generators with null capacity (sub-scenario denoted “-N"), while the second sub-scenario run treats storage units as generators with equivalent full nameplate capacities (sub-scenario denoted “-F").

In addition to the strategies implemented for power dispatch, other choices about how to represent different aspects of the power system may also affect the results of RA assessments. For the purposes of our analysis, we choose to explore the effects of two modeling choices that are easy to implement in our modeling framework and relevant to system operators: accuracy of short-term forecasts and consideration of transmission adequacy. We develop the following two system sensitivity scenarios to examine the impacts of these other factors on the characteristics of shortfalls.

- *PfSF*: Perfect short-term forecast. In this scenario, the short-term forecast for load and renewable generation is assumed to be a perfect forecast. In other words, the day-ahead time-series are the same as the real-time time-series. This differs from all of the scenarios introduced up to this point, in which the day-ahead forecasts included errors.
- *TxFr*: Transmission free. In this scenario, all transmission limits are omitted. This scenario has reduced fidelity compared to a scenario in which limits are enforced, but does represent a prevalent practice to ignore the transmission system when assessing adequacy in a regional power system. Consequently, by comparing it to the other scenarios in which transmission adequacy is considered, we will be able to determine how important it is to account for potential transmission limits when assessing system RA.

Table A.3 summarizes how the two system sensitivity scenarios described above differ from the benchmark scenario *RADp-4*.

Table A.3 System sensitivity scenarios designed to investigate how omitting day-ahead forecast errors or transmission limits affects estimates of RA metrics

Scenario	Short-term forecast errors	Transmission limit constraints
RADp-4	Included	Included
PfSF	Excluded	Included
TxFr	Included	Excluded

All scenarios except for *Conv* are developed based on our probabilistic RA assessment framework and thus are simulated using the constructed Monte Carlo simulation set. We record a series of contiguous hours when demand cannot be met as a shortfall. Small shortfall events with load shedding of less than 0.1% of hourly demand are screened and discarded, as we assume that they can be eliminated by demand-side management actions.

A.4 Simulation results and analysis

A.4.1 Dispatch schemes, RA metrics, and costs

In this subsection, we examine the results of the four scenarios from Table A.1, which were designed to help us investigate how the dispatch scheme implemented in the model (e.g., economic vs. non-economic) affects the resulting estimates of RA metrics. In addition to reporting the RA metrics themselves, we also report the total production costs that the model incurs in each scenario (not including penalty costs for unmet load). This provides information on how costly it is to operate the power system in a reliability-focused mode, and how much additional reliability is obtained from it.

Table A.4 reports reliability and production cost outcomes for scenarios *RADp-4*, *EcDp*, *AgDp*, and *CsDp*. We can observe that *AgDp* underestimates system reliability and it also fails to reduce the production costs due to the lack of robustness against unforeseen thermal generator failures. By contrast, *CsDp* slightly overestimates system reliability as a result of operating storage units during critical RA hours and not contributing to power dispatch during normal periods. However, *CsDp* is not well aligned with the economic reality since storage assets will be operated to earn profits for their owners instead of simply waiting to be called upon to avert shortfalls. This becomes evident when comparing production costs. *CsDp* operational costs are more than 50% higher than those of the economic dispatch reference scenario (*EcDp*), a premium that operators may not be willing to pay for a 10% improvement in LOLP or to serve 8 MWh of additional demand.

Table A.4 Reliability and production cost outcomes for least-cost and reliability-focused power system dispatch schemes

Scenario	LOLE (days/year)	LOLP	LOLEV (events/year)	EUE (MWh/year)	Production Cost (M $/year)
RADp-4	0.286	24.6%	0.288	41.92	96.3
EcDp	0.294	25.2%	0.296	43.25	76.9
AgDp	15.500	100%	17.762	5889.58	85.0
CsDp	0.250	22.6%	0.250	35.97	105.62

The comparison between *RADp-4* and *EcDp* reveals subtle differences in RA performance between reliability-focused and least-cost operations of power systems when the storage units are operated according to the predetermined SoC targets. The reliability-focused dispatch, which tries to minimize the impacts of thermal generator failures instead of total production costs, overestimates system reliability by reducing the LOLE by 2.7% and the EUE by 3.1%, compared to the least-cost dispatch, which reflects the dispatch logic applied in real-world power system operations. However, these artificial improvements are small and do not indicate a significant difference in RA assessment accuracy. Therefore, we find that the non-economic dispatch model, which only focuses on system reliability and ignores operational costs, can lead to fairly accurate RA assessments when coordinated with detailed operational strategies.

We further investigate the effects of the dispatch scheme on RA assessments by analyzing a sample instance in the Monte Carlo simulation set. Figure A.4 shows the production cost *and* shortfall cost in scenarios *RADp-4* and *EcDp*, by month. The shortfall cost is calculated by multiplying the shortfall in MWh by the VoLL as described earlier. We can observe that significant production cost differences between the two scenarios occur in July and August, when the net load reaches its maximum and the impacts of different dispatch strategies become more pronounced. During this period, the threat of a shortfall is higher and thus the reliability-focused dispatch *RADp-4* tends to underestimate reliability compared to the economic dispatch *EcDp* that better reflects the operation of real-world power systems. However, the economic impacts of system operations become more significant, since thermal generators are usually running at their maximum capacities with high marginal production costs.

