Western Interconnection Flexibility Assessment

Final Report

December 2015

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Energy+Environmental Economics

Western Interconnection Flexibility Assessment

Final Report

December 2015

Energy and Environmental Economics, Inc.
101 Montgomery Street, Suite 1600
San Francisco, CA 94104
415.391.5100
www.ethree.com

Project Team:
Nick Schlag, Arne Olson, Elaine Hart, Ana Mileva, Ryan Jones (E3)
Carlo Brancucci Martinez-Anido, Bri-Mathias Hodge, Greg Brinkman, Anthony Florita, David Biagioni (NREL)

Table of Contents

About this Study

This study was jointly undertaken by the Western Electricity Coordinating Council (WECC) and the Western Interstate Energy Board (WIEB) to investigate the need for power system flexibility to ensure reliable and economic operations of the interconnected Western electricity system under higher penetrations of variable energy resources. WECC and WIEB have partnered with Energy & Environmental Economics, Inc. (E3) and the National Renewable Energy Laboratory (NREL) to investigate these questions using advanced, stochastic reliability modeling and production cost modeling techniques. The study identifies and examines operational challenges and potential enabling strategies for renewable integration under a wide range of operating conditions, scenarios, and sensitivities across the Western Interconnection, with the goal of providing guidance to operators, planners, regulators and policymakers about changing system conditions under higher renewable penetration.

Funding for this project is provided by a number of sources. E3's role in the project is funded jointly by WECC and WIEB through grants received under the Department of Energy's American Recovery and Reinvestment Act (ARRA); NREL's role is funded directly by the Department of Energy.

Technical Review Committee

This study was overseen by a Technical Review Committee (TRC) comprising representatives from utilities, regulatory agencies, and industry throughout the Western Interconnection:

- Aidan Tuohy, Electric Power Research Institute
- Ben Kujala, Northwest Power Planning & Conservation Council
- Brian Parsons, Western Grid Group
- Dan Beckstead, Western Electricity Coordinating Council
- Fred Heutte, Northwest Energy Coalition
- James Barner & Bingbing Zhang, Los Angeles Department of Water & Power
- Jim Baak, Vote Solar Initiative
- Justin Thompson, Arizona Public Service
- Keith White, California Public Utilities Commission
- Michael Evans, Shell Energy North America
- Thomas Carr, Western Interstate Energy Board
- Thomas Edmunds, Lawrence Livermore National Laboratory
- Tom Miller, Pacific Gas & Electric Company

The TRC met a number of times throughout the course of the project to provide input and guidance on technical modeling decisions, to help interpret analysis, and to craft the study's ultimate findings and conclusions. The feedback of the TRC throughout the project was highly valuable to the project.

Executive Summary

Background

Over the past decade, the penetration of renewable generation in the Western Interconnection has grown rapidly: nearly 30,000 MW of renewable generation capacity—mostly solar and wind—have been built. By 2024, under current state policies, the total installed capacity of renewable generation in the Western Interconnection may exceed 60,000 MW. In front of this landscape of increasing renewable policy targets and declining renewable costs, interest in renewable generation and understanding its impacts on electric systems has surged recently. Regulators, utilities, and policymakers have begun to grapple with the potential need for power system "flexibility" to ensure reliable operations under high penetrations of renewable generation.

The Western Electricity Coordinating Council (WECC), in its role as the Regional Entity responsible for reliability in the Western Interconnection, is interested in understanding the long-term adequacy of the interconnected western grid to meet the new operational challenges posed by wind and solar generation across a range of plausible levels of penetration. WECC stakeholders have expressed a similar interest through study requests to examine high renewable futures that that implicate operational and flexibility concerns. The Western Interstate Energy Board (WIEB) seeks to understand these issues in order to inform policymakers about the implications of potential future policies targeting higher renewable penetrations.

With this motivation, WECC and WIEB collaborated to jointly sponsor this WECC Flexibility Assessment. The sponsors established three goals for this effort:

+ **Assess the ability of the fleet of resources in the Western Interconnection to accommodate high renewable penetrations while maintaining reliable operations.** Higher penetrations of renewable generation will test the flexibility of the electric systems of the West by requiring individual power plants to operate in fundamentally new ways, changing operating practices as well as the dynamics of wholesale power markets. This study aims to identify the major changes in operational patterns that may occur at such high penetrations and to measure the magnitude and frequency of possible challenges that may result.
+ **Investigate potential enabling strategies to facilitate renewable integration that consider both institutional and physical constraints on the Western system.** Existing literature has identified a wide range of possible strategies that may facilitate the integration of high penetrations of renewables into the Western Interconnection. These strategies comprise both institutional changes—for example, increased use of curtailment as an operational strategy and greater regional coordination in planning and operations—as well as physical changes to the electric system—new investments in flexible generating resources and the development of new demand side programs. This study examines how such measures can mitigate the challenges that arise with interconnection-wide increases in renewable penetration.
+ **Provide lessons for future study of system flexibility on the relative importance of various considerations in planning exercises.** The study

of flexibility and its need at high renewable penetrations is an evolving field. This effort is designed with an explicit goal of providing useful information to modelers and technical analysts to improve analytical capabilities for further investigation into the topics explored herein.

Assessment Methods

This study is organized into two sequential phases of analysis. The first phase, a resource adequacy assessment of the generation fleet, uses traditional loss-of-load-probability techniques to ensure that the electric system adheres to a traditional "one-day-in-ten-year" planning standard for loss of load. The second phase, the flexibility assessment, uses stochastic production cost analysis to examine the degree to which operational challenges are encountered with the additional of renewables to the system. By nesting the flexibility assessment within a traditional study of resource adequacy, the approach used in this study seeks to ensure that any challenges encountered in operations can be attributed to a lack of flexibility and are not simply the result of a system whose available capacity is inadequate to meet its peak demands.

The first phase of the analysis uses E3's **Renewable Energy Capacity Planning (RECAP)** model, a loss-of-load probability (LOLP) model designed to evaluate resource adequacy under high renewable penetrations. The RECAP analysis confirmed that the modeled resources across the study area were capable of meeting or exceeding the traditional one-in-ten loss of load frequency standard.[1] The second phase uses the **Renewable Energy Flexibility (REFLEX)** model for **PLEXOS for Power Systems**—an adaptation of traditional production cost modeling that incorporates principles from more traditional reliability analysis—to examine the impacts of high penetrations of renewable generation on the operations of these resources. Production simulation models are used for a variety of purposes, but the approach taken in this study has been tailored directly to the examination of flexibility challenges under high renewable penetrations.

While loss-of-load probability modeling is the de facto standard method for assessing conventional capacity adequacy, there is no analogous industry-standard approach for assessing the adequacy of flexibility in an electric system. A number of approaches have been explored, and a variety of metrics have been proposed as a means of measuring flexibility adequacy—for instance, "Expected Unserved Ramp."[2] While these metrics may be useful as *indicators* of power system inflexibility, they are not directly actionable because there is no standard for unserved ramping that power systems are required to meet. Rather than attempting to define a new metric, the REFLEX production cost modeling approach used in this study is designed to measure the *consequences* of inadequate flexibility, thereby illustrating a method through which actionable information can be provided to planners and decision-makers.

[1] The results of the Phase 1 analysis, summarized in the body of this report, were also published as a standalone report available at: **http://westernenergyboard.org/wp-content/uploads/2015/06/04-2015-WECC-WIEB-Flexibility-Assessment-Report-Interim.pdf**

[2] See, for example, EPRI's "Power System Flexibility Metrics: Framework, Software Tool and Case Study for Considering Power System Flexibility in Planning," available at: **http://www.epri.com/abstracts/Pages/ProductAbstract.aspx?productId=000000003002000331**.

In traditional LOLP analysis, the consequence of inadequate capacity is straightforward: the system is incapable of simultaneously meeting all loads, so some loads cannot be served. This is measured using conventional metrics such as LOLP, Loss of Load Expectation (LOLE), and Expected Unserved Energy (EUE). Inadequate flexibility can also lead to loss of load, even for systems that would otherwise be resource adequate. For example, if a large portion of the thermal fleet is cycled off to accommodate high output from renewable generators, the system might not have adequate resources committed to meet a large upward net load ramp (this is illustrated in Figure 1a). REFLEX therefore tracks EUE as one measure of power system inflexibility.

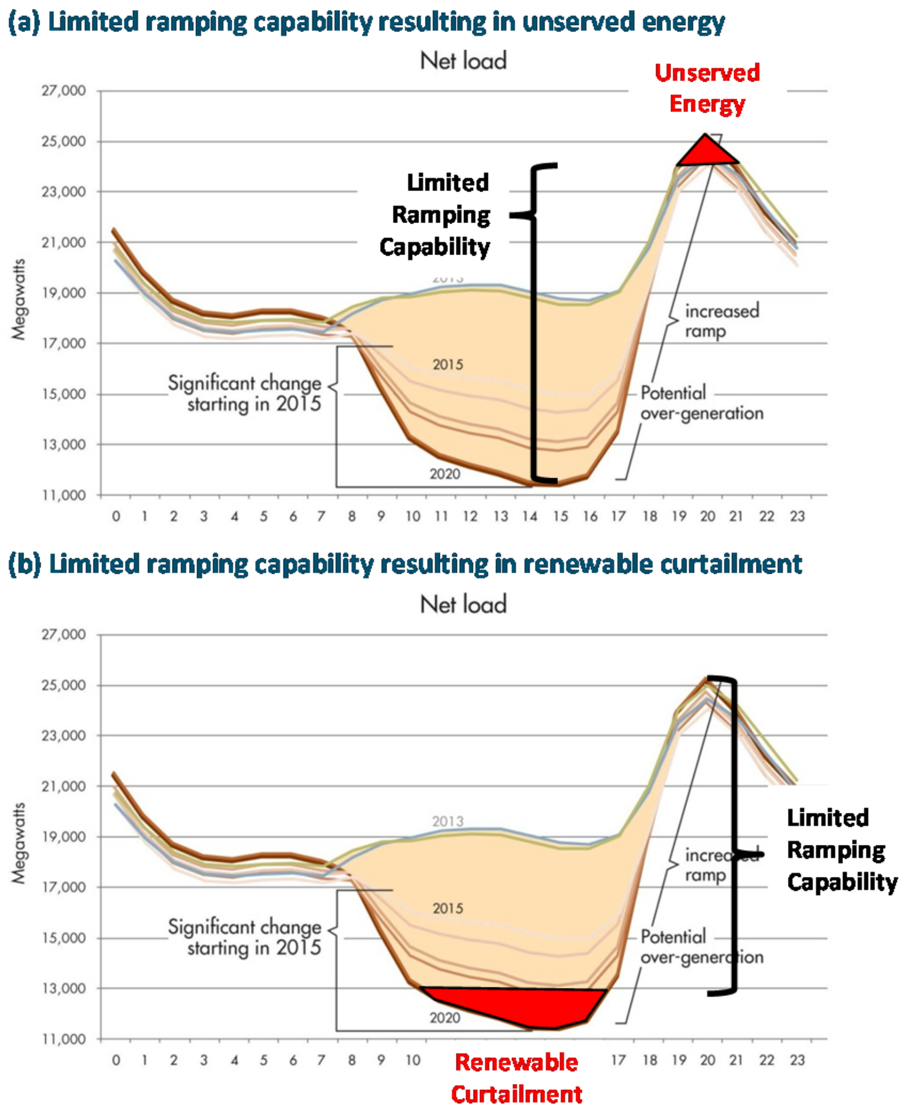
However, loss of load can be avoided through prospective curtailment of renewable generation to ensure that the thermal resources required to maintain reliability can remain online. This is illustrated in Figure 1b, where renewable generation is curtailed during the mid-afternoon in order to ensure that the system has sufficient operating capability to meet the evening ramp. In this way renewable curtailment is both a key operational strategy in flexibility-constrained systems and is also the primary consequence of inadequate flexibility.

Because the system operator has a choice between curtailing renewables and curtailing loads, it needs a method for determining which strategy to use on a given day. Additional information is therefore needed about the economic consequences of unserved energy and curtailed renewables. This information may come through market-based bids provided by loads and renewable project operators. In the absence of market information, deemed values are used based on available literature.

Studies typically, ascribe a high value of between \$10,000 and \$100,000/MWh of lost load. In contrast, the cost of renewable curtailment is effectively smaller by orders of magnitude. Out-of-pocket costs of curtailment include O&M costs as well as any lost production tax credits. In addition, under a production quota such as an RPS, curtailed renewable generation must typically be replaced like-for-like in order to ensure compliance. Thus, the cost of curtailment is determined by the "replacement cost" of renewable generation (the cost of procuring an additional MWh of generation to "replace" the curtailed energy).

In this study, the penalty prices assumed for loss of load and renewable curtailment are **\$50,000** and **\$100/MWh**, respectively. The asymmetry between these penalties ensures that REFLEX will steer the system away from loss of load if at all possible, allowing some amount of curtailment when necessary to ensure that the system has adequate operating flexibility. In this respect, renewable curtailment serves as the "default solution" for system operators facing flexibility constraints, the relief valve with which an operator can ensure the electric system remains within an operable range.

Figure 1. Tradeoff between upward and downward flexibility challenges



In order to capture an adequately broad sample of operating conditions and possible flexibility constraints using this framework, this study uses Monte Carlo sampling of load, wind, solar, and hydro across a wide range of historical conditions. Production cost model are typically used to analyze a single "typical" year with its unique set of load, wind, solar, and hydro profiles. This study makes use of much longer historical records for each of these variables, as summarized in Figure 2. Inputs for these variables are derived from a variety of sources: load shapes based on historical weather conditions spanning a thirty-year period are created using a neural network regression model; wind and solar profiles are derived from NREL's WIND and SIND Toolkits, respectively; and hydro data is based on a combination of actual and simulated historical data for Western hydro systems obtained from the Energy Information Administration (EIA) and the Northwest Power and Conservation Council (NWPCC). Profiles from different periods are combined using a stratified sampling methodology to produce Monte Carlo "day draws" for production cost modeling. By adapting a technique that has historically been used in LOLP analyses, this approach captures a more robust distribution of expected long-run conditions than any single historical year.

Figure 2. Historical conditions incorporated into draws

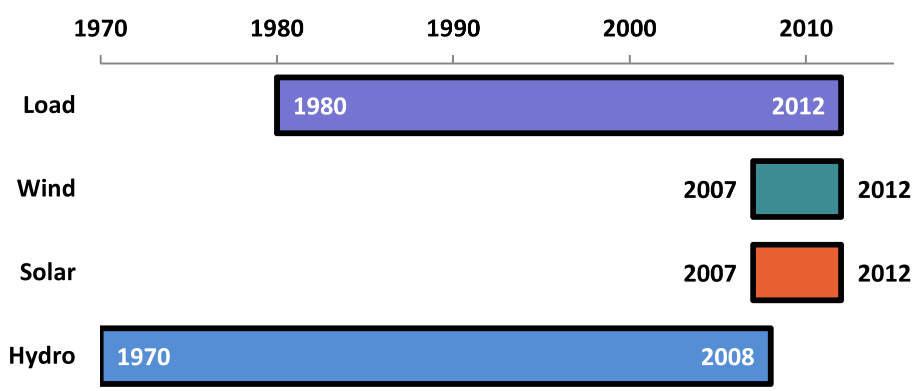
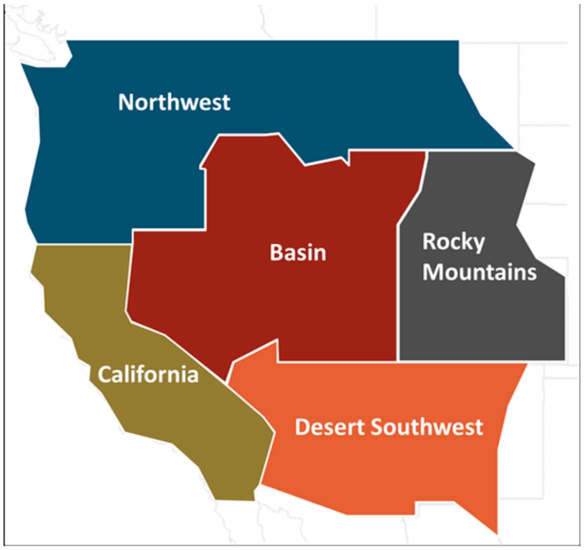


Figure 3. Regions included in analysis.



The analytical framework described above is used to evaluate the flexibility of five regions within Western Interconnection, shown in Figure 3. Each region represents a portion of the Western Interconnection with relatively homogeneous loads and resources, some existing degree of coordination in resource planning and/or operations, and limited internal transmission constraints.[3] In the study, each region is linked to its neighbors with a zonal transmission model. One of the key simplifying assumptions of this approach is that each region is perfectly coordinated in operations internally, effectively pools all of its generation resources to balance net load without regard for existing contracts, ownership, or operational conventions.

