Title

The importance of capturing power system operational details in resource adequacy assessments

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

Publication Date
2024-03-01

DOI
10.1016/j.epsr.2023.110057

The importance of capturing power system operational details in resource adequacy assessments

Benjamin D. Leibowicz[a], Nan Zhang[a], Juan Pablo Carvallo[b], Peter H. Larsen[b], Thomas Carr[c], Sunhee Baik[b]

[a] *Operations Research and Industrial Engineering, The University of Texas at Austin, ETC 5.128D, 204 E. Dean Keeton St., Austin, TX, 78712, United States*
[b] *Electricity Markets and Policy, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA, 94720, United States*
[c] *Western Interstate Energy Board, 1600 Broadway, Suite 1020, Denver, CO, 80202, United States*

Abstract

Traditional methods for assessing the resource adequacy (RA) of a power system are becoming obsolete due to emerging trends such as the increasing deployment of variable renewable energy and storage. Consequently, analysts are recommending that RA be assessed using a Monte Carlo simulation approach that models chronological power system operations over many instances of possible operating conditions. However, this approach is necessarily more complex and computationally demanding, which is an obstacle to real-world implementation. In this study, we investigate which operational details of power systems are important to capture in order to accurately evaluate a system's RA, versus details that add complexity but do not meaningfully affect RA results. To do so, we develop a probabilistic RA assessment framework by adapting an existing production cost model and apply it to a case study based on the IEEE Reliability Test System. Our results indicate that multi-year data, storage dispatch, and transmission limits are key details to incorporate. Accurate RA results can be obtained using non-economic dispatch strategies as long as they are coordinated with detailed operational strategies. We also demonstrate how popular expectation-based RA metrics can mask important differences in the characteristics of loss of load events.

Keywords: resource adequacy, long-term planning, probabilistic analysis, reliability, power system modeling

1. Introduction

1.1. Background on resource adequacy assessment

One formal definition of resource adequacy (RA) is provided by the North American Electric Reliability Corporation (NERC) [1], which defines RA as "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." The National Renewable Energy Laboratory (NREL) offers a more precise definition in the documentation of its Probabilistic Resource Adequacy Suite (PRAS). "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" [2]. This definition acknowledges that shortfalls may occur and suggests that achieving RA means keeping the risk of potential shortfalls within tolerable limits.

RA is only one dimension of power system reliability. The RA concept is employed for long-term planning and focuses on the balance between supply and demand over large areas. Reliability threats that are within the scope of RA include uncertainties to which probabilities can be assigned, such as variations in loads and variable renewable energy (VRE) generation as well as thermal generator outages. By contrast, low-probability, high-impact events such as natural disasters are typically considered through the lens of resilience rather than RA. While power deliverability via the transmission network is part of RA, transmission line outages are usually omitted from RA analysis (they are omitted in this study; [3] is one counterexample that includes them) and distribution line outages are always omitted. In this sense, RA always encompasses reliability threats in hierarchical level 1 (generation) of the power system according to the classification of Billinton and Allan [4, 5], but rarely threats in hierarchical level 2 (transmission) and never threats in level 3 (distribution). NERC explicitly distinguishes RA from operating reliability, which refers to reliability risks that are more granular in space and time [1]. Many of these more granular threats, such as short circuits or transformer overloads, are subsumed under power system security but not under RA [6].

Traditionally, RA has been assessed using simple methods and metrics [7]. Traditional RA analysis focuses on the peak load hour and develops a load forecast for some future year. Installed generation capacities are summed, perhaps weighted by coefficients less than one to reflect imperfect availability. As long as the anticipated capacity exceeds the projected load by more than some percentage, known as the planning reserve margin (PRM), then the system is deemed resource adequate. A slightly more complex yet still traditional approach is the convolution method. This probabilistic form of RA assessment assigns probabilities to power plant outage events, which are considered independent. By combining the resulting outage probability table with a load forecast, it is possible to estimate the probability that the system will have sufficient available capacity to serve the load [7, 8, 9].

Several emerging trends in power systems are making these traditional RA assessment methods obsolete. Many dispatchable generators are retiring and being replaced by VRE resources such as wind turbines and solar panels. VRE outputs during hours when the grid is stressed are highly uncertain and exhibit correlations among themselves and with loads that are not usually associated with thermal plants. Additionally, in systems with high VRE penetration, the peak load hour may no longer be the time when reliability is most at risk, since net loads (load minus VRE generation) could be highest during different seasons and hours that are hard to pinpoint in advance. Significant energy storage (e.g. batteries) may be added to power systems in the coming years and assessing its contributions to RA is difficult. In order for storage to contribute to satisfying the load in one hour, it must charge in the preceding hours and maintain sufficient energy until it is needed.

As a result of these trends and the problems they pose for traditional RA assessment methods, organizations such as the Energy Systems Integration Group (ESIG) have called for the development and implementation of more sophisticated approaches [7]. These approaches should model chronological power system operations, spanning the entire year and updating the state of the system (e.g. energy in storage) from one hour to the next. RA assessment should be probabilistic, with system operations simulated over many scenarios with different realizations of loads, VRE outputs, and generator outages. This Monte Carlo simulation enables analysts to compute probabilistic RA metrics that more intuitively correspond to notions of a reliable power system than a deterministic metric like the PRM. Common examples of probabilistic RA metrics include loss of load expectation (LOLE), loss of load probability (LOLP), loss of load hours (LOLH), loss of load events (LOLEV), and expected unserved energy (EUE) (see [10] for definitions of these metrics).

1.2. Motivation and contributions

In a Monte Carlo-style probabilistic RA assessment, there is an obvious tradeoff concerning the level of detail used in modeling power system operations. On the one hand, incorporating greater detail should allow the model to more closely resemble real-world operations and hence lead to more accurate estimates of RA metrics such as LOLE and EUE. Operational details that the model could incorporate include unit commitment and economic dispatch of thermal generators, storage dispatch, transmission limits, and day-ahead forecast errors (for loads and VRE outputs). On the other hand, representing power system operations in greater detail increases the complexity and computational burden of the RA assessment, especially because the power system operational model needs to be solved hundreds or thousands of times in a Monte Carlo analysis. High complexity and computational burden would be formidable obstacles to real-world implementation of probabilistic RA assessments, as traditional approaches are much simpler. The research gap that we address in this paper is that little is currently known about which operational details of power systems are critical to capture in RA assessments, as opposed to details that are not influential enough to justify their additional complexity. By providing evidence that certain operational details can be omitted without any notable effect on RA assessment accuracy, we help reduce barriers to real-world adoption of probabilistic RA assessment methods.