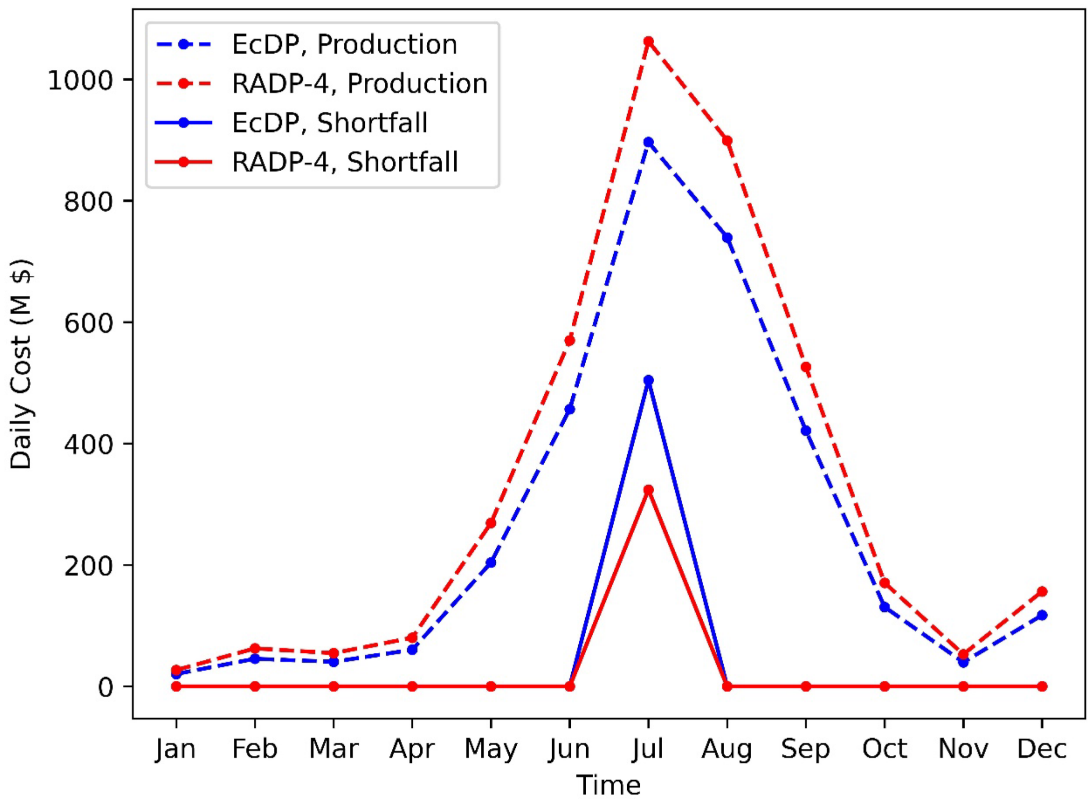

Figure A.4 Economic analysis of an instance in RA simulation

A.4.2 Non-economic dispatch strategies

We also report the impacts of different power dispatch strategies on RA assessments using the non-economic dispatch models from Table A.2. Our goal with this scenario analysis is to explore whether making simplifications to the non-economic dispatch scheme is able to reduce computational complexity without sacrificing accuracy when estimating RA metrics. Compared to the *RADp-4* benchmark scenario for non-economic dispatch (and especially the *EcDp* scenario for economic dispatch), these other non-economic dispatch scenarios represent power system operations in less detail.

Table A.5 reports the RA outcomes of these non-economic dispatch scenarios and benchmarks, which are evaluated using traditional expectation-based RA metrics including LOLE, LOLP, and EUE. The standard errors for the Monte Carlo estimates are also presented. We can observe that both *Conv* sub-scenarios, which represent the traditional convolution RA method with both extreme assumptions about storage capacity credit, fail to accurately assess the system's RA status. Results for both *Conv-N* – which ignores storage capacity contributions – and *Conv-F* – which assumes a capacity credit for storage equal to its total discharge capacity — deviate significantly from the RA outcomes of *RADp-1-F*. This result highlights the importance of making educated assumptions about the real-time availability of short-duration storage. Therefore, including a storage dispatch model that tracks the system's real-time state is evidently crucial for conducting RA assessments of modern power systems.

Table A.5 RA outcomes across dispatch strategies

Scenario	LOLE (days/year)	LOLP	EUE (MWh/year)
Conv-N	13.03	100%	2593.2
Conv-F	0.043	4.3%	6.3
RADp-1-N	18.378±0.199	100±0.0%	6819.11±128.12
RADp-1-F	0.368±0.027	31.2±2.1%	59.48±6.10
RADp-2	0.458±0.030	37.6±2.2%	134.34±12.75
RADp-3	0.416±0.029	34.6±2.1%	126.06±12.38
RADp-4	0.286±0.024	24.6±1.9%	41.92±4.84
EcDp	0.294±0.025	25.2%±1.9%	43.25±5.02

The dispatch model in *RADp-1*, which omits the operations of storage units, provides limited information on the system's RA status due to the exaggerated differences between its sub-scenarios that model storage units as firm generators with different equivalent capacities. This result suggests that as energy storage becomes more widespread in modern power systems, identifying the real-time status of storage units has substantial impacts on the performance of RA assessments, and embedding the storage operation problem into dispatch models is necessary to reflect a system's real-time status. A dispatch model that represents the chronological operations of both thermal generators and storage units is required to accurately measure RA as the electricity sector evolves to incorporate more intermittent resources and storage.

The comparison among scenarios *RADp-2*, *RADp-3*, and *RADp-4* shows that more sophisticated non-economic dispatch strategies could improve the performance of the system's RA assessment. The comparison between *RADp-3* and *RADp-2* shows that utilizing the reliability-focused thermal dispatch strategy in *RADp-3* leads to 15 fewer instances of shortfalls in the 500 instances in the Monte Carlo simulation set. These results still underestimate the system's reliability – as compared against *RADp-4* –

but improve the assessment in the right direction. Furthermore, the omission of active storage dispatch with optimized day-ahead SoC targets in *RADp-3* results in considerable increases from *RADp-4* in all three RA metrics: +0.132 days/year LOLE, +10% LOLP, and +84.74 MWh/year EUE. These changes are likely significant enough to affect the judgement about whether a power system is resource adequate. We find that a reliability-focused thermal dispatch yields minor improvements over a random-priority dispatch, but that a more accurate characterization of short-duration storage dispatch is essential to accurately evaluate the system's RA status.

The histograms of the shortfalls observed in *RADp-2* and *RADp-4* are shown in Figure A.5. The *RADp-2* scenario produces shortfalls with magnitudes higher than 400 MWh that do not appear in *RADp-4* when a more sophisticated storage dispatch strategy is used. This comparison indicates that the active dispatch strategy using the day-ahead forecast is close to the theoretical optimal for RA purposes, where a storage unit can discharge its maximum power during critical hours.