[3] The Alberta Electric System Operator (AESO) and BC Hydro (BCH) are excluded from this analysis due to lack of data availability; their exclusion from the study may overstate the magnitude of flexibility challenges, particularly if the flexibility of the hydro system in British Columbia could be used to facilitate renewable integration

This study begins with a conservative assumption that exchange between any two regions is limited to the range that has been observed historically rather than the full physical capability of the transmission system that links them. This approach serves three purposes: (1) it serves as a proxy for the many institutional constraints that exist in today's system, in which power exchange between regions generally occurs on a limited and bilateral basis; (2) it helps to isolate the renewable integration challenge in each region in order to characterize each fleet's ability to integrate its own renewable portfolio rather than relying on the flexibility and diversity of the entire Western system; and (3) it provides a useful counterfactual to a scenario in which limits on the transmission system are relaxed to their full physical capability, allowing this study to highlight the value that could be achieved through centralized dispatch of all resources in the WECC region.[4]

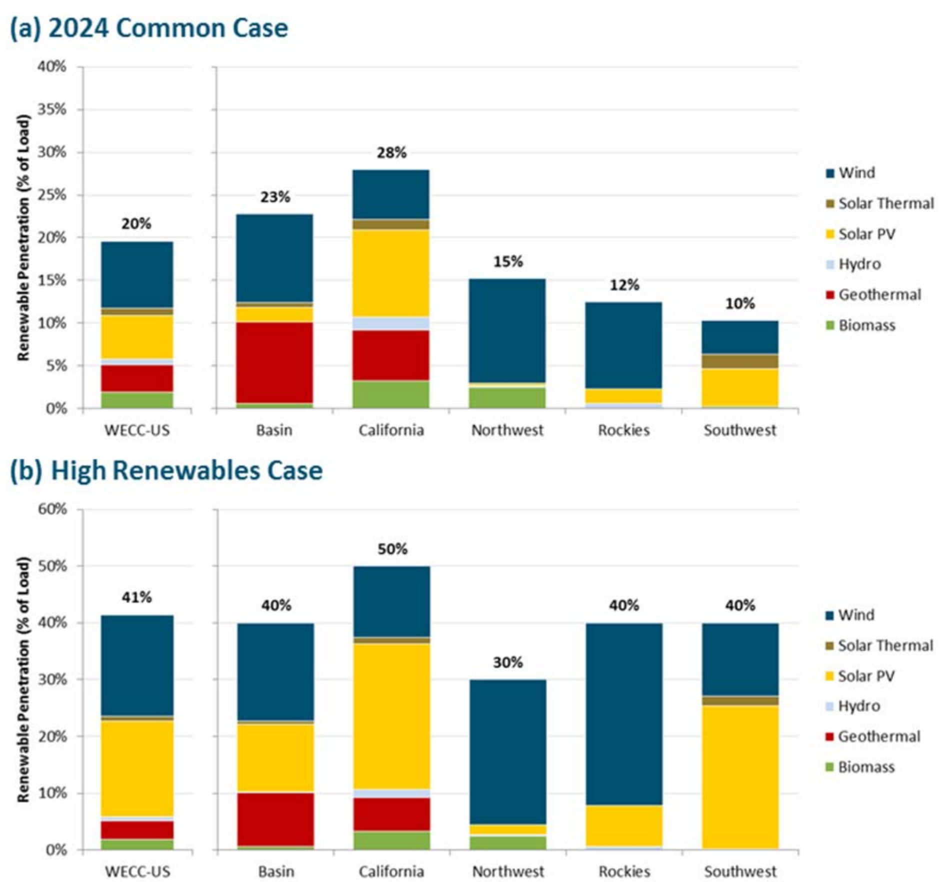
Two renewable portfolios are examined within the flexibility assessment for each region. The first, shown in Figure 4a, is based on the 2024 Common Case developed by WECC's Transmission Expansion Planning Policy Committee (TEPPC) and represents a penetration of renewable generation that is largely consistent with state RPS targets current as of 2014;[5] analysis of this case

[4] Increased regional coordination is one of the integration "solutions" examined in this report, and this contrast provides the underlying method through which its value is characterized.

[5] At the time of the development of the 2024 Common Case database and the start of this study, California's RPS policy remained at 33%. With the increase to a 50% target by 2030, the Common Case is no longer indicative of current policy in that region but still provides a useful point of reference for its near-term relevance.

provides a means of validating the model as well as an indication of potential flexibility challenges that could emerge under existing policy. The second portfolio studied is a “High Renewables Case,” whose composition is shown in Figure 4b. The penetrations chosen for study in the High Renewables Case were intentionally chosen to be high enough to cause significant changes in operations, and to ensure that flexibility challenges would result, allowing this study to characterize renewable integration challenges that may emerge above current policy levels.

Figure 4. Renewable portfolios analyzed in the flexibility assessment



Nature of Regional Renewable Integration Challenges

Because the penalty prices for unserved energy and curtailment prioritize load service over the delivery of renewable generation, renewable curtailment is the key indicator of a system that is constrained in its ability to integrate renewables. Curtailment can occur due to simple oversupply—where the generation available exceeds the load—or as a means to ensure reliable operations in the presence of dispatch flexibility constraints. REFLEX produces many metrics that are useful for understanding how high penetrations of renewables impact a system; however, this study uses renewable curtailment as the key indicator of flexibility constraints for each case.

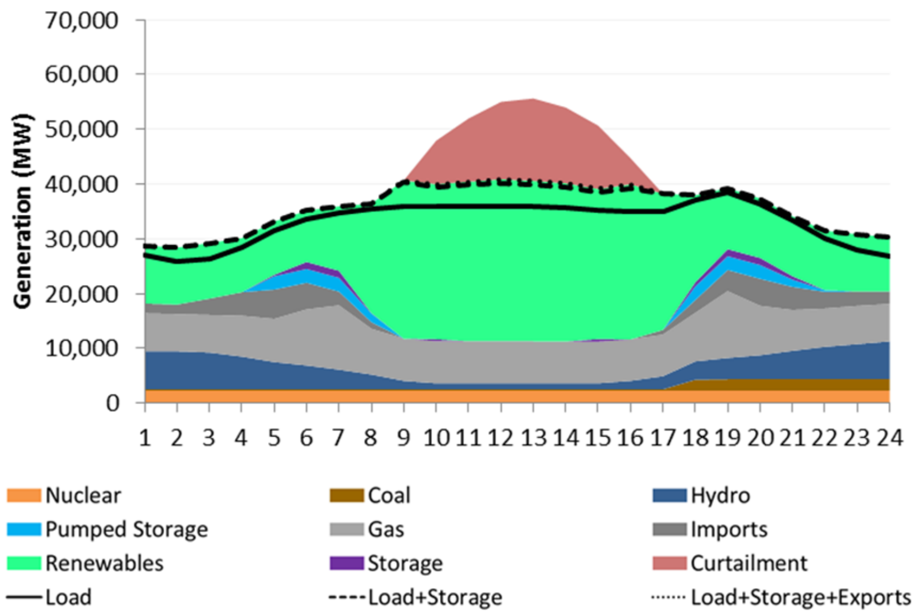
Table 1 summarizes the renewable curtailment observed in each region for the renewable portfolios examined. In the Common Case, the extent of renewable curtailment experienced across the Western Interconnection is limited—it constitutes less than 0.1% of the available renewable energy. In contrast, in the High Renewables Case, renewable curtailment appears routinely and represents a large share of the available renewable generation. The specific nature of the challenges differ across regions—each a unique result of the region's distinct renewable portfolio, the characteristics of its conventional generators, and the profile of its loads.

Table 1. Renewable curtailment observed in the Common Case and High Renewables Case (% of annual generation)

Scenario	Basin	California	Northwest	Rockies	Southwest	WECC-US
Common Case	0.01%	0.00%	0.06%	0.13%	0.00%	0.02%
High Ren Case	0.4%	**8.7%**	**5.6%**	0.6%	**7.3%**	**6.4%**

In the High Renewables Case, renewable curtailment is most prevalent in California and is driven by a phenomenon that has been well-characterized in past studies of renewable integration: because solar PV resources produce only during daylight hours, a large share of the generation in California's renewable portfolio is concentrated in the middle of the day. The high daytime renewable production exerts pressure on the non-renewable fleet to reduce its output to low levels, but non-renewable generation is limited in its ability to do so by inflexibility (especially for must-run resources like nuclear & cogeneration), minimum generation constraints, and the need to carry contingency and flexibility reserves. The result of this dynamic is a regular daily pattern of midday renewable curtailment illustrated in Figure 5, which shows a typical spring day in California in the High Renewables Case.

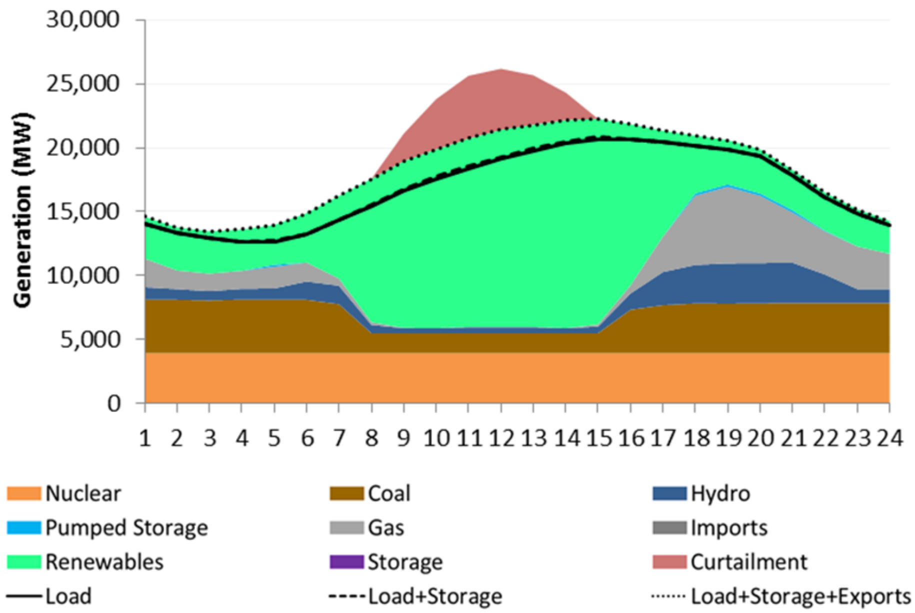
Figure 5. Typical spring day, California, High Renewables Case[6]



The Southwest also experiences a large volume of renewable curtailment in the High Renewables Case. Much like California, the Southwest relies heavily on solar PV resources in the High Renewables Case and, as a result, experiences a similar diurnal oversupply phenomenon. Figure 6 presents a typical spring day in the Southwest in which many of the same elements that characterize the California challenge are apparent: large diurnal net load ramps that coincide with sunrise and sunset and a mid-day period of renewable curtailment when all dispatchable resources have been reduced to their lowest output.

[6] Note that this study reports all results in Pacific Standard Time; all plots that show hourly results use the PST convention.

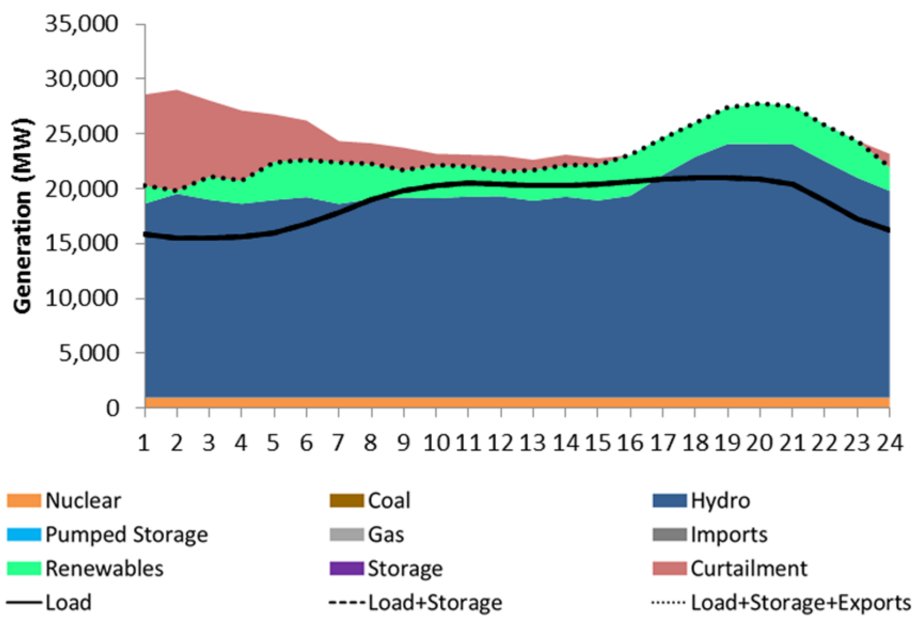
Figure 6. Typical spring day, Southwest, High Renewables Case



Notwithstanding these similarities, the Southwest does have an important distinguishing characteristic from California: its reliance on coal generation for a large share of its installed capacity. Historically, the coal fleet in this region has operated largely in a baseload capacity, ramping and cycling with limited frequency. The ability to operate the coal fleet more flexibly during high renewable output days is a key strategy to facilitate renewable integration in the Southwest and in other regions that rely on coal generation today.

The third region that experiences significant quantities of renewable curtailment in the High Renewables Case is the Northwest; though oversupply is its primary cause, the nature of the oversupply is entirely unique to that region. The Northwest is characterized by its large hydroelectric fleet, and during the spring runoff season, during average and wet hydro years, the Northwest hydro system is capable of meeting most or all of the Northwest's energy needs on a day-to-day basis. Adding wind generation to that system during that time of year results in an excess supply of zero-marginal-cost generation. Such oversupply events have already been experienced in parts of the Northwest in 2010-2012 when surplus hydro conditions, combined with dispatch inflexibility, led to wind curtailment.[7] With the addition of large amounts of wind generation in the High Renewables Case, the oversupply during the spring runoff is intensified. An example of such a day is shown in Figure 7.

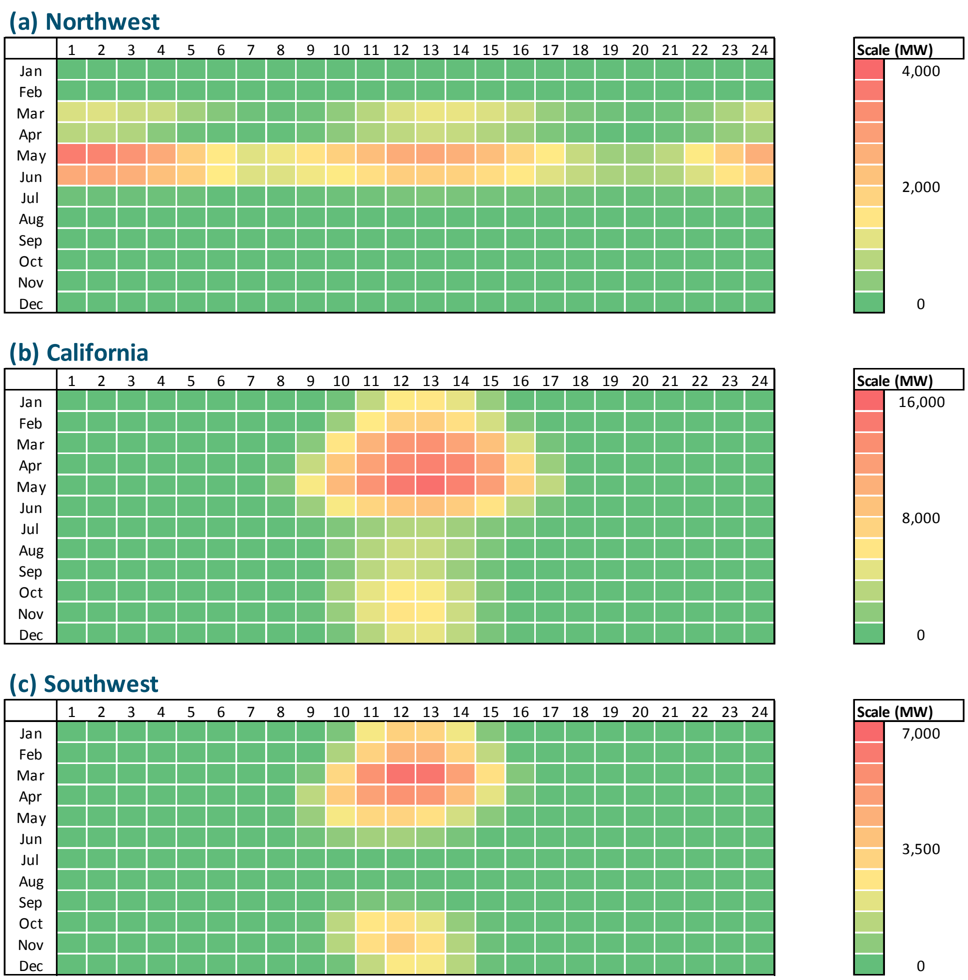
Figure 7. Typical spring day, Northwest, High Renewables Case



[7] In recognition of this phenomenon, the Bonneville Power Administration (BPA) has developed an "Oversupply Management Protocol" to allow for prudent management of generation resources during oversupply conditions.

Spring days are chosen to highlight flexibility challenges because the spring months (March through June) represent the most challenging periods for renewable integration throughout the West. The combination of relatively low loads, high hydro system output, and high renewable output results in a concentration of oversupply during this period of the year. Figure 8 illustrates the diurnal and seasonal trends in renewable curtailment for these three regions through heat maps that display the average amount of renewable curtailment in each month-hour. The similarities between California and the Southwest are immediately evident: throughout the year, curtailment is concentrated in the middle of the day, driven by the solar PV oversupply in each region. This oversupply is greatest in the spring and smallest in the summer, when each region's loads are highest due to high cooling loads and can absorb higher quantities of solar PV output. The seasonal trend in the Northwest, by contrast, is more closely linked to the spring runoff from the hydro system and its interaction with a wind-heavy renewable fleet, which leads to renewable curtailment at all hours in the spring but is most concentrated at night when wind output is high and loads are low.