Our contributions in this paper are summarized as follows:

1. We vary choices about how to represent (or omit) certain operational details of the

power system and evaluate how much influence each one exerts on the outcomes of a probabilistic RA assessment.
2. We find that multi-year data, storage dispatch, and transmission limits are key features to incorporate into an RA assessment.
3. On the other hand, economic dispatch appears to be less necessary to represent. Accurate RA results can be obtained using non-economic dispatch as long as resources are dispatched according to detailed operational strategies.
4. We propose and calculate new RA metrics that reflect additional properties of shortfall events that are obscured by the most common expectation-based metrics, such as the number of shortfalls and their average magnitude.

1.3. Literature review

Several recent publications have criticized traditional RA assessment methods and offered guidelines for modern approaches that are better equipped to handle emerging trends in power systems. [7] and [11] contend that RA analysis should model power system operations chronologically, simulating hour-by-hour operations for an entire year. This is necessary to investigate all hours with potential adequacy risks and capture the RA contributions of energy-limited resources such as storage. They also recommend stochastic analysis that simulates operations over many years of hourly VRE output and load data as well as random plant outages. However, they acknowledge that it is difficult to generate credible time series data that reflect inter-annual variability and correlations among variables because historical records are limited [12]. Furthermore, due to climate change and technology adoption (e.g. electric vehicles), future conditions may differ considerably from historically observed patterns. [11] acknowledges that RA assessment methods that conform to its recommendations will necessarily be more complex and challenging to implement. In this paper, we seek to distinguish power system operational details that are important to represent in RA assessments from those details that can be safely omitted.

RA metrics are a common theme in the literature on RA. The high-level publications [7] and [11] point out that individual RA metrics measure different characteristics of shortfalls. For example, LOLE measures the frequency of shortfalls, LOLH measures their total duration, and EUE combines aspects of frequency and duration to measure the total magnitude of unserved energy (in MWh) over the year. In [13], Fazio and Hua conduct a probabilistic RA assessment of the Pacific Northwest power system and compute the LOLEV, LOLH, and normalized EUE metrics. They find that the value of one metric does not uniquely determine the values of the others. Since each metric captures a different dimension of RA risk, they suggest that power system planners specify independent thresholds for all three metrics to ensure acceptable overall RA performance. One limitation of all of these "expected value" RA metrics is that they are averages. According to [11], it is important that planners look beyond these expected value metrics to understand the timing, frequencies, magnitudes, and durations of the individual shortfall events that may occur so that they can choose the most appropriate mitigation strategies. In this paper, we calculate several of the most common RA metrics, investigate the characteristics of individual shortfall events, and summarize them using some new metrics that we develop.

In addition to [13], several other papers in the literature conduct probabilistic RA assessments using power system operation models, as we do in this study. In [14], Simoglou et al. developed the Long-Term Scheduling (LTS) tool that determines the hourly unit commitment and energy injection decisions that minimize total operating costs. They simulated the operations of the Greek power system from 2018-2027 under 12 scenarios with different loads, natural gas prices, CO_2 prices, and thermal unit retirements. On the basis of their computed values of LOLE and EUE, they conclude that thermal plant retirements pose considerable RA risks, especially from 2025 onwards. Simoglou and Biskas [15] use an extended version of the LTS platform to analyze the effects of Greece's National Energy and Climate Plan, which will shift the resource mix further toward renewables, on RA. Results show that LOLE and EUE generally remain within acceptable ranges in all scenarios, but that the planned retirement of a lignite unit by 2028 would cause a notable increase in LOLE. The European Resource Adequacy Assessments performed by ENTSO-E provide examples of state-of-the-art probabilistic RA assessments occurring in industry [3]. ENTSO-E employs a pan-European power system model and a Monte Carlo-style simulation, with many scenarios of VRE outputs, demand profiles, forced outages, and other uncertain parameters, in order to estimate the LOLE for each European country. In the Central Reference Scenario Without Capacity Mechanism, their assessment finds LOLE values above 20 hours/year in Germany and Luxembourg by 2030, much higher than in most other nations. ENTSO-E acknowledges that its analysis requires "computationally heavy models" and "thousands of computing hours," even without representing the details of each country's power system [3].

The most similar paper in the literature to our own is [16], in that Sun et al. systematically explored the impacts of various operational details on probabilistic RA assessment results. They considered different dispatch objectives and day-ahead forecast errors, as we do in this study. Sun et al. examined the impacts of certain details that we do not investigate, such as solving a unit commitment problem at the day-ahead stage. Meanwhile, we consider the effects of additional modeling choices that were not examined in [16], such as transmission limits and a variety of non-economic dispatch strategies for thermal generators and storage. We also propose and calculate new RA metrics that capture additional features of shortfall events that are obscured by popular metrics such as LOLE and EUE. [16] adds to the motivation for our work because they found that representations of operational details can strongly affect RA outcomes, with RA performance generally declining as the system is represented in more detail. They acknowledged that their results are specific to their case study based on the Electric Reliability Council of Texas (ERCOT) grid. Our IEEE Reliability Test System case study is based on the Southwestern United States, and thus it provides an opportunity to see whether some of the findings from [16] hold in a different context.

The remainder of this paper is organized as follows. Section 2 describes our methodology and Section 3 explains how we develop the data for our case study. In Section 4, we outline our scenarios that reflect different representations of power system operations for RA assessment. Section 5 presents and discusses our results. We conclude in Section 6 with a summary of our key findings and directions for future work.