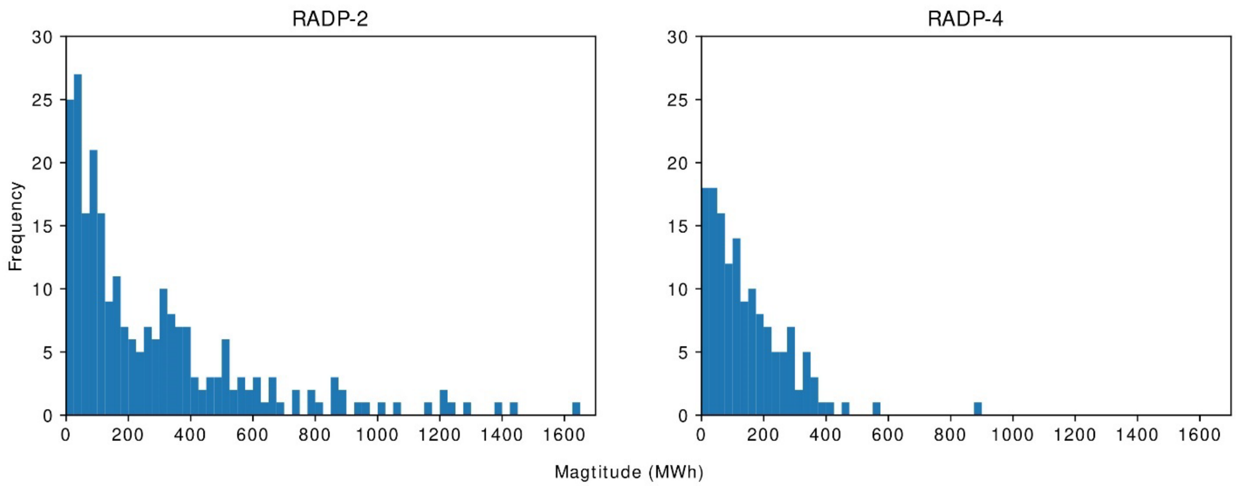

Figure A.5 Shortfall magnitude histograms for the whole simulation set

Figure A.6 shows the total number of shortfall hours observed across *RADp-2*, *RADp-3*, and *RADp-4*. The loss of load events in the power system happen exclusively in summer when the demand is significantly higher than the demand in other seasons. Though they have different LOLH in the simulation, the three scenarios show consistent temporal shortfall occurrences during the same range of days and hours. This suggests that though dispatch models with simplified operational details fail to assess the system's RA status accurately, they may still be useful for screening simulation periods, identifying critical hours, or providing bounds for target RA metrics due to their low computational complexity. This possible application is particularly useful for renewable effective load carrying capability (ELCC) studies in power systems without fuel- or energy-constrained resources, since it can make it more efficient to carry out the multi-round simulation process to find the equivalent amount of incremental load.

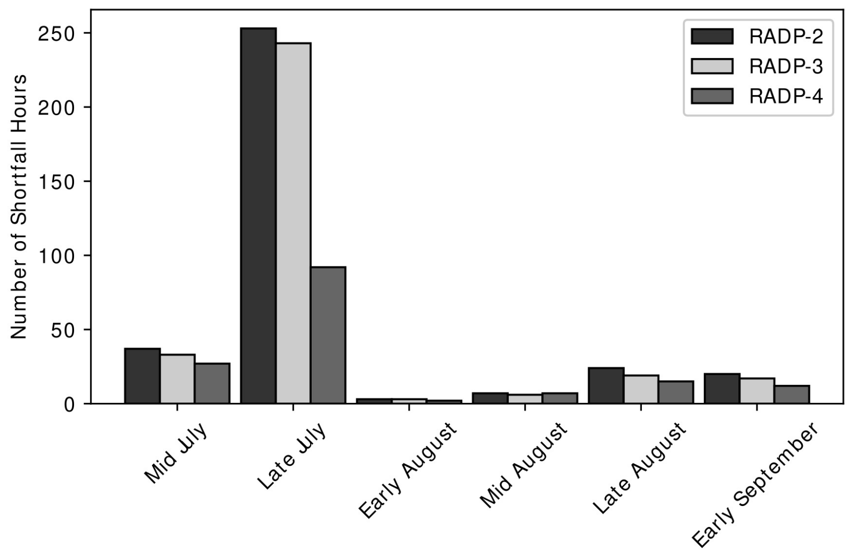

Figure A.6 Number of shortfall hours observed in the whole simulation set

A.4.3 System representation sensitivities

The results in the preceding subsections showed that the strategies used to dispatch thermal generators and storage units have an important influence on RA metric values. In this subsection, we turn to studying how several other choices about how to model power system operations may affect the outcomes of RA assessments. Table A.6 summarizes the impacts of two specific modeling assumptions on RA assessments using traditional RA metrics. *PfSf* assumes perfect day-ahead foresight instead of the default day-ahead forecast errors, while *TxFr* ignores transmission limits. Our goal is to understand how important it is to consider day-ahead forecast errors and transmission limits when attempting to determine a system's RA status.

Results show that the computed RA metrics in *PfSF* are the same as those from the benchmark scenario *RADp-4*. It suggests that short-term forecast errors have limited impacts on the accuracy of RA assessments, although this outcome may be different if unit commitment was actually being simulated.

Investigating the effects of neglecting transmission constraints is relevant because most current RA assessments do not represent transmission limits. It follows that measuring the impact of this assumption within our model can shed light on the potential consequences of this common modeling simplification. We find that relaxing the transmission limits in *TxFr* leads to a 23% underestimation of LOLE and LOLP, but only a 13% underestimation of EUE. These results suggest that incorporating transmission adequacy into RA assessments has a substantial impact on measuring shortfall frequency, but a weaker impact on estimating shortfall magnitudes.

Table A.6 Effects of ignoring day-ahead forecast errors or transmission limits on RA outcomes

Metric (Unit)	*RADp-4*	*PfSF*	*TxFr*
LOLE (days/year)	0.286	0.286	0.220
LOLP	24.6%	24.6%	19.0%
EUE (MWh/year)	41.92	41.92	36.49

We have demonstrated that different modeling assumptions can lead to substantial changes in the results of RA assessments. However, the effects of different modeling choices may not be fully captured by traditional expectation-based RA metrics like LOLE and EUE. As discussed in Section 3.2, researchers have suggested that a combination of metrics may be needed to properly capture the frequencies, durations, and magnitudes of shortfalls. In addition, it is possible that new metrics may need to be used as supplements to traditional metrics.