Figure 8. Seasonal and diurnal patterns of renewable curtailment, High Renewables Case

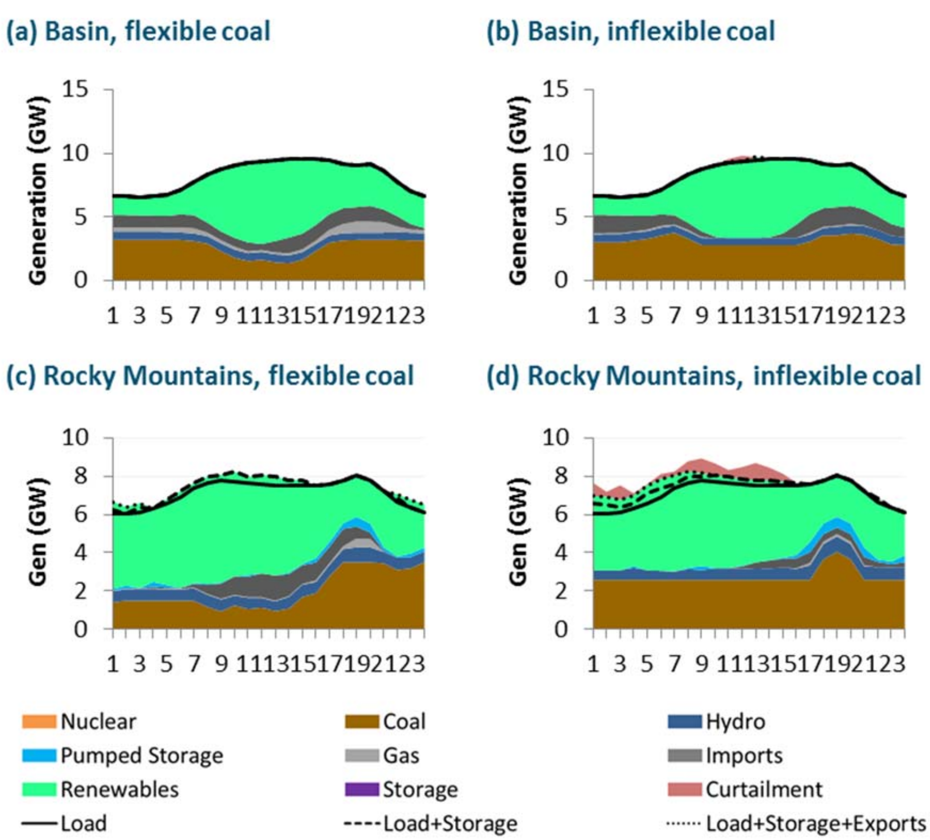


The dramatic increase in curtailment observed in these three regions between the Common Case and the High Renewables Case is indicative of a phenomenon that is highly nonlinear. In the absence of mitigation measures, curtailment grows at an increasing rate as renewable penetration climbs. For example, the marginal curtailment for a new solar PV resource (i.e. the next solar resource built on top of the portfolios described here) in California is between 50-60%—that is, in excess of half of its available energy would not be able to be delivered to the California system—a figure substantially higher than the average curtailment of 9%.

In contrast to California and the Northwest and Southwest regions, curtailment is observed in much lower quantities in both the Basin and Rocky Mountain regions. What distinguishes the Basin and Rocky Mountain regions from California and the Southwest—all four fleets rely primarily on dispatchable thermal resources to serve net load—is the amount of diversity assumed in their renewable portfolios. Whereas the concentration of solar PV during the middle of the day drives the oversupply phenomenon in California and the Southwest, production is far less concentrated during specific periods of the year in the Basin and Rocky Mountain regions. In the Basin, this is a result of technological diversity that combines geothermal, solar and wind resources; in the Rocky Mountains, it is the result of a portfolio comprising primarily wind resources whose scale and geographic dispersion result in production spread throughout the year.

While curtailment is observed in limited quantities in these regions, both the Rocky Mountains and the Basin region show dramatic changes in how the generation fleets operate to balance higher penetrations of renewable generation. A key driver of the magnitude of operational challenges in these two regions is the extent of the ability of thermal generators to ramp and cycle. Historically, coal generators have operated as baseload resources, running at high capacity factors throughout the year. Higher penetrations of renewable generation will exert pressure on operators to use these units more flexibly than they have historically; however, there are technical, economic, and institutional factors that might limit their ability to do so. The amount of flexibility that can be harvested from the aging coal fleet is uncertain and deserves further attention, but its implications for renewable integration are clear: an electric system with coal generators whose flexibility is limited in day-to-day operations will experience renewable curtailment in larger quantities and more frequently than one in which coal generation operates flexibly. This tradeoff is illustrated for both the Basin and Rocky Mountain regions in Figure 9.

Figure 9. Typical spring days, Basin and Rocky Mountains, High Renewables Case (with and without limits on coal flexibility)



Enabling Strategies for Renewable Integration

This study treats renewable curtailment as the “default solution” to flexibility challenges as it represents a last recourse for system operators—prior to shedding load—once the flexibility of the existing traditional dispatchable system has been exhausted. However, there are many options for alternative strategies that, if identified and deployed in advance, can provide operators with additional flexibility and mitigate the need to curtail renewable generation. These “solutions” include improvements in scheduling and dispatch, investments in new flexible generation, and new demand-side programs. Investigating the full potential list of integration solutions is beyond the scope of this study; the solutions chosen for study herein are intended to help illustrate the types of attributes that flexibility planners may wish to consider. This study considers three specific measures to facilitate renewable integration:

+ Increased regional coordination;
+ Investments in energy storage technologies; and
+ Investments in flexible gas generation.

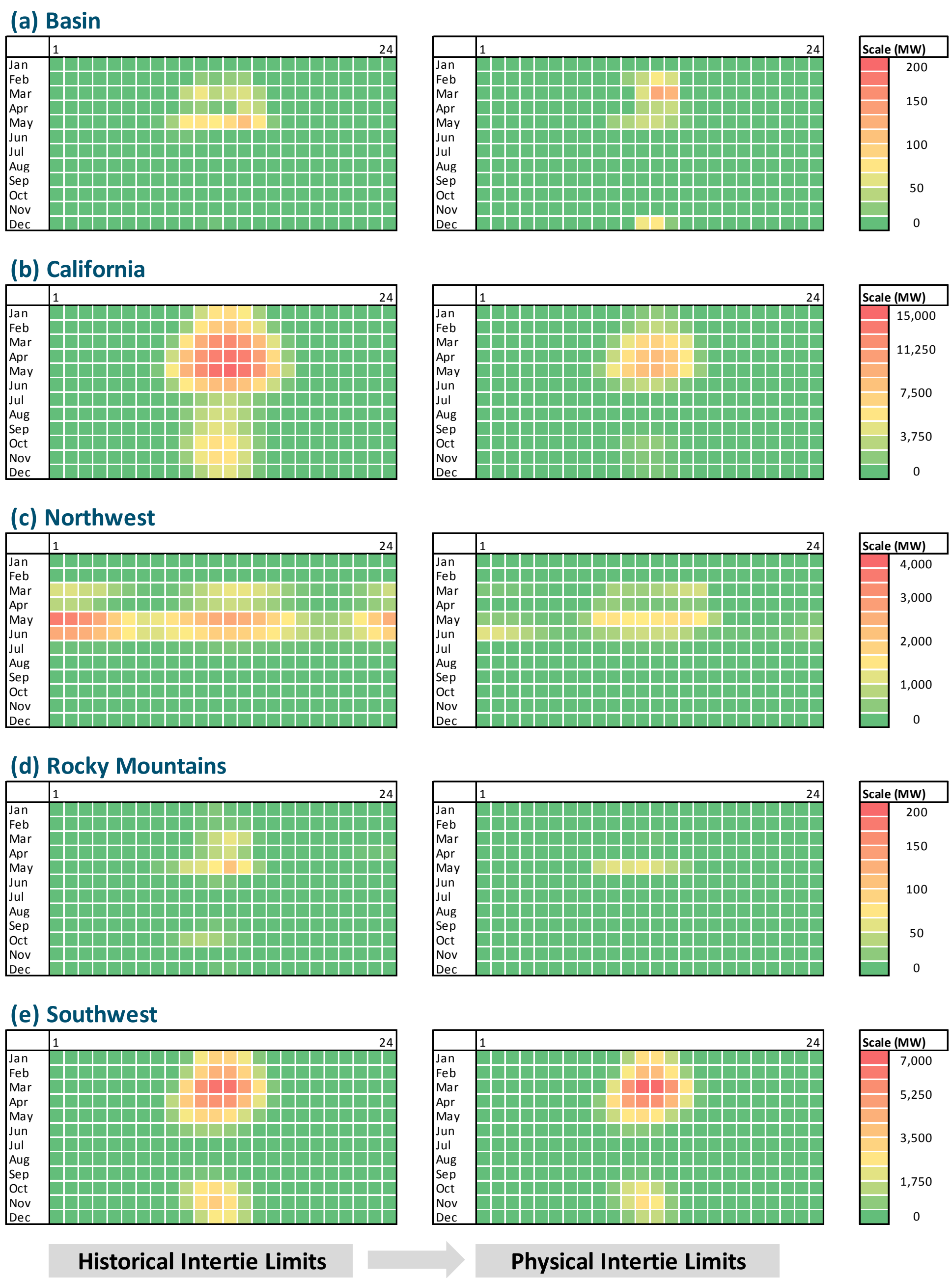
The first among these—improving regional coordination in operations—represents an institutional improvement to facilitate renewable integration. Historically, the balkanized Western Interconnection has operated largely according to long-term contractual agreements and through bilateral power exchange. “Regional coordination” could represent a step as incremental as improving upon existing scheduling processes or as comprehensive as the consolidation of the Western Interconnection under a single centralized operator. This study does not presume the mechanism through which coordination occurs, but demonstrates the value of achieving full coordination by contrasting a future in which interregional exchange is limited to the historical range with one in which it is limited by the ratings of interregional transmission paths, allowing fuller utilization of load and resource diversity.

Relaxing the constraint on interregional exchange to allow the use of the transmission system to its physical limits results in a reduction of renewable curtailment from 6.4% to 3.0% (see Table 2). A significant share of this value is derived from regions that, facing an oversupply condition that might otherwise require renewable curtailment, are able to use the full capability of the transmission system to find an alternative market for their power. This impact is shown in Figure 10. The reduction of curtailment is largest in California—which is both electrically and institutionally close to a market in the Northwest that can accommodate midday exports outside of the spring season—and in the Northwest—whose nighttime oversupply finds a destination in both California and the Southwest markets. Notably, the observed impact on the Southwest is lower than in the prior two regions, as connections with the Basin and Rocky Mountains provide it with limited export markets and California's frequent simultaneous oversupply makes it an impractical market for solar surplus.

Table 2. Comparison of regional renewable curtailment with historical and physical intertie limits imposed, High Renewables Case

Scenario	Basin	California	Northwest	Rockies	Southwest
Historical Intertie Limits	0.4%	8.6%	6.1%	0.1%	7.3%
Physical Intertie Limits	0.5%	3.1%	1.6%	0.5%	6.0%
Difference	**+0.1%**	**-5.6%**	**-4.5%**	**-0.0%**	**-1.3%**

Figure 10. Impact of increased regional coordination upon seasonal curtailment patterns.



While desirable from a technical and economic perspective, harvesting the full diversity of the Western Interconnection through coordination of operations to integrate renewable generation presents a significant institutional challenge. Individual utilities, balancing authorities, and planning entities may be justified in examining potential investments or demand-side programs to provide greater local or regional operational flexibility. To examine the potential flexibility benefits of investments in new flexible resources, 6,000 MW of energy storage[8] and new fast-starting, fast-ramping gas CCGTs are added to the High Renewables Case in separate scenarios to quantify their impact on operations.[9] The impact of each additional resource on renewable curtailment is summarized in Table 3.

Table 3. Impacts of new investments on regional renewable curtailment, High Renewables Case.

Scenario	Basin	California	Northwest	Rockies	Southwest
Reference Grid	0.4%	8.7%	5.6%	0.6%	7.3%
+6,000 MW Storage (2hr)	0.4%	7.2%	5.7%	0.6%	6.2%
+6,000 MW Storage (6hr)	0.4%	5.8%	5.8%	0.6%	5.1%
+6,000 MW Storage (12hr)	0.4%	5.8%	5.7%	0.6%	5.1%
+6,000 MW Flex CCGT	0.4%	8.7%	5.6%	0.6%	7.3%

The primary result this analysis indicates is that the addition of new "downward" flexibility (e.g. the ability to charge energy storage) provides substantially greater benefits than the addition of new "upward" flexibility (e.g.

[8] Multiple durations of energy storage resources were modeled, including 2-hr, 6-hr, and 12-hr.

[9] In each scenario, 4,000 MW of the candidate resource was located in California and 1,000 MW was located in both the Northwest and the Southwest. The choice of where to add storage was made on the basis of where the largest flexibility challenges were identified.

the ability to ramp from Pmin to Pmax very quickly), as it expands the net load range across which a system can operate. This analysis indicates several findings on the impacts of new flexibility investments:

+ Energy storage provides a clear benefit through curtailment mitigation in the Southwest and California. This is largely due to the fact that each region's diurnal solar oversupply allows storage resources to cycle almost daily, charging during the middle of the day during curtailment hours and discharging in the early morning and/or evening to help meet solar-driven net load ramps.
+ The value of energy storage in the Northwest is limited by comparison. Here, where oversupply events during the spring runoff persist much longer, often throughout the day, there is a much more limited opportunity to shift generation within the day.[10] Also, the Northwest already has significant intra-day energy storage capability through its existing system of hydroelectric resources.
+ In all regions where new flexible gas CCGT capacity is added, the impact is minimal. While further study is needed, it appears that all regions have sufficient upward ramping capability in the base system. These new, efficient gas resources may reduce system *cost* due to their efficiency, but they do little to reduce curtailment because they do not appreciably improve the system-wide ratio of minimum to maximum generation capability.

[10] Note that the REFLEX approach of using day draws likely understates the value of storage in the Northwest, as it does not capture value that could be provided through inter-day charging and discharging patterns.

While the general findings of these cases help to illustrate the attributes of new flexible resources that provide significant benefits, the limitations of their applicability must also be understood:

+ One of the underlying assumptions of the regional approach is that each constituent utility and balancing authority of a region makes its resources fully available for optimal dispatch within that region. However, this assumption is optimistic for today's electric system—one that is dominated by bilateral transactions—and findings that apply to a wholly optimized region may not be applicable to individual entities within it.
+ The starting case to which the flexible resources are added assumes a significant degree of flexibility in the existing thermal system. Production simulation models typically do not consider increased O&M costs incurred as units are asked to cycle more frequently. As the sensitivity on coal flexibility demonstrates, there is value to a system that can operate its thermal units flexibly (i.e. can reduce their output to very low levels or turn off), and to the extent that technical, economic, or institutional factors prevent the coal fleet from operating as this study assumes, additional flexible thermal generation could have more value than this study identifies.
+ A critical qualifier for this analysis is that each regional fleet already meets traditional resource adequacy needs prior to the addition of incremental flexible capacity. While the value of flexibility alone may not be enough to justify procuring a specific resource, entities should carefully consider the benefits that additional flexibility could provide when new system capacity is needed due to load growth, retirement of aging coal generation, or other factors.

Implications and Next Steps

The technical findings and conclusions reached through this study have a number of implications that are relevant for regulators and policymakers seeking to enable higher penetrations of renewable generation on the system and to ease the associated challenges.

1. The analysis conducted in this study identifies no technical barriers to the achievement penetrations of renewable generation of up to 40% of total supply in the Western Interconnection.

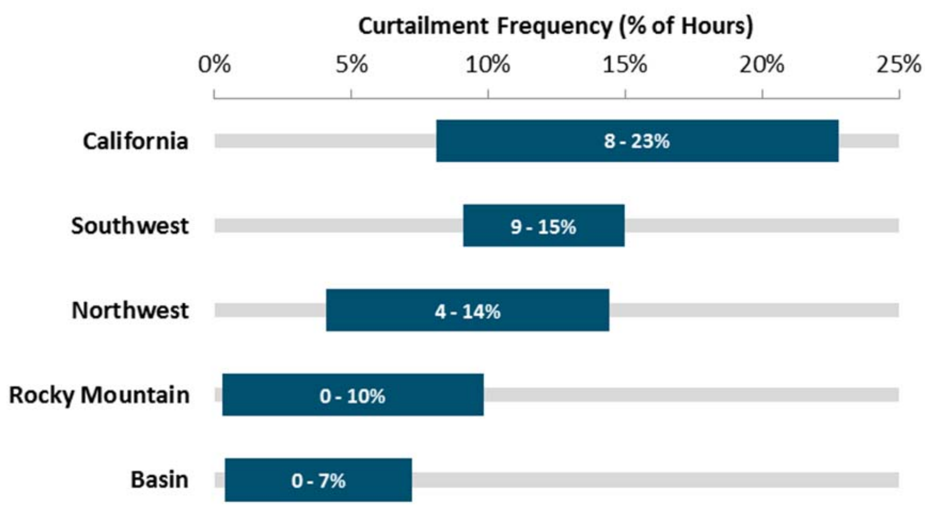
In both the Common Case and the High Renewables Case, the flexibility assessment demonstrates capability of the electric system of the Western Interconnection to serve loads across a diverse range of system conditions. Further, no “need” for additional flexible capacity beyond existing and planned resources is identified in either case. Examination of issues relating to stability, adequacy of frequency response, and the potential impact of major contingencies is beyond the scope of this study; however, other studies to examine these topics are have come to similar conclusions regarding the technical feasibility of integrating high penetrations of renewable generation.[11]

What does distinguish the High Renewables Case from the Common Case is the significant quantity of renewable curtailment observed. At relatively low penetrations, renewable generation can be integrated into the system

[11] See the Western Wind and Solar Integration Study – Phase 3 (NREL & GE), available at: http://www.nrel.gov/electricity/transmission/western-wind-3.html

effectively with limited curtailment; however, once a region's penetration surpasses a certain threshold, curtailment occurs with increasing frequency. Renewable curtailment is characteristic of all of the High Renewables scenarios investigated: Figure 11 summarizes the range of curtailment frequency across all of these scenarios.

Figure 11. Observed range of hourly curtailment frequency across all High Renewables scenarios in each region.



2. Routine, automated renewable curtailment is a fundamental necessity to electric systems at high renewable penetrations, as it provides operators with a relief valve to manage net load conditions to ensure a system can be operated reliably at increasing penetrations.

Historically, the idea of "curtailment" has had negative connotations, but at high penetrations of renewable generation, it is a fundamental necessity.

Notwithstanding its cost to ratepayers, renewable curtailment provides operators with a tool with which to manage the net load to ensure reliable service when the flexibility of all other existing resources has been exhausted; in this respect, it serves as the "default solution" for a system constrained on flexibility.

The role of curtailment in the operations of an electric system at high renewable penetrations is multifaceted. In addition to allowing operators to mitigate oversupply events when generation would otherwise exceed loads, renewable curtailment can be used to mitigate the size of net load ramps across one or more hours, to adjust for forecast error between day-ahead and other scheduling processes, and to substitute for traditional flexibility reserve services (including regulation).