2. Methodology

Our methodology adapts the open-source production cost model Prescient to the application of probabilistic RA assessment. Prescient is developed by Sandia National Laboratories to simulate electric grid operations [17]. Note that we choose to streamline a production cost model to run faster, many times within a Monte Carlo simulation, rather than begin with an existing probabilistic RA assessment tool and augment it with a sophisticated representation of economic dispatch. The latter is already available in Prescient, so we choose to modify it for our purposes and embed it within a Monte Carlo simulation.

Prescient 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. We make several modifications to Prescient to make it suitable for probabilistic RA assessment.

2.1. 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 [18], 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. The unit commitment and linearized merit order models both determine the optimal generation schedule to minimize the 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 [19]. Though we replace the unit commitment problem with the merit order model in the day-ahead stage, we keep the two-stage framework with day-ahead and real-time simulations. The day-ahead solution is used to set state of charge targets for storage in the real-time problem but is based on an imperfect forecast of operating conditions and does not anticipate thermal generator failures. For transmission, we apply a simplified transportation model to represent power flows between zones in order to maintain computational tractability.

2.2. Storage operations

An accurate representation of storage status and operations could be particularly important for RA assessment. If storage operations are neglected or oversimplified, such as by treating storage as a firm resource that can always provide maximum power, then system reliability would likely be overestimated since the actual energy in storage is not being monitored. This would not capture energy limitations, which are identified as being increasingly relevant for RA [11]. 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. 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 [20], 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 induces 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). See Appendix A for mathematical details about how SoC deviations are penalized in the objective and how storage operations are constrained.

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. Note that we do not incorporate the operations of long-duration storage – including reservoir hydropower and pumped storage – into our framework, as they are not relevant to our specific case study. However, incorporating 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.

2.3. Model objective functions

Power systems in the real world are dispatched economically with the goal of minimizing operational costs. 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 cost. Incorporating economic dispatch into RA assessment modeling is appealing because this is how power systems are 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 RA results.

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 either inspired by other RA models [2] or designed with strong RA performance as the goal. These non-economic formulations 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 rules developed for this work are described as follows:

- Minimize shortfalls: 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.

- Minimize renewable curtailment: Renewable generation in our framework is represented as having no variable 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.

- Minimize storage SoC deviation: In the real-time stage, each storage unit is dispatched according to its 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.

- Minimize expected power reduction due to thermal generator failures: 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 outlined above are integrated into a single objective function by assigning properly scaled penalty coefficients, as seen in Appendix A. We will compare scenarios based on these non-economic dispatch schemes to other scenarios in which the system is economically dispatched. Our scenarios are described in detail in Section 4.

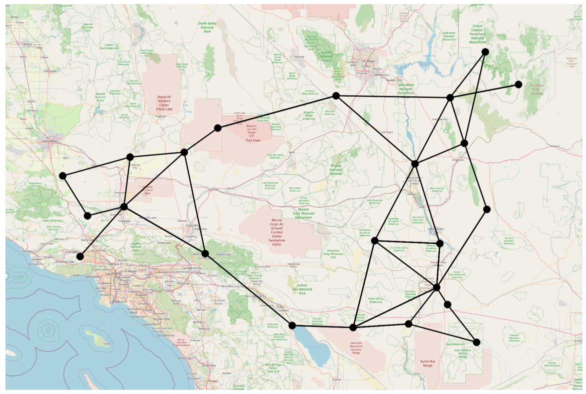

Figure 1: Zones and transmission network in our case study based on the IEEE RTS

3. Data preparation

We develop our case study based on the IEEE Reliability Test System - Grid Modernization Laboratory Consortium (RTS-GMLC) [21]. 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.

3.1. 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 these zones and aggregate them to produce a simplified transmission network with 35 lines. The zones and the simplified transmission network are depicted in Figure 1.

IEEE RTS represents a power system with strong RA performance. Computing RA metrics for this system would not clearly reveal the potential differences in RA assessment outcomes stemming from different modeling choices 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 from the system. In addition, we resdesign the system's short-term storage resources. 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 emphasized that the simplified test power system does not represent any real-world 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, our case study 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. However, we encourage readers to be cautious about extrapolating our conclusions to power systems and regions that differ in significant ways from our case study. Our probabilistic RA metrics computed via numerical simulation must be interpreted as specific to the IEEE RTS with our modifications outlined above.

3.2. Time series data expansion

IEEE RTS only includes a single year time series for loads and VRE generation. This is insufficient for conducting a probabilistic RA assessment, which requires simulating system operations over many years of data with different operating conditions that may arise. To address this shortcoming, we expand the one year of raw data into five years of load and renewable generation time series data in order to capture inter-annual variability.

Many methods have been proposed to capture the correlation among electricity time series [22, 12] and generate synthetic electricity data [23, 24]. 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 using the formula

$$X_t = T_t + W_t + D_t + R_t,$$

where X_t represents the triplet of raw load time series across three RTS regions, T_t represents the trend, W_t represents the weekly periodicity, D_t represents the daily periodicity, and R_t represents the residual. The residual is assumed to be a multivariate stationary time series, and we deploy the Probabilistic AutoRegressive model to generate synthetic residuals [25]:

$$\hat{R}_t = PAR(R_t).$$

The synthetic load time series are then recovered via

$$\hat{X}_t = T_t + W_t + D_t + \hat{R}_t.$$

We apply the above procedure to each four-week period of raw load data to capture differences in profiles throughout the year while maintaining a sufficiently large training dataset. A sample of synthetic load that we generate for one region is shown in Figure 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 output and load. By deploying this method, we expand the load and solar generation time series from one year to five years.