We use the simulation results of the benchmark scenario *RADp-4* to diagnose whether traditional metrics are sufficient to describe the RA status of a power system. Figure A.7 shows the distribution of hourly EUE over 100 random thermal generator failure instances across the five weather patterns in scenario *RADp-4*. We find that most shortfalls happen between 16:00 (4pm) and 20:00 (8pm) in the summer, when PV generation is declining and the load is increasing. In addition, the whole year's aggregated EUE is similar across weather patterns. However, the specific hourly distribution of shortfalls is significantly different. The critical hours (with an EUE above the 90th percentile EUE across simulations) occur on different days across the five weather patterns, and the hour with the highest EUE in one weather pattern may not have a shortfall in another weather pattern. This finding has two implications. First, it shows that using a single year's simulation may not necessarily lead to a significant bias in assessing the aggregate shortfall magnitude. Second, it shows that the timing of shortfalls will differ and none of the traditional expectation-based metrics will identify this difference. Our analysis highlights the relevance of considering multiple probabilistic RA metrics, or even the full distributions of outcomes instead of just their descriptive statistics, in RA assessments to describe those differentiated shortfall characteristics.

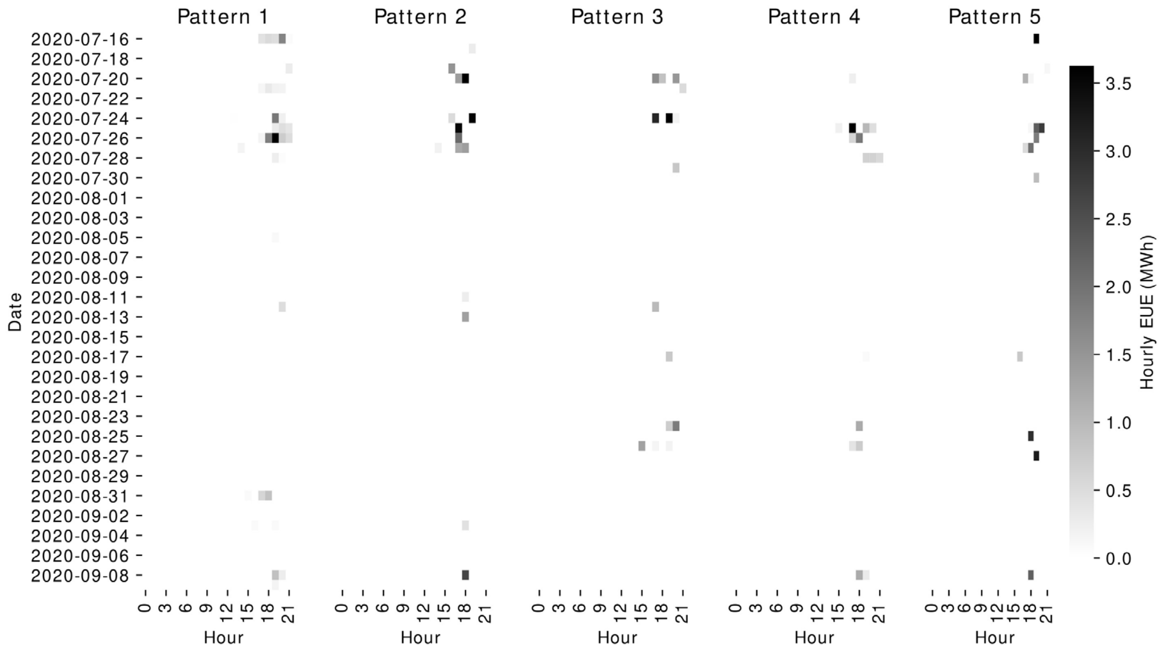

Figure A.7 Heatmap of hourly EUE across the five weather patterns for scenario *RADp-4*

The weather pattern analysis shows that differences in timing can arise within RA assessments, but there are no suggested metrics to capture these differences. If there is a reasonable belief that the VoLL would differ substantially across hours within the 16:00 to 20:00 hour range, then capturing these timing differences may be relevant for improved RA assessments. We study whether new RA metrics can better capture the nuanced effects of power system modeling choices on RA outcomes. Table A.7 reports several RA assessment statistics from the shortfall records in the simulations of system sensitivity scenarios, evaluated using alternative metrics suggested in NWPCC (2018) that capture the magnitudes and durations of shortfalls. These metrics provide more information about individual shortfall events, including the maximum all-hour and peak-hour magnitude, maximum duration, and total number of shortfalls. These metrics provide some insights on the *shape of the shortfall distribution*, as opposed to traditional aggregate metrics that focus on the mean or expected value of the distribution. The outcomes in *PfSF* are still the same as those in *RADp-4*, which is consistent with the previous finding that short-term forecast errors are unlikely to be a key operational detail to accurately evaluate RA when unit commitment is not considered.

Table A.7 RA statistics using alternative metrics across scenarios

Metric (Unit)	*RADp-4*	*PfSF*	*TxFr*
Number of Shortfalls	144	144	111
Max Shortfall Magnitude (MWh)	877.25	877.25	877.25
Average Shortfall Magnitude (MWh)	145.55	145.55	164.36
Max Shortfall Peak-hour Load Shedding (MW)	603.34	603.34	603.34
Average Shortfall Peak-hour Load Shedding (MW)	137.45	137.45	153.41
Max Shortfall Duration (h)	2	2	2
Average Shortfall Duration (h)	1.076	1.076	1.108

The comparison between scenarios *RADp-4* and *TxFr* shows that ignoring transmission constraints results in the underestimation of LOLE and EUE but does not lead to inaccuracies in the magnitude of peak-hour load shedding. This is because when the bulk system is modeled as a single zone without the transmission network, the spatial imbalance between the zones with large net load requirements and zones that have excess generation is eliminated. In other words, the presence of transmission limits creates smaller sized shortfalls that are "hidden" when removing the limits. These smaller shortfalls decrease the average MWh magnitude, as the ones remaining when limits are not enforced are larger. The reduced number and higher average duration of shortfalls in *TxFr* compared to *RADp-4* are evidence of this explanation. Choices about whether and how to model the transmission network will affect RA assessments by not only determining whether the power can be delivered to the load points, but also changing the dispatch logic of thermal generators and storage units that are geographically scattered across the represented region.