Because of the crucial role of renewable curtailment in operating electric systems at high penetrations of renewables, ensuring that curtailment is available and can be used efficiently in day-to-day operations requires a number of steps:

+ **Market structures and scheduling processes must be organized to allow participation by renewable generators.** Within organized markets, this means ensuring that utilities can submit bids into the market on behalf of renewable generators that reflect the opportunity cost of curtailing these resources as well as ensuring that renewable plants are not excessively penalized for deviations from their schedules due to forecast errors. In environments in which vertically integrated utilities or another type of scheduling coordinator is responsible for determining system dispatch, the operator must begin to consider the role of renewable curtailment in scheduling and dispatch decisions for both renewable and conventional resources.

+ **Contracts between utilities and renewable facilities must be structured to allow for economic curtailment.** Historically, many power purchase agreements have been set up to pay renewables for the generation that they produce and have included provisions limiting curtailment under the premise that limiting risk and ensuring an adequate revenue stream to the project are necessary to secure reasonable financing. **Compensated curtailment,** under which developers are paid a PPA price both for generation that is delivered to the system as well for estimated generation that is curtailed, would be one means of achieving this goal.

To the extent that additional institutional barriers that would limit operators from effectively dispatching renewable resources exist, these barriers must also be addressed to allow for renewable integration at higher penetrations.

3. Another key step to enabling reliable and efficient operations under high penetrations is ensuring operators fully understand the conditions and circumstances under which renewable curtailment is necessary or desirable.

In some instances—namely, during oversupply conditions—the need to curtail is relatively intuitive; however, in other instances, the important role of curtailment may not be so obvious. For example, an operator faced with a choice between keeping a specific coal unit online and curtailing renewables or decommitting that coal unit to allow additional renewable generation should make that decision with knowledge of the confidence in the net load forecast as well as an understanding of the consequences of possible forecast errors. Similarly, an operator anticipating a large upward net load ramp may decide to curtail renewable generation prospectively to spread the ramp across a longer duration if the ramp rates of conventional dispatchable units are limited. Additional work is necessary to identify such operating practices and conditions in which renewable curtailment may be necessary outside of oversupply conditions to ensure reliable service.

The role of operating reserves at avoiding unserved energy under unexpected upward ramping events must also be considered. Resources under governor response or Automated Generation Control (AGC) respond quickly to small deviations in net load. Contingency reserves (spinning and supplemental or "non-spin" reserves) are used to manage large disturbances such as the sudden loss of a generator or transmission line. Additional categories of reserve products—for instance, "load following" or "flexibility" reserves—have been contemplated at higher renewable penetration, but have not yet been formalized. How these reserves are deployed will impact the magnitude of challenges encountered at higher renewable penetrations.

4. The consequences of extended periods of negative pricing must be examined and understood.

Historically, the centralized markets and bilateral exchanges of the Western Interconnection have, for the most part, followed the variable costs of producing power—most often the costs of fuel and O&M for coal and gas plants. In a future in which renewable curtailment becomes routine, forcing utilities to compete to deliver renewable generation to the loads to comply with RPS targets, the dynamics of wholesale markets will change dramatically. How the dynamics of negative pricing ultimately play out remains a major uncertainty; nonetheless, with frequent low or negative prices in a high renewables future, utilities, other market participants, and regulators will be confronted by a host of new questions:

+ How should generators that provide other services to the system during periods of low negative prices be compensated?
+ How can the proper signal for investment in generation resources be provided as frequent negative prices further erode margins in energy markets?
+ Do negative prices create new issues for loads, who, rather than paying for power from the wholesale market during periods of curtailment, would be paid to consume?
+ At what point does the prevalence of negative prices lead to new policy mechanisms other than production quotas to promote the development of new renewable energy?
+ How should future retail tariffs be designed to balance considerations of equity and cost causation with the radical changes in wholesale market signals?

These and other questions will require consideration as penetrations of renewables continue to increase.

5. While renewable curtailment is identified as the predominant challenge in operations at high renewable penetrations, its magnitude can be mitigated through efficient coordination of operations throughout the Western Interconnection.

The balkanization of operations today presents an institutional barrier to efficient renewable integration; by allowing full utilization of the natural diversity of loads and resources throughout the Western Interconnection, regional coordination offers a low-hanging fruit to mitigating integration challenges. A number of studies have identified the significant operational benefits that can be achieved through balancing authority consolidation, a conclusion that is supported by the reduction in renewable curtailment at high penetrations identified in this study.

6. Many supply- and demand-side solutions merit further investigation to understand their possible roles in a high renewable penetration electric system.

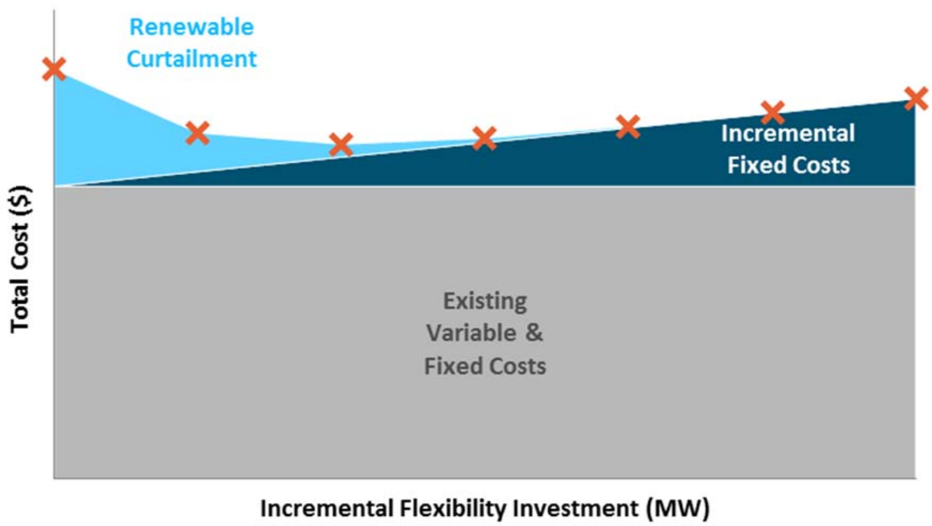
This study examines a select few of the multitude of possible supply- and demand-side portfolio measures available to utilities to illustrate how different attributes do (or do not) provide value to electric systems at high penetrations of renewable generation. The solutions examined within this study illustrate how different "types" of flexibility impact a system to differing degrees: whereas storage effectively mitigates renewable curtailment through its ability to charge during periods of surplus, fast-ramping flexible gas resources have a comparatively limited impact on operations, displacing less efficient gas generation resources but effecting minimal changes in curtailment.

7. The ability of renewable curtailment to serve as an "avoided cost" of flexibility points to an economic decision-making framework through which entities in the Western Interconnection can evaluate potential investments in flexibility and ultimately rationalize procurement decisions.

As the need for operational flexibility has grown, a number of efforts have explored whether additional planning standards—analogous to those used for resource adequacy today—are necessary to ensure that when the operating day comes, the generation fleet is sufficiently flexible to do serve load reliably. As this study demonstrates, so long as (1) the generation fleet is capable of meeting extreme peak demands, and (2) the operator can use curtailment as a relief valve for flexibility constraints, the operator can preferentially dispatch the system to avoid unserved energy. Thus, the consequence of a non-renewable fleet whose flexibility is inadequate to balance net load is renewable curtailment, whose implied cost is orders of magnitude smaller than the cost of unserved energy. In this respect, the determination of flexibility adequacy is entirely different from resource adequacy: for resource adequacy, conservative planning standards are justified on the basis of ensuring that costly outages are experienced exceedingly rarely; for flexibility adequacy, the appropriate amount of flexibility for a generation system is instead an economic balance between the costs of "inadequacy" (renewable curtailment) and the costs of procuring additional flexibility.

Because renewable curtailment can serve as an "avoided cost" of flexibility, the question of "flexibility adequacy" is economic, rather than technical. Renewable curtailment imposes a cost upon ratepayers, reflected in this study by the idea of the "replacement cost," and, to the extent it can be reduced through investments in flexibility, its reduction provides benefits to ratepayers. At the same time, designing and investing in an electric system that is capable of delivering all renewable generation to loads at high penetrations is, itself, cost-prohibitive. Between these two extremes is a point at which the costs of some new investments or programs that provide flexibility may be justified by the curtailment they avoid, but the cost of further investments would exceed the benefits. This idea is illustrated in Figure 12, which shows the tradeoff between the costs of renewable curtailment with the costs of a possible theoretical measure undertaken to avoid it. This study provides both an example of the type of analytical exercise that could be performed to quantify the operational benefits of flexibility solutions as well as a survey of the analytical considerations and tradeoffs that must be made in undertaking such an exercise.

Figure 12. Illustration of an economic framework for flexibility investment.



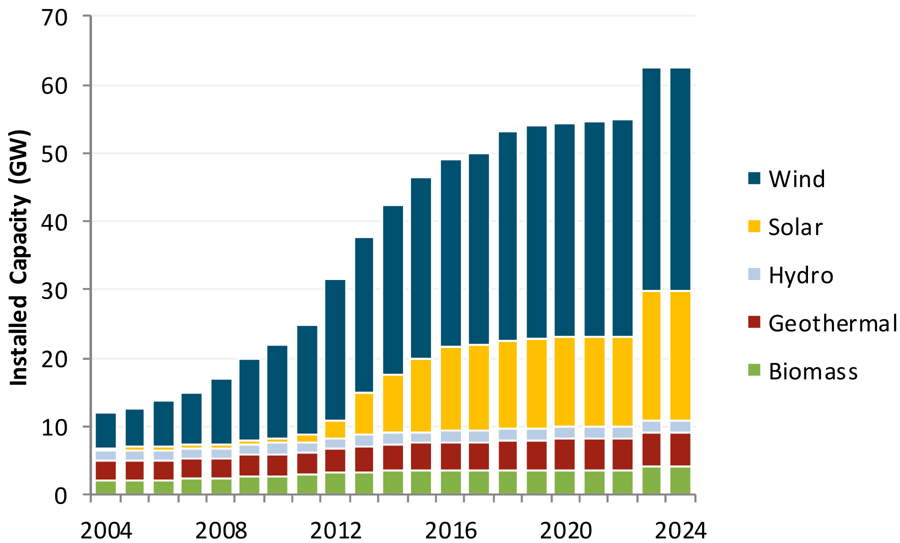
While not performed in the context of this study, this type of economic assessment of flexibility solutions to support renewables integration will depend on rigorous modeling of system operations combined with accurate representation of the costs and non-operational benefits of various solutions. The specific types of investments to enable renewable integration that are found necessary will vary from one jurisdiction to the next, but the overarching framework through which those necessary investments are identified may be consistent. Implementation of such an economic framework for decision-making for flexibility will foster the transition to high renewable penetration, enabling the achievement of policy goals and decarbonization while mitigating the ultimate impacts of those changes to the quality and cost of service received by ratepayers.

1 Introduction

1.1 Study Motivation

Over the course of the past decade, the penetration of renewable generation in the Western Interconnection has grown rapidly: during this time, nearly 30,000 MW of renewable generation capacity—mostly solar and wind—have been built within the footprint of the Western Interconnection (see Figure 13), increasing the total amount of renewable installed capacity by a factor of four. Much of this increase has been driven by the states' pursuit of Renewables Portfolio Standards (RPS), which specify a minimum share of load that each utility must serve with renewable generation. By 2024, under current state policies, the total installed capacity of renewable generation in the Western Interconnection may exceed 60,000 MW.

Figure 13. Historical & projected growth of renewables in the Western Interconnection under RPS policies in place through 2014.



Data Source: WECC 2024 Common Case

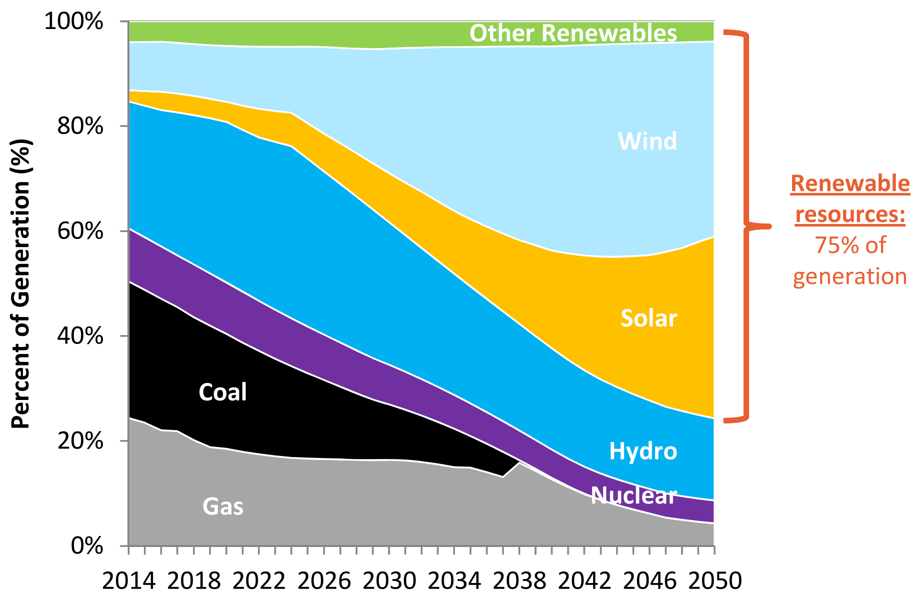
Yet even as renewable development continues at unprecedented rates in pursuit of current policy targets, many have begun to look beyond these levels, anticipating an even more central role for renewables in the Western Interconnection:

+ California's SB350, enacted in 2015, established that a 50% RPS target by 2030, a substantial increase from the prior RPS goal of 33% by 2020.
+ Elsewhere in the Western Interconnection, technology cost reductions have allowed renewable resources to compete with conventional resources on an economic basis, and utilities in the Rocky Mountain region have signed power purchase agreements for wind generation in excess of current RPS targets on the basis of this advantage.
+ The Environmental Protection Agency's proposed regulations on greenhouse gas emissions under Section 111(d) of the Clean Air Act

would require individual states to develop implementation plans to meet 2030 emissions standards; one of the key “building blocks” of the proposed legislation is the displacement of existing fossil resources with new renewable generation.

+ A burgeoning interest in policies to address climate change mitigation has intensified the focus on the long-term role of renewables in the West, as a number of studies have highlighted the central role that renewable generation may play in the decarbonization of the electric sector, a crucial step in the achievement of long-term carbon goals (see Figure 14, based on analysis in from IDDRI’s *Pathways to Deep Decarbonization in the United States*). By 2050, achieving deep decarbonization of the economy could require low-carbon generation to account for 80-90% of total power supply, which would imply a dramatic acceleration of current renewable policy. In the Western Interconnection, aggressive decarbonization could result in renewable penetrations between 75-80%.

Figure 14. Potential role of renewable generation in long-term decarbonization of electric sector.[12]



Data source: **E3 PATHWAYS High Renewables Case**

With growing penetrations of renewable resources, new challenges will arise for resource planners and operators. Wind and solar resources, which will likely account for a significant share of the additional renewable generation in the Western Interconnection, are characterized by three key attributes that have important implications for power system operations:

+ **Variability:** production changes from moment to moment, and from hour to hour;

[12] Figure shows the mix of generation in the Western Interconnection for the "High Renewables" deep decarbonization pathway. Available at: http://unsdsn.org/wp-content/uploads/2014/09/US_DDPP_Report_Final.pdf

- **Uncertainty:** production over a given period of time cannot be predicted with perfect accuracy; and
- **Concentration:** production is highly concentrated during certain hours of the year in which the resource is available.

As the penetrations of variable renewable resources in the Western Interconnection continue to increase, planners must confront the question of how to build and operate a reliable system in which a large portion of the energy available has these qualities.

The Western Electricity Coordinating Council (WECC), in its role as the Regional Entity responsible for reliability in the Western Interconnection, is interested in understanding the long-term adequacy of the interconnected western grid to meet the operational challenges posed by wind and solar generation across a range of plausible levels of penetration. This interest is echoed by many of WECC's stakeholders, whose study requests focus on futures in which renewable generation continues to expand beyond the levels anticipated by current policy. The Western Interstate Energy Board (WIEB) is interested in understanding these issues in order to inform policymakers about the implications of potential future policies targeting higher renewable penetrations.

In order to explore the implications of such changes on generation planning and system operations, this study presents a comprehensive framework using a combination of loss-of-load-probability and production cost modeling techniques through which the reliability and flexibility of a generation fleet may be evaluated under increasing penetrations of renewable generation. This framework is applied to two scenarios: a Base Case that captures current state renewable policies and a High Renewables Case that includes additional renewable generation throughout the Western Interconnection. In doing so, this study seeks to achieve three goals:

+ **Assess the ability of the fleet of resources in the Western Interconnection to accommodate high renewable penetrations while maintaining reliable operations.** Current state policies are expected to drive substantial change in the electric sector in the coming decade, but greater changes still may be on the horizon. Higher penetrations of renewable generation will test the flexibility of the electric systems of the West, requiring individual power plants to operate in ways that they have not historically and changing the dynamics of wholesale power markets. This study aims to identify the major changes in operational patterns that may be experienced at such high penetrations and to measure the magnitude and frequency of possible challenges that may result.
+ **Investigate potential enabling strategies to facilitate renewable integration that consider both institutional and physical constraints on the Western system.** Existing literature has identified a wide range of possible strategies that may facilitate the integration of high penetrations of renewables into the Western Interconnection. These strategies comprise both institutional changes—increased use of curtailment as an operational strategy and greater regional coordination in planning and operations—as well as physical changes—new investments in flexible generating resources and the development of novel demand side programs. This study examines how such strategies can play an enabling role as the penetration of renewable generation continues to increase.
+ **Provide lessons for future study of system flexibility on the relative importance of various considerations in planning exercises.** The study

of flexibility and its need at high renewable penetrations is an evolving field. This effort is designed with an explicit goal of providing useful information to modelers and technical analysts to improve analytical capabilities for further investigation into the topics explored herein.