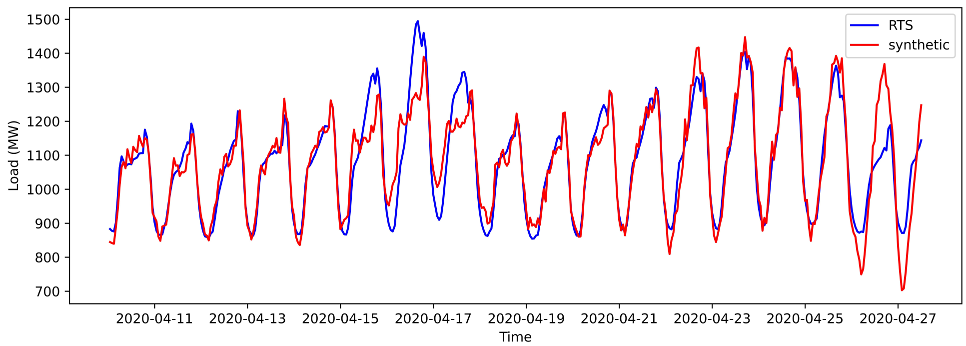

Figure 2: Synthetic load sample

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 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 [26]. 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. These four historical wind time series, plus the one-year time series included in the RTS dataset, combine to give us five yearly profiles of wind generation. Samples from two of the wind generation profiles, the RTS profile and one empirical profile, are shown in Figure 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.

We conclude the data preparation by constructing a Monte Carlo simulation set that consists of 500 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. While it would be ideal for accuracy to include even more than 500 instances, this number strikes

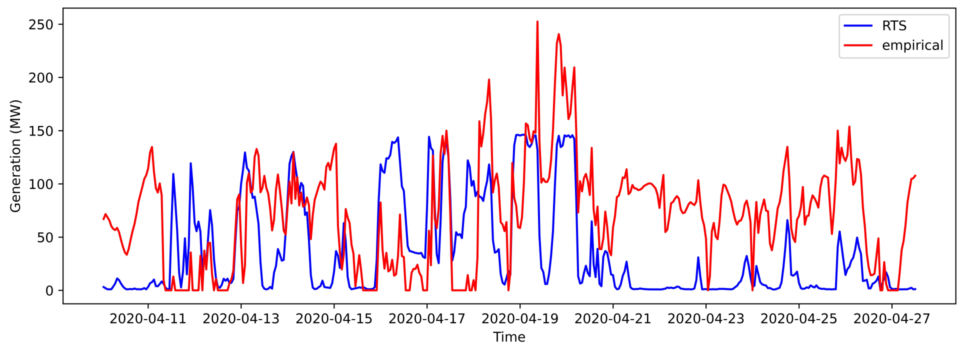

Figure 3: Samples from two wind generation time series, one from the RTS and one based on the WIND Toolkit data

a balance between the desire to obtain accurate results and the time required to solve our more complex dispatch formulations many times.

4. 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. We construct non-economic dispatch scenarios according to the following three strategies for thermal dispatch and four strategies for storage dispatch:

- 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 SoC target at that time. The target is given as an exogenous constant (50% charged).
  - Active strategy: Each storage unit will be dispatched to meet its predetermined 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.[1]

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 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 to explore how other power system modeling choices influence RA outcomes.

The four scenarios in this first set 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.

[1]Note that our Reserve storage dispatch strategy mirrors the strategy implemented in NREL's PRAS [2], in which "resources are dispatched conservatively so as to approximately minimize unserved energy over the full simulation horizon, charging from the grid whenever surplus generating capacity is available, and discharging only when needed to avoid or mitigate unserved energy."

Table 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

- *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. Note that the Sequential Monte Carlo simulation option in the existing PRAS software [2], which is its most detailed option for modeling system operations, essentially follows the logic of our *CsDp* scenario in that it dispatches generators and storage solely to maximize system reliability. Therefore, by comparing results from our *CsDp* scenario to those from *EcDp*, we will observe how important it is for RA outcomes to model economic instead of reliability-focused dispatch logic.

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 2, 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 2.

Table 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 and real-time	Reliability-focused	Active	Difficult

The *Conv* scenario is constructed according to [9]. 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 day-ahead forecasts and consideration of transmission adequacy. We develop the following two system sensitivity scenarios to examine the impacts of these other factors:

- *PfDF*: Perfect day-ahead forecast. In this scenario, the day-ahead forecast for load and VRE 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 include errors.[2]

- *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

[2]Day-ahead forecast errors in all scenarios other than *PfDf* are included in the IEEE RTS data and incorporated directly. The RTS data include day-ahead forecasts of hourly generation as well as actual generation profiles with 5-minute resolution, as described in [21, 27, 28].

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

Scenario	**Day-ahead forecast errors**	**Transmission limits**
RADp-4	Included	Included
PfDF	Excluded	Included
TxFr	Included	Excluded

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 can determine how important it is to account for potential transmission limits when assessing system RA.

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

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.

Before presenting our results, we conclude this section by summarizing the three scenario sets that we run, and the primary questions they are designed to investigate, in Table 4.

Table 4: Summary of our three sets of scenarios and the primary questions they are designed to investigate. Note that *RADp-4* appears in multiple scenario sets as a benchmark scenario.

Questions	What is the impact of using economic vs. non-economic dispatch strategies to assess the RA of a system?	How do less complex implementations of non-economic dispatch affect outcomes of RA assessments?	How does omitting day-ahead forecast errors or transmission limits affect outcomes of RA assessments?
Scenarios	*RADp-4* *EcDp* *AgDp* *CsDp*	*Conv* *RADp-1* *RADp-2* *RADp-3* *RADp-4*	*RADp-4* *PfDF* *TxFr*

5. Simulation results and anaylsis

5.1. Dispatch schemes, RA metrics, and costs

In this subsection, we examine the results of the four scenarios from Table 1, which were designed to help us investigate how the dispatch scheme implemented in the model (e.g.,

Table 5: Reliability and production cost outcomes for economic and reliability-focused 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

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 5 reports reliability and production cost outcomes for scenarios *RADp-4*, *EcDp*, *AgDp*, and *CsDp*. We observe that *AgDp* underestimates system reliability (relative to the *EcDp* baseline) and it also fails to reduce the production costs due to the lack of robustness against unforeseen thermal generator failures. RA performance in the *AgDp* scenario is very poor because it dispatches storage as a cost-free resource, meaning that there tends to be little energy in storage to cope with hours when load exceeds available generation capacity. By contrast, *CsDp* slightly overestimates system reliability as a result of operating storage units only 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 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.

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 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 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 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*, which better reflects the operation of real-world power systems. However, the economic impacts of operations become more significant, since thermal generators are usually running at their maximum capacities with high marginal production costs.