A.5 Summary

Traditional RA assessments that do not explicitly model power system operations have become obsolete for evaluating the reliability performance of modern power systems due to several rapid changes in the electricity industry, and there is general agreement about the urgent need to develop a new RA framework with respect to metrics, data, modeling approaches, and institutional analysis. In this analysis, we have created a technical framework for probabilistic RA assessment and used it to study how key choices about how to model power system operations affect the values that are obtained for RA metrics. As summarized in Table A.8, the results helped us distinguish operational details that are critical to include in any accurate RA assessment from details that are computationally burdensome but do not significantly affect the evaluation of RA.

Table A.8 Impacts of operational details on RA assessments

Operational or simulation characteristic	Impacts on RA assessment accuracy	Level of effort to represent in models
Multi-year data	High	Medium
Transmission limits	High	Medium
Storage dispatch	High	Medium (short duration) High (long duration)
Non-economic thermal dispatch	Medium	Low
Operational cost	Medium	High
Short-term forecast error	Low	High

Several high-level findings emerged from the case study explored using our technical RA framework.

- First, non-economic dispatch schemes that ignore economic objectives can lead to fairly accurate RA assessments when coordinated with detailed operational strategies.
- Second, representing the detailed chronological operations of thermal generators and storage units is essential for accurate RA assessments, but simplified dispatch models can be used as screening tools (e.g., to identify critical hours) due to their low computational complexity.
- Multi-year data is important to include to capture inter-annual variations in system conditions.
- Neglecting to incorporate transmission limits into RA assessment could lead to substantial underestimation of traditional "expected value" RA metrics. However, experiments with alternate metrics that focus on individual event characteristics show that neglecting transmission limits will not mask the most critical shortfall events.
- New RA metrics that capture event-specific shortfall characteristics should be used as supplements to traditional metrics to better capture the impacts of different modeling assumptions on RA outcomes, as well as better describe the ability of the system to prevent specific high-impact shortfalls.

This study has been motivated by the consensus that current RA assessments are inadequate for modern power systems, and with the goal of prioritizing ways to improve them. We believe that our findings provide several useful insights to stakeholders to support decision-making in RA assessments. Future work can consider incorporating the operations of energy-constrained resources (e.g., hydroelectric resources) and long-duration storage to expand the capabilities of our technical framework. In addition, thermal generator failures that may lead to low-probability, but high-impact shortfalls can be included in the economic dispatch model, which can potentially bring resilience analysis into RA modeling. Finally, our framework can also be deployed alongside capacity expansion models that enable planners to optimize their least-cost resource portfolios while maintaining robust RA levels. The developed power system operation model may have a promising capability to support decision-making in planning processes, such as the capacity accreditation of variable renewable resources or the location-sizing problem for battery storage investments.

Appendix B. List of entities surveyed for resilience integration into planning processes

Table B.1 List of entities surveyed for resilience integration into planning processes

Entity	Type	Planning process	Year
Avista	Utility	Integrated resource plan	2021
Puget Sound Energy	Utility	Integrated resource plan	2021
PacifiCorp	Utility	Integrated resource plan	2021
Seattle City Light	Utility	Integrated resource plan	2021
Tacoma Power	Utility	Integrated resource plan	2021
Public Utility of New Mexico	Utility	Integrated resource plan	2021
Public Service Company of Colorado	Utility	Electric resource plan	2021
Pacific Gas & Electric	Utility	Integrated resource plan	2020
San Diego Gas & Electric	Utility	Integrated resource plan	2021
Southern California Edison	Utility	Integrated resource plan	2021
PSE&G Long Island	Utility	Integrated resource plan	2021
Florida Power and Light	Utility	10-Year Power Plant Site Plan	2020
MISO	ISO/RTO	Resource adequacy assessment	2021
PJM	ISO/RTO	Resource adequacy assessment	2021
ISO New England	ISO/RTO	Resource adequacy assessment	2021
CAISO	ISO/RTO	Resource adequacy assessment	2021
NWPCC	Regional organization	Resource adequacy assessment	2021
WECC	ERO	Resource adequacy assessment	2021
NERC	ERO	Resource adequacy assessment	2021

Appendix C. Detailed reviews of RA assessments in IRP and by ISO/RTO

C.1 Reviews of RA assessments in IRP

This subsection reports on the RA assessment approaches implicit in the IRPs filed by six LSEs located in the Pacific, Mountain West, and Southeast regions of the U.S.: PacifiCorp, Puget Sound Energy, Portland General Electric, Tucson Electric Power, Public Service Company of Colorado, and Duke Energy Carolinas.

C.1.1 PacifiCorp

PacifiCorp is one of the larger IOUs in the West with 1.7 million customers across 136,000 square miles in six Western states (Utah, Oregon, Wyoming, Washington, Idaho, and California). PacifiCorp has generating capacity of 10,326 MW and about 15,800 miles of bulk transmission lines.

PacifiCorp developed its 2021 IRP using a sophisticated set of three interrelated Plexos models: (1) the Plexos Long-Term planning model (LT model), which is a capacity expansion model; (2) the Plexos Medium-Term model (MT model), which is a stochastic model; and (3) the Plexos Short-Term model (ST model), which is an hourly production cost model (Pacificorp, 2021).

The modeling process begins with the capacity expansion LT model that develops an initial set of resource portfolios consistent with different planning cases over a 20-year time horizon. The second step is to evaluate the initial portfolios with the ST model to identify and quantify hourly periods of unserved load and to develop the net system value of resources. The portfolio is adjusted to address any shortfalls and to produce a reliable dispatch for every hour over 20 years. The third step is to use the MT model to perform stochastic risk analysis of the portfolios. The MT model performs Monte Carlo sampling of variables across the 20-year period for loads, resource prices, hydro conditions, and thermal unit outages. The MT risk analysis is combined with the ST model system cost information to derive a final risk-adjusted present value of revenue requirement (PVRR) for each portfolio. PacifiCorp's portfolios are then ranked based on a risk-adjusted PVRR.