1.2 Prior Renewable Integration Studies

There have been a number of prior efforts to examine the impacts of high penetrations of renewables on the Western Interconnection. Each study examined this question through its own lens:

+ The California Independent System Operator (CAISO) conducted a technical analysis of the ability of the CAISO system to integrate renewables to meet a 20% RPS: **Integration of Renewable Resources: Operational Requirements and Generation Fleet Capability at 20% RPS (CAISO 2010)**. This analysis identified increased ramping, increased operating reserve requirements, and overgeneration as potential consequences of increased wind and solar penetration. This study also flagged out-of-market scheduling as a potential constraint in efficiently integrating renewables and inadequate compensation for generators through energy markets as an emerging challenge under higher wind and solar penetrations.
+ As part of its ongoing engagement in the California Public Utilities Commission's (CPUC) Long-Term Procurement Proceeding, the CAISO has continued to investigate the implications of increased renewable penetrations with respect to the need for new flexible generation capacity to facilitate renewable integration through production cost modeling. The CAISO has studied a number of portfolios at 33% and 40% penetration, identifying shortfalls in operating reserves and the

need to curtail renewable generation as possible consequences of higher renewable penetrations.

+ The **Western Wind and Solar Integration Study Phase 1 (NREL 2010)** examined operations across the Western Interconnection with 35% penetration of wind & solar energy west-wide. This study identified geographical diversity, renewable forecasting for operations, and subhourly generation and interchange scheduling as key strategies to alleviating integration challenges across the West. Subsequent phases of the WWSIS delved into specific topics: **WWSIS Phase 2 (2013)** focused on quantifying the wear-and-tear costs and emissions impacts of cycling, finding a small impact; and **WWSIS Phase 3 (2015)** examined large-scale transient stability and frequency response under high penetrations, finding no indication that the Western system would be inadequate in its ability to respond to disturbances. Each of the phases of this work underscores the technical feasibility of operating a system at high renewable penetrations.
+ **Investigating a Higher Renewables Portfolio Standard in California (E3 2013)** conducted on behalf of the five largest utilities in California investigated the operational challenges of meeting a 50% RPS. This study found that between 33% and 50% RPS, oversupply would emerge as a new operational challenge (in some hours, renewable supply plus must-run generation exceeds load). This study identified renewable portfolio diversity, regional coordination, and resources that could provide downward flexibility as most promising for meeting a 50% RPS in California.
+ The **Low Carbon Grid Study – Phase 1 (NREL 2014)** showed that a diverse renewable portfolio, regional coordination, flexible loads, and energy storage could be effective at alleviating renewable integration challenges in California at 50-55% renewables.

- **Regional Transmission Plan (TEPPC).** TEPPC has historically studied a number of scenarios that posit the expansion of renewable resources in the West to levels at and above current policy. The analysis in these studies has focused on identifying potential changes in interregional transmission flows as well as potential interregional transmission upgrades that might facilitate the delivery of renewable generation from different parts of the Western Interconnection to loads.

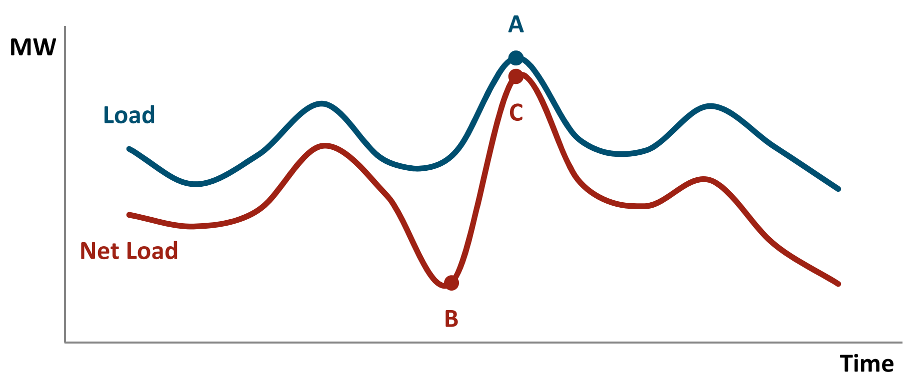
While these studies had different focuses, they have identified similar challenges and opportunities around the level of coordination in operations across the West and the extent to which renewables could be integrated efficiently by electricity markets. Because the Western Interconnection represents not a single market, but a conglomerate of independently operated investor-owned utilities, public utilities, and the CAISO, some level of operational coordination between these entities will be required to efficiently integrate renewables at higher penetrations. What has not yet been explicitly investigated is how this coordination (or lack thereof) might impact the planning of electricity systems across the West, a question of interest to WECC, WIEB, and their stakeholders.

1.3 Flexibility Planning Paradigm

While operational simulations are necessary to address renewable integration challenges, this assessment differs from prior studies in that it takes the perspective of the electric system resource planner. It examines not only how technical constraints on operations may impact the integration of renewables, but how institutional barriers and uncertainties may affect planning decisions as well.

Historically, resource planning for system reliability has generally focused on ensuring the adequacy of generation supply to meet loads during peak periods. A large majority of the capacity that exists today in the West is "dispatchable": available to produce power at a level prescribed by the system operator on demand (within a given range). In this type of system, the peak load period corresponds to the time of year when the reliability of the system is most vulnerable (point A in Figure 15), as either a shortage of supply or an extreme peak event could trigger a loss of load. The focus on planning for the peak period was justified in the sense that a fleet capable of meeting the highest possible load throughout the year would be able to serve lower loads throughout the rest of the year reliably as well.

Figure 15. Illustrative diagram of the need for flexibility.



With increasing penetrations of variable resources, system operators will still face the task of ensuring reliability during peak periods but will face an additional challenge of ensuring sufficient operational flexibility to meet load under rapidly changing conditions. In these systems, operators must respond to "net load"—demand net of variable resources—rather than that of simply load itself. Higher penetrations of variable resources magnify both the variability and uncertainty of the net load; in such an environment, systems must be capable of operating with sufficient flexibility to respond to potentially large changes in the net load (i.e. move from point B to point C in Figure 15, or, more generally, between any two plausible sequential states of net load).

The need for flexibility in power system operations is not entirely new: operators have historically dispatched generators to respond to hourly changes in load and have held operating reserves to accommodate errors in load forecasts and subhourly variability. The novelty of the flexibility challenge at high penetrations of renewables is in the increased amount of flexibility needed: variable resources increase the magnitude and frequency of large ramps in net load and the uncertainties that an operator must accommodate. In such an environment, the question of how to design and operate a system with adequate flexibility while limiting costs and ensuring reliability becomes a central focus rather than a secondary goal to traditional resource adequacy.

Anticipating the growth of variability and uncertainty in net load, a planner seeking to ensure flexibility adequacy must consider how to ensure that the necessary operational flexibility is available. Flexibility in operations—and by extension, the reliable service of load—may be provided by a number of sources:

+ **Utilize flexibility offered by existing dispatchable resources.** A large portion of the existing resources in the Western Interconnection are dispatchable, and their output may be varied over time within operating limits (minimum and maximum output, minimum up and down times, start times, and ramp limitations). In some cases, existing resources may be capable of operating flexibly; in other cases, technical changes and/or investments may be needed to enable greater flexibility from existing resources.

+ **Use interties with neighboring areas to import and/or export energy as needed.** The Western Interconnection has neither a centralized planning authority nor a central optimized dispatch of resources across its footprint; most of the interregional trade in the West occurs through bilateral transactions. Planners in the future will be faced with the policy question of the extent to which they are willing to rely on the flexibility of their neighbors rather than that provided by their own resources.

+ **Curtail output of renewable generators.** If the power system is insufficiently flexible to accommodate all net load conditions, the system operator may be forced to curtail renewable energy output in order to preserve bulk system reliability. This may need to be accomplished in real time in response to oversupply conditions, or may need to occur prospectively in order to ensure that the system can meet potential upward ramping needs. For example, in the illustrative example of Figure 15, if the generation fleet does not have the capability to ramp from B to C in the required time, curtailment of renewable generation could increase the net load at point B and may help the system meet the upward ramp to point C. While the curtailment of renewable generation results in some lost value to ratepayers, there will be instances—especially as penetrations increase—where the operational value of curtailing renewables outweighs the lost value to ratepayers. Efficient use of curtailment to may require new market structures and contractual agreements.

+ **Invest in new flexible generation infrastructure and/or demand-side programs.** Investments in new flexible resources—generation resources such as energy storage and flexible gas resources as well as new demand-side programs such as load shifting, demand response, and load flexibility—may help to facilitate renewable integration.

This study intends to highlight the questions and decisions that resource planners must consider as power systems move toward high renewable penetrations.

1.4 Organization of Report

This report is organized as follows:

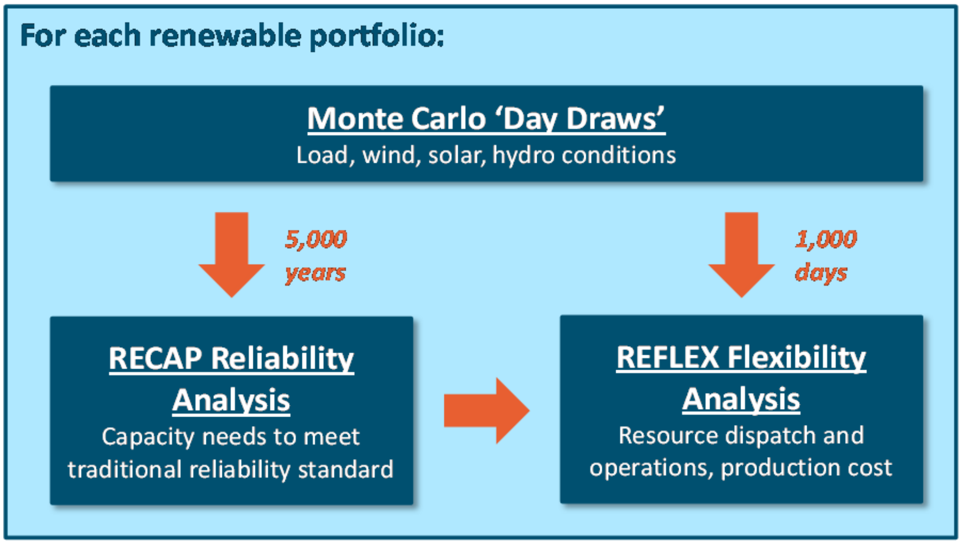
+ **Section 2** provides an overview of the scope of analysis and modeling approach used in this study;
+ **Section 3** presents the data and methods used in the flexibility assessment of the Western Interconnection;
+ **Section 4** presents the results of the resource adequacy assessment (Phase 1);
+ **Section 5** presents and discusses key results from the flexibility assessment (Phase 2);
+ **Section 6** presents key conclusions based on the study's technical findings and discusses the implications of these findings; and
+ **Section 7** is an appendix that discusses technical lessons learned in the modeling process.

2 Modeling Approach

2.1 Overview

The framework for the flexibility assessment divides the analysis into two phases, as outlined in Figure 16: first, the resource adequacy of the generation fleet is studied using a traditional loss-of-load probability modeling framework; subsequently, the operations of the system are modeled using an adaptation of traditional production cost analysis. The two-phase approach to the flexibility assessment is used to isolate challenges that are related to limits on the system's operational flexibility from reliability events that result from a pure shortage of capacity.

Figure 16. Two phases of analysis in the flexibility assessment framework.



The first phase of the analysis uses the **Renewable Energy Capacity Planning (RECAP)** model to assess the resource adequacy of the various regions of the Western Interconnection using an industry-standard loss-of-load-probability modeling framework. In order to compare the reliability of a system with a minimum reliability threshold, RECAP uses a probabilistic, time-sequential model of both loads and resources to simulate thousands of years of possible conditions in the electric sector; examining such a breadth of conditions is needed because loss-of-load events are exceptionally rare. The result of the first phase is an identification of "pure" capacity needed to avoid loss of load that results purely from a shortage of resources.

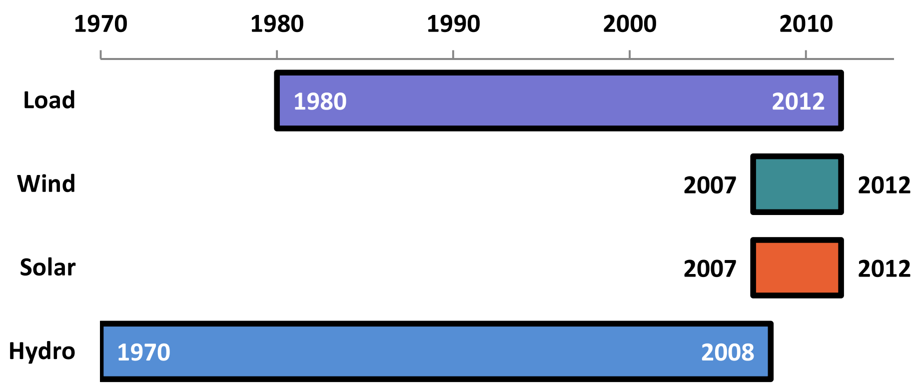
The second phase uses the **Renewable Energy Flexibility (REFLEX)** model for **PLEXOS for Power Systems**—an adaptation of traditional production cost—to evaluate the operations of the electric system studied in the first phase. The computational complexity of this question is, by necessity, substantially greater than the reliability analysis of the first phase; production cost models such as the one used herein use a least-cost unit commitment and dispatch algorithm to minimize the costs of operating an electric system subject to a broad set of constraints. The complexity of the unit commitment problem necessitates the analysis of a reduced set of conditions in comparison to the first phase; instead of simulating thousands of years, operations of the system are simulated for thousands of plausible days in a search for conditions that may challenge the flexibility of the system.

In both of the phases of analysis, this study uses random sampling of weather-correlated load, wind and solar conditions as well as a range of historically-based hydroelectric conditions in order to capture a breadth of plausible system conditions. Each "draw" analyzed in this study—a twenty-four hour period—consists of hourly profiles for load, wind, and solar; as well as a daily budget and operating constraints for hydroelectric generators in the Western system. Because data for these four variables is rarely available over a long and consistent time period, time-synchronous data, which captures the proper correlations, does not typically represent the full distribution of conditions a system may experience. To better represent the distribution of possible system conditions, this study creates a set of random load, wind, solar, and hydro draws by matching the conditions from large historic libraries of each while preserving underlying seasonal and weather-related correlations among them.[13] In order to capture as diverse a set of conditions as possible, draws are created from extensive sets of load, wind, solar, and hydro performance data; the extent of

[13] A more detailed description of the methodology used to match load, wind, solar, and hydro draws while preserving key weather-driven and seasonal relationships among them is presented in Section 3.1.3.

each data set used in this study is shown in Figure 17. These datasets serve as the backbone for analysis of both resource adequacy and operational flexibility presented herein.

Figure 17. Historical conditions used to develop load, wind, solar, and hydro data for draws.



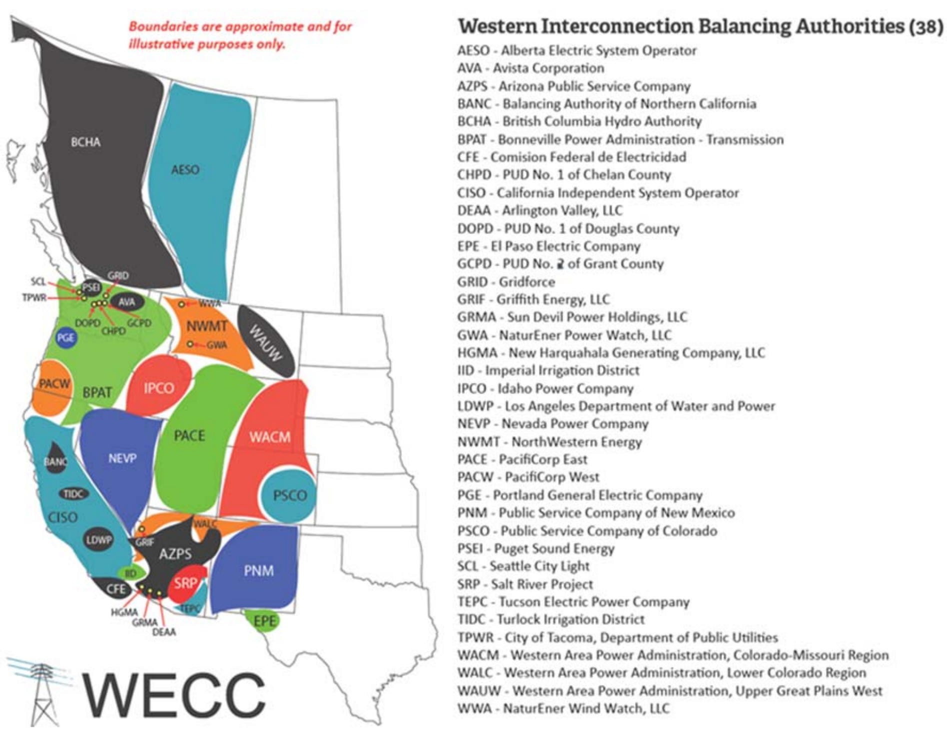
To simulate detailed operations of generators across the Western Interconnection, this study relies heavily on the production cost platform PLEXOS for Power Systems and the technical constraints on generators developed for TEPPC's 2024 Common Case. This analysis deviates from the Common Case with respect to the treatment of loads, renewable resources, hydro resources, and transmission constraints, all of which are discussed herein.

2.2 Study Scope

2.2.1 GEOGRAPHIC SCOPE

The Western Interconnection, shown in Figure 18, is composed of 38 balancing authority areas (BAs), each responsible for ensuring the balance between load and generation within its footprint on an instantaneous basis. This is achieved through the scheduling and dispatch of generation resources, the scheduling of interchange with neighboring balancing authorities, and the reservation of generation capacity to meet ancillary services needs that allow the system to balance subhourly fluctuations and respond to contingencies.

Figure 18. Balancing authorities of the Western Interconnection.