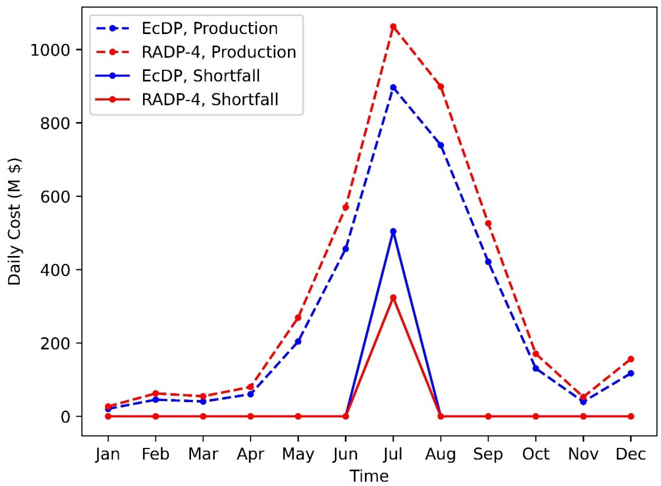

Figure 4: Production and shortfall costs in the *RADp-4* and *EcDp* scenarios, by month, in one simulation instance

5.2. Non-economic dispatch strategies

We analyze the impacts of different non-economic power dispatch strategies on RA assessments using the scenarios defined in Table 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 6 reports the RA outcomes of these non-economic dispatch scenarios as well as a few benchmarks. 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

Table 6: RA outcomes across non-economic dispatch strategies and several benchmarks

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

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*.[3] 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 RA assessments of modern power systems.

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, tracking the real-time status of storage units has substantial impacts on RA assessment results, 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 accuracy 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 occurrences 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 to some degree. 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

[3] See [29, 30, 31] for references on calculating storage capacity credit and the implications of representing storage's contributions to RA via its capacity credit.

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.

Histograms of shortfall magnitudes observed in *RADp-2* and *RADp-4* are presented in Figure 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 (featured in *RADp-4*) is close to the theoretical optimal for RA purposes, where a storage unit can discharge its maximum power during critical hours.

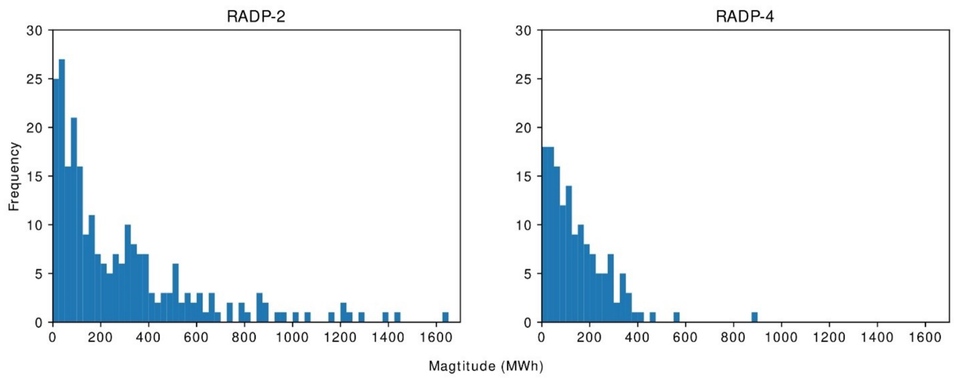

Figure 5: Shortfall magnitude histograms for *RADp-2* and *RADp-4* over the whole simulation set

Figure 6 shows the total number of shortfall hours observed in each month for the *RADp-2*, *RADp-3*, and *RADp-4* scenarios, over the whole simulation set. Loss of load events in the case study system occur exclusively in summer when the demand is significantly higher than in the other seasons. Though they have different LOLH in the simulation, the three scenarios show consistent temporal shortfall patterns that span the same range of days and hours. This suggests that even though models with simplistic operations 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.

5.3. System representation sensitivities

The results in the preceding subsections showed that the strategies used to dispatch thermal generators and storage units have a strong 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 7 summarizes the impacts

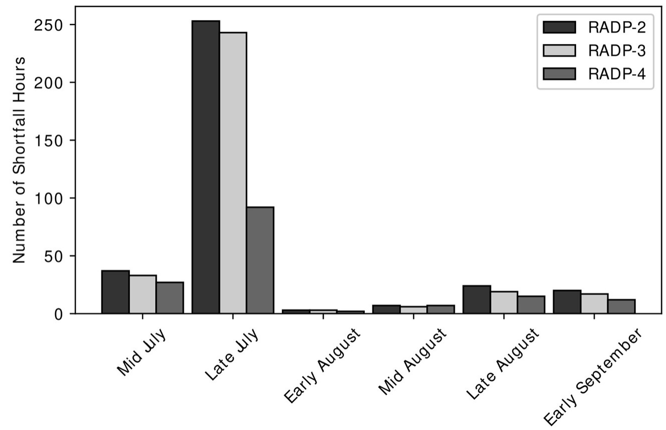

Figure 6: Number of shortfall hours in each month over the whole simulation set

Table 7: Effects of ignoring day-ahead forecast errors or transmission limits on RA outcomes

Scenario	**LOLE (days/year)**	**LOLP**	**EUE (MWh/year)**
RADp-4	0.286 ± 0.024	$24.6 \pm 1.9\%$	41.92 ± 4.84
PfDF	0.286 ± 0.024	$24.6 \pm 1.9\%$	41.92 ± 4.84
TxFr	0.220 ± 0.022	$19.0 \pm 1.7\%$	36.49 ± 4.87

of two specific modeling assumptions on RA assessment results. *PfDF* 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 *PfDF* are the same as those from the benchmark scenario *RADp-4*. It suggests that day-ahead forecast errors have limited impacts on the accuracy of RA assessments, although we acknowledge that this outcome could be different if unit commitment is fully represented.

Investigating the effects of neglecting transmission constraints is relevant because many current RA assessments do not represent them. It follows that measuring the impact of this assumption within our model can shed light on the potential consequences of this common 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. We will examine the effects of representing transmission constraints on more detailed characteristics of shortfalls in the following subsection.