PacifiCorp adopted the capacity factor approximation method (CF Method) for calculating the capacity contributions of resources. The CF Method, developed by NREL, calculates the marginal resource contribution based on the resources expected availability during periods when the risk of loss of load events is highest. A final CF Method analysis was performed using small incremental changes from the preferred portfolio.

C.1.2 Puget Sound Energy

Puget Sound Energy (PSE) is an electric and gas utility in the state of Washington. PSE has 1.1 million electricity customers in a service area spanning 6,000 square miles. PSE has a generating capacity of 5,044 MW that includes ownership of 3,597 MW of capacity and long-term contracts equal to 2,447 MW of capacity. Hydropower is an important resource in the Pacific Northwest and PSE derives about 30% of its electricity from hydro resources.

PSE performed resource adequacy analysis for its 2021 IRP that integrated three distinct modeling tools: (1) PSE's Resource Adequacy Model (RAM); (2) the Genesys regional model developed by the Northwest Power and Conservation Council (NWPCC) and Bonneville Power Administration; and (3) the Wholesale Purchase Curtailment Model (WPCM) developed by PSE (PugetSound, 2021).

PSE begins its modeling process with the Genesys model and the NWPCC's 2023 assessment to obtain a probabilistic regional resource adequacy assessment. PSE uses Genesys to generate over 7,000 simulations based on historical data that include 80 different hydro years, 88 years of temperature data, and several PSE specified updates. PSE adopted the 5% loss of load probability (LOLP) reliability metric which is consistent with the NWPCC's practice. The outputs from Genesys modeling are used as inputs to PSE's reliability analysis in the IRP.

PSE has historically relied on short-term wholesale energy markets as an important resource to meet loads. PSE developed its own model, the WPCM, to better anticipate the behavior of the short-term wholesale market. Key data inputs for the model are the loads and resources for the Northwest region and data on individual utilities that participate in the short-term wholesale market. The model performs simulations that inform PSE's preferred strategy to rely on the short-term wholesale market.

PSE uses its RAM model to assess the physical supply risks over time, guide peak load planning standards, and quantify the peak capacity contributions of renewable and energy-limited resources. The RAM model enables PSE to derive multiple probabilistic metrics including the following: Loss of Load Probability (LOLP), Expected Unserved Energy (EUE), Loss of Load Hours (LOLH), Loss of Load Expectation (LOLE), and Loss of Load Events (LOLEv). The output of the RAM model is used to determine the amount of "perfect capacity" that is needed to cure deficiencies by 2027 and 2035, the planning horizons in the PSE model. PSE calculates a planning reserve margin using this perfect capacity, and then uses the ELCCs of different resources to determine their contributions to resource adequacy.

Stakeholder feedback in PSE's IRP process has prompted PSE to integrate the impacts of climate change into load forecasts for its 2023 IRP. PSE plans to integrate climate change into its load forecasts using a method that the NWPCC developed for its 8th Power Plan released in 2021. Three different climate change models were selected that provide regional temperature forecasts of temperatures across seasons. PSE will integrate existing historical load data with the climate model forecasts to create an adjusted set of forecasts for heating degree days and cooling degree days.

C.1.3 Portland General Electric

Portland General Electric (PGE) is an electric utility based in Portland, Oregon with a service area consisting of 7 counties and 51 cities in Oregon. PGE has roughly 900,000 customers. PGE's historic peak load is 4,447 MW, which occurred in June 2021.

PGE uses two specific models to develop its reliability assessment within its 2019 IRP: (1) the Renewable Energy Capacity Planning model (RECAP) developed by Energy and Environmental Economics (E3); and (2) the production cost model Aurora developed by Energy Exemplar (PGE, 2019).

RECAP is a capacity planning and loss of load probability (LOLP) model. The model calculates the LOLP for each month, day type, and hour for a test year. RECAP also calculates the capacity needed to achieve a reliability target and the marginal capacity contributions for additional resources.

PGE adopted a reliability target defined by a loss of load expectation (LOLE) of no more than 2.4 hours per year, or one day in 10 years. RECAP was also used to derive the ELCCs of incremental resource changes to the portfolio.

PGE used the Aurora model to derive forecasts of wholesale electricity prices and economic dispatch across the entire Western Interconnection footprint. The consulting firm Wood Mackenzie developed a database of resources, loads, fuel prices and zonal topology, and transmission constraints. This model was used to develop hourly wholesale prices and alternative scenarios including changes to the price of carbon and different interconnection-wide resource buildouts. The modeling period for these scenarios was from 2020 to 2040.

C.1.4 Tucson Electric Power

Tucson Electric Power (TEP) developed its 2020 IRP with a resource adequacy analysis that evaluated the goal of significant increases in renewables and reductions in CO_2 emissions to 80% below 2005 levels by 2035 (TEP, 2020). The first step in TEP's RA analysis was to check whether TEP has resources with firm capacity to meet firm load obligations at the system peak with a 15% planning reserve margin. TEP expanded its RA analysis to evaluate system flexibility for meeting higher levels of renewable generation in its fleet. TEP hired a consultant, Siemens Industry Inc., to develop a flexible resource adequacy study. The study addressed two questions: (1) does TEP have adequate flexible capacity to meet loads with 30% renewable resources? (2) If the 30% target is feasible, how much more renewable energy can be integrated before additional flexibility resources are needed?

The TEP flexibility study developed six scenarios with a range of renewable energy penetration from 28%, 35%, and 50% and varying mixes of solar and wind. The study performed a Monte Carlo statistical analysis with 250 iterations of net load over one year to determine the 99th percentile values of peak net load for 3-hour net load ramps, 10-minute net load ramps, and curtailment levels of renewable generation. The maximum flexibility requirements were compared to the capacity and flexibility of TEP's 2024 portfolios.

A second modeling approach used the Aurora production cost model to simulate loads, renewable generation, and dispatch on a 10-minute interval for one year. This modeling effort internalized important resource constraints such as minimum generation limits, minimum up and down times, maximum ramp rates, outages, and transmission constraints. The model output was analyzed for flexibility shortfalls across the different portfolios. The results identified the amount of generation from each resource and the levels of curtailment under the different portfolios.