While each individual BA is responsible for ensuring the adequacy of its resources to serve load, examination of resource adequacy and operational flexibility at this granular level ignores the interconnectedness of many BAs and the geographic diversity of load and resources that mitigate reliability challenges. Consequently, this study adopts a regional perspective, examining reliability and flexibility in the five regions shown in Figure 19, each representing an aggregation of BAs in the West summarized in Table 4. The choice of regional boundaries was based on a number of considerations, including:

+ **Existing institutional infrastructure supporting coordination in planning and/or operations.** With this study's focus on questions related to flexibility resource planning—and, relatedly, on operations—the regional framework is useful because most of the regions evaluated in this study have some existing degree of coordination in planning and/or operations.
+ **Homogeneity of loads and renewable resources.** Across the WECC, there is a high degree of diversity in loads and resources. The boundaries chosen for this study divide the WECC into regions with relatively uniform load patterns—that is, loads with similar seasonal and daily patterns. Similarly, each region is also characterized by its locally available renewable resources and its relative preferences for those resources.
+ **Limited major internal transmission constraints.** Regional boundaries are chosen to highlight major existing transmission paths in the WECC that act as conduits for power exchange in the West.
+ **Consistency with TEPPC conventions.** The division of the Western Interconnection into the five regions shown in Figure 19 is largely consistent with the groupings of balancing authorities used by TEPPC in modeling resource adequacy and the sharing of reserves in operations.

Figure 19. Five regions of focus for this study.

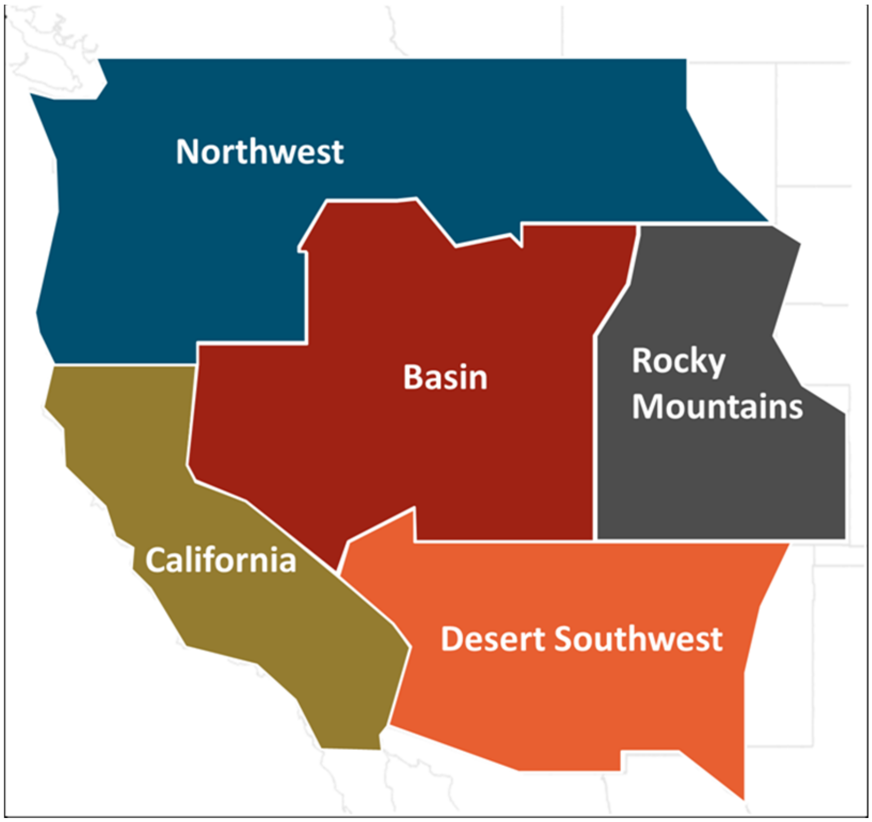


Table 4. Balancing authorities included in each region in the study.

Region	Constituent Balancing Authorities
Basin	IPCO, PACE, NEVP[14]
California	BANC, CISO, IID, LDWP, TIDC
Northwest	AVA, BPAT, CHPD, DOPD, GCPD, NWMT, PACW, PGE, PSEI, SCL, TPWR, WAUW
Rockies	PSCO, WACM
Southwest	AZPS, EPE, NEVP[14], PNM, SRP, TEPC, WALC
Excluded	*AESO, BCHA, CFE*

[14] In this study, the Nevada Power BAA is bifurcated between two regions; Northern Nevada (formerly SPP) is included with the Basin, and Southern Nevada (formerly NEVP) is included in the Southwest.

Notably, this study omits the Canadian balancing authorities (BC Hydro and the Alberta Electric System Operator) as well as the Comisión Federal de Electricidad (CFE). For a variety of reasons, the same records of wind, solar, and hydro data used to model the rest of the Western Interconnection are not available outside the United States, so their interactions with the rest of the West are not considered within the scope of this study. This may have the effect of overstating flexibility challenges in the five regions considered, especially because BC Hydro's large hydroelectric generation fleet has often been operated in response to market conditions and could, in theory, be used to facilitate the integration of renewables in the future. The extent to which this is true, of course, would depend on load and resource conditions in British Columbia as well as the willingness of the operator to use the flexibility of that system in such a fashion.

Through its organization as a collection of regional studies, this study is intended to mimic the perspectives of resource planners in each region seeking to understand future renewable integration challenges while recognizing that the Western system is not centrally and optimally dispatched. While some degree of coordination exists among balancing authorities in today's system, power exchange in the Western Interconnection is generally conducted through long-term contractual arrangements or bilateral agreements rather than through a centralized optimal dispatch of resources. As a result, the ability of the system to integrate higher penetrations of renewables successfully depends not only on the technical capabilities of the generating fleet but on the scheduling practices and conventions of neighboring balancing authorities and regions.

The current practice of bilateral transactions and scheduling of power exchange does not result in least-cost dispatch across the entire interconnection, as is assumed in conventional production cost modeling. The regional approach taken in this study provides the ability to mimic the "friction" that prevents optimal dispatch in today's operations. This approach, which allows this study to examine both technical and institutional barriers to renewable integration in the Western Interconnection, distinguishes this study from a number of prior technical analyses of high renewable penetrations that have examined the ability of the Western Interconnection to balance high penetrations of renewable generation as an integrated whole.

Using the regional approach, this study identifies and characterizes challenges that may accompany higher renewable penetrations in each region. "Challenges" may appear in a number of forms, including the need to curtail renewable generation due to oversupply or ramping constraints; substantial changes in the dispatch patterns of coal, gas, and hydro generating resources from their historical utilization; and constraints on interregional power exchange. The degree to which each of these issues may materialize at higher renewable penetrations will vary from one region to another depending on the characteristics of loads and renewable generation as well as the composition of the non-renewable generation fleet. By studying each region individually, this study seeks to highlight both the nature of the integration challenges that each region may face as well as the measures and/or steps that may prove most effective in facilitating renewable integration.

The five regions examined in this study encompass the U.S. portion of the Western Interconnection. Exclusion of the AESO, BC Hydro, and CFE BAs is due to the dearth of publicly available operational data with which future flexibility issues in each of these jurisdictions may be examined; however, the approach and techniques used in this assessment could be applied just as validly to each of these areas if such data becomes available.

2.2.2 RENEWABLE PORTFOLIOS

This analysis examines reliability and the need for flexibility for two future possible portfolios of renewable resources in the Western Interconnection:

+ A **Common Case**, which reflects current state RPS targets as captured in TEPPC's 2024 Common Case; and
+ A **High Renewables Case**, which includes additional wind and solar generation throughout the footprint of the Western Interconnection in excess of current policy goals.

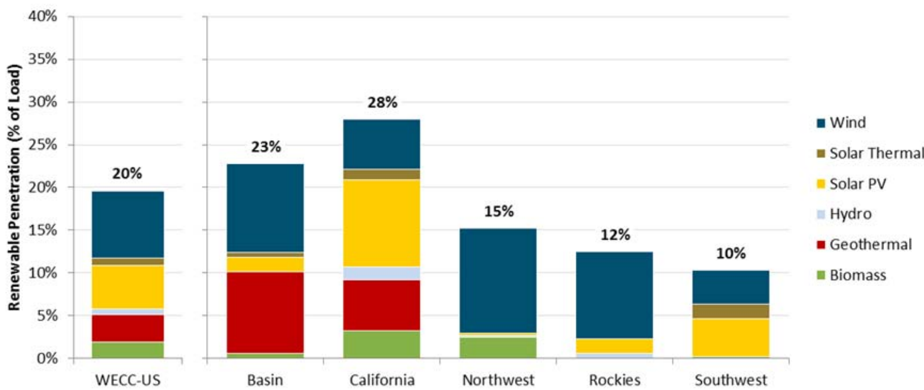
Each portfolio contains a geographically and technologically diverse set of renewable resources intended to represent a plausible future for renewable development in the Western Interconnection.

The Common Case analyzes in this study reflects current renewable portfolio standard (RPS) goals in most states as captured in TEPPC's 2024 Common Case.[15] The assumptions of future renewable development in accordance with these policies were specified by TEPPC with input from stakeholders during the development of the 2024 Common Case. Across the study footprint, renewable

[15] At the time this investigation was undertaken, California's RPS target was still 33%, as SB350 had not yet been passed by the legislature.

generation serves approximately 20% of load in the Common Case. The mix of renewable resources in each region, expressed as a percentage of total electric load, is shown in Figure 20.[16]

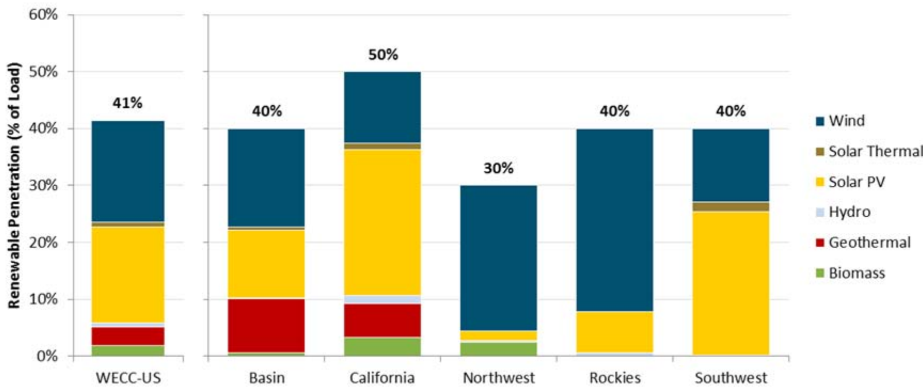
Figure 20. Renewable penetrations studied in the Common Case (based on TEPPC's 2024 Common Case)



In addition to the Common Case, this study explores flexibility challenges and integration strategies for a High Renewables Case developed specifically for analysis in the Flexibility Assessment. The mix and penetration of renewable resources in each region in the High Renewables Case is shown in Figure 21.

[16] In reporting the penetration of renewable generation, this study uses the convention of calculating the combined penetration of behind-the-meter and wholesale renewable generation as a percentage of total load at the transmission level. It therefore differs from the accounting used in most state renewable portfolio standards in three respects: (1) behind-the-meter solar PV is included in the renewable penetration; (2) the penetration is expressed as a percentage of load a the transmission level rather than at the customer meter; and (3) it does not account for the contractual entitlement of loads in one region to renewable generation in another based on existing long-term contracts (e.g. wind generation in the Columbia River Gorge under contract to utilities in California is shown in the Northwest, not in California).

Figure 21. Renewable resources included in the High Renewables Case



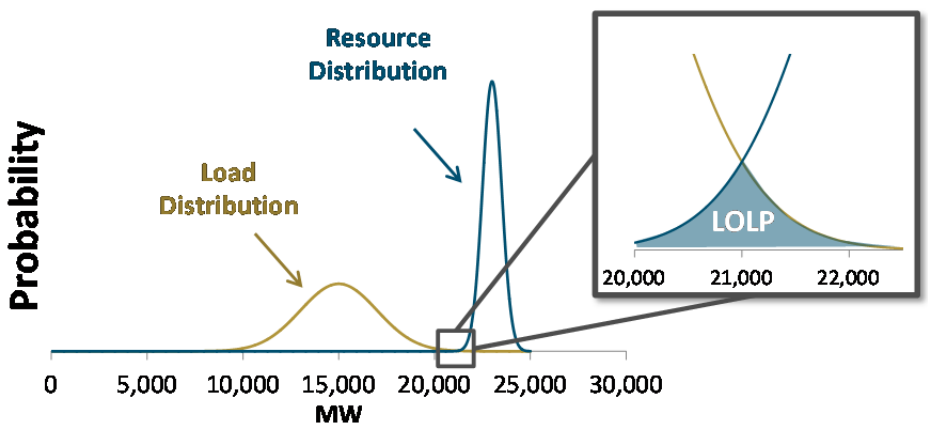
In its choice of renewable portfolios for study in the High Renewables Case, this study seeks to examine a level of penetration that will present challenges to the flexibility of the system. Each region's portfolio is chosen to examine a pressing question related to the interaction between increased renewable penetration and the region's existing resources. In regions with relatively high penetrations of solar (Basin, California, and Southwest), the primary question investigated in this study is the extent to which the non-renewable fleet in each region can accommodate the large concentration of solar generation during the middle of the day. In regions with high wind penetrations (Northwest and Rocky Mountains), the motivating question is how the hour-to-hour, day-to-day, and season-to-season variability (and associated forecast uncertainty) of wind generation may be balanced by non-renewable resources.

2.3 Phase 1: Resource Adequacy Assessment

2.3.1 HISTORICAL APPROACH TO RELIABILITY PLANNING

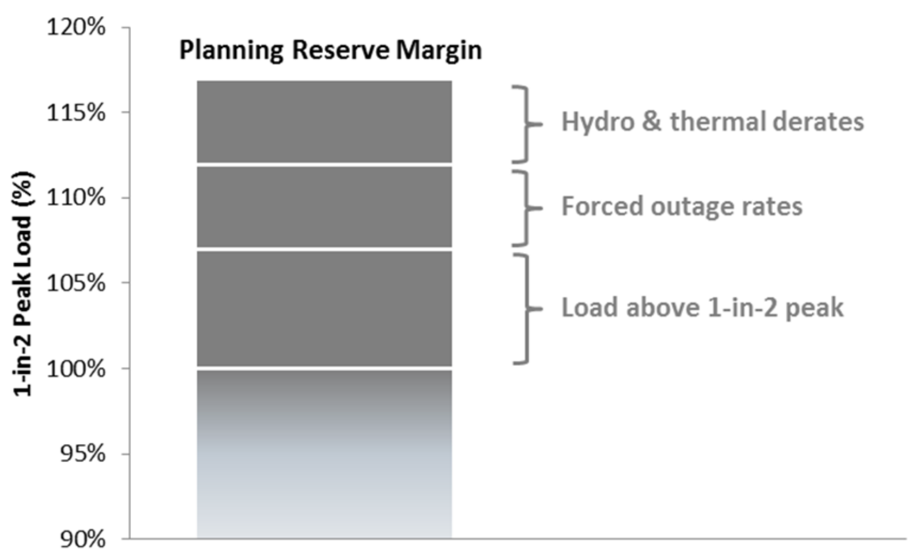
Reliability modeling has a long history in electric sector resource planning. Loss-of-load-probability (LOLP) modeling, a modeling framework in which the availability of generation resources is compared against potential system load across a broad range of possible conditions, has been established as the industry standard. Because tolerance for loss of load due to generation inadequacy is typically very low—a common standard is "one day in ten years"—such an approach is necessary to capture the tails of the distribution during which loss of load may occur (see Figure 22).

Figure 22. Loss of load probability modeling framework.



Because of the computational complexity of loss-of-load-probability modeling, planners commonly use a "planning reserve margin" (PRM) target as a simpler benchmark for generation adequacy. Traditionally, a system's PRM has been defined as the amount of nameplate generation capacity (including available imports) in excess of the system's expected 1-in-2 peak demand; many utilities and generation planners have established PRM targets of 13-17%. The reserve margin above the 1-in-2 peak demand may be understood as the additional capacity needed to ensure reliability while accounting for a number of inherently uncertain factors in an electric system, including variations in annual peak demand from the forecast 1-in-2 peak, forced outages of generators and/or transmission lines, and thermal derates of generation capacity. While the PRM target itself should be derived through loss-of-load-probability modeling, these factors can be understood as the "building blocks" that contribute to the need to hold a reserve margin above the 1-in-2 peak, as illustrated in Figure 23.

Figure 23. Planning reserve margin "building blocks"



2.3.2 RELIABILITY PLANNING WITH VARIABLE GENERATION

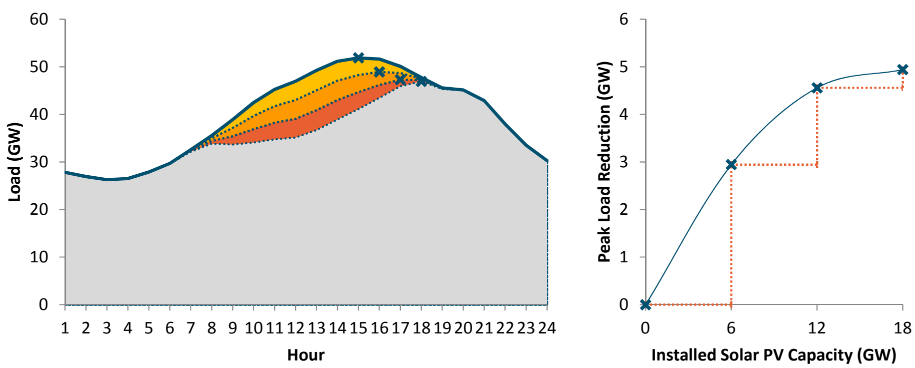
The simple planning reserve margin framework has proven adequate for reliability planning when most of the capacity in a system is dispatchable and can be called on to serve loads during times of peak; however, with the addition of increasing amounts of variable generation from wind & solar technologies in many areas in the Western Interconnection, the simple conventions of the planning reserve margin are challenged. Because such resources cannot be controlled by the operator and often produce at levels below their nameplate capacity during times of peak load, variable generation technologies do not contribute the same amount of reliable capacity to a system as dispatchable generators.