5.4. Shortfall characteristics and additional RA metrics

We have demonstrated that different modeling assumptions can lead to substantial changes in RA assessment results. 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 1.3, 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 7 shows the distribution of hourly EUE over 100 random thermal generator failure instances for each of 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. Understanding the times and durations of shortfalls could be important for identifying appropriate RA enhancement strategies (e.g., solar PV will contribute more to daytime RA than nighttime RA) and because the costs of power interruptions depend on their timing and are not necessarily proportional to their duration. 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.

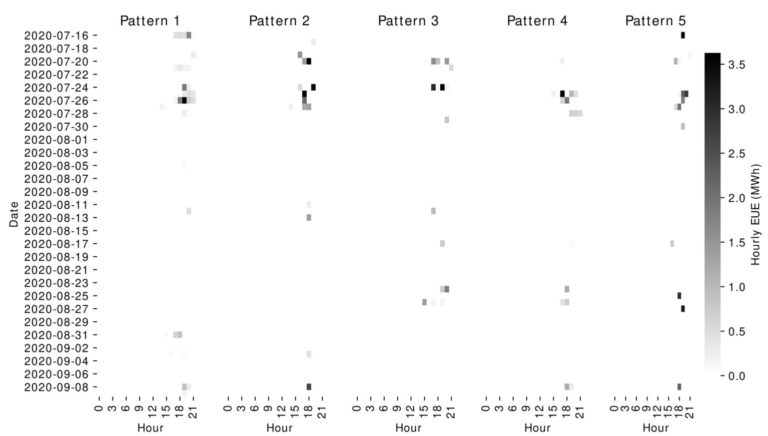

Figure 7: Heat map of hourly EUE in scenario *RADp-4* for each of the five weather patterns

We now explore whether additional RA metrics can better capture the nuanced effects of power system modeling choices on RA outcomes. Table 8 reports several RA statistics based on the shortfall records in the system sensitivity scenarios, evaluated using the alternative metrics suggested in [32] 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 *PfDF* are still the same as those in *RADp-4*, which is consistent with the previous finding that day-ahead forecast errors are unlikely to be a key operational detail to accurately evaluate RA, at least when unit commitment is not considered.

Table 8: Additional RA metrics for the *RADp-4*, *PfDF*, and *TxFr* scenarios

Metric (Unit)	***RADp-4***	***PfDF***	***TxFr***
Number of shortfalls	144	144	111
Maximum shortfall magnitude (MWh)	877.25	877.25	877.25
Average shortfall magnitude (MWh)	145.55	145.55	164.36
Maximum 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
Maximum shortfall duration (hours)	2	2	2
Average shortfall duration (hours)	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, transmission limits can create smaller shortfalls that are "hidden" when these limits are neglected. These smaller shortfalls reduce 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.

6. Conclusions

Traditional RA assessments that do not model chronological power system operations have become obsolete for evaluating the reliability performance of modern power systems due to several rapid changes in the electricity industry. In this analysis, we 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 9, 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 9: 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
Day-ahead forecast error	Low	High

Several high-level findings emerged from the case study explored using our probabilistic RA assessment framework:

- Non-economic dispatch schemes that ignore economic objectives can lead to fairly accurate RA assessments when coordinated with detailed operational strategies.
- Representing the detailed chronological operations of thermal generators and storage units is essential for accurate RA assessments, but simplified dispatch models could have value as screening tools (e.g., to identify critical hours for RA) due to their low computational complexity.
- Multi-year data is important to include to capture inter-annual variations in system operating conditions.
- Neglecting to incorporate transmission limits into RA assessment could lead to substantial underestimation of traditional "expected value" RA metrics. However, neglecting transmission limits may not mask the largest 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.

Acknowledgments

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

Appendix A. Mathematical formulations for selected model components

Nomenclature

Sets and Indices:

D, d	days
H, h	hours
N, n	nodes
T, t	transmission lines
G, g	thermal generators
R, r	renewable generators
X, x	storage units

Parameters:

I_g, I_r	installed capacity of thermal generator g or renewable generator r
I_g^{min}	minimum run rate for thermal generator g
λ_g	forced outage rate for thermal generator g
c_g	convex production cost for thermal generator g
I_x	installed power capacity for storage unit x
η_x	duration for storage unit x
I_x^c, I_x^d	charging and discharging limits for storage unit x
I_t	transmission limit for transmission line t
$\rho_x^c, \rho_x^d, \rho_x^s$	charging, discharging, and storage efficiency for storage unit x
$l_{n,d,h}$	load at node n at a given time (day d, hour h)
$k_{r,d,h}$	capacity factor for renewable generator r at a given time
$\sigma_{g,d,h}$	active/outage indicator ($\{1, 0\}$) for thermal generator g at a given time

Decision Variables:

$L_{n,d,h}^{srv}, L_{n,d,h}^{umt}$	served load and shortfall at node n at a given time
$P_{g,d,h}$	power dispatch for thermal generator g at a given time
$P_{r,d,h}, P_{r,d,h}^{crt}$	power dispatch and curtailment for renewable generator r at a given time
$P_{x,d,h}^c, P_{x,d,h}^d$	storage charge and discharge for storage unit x at a given time
$S_{x,d,h}$	SoC for storage unit x at a given time
$F_{t,d,h}$	transmission flow on line t at a given time

Other Elements:

$l_{r,d,h}^{fct}$	forecast load and shortfall at node n at a given time
$k_{r,d,h}^{fct}$	forecast capacity factor for renewable generator r at a given time
$S_{x,d,h}^{tgt}$	SoC target for storage unit x at a given time

Hourly dispatch problem for day d, hour h

- Objective function

$$Minimize \quad \omega_1 \sum_{n \in N} L_{n,d,h}^{umt} + \omega_2 \sum_{r \in R} P_{r,d,h}^{crt} + \omega_3 \sum_{x \in X} |S_{x,d,h} - S_{x,d,h}^{tgt}| + \omega_4 \sum_{g \in G} \lambda_g P_{g,d,h}$$

Notes about the objective function:

- ω_1, ω_2, ω_3, and ω_4 are the penalty coefficients.
- The four terms are total shortfalls, total renewable curtailment, total storage SoC deviation, and total expected power reduction due to thermal generator failures.
- Regarding storage dispatch strategies, $S^{tgt}_{x,d,h}$ is set to 50% for the Passive strategy, set to 0% for the Cost-free resource strategy, set to 100% for the Reserve strategy, and obtained from the day-ahead dispatch solution in the Active strategy.
- Regarding thermal dispatch strategies, λ_g is replaced by 0 in the Random-priority strategy, and it is replaced by c_g in the Economic strategy.