C.1.5 Public Service Company of Colorado - Xcel Energy

Public Service Company of Colorado (PSCo) is an operating subsidiary of Xcel Energy, a utility holding company based in Minneapolis, Minnesota. PSCo is the largest IOU operating in Colorado, providing electricity and natural gas to customers. PSCo serves the Denver metro area that includes 1,540,082 customers. Xcel Energy has three other operating companies: Northern States Power-Minnesota, Northern States Power in Wisconsin, and Southwestern Public Service Co. that operates in Texas and New Mexico.

Colorado policy requires PSCo to develop an Electric Resource Plan (ERP) and a Clean Energy Plan (CEP). PSCo's 2021 ERP/CEP had a planning period from 2021 to 2055. A key policy driver in the 2021 ERP/CEP was to meet a carbon reduction target of 80% by 2030 from 2005 levels (PSCo, 2021).

PSCo and the other Xcel subsidiary utilities recently adopted a new planning model called EnCompass that was developed by Anchor Power Solutions. Prior to Encompass, the Xcel utilities had used the Strategist planning model for 20 years. Xcel Energy switched planning models because of the greater complexity of the electricity system driven by the expansion of variable generation resources such as solar and wind, and battery energy storage. Xcel Energy intends to use EnCompass to develop and analyze capacity expansion plans and the corresponding production costs of those plans under a variety of scenarios and sensitivities. This model simultaneously solves the capacity expansion plan, production costs, environmental constraints, and ancillary service markets in a single co-optimization process. The model performs the production cost calculations in a chronological manner that enables it to account for operational constraints such as start times, ramp rates, and unit commitment.

PSCo developed resource portfolios in an iterative process using EnCompass. Initial portfolios were developed with various reliability requirements upfront in the model. A review of the modeling output would lead to adjustments that would be incorporated into the model for another round. This iterative process effectively enforces numerous reliability requirements such as reserve requirements, flex reserve requirements, and ELCC requirements.

PSCo hired a consultant, Astrape Consulting, to provide a planning reserve margin target for the 2021 ERP. The consultant used its Strategic Energy and Risk Model to perform 9,500 yearly simulations with one hour granularity at different reserve margins. The study found that for a LOLE of 0.1 standard, the corresponding reserve margin for the PSCo system would be 17.4%.

C.1.6 Duke Energy Carolinas

Duke Energy Carolinas (DEC) is a subsidiary of Duke Energy that operates a single utility system across both North Carolina and South Carolina. DEC owns about 23,200 MW of capacity that serves 2.7 million customers over 24,000 square miles of service territory.

In 2019, Duke Energy announced a corporate goal to reduce CO_2 emissions 50% from 2005 levels by 2030 and achieve net-zero emissions by 2050. In September of 2020, DEC released an IRP that proposed six portfolios that would lower carbon emissions over the 15-year time horizon of the IRP (DEC, 2020). These portfolios contemplated various combinations of coal plant retirements and expanding cleaner resources like solar, onshore and offshore wind, pumped hydro storage, energy efficiency, and small modular nuclear reactors. The North Carolina Public Utility Commission responded to stakeholder concerns and ordered DEC to modify the IRP with different modeling assumptions and nine supplemental portfolios. DEC's modified IRP was released in December 2020 (DEC, 2021).

Similar to Public Service Company of Colorado, DEC hired Astrape Consulting to perform the resource adequacy analysis for its 2020 IRP. Astrape recommended and DEC agreed to adopt the one day in 10 year LOLE standard. Astrape performed probabilistic modeling to derive a planning reserve margin under several different assumptions and cases. The first case (the “Island Case”) assumed that DEC would operate as an island and would not be able to buy or sell power from neighboring utilities. Modeling results of the Island Case showed that DEC would need a planning reserve margin of 22.5% to attain the 0.1 LOLE requirement. The second case (the “Base Case”) assumed that DEC would operate interconnected to other utilities in the region. DEC’s required reserve margin dropped to 16.0 % under a 0.1 LOLE standard when connected to a regional system. A third case (the “Combined Case”) assumed that DEC would operate within the broader Duke Energy utility system as a single balancing authority. DEC’s required margin was 16.75% under a 0.1 LOLE standard. After considering several other sensitivities, Astrape and DEC agreed to a 17.0 % target planning reserve margin for IRP purposes.

C.2 Reviews of RA assessments in IRP by ISO/RTO and regional organizations

This subsection reports a summary of RA practices for three ISOs: MISO, ISO-NE, and PJM, in addition to the non-ISO Northwest Power and Conservation Council.

C.2.1 MISO

The Midcontinent Independent System Operator (MISO) manages the high-voltage transmission of electricity across 15 states in the Midwest and the South (Arkansas, Illinois, Indiana, Iowa, Kentucky, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Montana, North Dakota, South Dakota, Texas, and Wisconsin) and the Canadian province of Manitoba. MISO serves 45 million customers covering 900,000 square miles. MISO has a generation capacity of 190 GW and about 68,000 miles of bulk transmission lines (MISO, 2023).

MISO first developed its Regional Resource Assessment (RRA) in 2021 to explore how the region’s generation portfolio might evolve over a 20-year time horizon to meet the MISO members’ and states’ stated objectives. In its 2022 RRA, MISO extended its region-wide basis analysis to local resource zones and provided more granular results (MISO, 2022a).

To run the resource assessment models, MISO collected the members’ carbon reduction and renewable target goals, planned additions, and planned retirements and ran optimizations (using the model called Electric Generation Expansion Analysis System (EGEAS)) to find the least-cost resource mix under various resource forecasts (MISO, 2022b). The model examines the flexibility needs of the region (i.e., flexibility assessment) and its ability to meet the established LOLE requirement (i.e., resource adequacy assessment) from 2022 to 2041.

In the resource adequacy analysis, MISO reviewed the potential impacts of changes in the planning reserve margin requirement and capacity contributions of wind, solar, solar plus storage, and battery and energy storage systems. The analysis requires a set of inputs, including load forecasts and renewable energy production profiles created for different historical years (2007 to 2012 and 2014 to 2018) under the pre-defined electrification and weather scenario (MISO Futures), generator forced outage assumptions, planned retirements and maintenance schedules, energy storage operation, and hybrid plant design and operation. The analysis was conducted using Plexos, which calculates the LOLE using the chronological Monte Carlo approach. MISO used an average of 0.1 days/year LOLE target in the analysis.