Across the Western Interconnection, planners have taken a variety of approaches to adapt the simple PRM framework for continued use with increasing penetrations of variable generation. The most common approach has been to develop rules of thumb by which the nameplate capacity of wind and solar resources may be adjusted in the calculation of the PRM to reflect their limited contributions to system reliability; such adjustments are commonly derived through an analysis of historical and/or simulated output during peak load periods. This approach typically yields multipliers of 40-70% for solar technologies and 5-30% for wind technologies, and values in this range are currently used in many reliability planning exercises across the Western Interconnection.

While this rule-of-thumb approach may provide a reasonable approximation of the reliable capacity for variable generation at low penetrations, it fails to capture the impact of variable generation on the timing of the net peak:

increasing penetrations of variable generation will cause the timing of the net peak to change, a concept illustrated in Figure 24 for increasing penetrations of solar PV. As the net peak shifts away from periods in which output from variable resources is relatively concentrated to periods in which it is less concentrated, the marginal contribution of those variable resources is reduced, resulting in declining returns to scale with each renewable technology.

Figure 24. Illustrative reduction of peak load impact of increasing penetrations of solar PV.



Incorporating this effect into the assessment of system reliability is made possible by a return to the original loss-of-load probability modeling framework, from which a measure of the "effective load carrying capability" (ELCC) of variable generation may be derived.[17] In the context of LOLP modeling, ELCC is defined as the additional firm load that can be met by a generation resource while maintaining the same level of reliability as the base system. Notably, ELCC

[17] Garver, L.L., "Effective Load Carrying Capability of Generating Units," Power Apparatus and Systems, IEEE Transactions on , vol.PAS-85, no.8, pp.910,919, Aug. 1966 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4073133&isnumber=4073117

captures the interactive effects between renewable generators as penetrations increase, and is a suitable measure of capacity for use in traditional planning reserve margin calculations. Robust generation adequacy planning under high penetrations of variable generation therefore requires a reestablishment of the linkage between the conventions used in the planning reserve margin framework and the underlying stochastic reliability modeling through which those conventions were originally derived.

2.3.3 RECAP MODELING FRAMEWORK

In order to provide a platform for the reliability analysis of electric systems with high penetrations of renewable generation, E3 developed the **Renewable Energy Capacity Planning (RECAP)** model, a stochastic loss-of-load-probability model that uses datasets of load, wind, and solar conditions to provide a forward-looking assessment of system reliability. The RECAP model relies on industry standard methods for assessing power system reliability[18,19] with additional features that allow it to make better use of often limited datasets for wind and solar (see Section 3.1.3 on draw methodology). The RECAP model simulates years by randomly sampling loads, renewables, monthly hydroelectric energy budgets, generator outages and maintenance. Outage conditions occur when load exceeds the sum of available supply resources, also accounting for operating reserves.[20]

[18] R. Billinton and R. N. Allan, Reliability Evaluation of Power Systems, Second ed. New York: Plenum Press, 1996.
[19] R. Billinton and W. Li, Reliability Assessment of Electric Power Systems Using Monte Carlo Methods. New York: Plenum Press, 1994.
[20] Operating and contingency reserves are often ignored in LOLP modeling and are not included in results for this report except where explicitly stated

As already noted, the conditions under which loss of load occurs are exceedingly rare and it becomes necessary to analyze a large number of Monte Carlo draws before reliability statistics converge. This study simulates plausible combinations of load and resources across 5,000 years, to produce the metrics intended to help inform reliability planning. Key probabilistic outputs from the RECAP model, each of which depends on the characteristics of loads and generation fleet, include:

+ **Loss of load frequency (LOLF):** the expected frequency of reliability events.
+ **Loss of load expectation (LOLE):** the number of hours of expected lost load.
+ **Expected unserved energy (EUE):** the amount of load that is unserved during reliability events.
+ **Normalized expected unserved energy (EUE-Norm):** Expected unserved energy divided by the sum of annual load.

In this study, the "1-in-10" standard is interpreted to mean an average of one loss of load event every ten years, alternatively stated as a loss of load frequency of 0.1. Among the common interpretations, this is the most stringent[21] metric for gauging reliability; the use of a conservative standard helps to ensure that reliability challenges identified in the flexibility assessment may be attributed to a lack of flexibility rather than a lack of capacity.

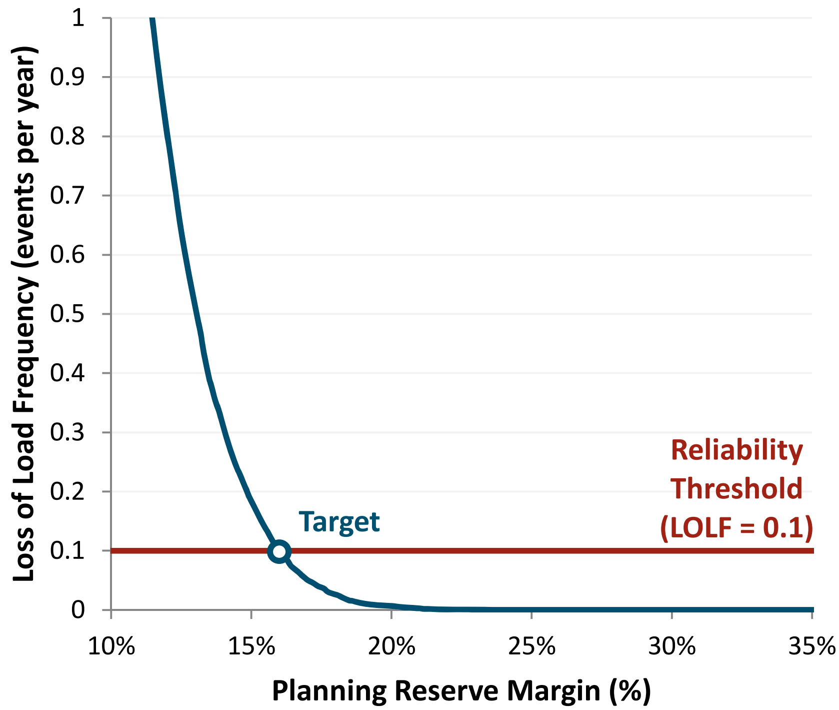
[21] The Brattle Group & Astrape Consulting, *Resource Adequacy Requirements: Reliability and Economic Implications*. Available at: **http://www.ferc.gov/legal/staff-reports/2014/02-07-14-consultant-report.pdf**

A second element of core functionality is the capability to calculate the **effective load carrying capability (ELCC)** of a resource (or portfolio thereof), a measure of its contribution to system reliability, expressed in terms of firm megawatts of demand served. This type of metric is particularly useful for reliability planning with high penetrations of variable resources, as it provides a rigorous analytical foundation with which to measure the contribution of variable energy resources to system reliability in the context of a traditional planning reserve margin calculation.

Third, RECAP evaluates the **planning reserve margin** of the system in order to establish the link between the loss-of-load-probability analysis and the simpler conventions that have become common practice. The calculation of a planning reserve margin allows RECAP not only to link the reliability of a given system to its current planning reserve margin, but to derive each system's target PRM based on a loss-of-load-frequency threshold of one event in ten years. The target PRM represents the minimum planning reserve margin for a system for a system to adhere to acceptable standards of system reliability.

The functional relationship between the stochastically derived loss-of-load frequency and the simpler planning reserve margin is shown in Figure 25. The shape of the curve that defines the relationship between LOLF and PRM will reflect the characteristics of the electric system analyzed; the curve itself and the resulting target planning reserve margin depend on such factors as the size (and amount of diversity) in the system, the inter-annual variability in weather conditions, and the probabilities of forced outages on individual generators.

Figure 25. Illustration of relationship between loss of load frequency and planning reserve margin



2.4 Phase 2: Flexibility Assessment

2.4.1 BACKGROUND

Flexibility analysis has a much shorter history than capacity-based reliability analysis, as it has only become a relevant question in resource planning in the context of higher renewable penetrations in recent years. Because of this relatively short history, there is currently no industry-standard methodology for assessing the flexibility of a power system. One of the goals of this study is therefore to investigate the key drivers of the need for flexibility in an electricity system and to identify critical modeling considerations in flexibility planning analyses going forward.

The approach used in this study builds on production cost modeling. Production cost models, which simulate the optimal unit commitment and economic dispatch of an electric system subject to a set of constraints, are commonly used to model the operations of electric systems. Such models are employed across a diverse range of applications and types of analyses, including transmission planning studies, renewable integration studies, asset valuation exercises, market price forecasts, and integrated resource plans. Depending on the purpose of the modeling exercise, production cost models can produce a range of different outputs, including flows on transmission lines, operating behavior of individual power plants, and market prices for energy and ancillary services over various timescales and at different locations on an electric system.

The computational complexity of production cost models typically far exceeds that of loss-of-load-probability modeling; while the latter simply compares total available generation capacity against load, the former provides a least-cost solution for how those resources should be optimally committed and dispatched subject to a large number of additional constraints. The types of constraints imposed on production cost models—which include such parameters as technical constraints on the operations of individual power plants, the need for individual entities to hold operating reserves in addition to meeting load, and a representation of the underlying transmission network—may range in their complexity depending on the particular application of the model. The computational complexity of the simulation depends on the level of detail included with respect to each of these classes of constraints and on the chronological extent of the simulation (e.g. one day versus several years). With today's computational resources, full representation of all possible constraints over a wide range of conditions is not yet practical. Accordingly, the level of detail is often tailored to the underlying purpose of the particular application in order to focus on key outputs while allowing for simplifications with respect to less important constraints. For example:

+ TEPPC's annual transmission planning studies, which study future flows on transmission lines in the West under a variety of future scenarios, include a full representation of the high-voltage transmission network of the Western Interconnection—over 25,000 nodes are included in a nodal transmission network—but makes simplifications in the unit commitment logic used to dispatch generators.
+ NREL's Western Wind and Solar Integration Study (Phase 1), which studied the impact of high penetrations of variable generation on the integrated Western Interconnection, used a multi-stage unit commitment model as well as a five-minute real-time dispatch of generation resources but used a simplified zonal representation of the transmission network.

In order to facilitate effective planning for flexibility, the modeling approach used should prioritize detail with respect to the key constraints on a system that might limit its operational flexibility.

2.4.2 REFLEX MODELING FRAMEWORK

This study uses the **Renewable Energy Flexibility (REFLEX) model** for **PLEXOS for Power Systems**, an application of production cost modeling that has been tailored specifically to use in the context of informing flexibility planning decisions. While much of the data and substantial portions of the modeling approach utilized in this study can be and has been used for operations modeling, the approach taken for this analysis is differentiated from an operational study that seeks to predict how a system may operate under higher penetrations of renewables. This study instead seeks to understand more specifically how a planner might use simulated operational information to inform a flexibility planning decision. In this planning context some operational and institutional constraints are designed not to predict how a system may operate, but are instead designed to address, in an analogous way to reliability planning, the extent to which a planner chooses to reliably count on a resource or transaction to alleviate potential flexibility challenges in the future. To address these questions, E3 has developed the REFLEX production cost modeling framework, which is implemented within PLEXOS for Power Systems.

With its off-the-shelf modeling capabilities, PLEXOS for Power Systems provides a strong foundation for the assessment of generation flexibility. PLEXOS enables the use of a multistage unit commitment and dispatch simulation—used to model a day-ahead commitment cycle as well as the subsequent actual hourly dispatch—to capture the effects of scheduling based on imperfect forecasts on operations as well as the differing behaviors of inflexible units with longer start-up times. In this simulation, PLEXOS enforces a variety of security constraints that limit the flexibility of conventional resources, including maximum output, minimum stable level, minimum up and down times, and maximum ramp rates. In addition, PLEXOS allows for cooptimization of energy and ancillary services dispatch, ensuring that the dispatch of units on the system reflects the need to hold operating reserves to accommodate the subhourly variability and forecast uncertainty of load, wind, and solar generation.

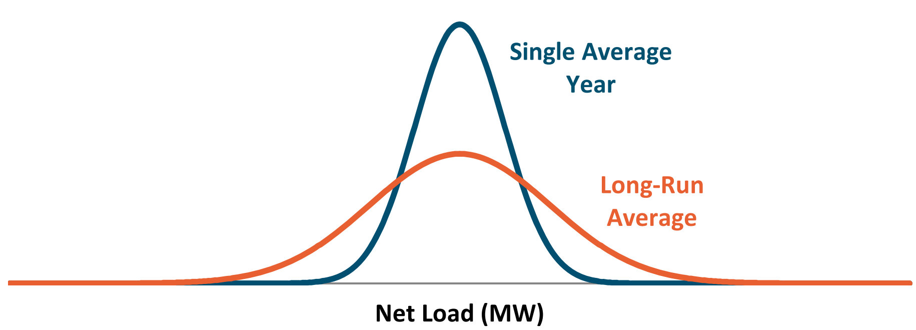
Table 5. Key characteristics of operational flexibility modeling framework.

Flexibility Assessment Goal	Modeling Approach
Reflect lumpiness of unit commitment decisions in dispatch	Use mixed integer program (MIP) logic to model unit commitment decisions
Account for the impact of uncertainty inherent in wind & solar production on operations	Model two dispatch cycles (day-ahead and real time) with associated profiles for output from variable energy resources
Allow for economic assessment of curtailment as an operational strategy	Allow for renewable curtailment at a penalty price that reflects the lost value of the renewable attribute, captured by the "replacement cost" of renewable energy
Capture a broad range of load, wind, solar, and hydro conditions	Construct draws from load, renewable, and hydro conditions that reflect a much longer time horizon than a single year of data
Capture appropriate start-up decisions in dispatch	Simulate each draw as a three-day period, including the neighboring days in the simulation to allow proper characterization of long-start units and avoid edge effects in the simulation but considering only the middle day for results
Quantify consequences of holding inadequate reserves	Determine the provision of flexibility reserves endogenously within the economic dispatch of REFLEX by parameterizing the relationship between a reserve shortfall and the expectation of loss of load and/or curtailment
Allow for limits on imports/exports as a strategy for renewable integration	Impose limits on flows and ramping limits on major WECC transmission paths in order to control the influence of changing net load patterns between regions
Represent the flexibility of the hydroelectric fleet as historically observed	Constrain the operating range and ramping capability of the hydroelectric fleet based on historical hourly operations

With these attributes, the REFLEX modeling framework aims to provide answers to three questions to inform planning for flexibility:

+ **How often is the system expected to encounter flexibility constraints?** This requires simulation of a more complete collection of system conditions than are typically encountered in a single year (the standard study period for production cost analyses). The REFLEX approach addresses this need by combining the Monte Carlo framework of reliability modeling with the analytical rigor of production cost modeling to provide a view of plausible operating patterns and needs across a broad range of conditions, incorporating the full possible distribution of conditions on the electric system (Figure 26). In this study, the same probabilistic draw approach used to generate plausible days of conditions for the capacity analysis is used to create inputs for the flexibility assessment, allowing for statistical analysis of the observed operational challenges.

Figure 26. Difference between studying a single year and a long-run distribution.



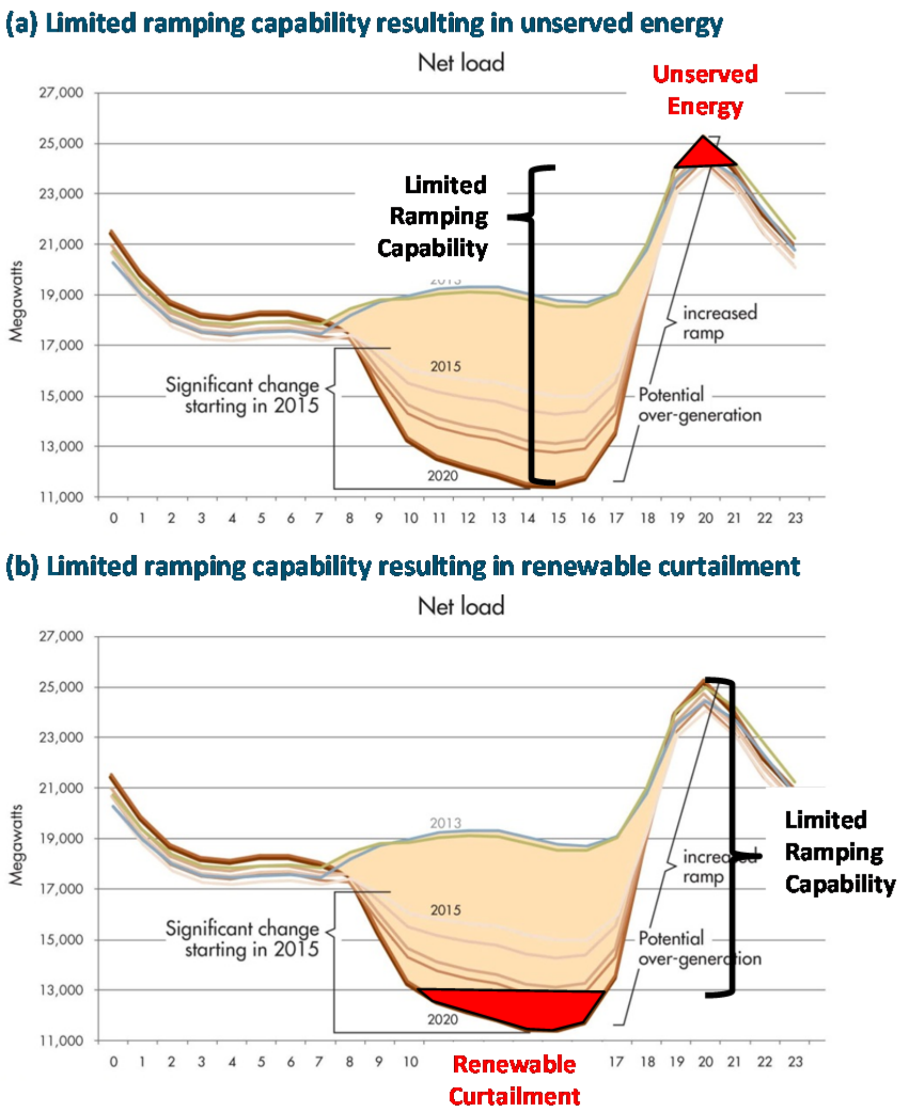
+ **How responsive should the planner assume that the system's resources will be in the event that operations become flexibility**

constrained? Flexibility is limited on an electricity system by two types of constraints: technical constraints related to the physical operating limits of the generators on the system and institutional constraints related to the ability of the system to respond to fluctuations in both load and generation given the limitations of market structures, information transfer, and bilateral arrangements. Just as today's system planner may determine a reliable level of imports from other systems to count toward a planning reserve margin, planners considering flexibility constraints may also need to decide how much flexibility to assume can be provided by neighboring balancing areas during flexibility constrained periods. A planner may conservatively assume that less flexibility is available from resources outside the footprint of the system than they can technically provide in order to ensure that internal flexibility challenges can be mitigated in the long term even if procurement plans or operational paradigms change in the neighboring systems.