- Main constraints
  - Satisfy the load with thermal and renewable dispatch, net storage discharge, and net incoming power from the transmission system at each node in each hour:

$$\sum_{g\in G(n)} P_{g,d,h} + \sum_{r\in R(n)} P_{r,d,h} + \sum_{x\in X(n)} P^d_{x,d,h} - \sum_{x\in X(n)} P^c_{x,d,h} + \sum_{ni\in T} F_{ni,d,h} - \sum_{no\in T} F_{no,d,h} = L^{srv}_{n,d,h}$$
$$L^{srv}_{n,d,h} + L^{umt}_{n,d,h} = l_{n,d,h}, \forall n \in N$$

  - Renewable generation for each renewable generator in each hour:

$$P_{r,d,h} + P^{crt}_{r,d,h} = k_{r,d,h} I_r, \forall r \in R$$

  - Thermal generation for each active thermal generator in each hour:

$$\sigma_{g,d,h} I^{min}_g \leq P_{g,d,h} \leq \sigma_{g,d,h} I_g, \forall g \in G$$

  - Transmission limit for each transmission line:

$$L_{t,d,h} \leq I_t, \forall t \in T$$

  - Storage operation for each storage unit in each hour (knowing $S_{x,d,h-1}$ from the previous hour's system state):

$$\eta_x S_{x,d,h-1} I_x = \rho^s_x \eta_x S_{x,d,h-1} I_x - \frac{1}{\rho^s_d} P^d_{x,d,h} + \rho^c_x P^c_{x,d,h}$$
$$\frac{P^d_{x,d,h}}{I^d_x} + \frac{P^c_{x,d,h}}{I^c_x} \leq 1, \forall x \in X$$

Day-ahead forecast problem for day d

- Objective function

$$Minimize \quad \sum_{h\in H} (\omega_1 \sum_{n\in N} L^{umt}_{n,d,h} + \omega_2 \sum_{r\in R} P^{crt}_{r,d,h})$$

- Main constraints
  - Satisfy the load with thermal and renewable dispatch, net storage discharge, and net incoming power from the transmission system at each node in each hour:

$$\sum_{g\in G(n)} P_{g,d,h} + \sum_{r\in R(n)} P_{r,d,h} + \sum_{x\in X(n)} P^{d}_{x,d,h} - \sum_{x\in X(n)} P^{c}_{x,d,h} + \sum_{ni\in T} F_{ni,d,h} - \sum_{no\in T} F_{no,d,h} = L^{srv}_{n,d,h}$$
$$L^{srv}_{n,d,h} + L^{umt}_{n,d,h} = l^{fct}_{n,d,h}, \forall n \in N, h \in H$$

  - Renewable generation forecast for each renewable generator in each hour:

$$P_{r,d,h} + P^{crt}_{r,d,h} = k^{fct}_{r,d,h} I_r, \forall r \in R, h \in H$$

  - Thermal generation forecast for each active thermal generator in each hour:

$$I^{min}_{g} \leq P_{g,d,h} \leq I_g, \forall g \in G, h \in H$$

  - Transmission limit for each transmission line:

$$L^{t,d,h} \leq I_t, \forall t \in T, h \in H$$

  - Daily storage operation forecast for each storage unit in each hour:

$$\eta_x S_{x,d,h-1} I_x = \rho^{s}_{x} \eta_x S_{x,d,h-1} I_x - \frac{1}{\rho^{s}_{d}} P^{d}_{x,d,h} + \rho^{c}_{x} P^{c}_{x,d,h}$$
$$\frac{P^{d}_{x,d,h}}{I^{d}_{x}} + \frac{P^{c}_{x,d,h}}{I^{c}_{x}} \leq 1, \forall x \in X, h \in H$$

The optimal solution $\{S^{*}_{x,d,h}\}$ will serve as the predetermined SoC targets $\{S^{tgt}_{x,d,h}\}$ for hourly storage dispatch with the Active strategy.

References

[1] NERC, 2022 Long-Term Reliability Assessment, Tech. Rep. December, North American Electric Reliability Corporation (NERC) (2022).

[2] G. Stephen, Probabilistic Resource Adequacy Suite (PRAS) v0.6 Model Documentation, National Renewable Energy Laboratory Technical Report (2021) NREL/TP–5C00–79698.

[3] ENTSO-E, European Resource Adequacy Assessment: 2022 Edition, Tech. rep., European Network of Transmission System Operators for Electricity (ENTSO-E) (2022).

[4] R. Billinton, R. N. Allan, Reliability Evaluation of Power Systems, Springer, New York, 1996.

[5] R. Allan, R. Billinton, Power system reliability and its assessment. Part 1 background and generating capacity, Power Engineering Journal 6 (4) (1992) 191–198.

[6] E. Bompard, T. Huang, Y. Wu, M. Cremenescu, Classification and trend analysis of threats origins to the security of power systems, International Journal of Electrical Power and Energy Systems 50 (1) (2013) 50–64. doi:10.1016/j.ijepes.2013.02.008.

[7] Energy Systems Integration Group, Redefining Resource Adequacy for Modern Power Systems, Tech. rep., A Report of the Redefining Resource Adequacy Task Force, Reston, VA (2021).
URL `https://www.esig.energy/wp-content/uploads/2022/12/ESIG-Redefining-Resource-Adec`

[8] R. Booth, Power System Simulation Model Based on Probability Analysis, IEEE Transactions on Power Apparatus and Systems PAS-91 (1) (1972) 62–69. doi:10.1109/TPAS.1972.293291.