C.2.2 ISO-NE

ISO New England (ISO-NE) operates a competitive wholesale electricity market and high-voltage transmission system. ISO-NE serves six states (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont) with 7.2 million customers covering about 72,000 square miles. ISO-NE has a generation capacity of about 33 GW and roughly 9,000 miles of bulk transmission lines (ISO-NE, 2023).

ISO-NE conducts a comprehensive analysis at least once every three years to forecast electricity demands over the next ten years, determine whether the region has enough capacity to satisfy the resource adequacy requirements, and identify generation capacity and transmission expansion plans to address any identified needs. The most recent study (2021 Regional System Plan) was released in November 2021.

As the first step of the regional system planning process, the forecast of peak demand and annual electricity consumption was projected using gross energy modeling and gross demand modeling tools (both are regression models constructed using weather, economic, and time variables) with forecast inputs related to the economy and weather (Black and Rojo, 2019).[9] The forecasted electricity demands – both the forecast of the 50/50 peak loads and 90/10 peak loads (available in (ISO-NE, 2022)) – were then used to evaluate the resource adequacy and transmission planning processes. The system-wide and local-area capacity needs were calculated using the General Electric Multi-Area Reliability Simulation Program (GE MARS), which conducts a chronological Monte Carlo simulation to incorporate uncertainties in loads and resources under a wide range of existing and future system conditions (ISO-NE, 2021). The analysis used an average of 0.1 days/year LOLE target.

C.2.3 PJM

PJM operates a competitive wholesale electricity market and high-voltage transmission system. PJM serves 13 states (Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, and West Virginia) and the District of Columbia. PJM serves more than 65 million customers covering about 370,000 square miles. PJM has a generation capacity of 185 GW and about 85,000 miles of bulk transmission lines (PJM, 2022).

PJM conducts resource adequacy planning studies to determine the capacity resources needed to serve forecasted loads while satisfying reliability criteria. The planning study consists of three components. First is the Generator Availability Data System (GADS), which collects the availability of generating units over the past five years and identifies the parameters for generation forecasting. Second is load forecasting, which projects electricity demands for the next 15 years at the monthly level generated using multiple regression analysis of the hourly metered load data with estimated load drops on calendar effects, weather conditions, economic drivers, and end-use characteristics (PJM, 2021a). The third is installed reserve margin (IRM) analysis, which determines whether PJM has enough generation

[9] The demand forecast "considers demand history as an input," thus is expected to capture the growth of PV DERs in the demand forecasting model.

capacity to serve forecasted peak and IRM while satisfying the reliability criteria. PJM used the Probability Reliability Index Study Model (PRISM) to calculate PJM's reserve requirement while incorporating uncertainties from generation performance, load forecasting (due to economic and/or weather conditions), and connections to its adjacent regions (PJM, 2021b). The study used a 1 day in 10 years LOLE reliability standard when determining the IRM and forecast pool requirement estimates. However, PJM conducts zonal or locational reliability analyses that require a 1 day in 25 year LOLE standard, in contrast to the overall 1 day in 10 years for the PJM region. There are separate capacity market demand curves defined for each zone and prices can separate in the capacity auction depending on the relative costs and quantities of internal and external capacity and the available transmission.

The final reserve margin value calculating in the planning study serves as an input to the Reliability Pricing Model (RPM), which ensures long-term resource adequacy while benefitting from competitive pricing. RPM determines the plans to procure the region's unforced capacity obligation through market mechanisms. RPM consists of three mechanisms: 1) base residual auction to procure resources three years ahead and allocate the cost through a locational reliability charge, 2) incremental auctions to adjust resource procurement plans of delivery following the changes to the markets and reliability requirement adjustments, and 3) the bilateral market to cover any auction commitment shortages. The output of the RA planning study determines the base residual action's reliability requirement (PJM, 2021b).

C.2.4 NWPCC

The Northwest Power and Conservation Council (NWPCC) is not an ISO, but rather a regional organization that is organized under the 1980 Northwest Power Act. The council serves several important roles, including revisiting the power plan once every five years to ensure that its member states (Washington, Oregon, Idaho, and Montana) have an adequate, efficient, economical, and reliable electricity supply.

The 2021 Power Plan presents an electricity demand forecast for 2021 to 2041 and develops a resource procurement plan that supplies customer demands in a reliable manner. The resource adequacy assessment process consists of three steps. First, a range of electricity demand forecasts are developed based on the predetermined sets of key economic assumptions and weather scenarios. The NWPCC uses the Energy2020 model to develop monthly average and monthly peak load forecasts for various scenarios with differing long-term economic and end-use assumptions (Systematic Solutions Inc., 2021). Second, hourly load forecasts are derived using the output of the Energy2020 model and the NWPCC's econometric Short Term Load model that, among other things, incorporates the effects of temperature on loads. The NWPCC's current power plan incorporates climate change projected temperatures (loads), precipitation (river flows) and wind speeds, derived from three separate General Circulation Models, instead of previously used historical data. (NWPCC, 2021b). Finally, the forecasted demand is used to develop the regional resource strategy. The council uses three models to inform decisions: Aurora, Genesys and the Regional Portfolio Model (RPM).

The Aurora model estimates electricity prices and projects the resource buildout of the western grid (NWPCC, 2021a). The Genesys model is a Monte-Carlo chronological hourly simulation model that optimizes resource dispatch given uncertainties in (i) generator performance (i.e., forced outages), (ii) load forecast, (iii) wind, and solar generation and (iv) natural river flow. It sets a policy-driven limit for hourly imports of available market supply and incorporates the effect of prices on market dynamics. The NWPCC uses this model to assess future resource needs based on its current 5% annual LOLP adequacy standard, although is now testing a newly proposed multi-metric standard. Quarterly capacity and energy planning reserve margins needed to maintain adequacy are derived from the Genesys output, along with the effective capacity contributions (termed the associated system capacity contribution) for a variety of new resource portfolios. (NWPCC, 2023).

The Regional Portfolio Model determines cost-effective regional capacity expansion decisions that meet the quarterly planning reserve margin adequacy targets derived from the outputs of the Genesys model. The NWPCC then develops its resource strategy based on the expansion decisions and potential likelihoods of analyzed future scenarios.