[9] E. Preston, C. Barrows, Evaluation of year 2020 IEEE RTS generation reliability indices, 2018 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2018 - Proceedings (x) (2018) 1–5. doi:10.1109/PMAPS.2018.8440394.

[10] J. P. Carvallo, N. Zhang, B. D. Leibowicz, T. Carr, M. Galbraith, P. H. Larsen, Implications of a regional resource adequacy program for utility integrated resource planning, The Electricity Journal 34 (2021) 106960. doi:10.1016/j.tej.2021.106960.

[11] D. Stenclik, A. Bloom, W. Cole, G. Stephen, A. F. Acevedo, R. Gramlich, C. Dent, N. Schlag, M. Milligan, Quantifying Risk in an Uncertain Future: The Evolution of Resource Adequacy, IEEE Power and Energy Magazine 19 (6) (2021) 29–36. doi:10.1109/MPE.2021.3104076.

[12] I. Vehvilainen, Joint assessment of generation adequacy with intermittent renewables and hydro storage: A case study in Finland, Electric Power Systems Research 199 (2021) 107385. doi:10.1016/j.epsr.2021.107385.

[13] J. Fazio, D. Hua, Three probabilistic metrics for adequacy assessment of the Pacific Northwest power system, Electric Power Systems Research 174 (2019) 105858. doi:10.1016/j.epsr.2019.04.036.

[14] C. K. Simoglou, E. A. Bakirtzis, P. N. Biskas, A. G. Bakirtzis, Probabilistic evaluation of the long-term power system resource adequacy: The Greek case, Energy Policy 117 (March) (2018) 295–306. doi:10.1016/j.enpol.2018.02.047.

[15] C. K. Simoglou, P. N. Biskas, Assessment of the impact of the National Energy and Climate Plan on the Greek power system resource adequacy and operation, Electric Power Systems Research 194 (2021) 107113. doi:10.1016/j.epsr.2021.107113.

[16] Y. Sun, B. Frew, S. Dalvi, S. C. Dhulipala, Insights into methodologies and operational details of resource adequacy assessment: A case study with application to a broader flexibility framework, Applied Energy 328 (2022) 120191. doi:10.1016/j.apenergy.2022.120191.

[17] J. P. Watson, R. Knueven, Bernard Concepcion, D. Melander, A. Short, D. Zhang, Pengchu Woodruff, USDOE, Prescient, Version 1.0 (2020).
URL `https://www.osti.gov//servlets/purl/1695703`

[18] L. Wu, M. Shahidehpour, T. Li, Cost of reliability analysis based on stochastic unit commitment, IEEE Transactions on Power Systems 23 (3) (2008) 1364–1374. doi:10.1109/TPWRS.2008.922231.

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

[20] Z. Shen, W. Wei, D. Wu, T. Ding, S. Mei, Modeling arbitrage of an energy storage unit without binary variables, CSEE Journal of Power and Energy Systems 7 (1) (2021) 156–161. doi:10.17775/CSEEJPES.2019.03340.

[21] C. Barrows, E. Preston, A. Staid, G. Stephen, J. P. Watson, A. Bloom, A. Ehlen, J. Ikaheimo, J. Jorgenson, D. Krishnamurthy, J. Lau, B. McBennett, M. O'Connell, The IEEE Reliability Test System: A Proposed 2019 Update, IEEE Transactions on Power Systems 35 (1) (2020) 119–127. doi:10.1109/TPWRS.2019.2925557.

[22] C. L. T. Borges, J. A. Dias, A Model to Represent Correlated Time Series in Reliability Evaluation by Non-Sequential Monte Carlo Simulation, IEEE Transactions on Power Systems 32 (2) (2017) 1511–1519. doi:10.1109/TPWRS.2016.2585619.

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

[24] P. W. Talbot, C. Rabiti, A. Alfonsi, C. Krome, M. R. Kunz, A. Epiney, C. Wang, D. Mandelli, Correlated synthetic time series generation for energy system simulations using Fourier and ARMA signal processing, International Journal of Energy Research 44 (10) (2020) 8144–8155. doi:10.1002/er.5115.

[25] N. Patki, R. Wedge, K. Veeramachaneni, The synthetic data vault, Proceedings - 3rd IEEE International Conference on Data Science and Advanced Analytics, DSAA 2016 (2016) 399–410doi:10.1109/DSAA.2016.49.

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

[27] R. N. Allan, R. Billinton, N. M. K. Abdel-Gawad, The IEEE Reliability Test System – Extensions to and Evaluation of the Generating System, IEEE Transactions on Power Systems 1 (4) (1986) 1–7. doi:10.1109/TPWRS.1986.4335006.

[28] D. Lew, G. Brinkman, E. Ibanez, A. Florita, M. Heaney, B.-M. Hodge, M. Hummon, G. Stark, J. King, S. Lefton, N. Kumar, D. Agan, G. Jordan, S. Venkataraman, The Western Wind and Solar Integration Study Phase 2, Tech. rep., National Renewable Energy Laboratory (2013).
URL https://www.nrel.gov/docs/fy13osti/55588.pdf

[29] G. Stephen, T. Joswig-Jones, S. Awara, D. Kirschen, Impact of Storage Dispatch Assumptions on Resource Adequacy and Capacity Credit, 2022 17th International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2022 (2022) 1–6doi:10.1109/PMAPS53380.2022.9810584.

[30] A. W. Frazier, W. Cole, P. Denholm, D. Greer, P. Gagnon, Assessing the potential of battery storage as a peaking capacity resource in the United States, Applied Energy 275 (June) (2020) 115385. doi:10.1016/j.apenergy.2020.115385.

[31] A. D. Mills, P. Rodriguez, A simple and fast algorithm for estimating the capacity credit of solar and storage, Energy 210 (2020) 118587. doi:10.1016/j.energy.2020.118587.

[32] Northwest Power and Conservation Council (NWPCC), Pacific Northwest Power Supply Adequacy Assessment for 2023, Tech. rep. (2018).
URL https://www.nwcouncil.org/sites/default/files/2018-7.pdf