+ **To what extent is the system operator (on behalf of the ratepayers) willing to accept the consequences of inadequate resource flexibility?** In traditional reliability analysis, the consequence of inadequate capacity is relatively straightforward: the system is incapable of simultaneously meeting all loads, so some loads are shed. Inadequate flexibility may also lead to loss of load, if for example a large portion of the thermal fleet is cycled off to accommodate renewable resource generation that then experiences a rapid downward ramp that cannot be met with the limited available upward ramping capability (this is illustrated in Figure 27a). However, considering the operator's strong preference for curtailment of generation over loss of load, a more likely (and more economic) scenario is that the system operator would curtail some portion of the renewable energy to ensure that the thermal resources required to maintain reliability can remain online (Figure

27b). In this way renewable curtailment is both a key operational strategy in flexibility-constrained systems and is the primary consequence of inadequate flexibility. The extent to which renewable curtailment is utilized in a flexibility planning analysis will depend on its assumed cost. This study analyzes the economic tradeoff between using renewable curtailment to provide operational flexibility versus relying on flexible thermal resources by testing sensitivities on the renewable curtailment cost penalty.

Figure 27. Illustration of the operational tradeoff between unserved energy & renewable curtailment.



The REFLEX analysis yields a number of metrics and outputs to help characterize flexibility planning challenges, including:

- **Distributions** of the hourly net load, hourly net load ramps, and net load ramps over longer durations (2-hour, 3-hour, etc.);
- **Annual expected values** of oversupply (which must be mitigated with either renewable curtailment or spilling hydropower), annual production cost, and annual CO_2 emissions; and
- **Month-hour expected values** of the above metrics and of operating reserve provisions and violations to illuminate seasonal and diurnal trends and to identify periods with the greatest flexibility challenges.

In addition to these probabilistic metrics, the REFLEX approach provides useful snapshots of challenging operating days that highlight the impacts of variable renewables on net load and the dispatch of available resources to meet the need for flexibility.

REFLEX provides a framework in which the effectiveness of new investments in enabling renewable integration may be tested: "in-and-out" cases allow for the identification of promising (high-value) integration solutions. A scenario can be tested, for example, with and without a new technology like energy storage. The difference in the expected annual cost (including production costs, curtailment cost, and fixed costs of any solutions) between the cases will determine if the new technology provides a cost-effective approach to mitigating integration challenges. The cost effectiveness of integration solutions will depend on a number of exogenous factors, including the level of coordination between planning entities across the West. If each region plans to mitigate its own renewable integration challenges without (or with limited) reliance on its neighbors, then integration challenges within each region will likely be greater and integration solutions will be more cost effective than in a scenario in which regional planners coordinate to take advantage of potential complementarity between regional load and renewable supply imbalances.

The REFLEX framework, through its focus on these key drivers of the need for flexibility and the related physical and institutional constraints, provides a platform upon which to study the potential realization of integration challenges at high penetrations of renewable generation. This study uses this platform to explore the possible magnitude and frequency of such challenges under a specific set of renewable portfolio assumptions, as well as potential steps that might be taken to relieve those challenges, in order to identify barriers and opportunities for renewable integration in the Western Interconnection.

2.4.3 FLEXIBILITY SCENARIOS

This study examines a broad range of factors that influence flexibility. The full range of cases is shown in the table on the next page. It is important to note that none of these cases is intended to be predictive of exactly how the system might look or operate under different renewable penetrations; rather, through pairwise comparisons among the manifold scenarios developed herein, this study intends to provide indicative conclusions on the importance and the impact of a variety of factors in renewable integration. Three types of scenarios are investigated:

+ **Reference Grid.** Both the Common Case and the High Renewables Case are simulated with a set of assumptions described as the 'Reference Grid.' The purpose of these cases is to serve as a central scenario against which the results of others can be compared.

- **Sensitivities.** A number of sensitivities on the Reference Grid scenario are analyzed in order to explore the impact of key assumptions. Key sensitivities studied include limitations on the flexibility of coal plants, variations on hydro constraints, and sensitivities on curtailment prices.
- **Integration solutions.** The study also examines a number of enabling strategies for renewable integration, including increased regional coordination (modeled by relaxing constraints on interties) as well as adding energy storage and new flexible gas generation capacity.

The collection of scenarios considered in the flexibility assessment is presented in Table 6.

Table 6. Scenarios modeled for analysis.

ID	Renewable Build	Intertie Flow Limits	Intertie Ramp Limits	Curtailment Price	Unserved Energy Price	Hydro Ramp	Coal Flexibility	Add'l Gen	Flexibility Reserves
1	**2024 CC**	**Historic**	**Historic**	**$100**	**$50,000**	**Historic**	**Physical**	**none**	**REFLEX**
2	2024 CC	Physical	--	$100	$50,000	Historic	Physical	none	REFLEX
3	2024 CC	Historic	Historic	$100	$50,000	--	Physical	none	REFLEX
4	**High Ren**	**Historic**	**Historic**	**$100**	**$50,000**	**Historic**	**Physical**	**none**	**REFLEX**
5	High Ren	Physical	--	$100	$50,000	Historic	Physical	none	REFLEX
6	High Ren	Physical	Historic	$100	$50,000	Historic	Physical	none	REFLEX
7	High Ren	Historic	Historic	$100	$50,000	--	Physical	none	REFLEX
8	High Ren	Historic	Historic	$100	$50,000	Historic	Physical	Storage (2hr)	REFLEX
9	High Ren	Historic	Historic	$100	$50,000	Historic	Physical	Storage (6hr)	REFLEX
10	High Ren	Historic	Historic	$100	$50,000	Historic	Physical	Storage (12hr)	REFLEX
11	High Ren	Physical	--	$100	$50,000	Historic	Physical	Storage (2hr)	REFLEX
12	High Ren	Physical	--	$100	$50,000	Historic	Physical	Storage (6hr)	REFLEX
13	High Ren	Physical	--	$100	$50,000	Historic	Physical	Storage (12hr)	REFLEX
14	High Ren	Historic	Historic	$100	$50,000	Historic	Limited	none	REFLEX
15	High Ren	Historic	Historic	$100	$50,000	Historic	Physical	Flex Gas CCGT	REFLEX
16	High Ren	Historic	Historic	$100	$50,000	Historic	Physical	none	REFLEX
17	High Ren	Historic	Historic	$30	$50,000	Historic	Physical	none	REFLEX
18	High Ren	Historic	Historic	$300	$50,000	Historic	Physical	none	REFLEX
19	High Ren	Historic	Historic	$1,000	$1,000	Historic	Physical	none	REFLEX
20	High Ren	Historic	Historic	$100	$50,000	Historic	Physical	none	No EFD

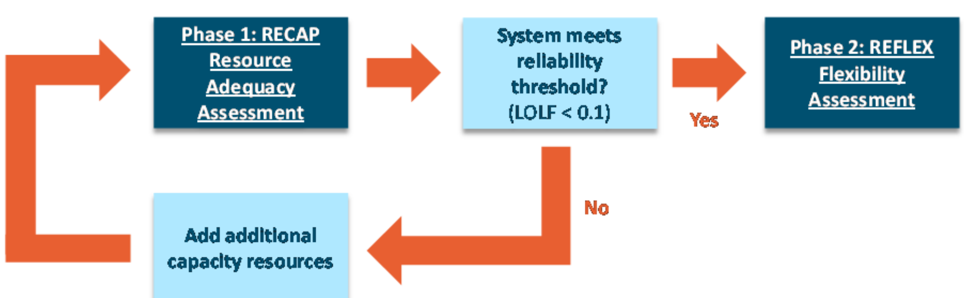
Green rows show modeling assumptions used in 'Reference' scenarios

Orange cells indicate assumptions that deviate from those used to model Reference scenarios

3 Data & Methods

While methodologically distinct, the resource adequacy and flexibility assessments described herein are designed to ensure consistency across input assumptions to the greatest extent possible. The RECAP analysis compares simulated loads and available generation resources stochastically across thousands of years of potential combinations of conditions, accounting for the variability of load, renewable output and hydro conditions, as well as the risk of outages of traditionally dispatchable generators. If the loss-of-load frequency exceeds a threshold of one day in ten years (LOLF > 0.1), the system's reliability is deemed insufficient, and capacity must be added to match the deficit identified. This is a crucial and natural prerequisite for the flexibility assessment, as it ensures that any challenges that arise in the operational modeling can be attributed to a lack of flexibility and are not simply a result of a need for "pure" capacity. This framework is illustrated in Figure 28.

Figure 28. Role of RECAP reliability assessment in the study.



The flexibility assessment then tests whether a system that is adequate from a pure capacity perspective has adequate flexibility to integrate its variable renewables. For internal consistency, the same datasets are used to characterize load, wind and solar availability, generator availability and operating constraints, and hydro availability across both analyses, though the specific applications of these data may differ. The data inputs needed for both the resource adequacy and flexibility assessments, shown in Table 7, are derived from the 2024 Common Case where possible; however, additional data sets are incorporated into the analysis as needed to improve the characterization of various key inputs. This section describes each of the categories of input data and the sources and assumptions used in the resource adequacy and flexibility assessments.

Table 7. Data needs for RECAP & REFLEX analyses.

Category	Data Needs	Units	RECAP	REFLEX
Load	Hourly profiles for multiple years	MW	✓	✓
	Corresponding day-ahead forecasts	MW		✓
Variable Renewable Generation	Hourly profiles for multiple years	MW	✓	✓
• *Solar PV/Thermal* • *Wind*	Corresponding day-ahead forecasts	MW		✓
Conventional Generation	Maximum output (monthly)	MW	✓	✓
• *Nuclear*	Minimum stable level	MW		✓
• *Coal*	Ramp rate	MW/hr		✓
• *Gas*	Forced outage rate	%	✓	✓
• *Biomass*	Maintenance rate	%		✓
• *Geothermal*	Heat rate	Btu/kWh		✓
	Fuel cost	$/MMBtu		✓
	Variable O&M	$/MWh		✓
	Start Cost	$		✓
	Minimum up & down time	Hrs		✓
Hydro	Hydro conditions for multiple years	GWh	✓	✓
• *Conventional hydro*	Sustained peaking capability	MW	✓	
• *Small hydro*	Maximum hydro output	MW		✓
	Minimum hydro output	MW		✓
	Hydro ramping limitations	MW/hr		✓
Interregional Power Exchange	Availability of imports during peak	MW	✓	
	Obligations to export during peak	MW	✓	
	Transfer limits between regions	MW		✓
	Ramp rate limitations	MW/hr		✓
Reserves	Contingency reserves	% of load		✓
	Regulation	% of load		✓
	Flexibility reserves	MW		✓
Penalty Prices	Unserved energy penalty price	$/MWh		✓
	Curtailment cost	$/MWh		✓

3.1 Load, Wind, and Solar Inputs

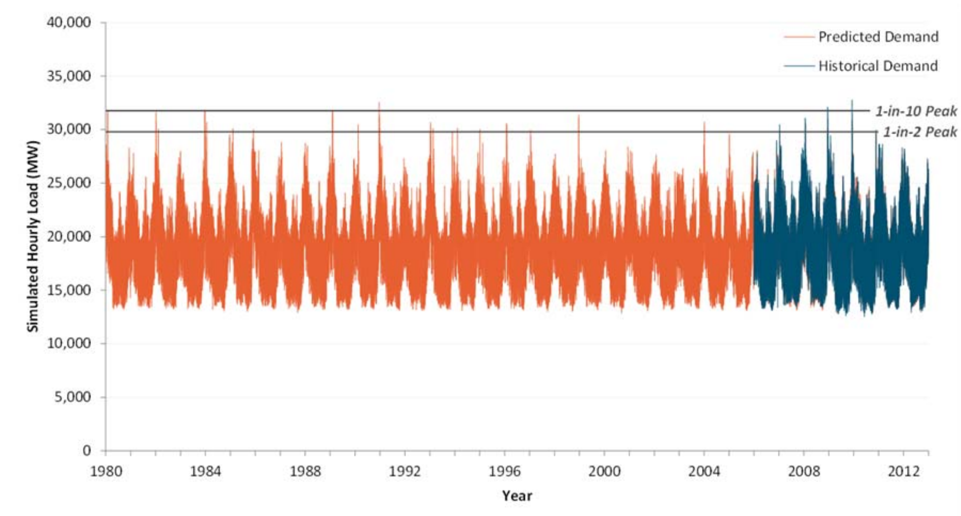
The Monte Carlo framework used by the RECAP model requires a robust characterization of possible load and renewable conditions. In order to represent the distribution of possible conditions, this study uses a stratified sampling methodology from multiple years of asynchronous load and renewable data to create “draws” of individual days. This section describes:

- How historical hourly profiles are developed for load, wind, and solar resources; and
- How individual day “draws” are created for both reliability and flexibility assessment using a stratified sampling methodology from each data set.

3.1.1 LOAD PROFILES

This analysis uses simulated load profiles that reflect expected patterns in hourly load given the meteorological conditions experienced over the historical period from 1980 to 2012. These profiles are created using a neural network regression that links load with daily weather indicators, using the observed relationship between the two from the period 2005-2012 to simulate load shapes consistent with the 2024 Common Case for each weather year between 1980 and 2012 and for each of the load areas modeled in the Common Case. Figure 29 shows an example of the hourly shape simulated for the aggregation of load areas in the Northwest.

Figure 29. Simulated hourly load shapes for the Pacific Northwest, 1980-2012.



This approach provides a rich dataset for both reliability and flexibility modeling that captures a wide range of possible weather—and resulting load—conditions. This approach is particularly useful in the stochastic assessment of resource adequacy, as the longer historical record establishes a probability distribution for extreme load events that can contribute to the risk of loss of load.

Corresponding forecasts for the day-ahead and hour-ahead timeframe are developed through an algorithm that pairs each day with another seasonally appropriate historical day in order to match aggregate assumed statistics for the mean absolute error (MAE) in each of these timeframes.

3.1.2 WIND & SOLAR PROFILES

The renewable portfolios studied in Base Case and High Renewables Case amount to penetrations of 20% and 42% of annual load across the study footprint, respectively. In both scenarios, a small share of this total—7% of load—is supplied by various biomass, geothermal, and small hydro resources; the remaining balance is supplied by variable and uncertain wind & solar resources. In the Common Case, shown in Table 8, wind and solar resources account for 13% of load. In the High Renewables Case, shown in Table 9, penetration of wind and solar is 34% across the study footprint.

Table 8. Wind & solar resources in the 2024 Common Case (GWh).

Technology		Basin	California	Northwest	Rockies	Southwest
Solar PV	Fixed Tilt	989	15,128	62	187	3,385
	Tracking	—	7,912	—	12	1,180
	Rooftop	351	8,257	326	1,027	3,217
Solar Thermal	No Storage	—	1,597	—	—	1,140
	Storage	440	2,189	—	—	1,659
Wind		8,143	18,474	23,805	7,478	6,451
Total		**9,922**	**53,558**	**24,194**	**8,705**	**17,032**
Total (% of load)		*12%*	*16%*	*13%*	*12%*	*10%*

Table 9. Wind & solar resources in the High Renewables Case (GWh).

Technology		Basin	California	Northwest	Rockies	Southwest
Solar PV	Fixed Tilt	2,933	25,819	1,278	1,401	13,247
	Tracking	4,245	33,578	929	1,829	16,667
	Rooftop	2,414	20,330	1,120	2,038	11,666
Solar Thermal	No Storage	—	1,597	—	—	1,140
	Storage	440	2,189	—	—	1,659
Wind		13,644	39,148	49,492	23,645	20,936
Total		**23,676**	**122,661**	**52,819**	**28,913**	**65,315**
Total (% of load)		*29%*	*37%*	*28%*	*40%*	*38%*

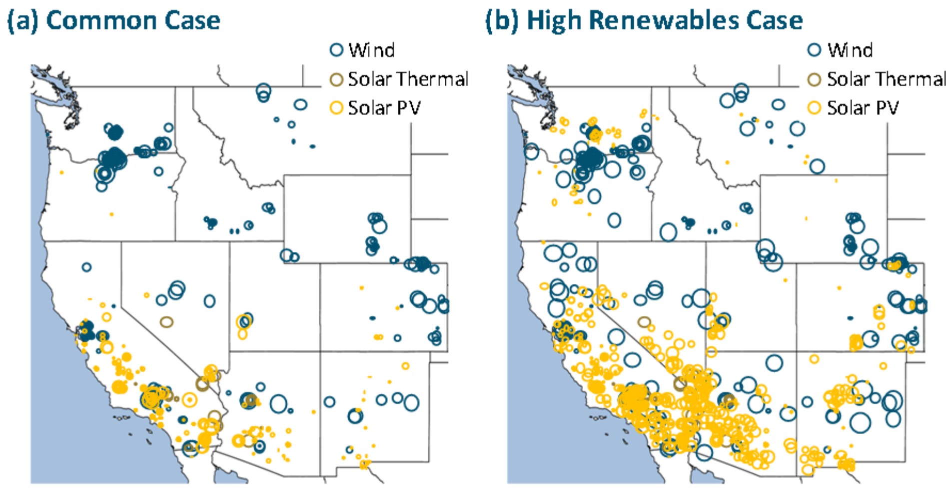
To represent the variable and uncertain output of these resources, this study uses several data sources to develop hourly profiles for the output of wind and solar resources throughout the Western Interconnection across multiple years:

- Wind profiles are from NREL's **Wind Integration National Dataset (WIND) Toolkit**, which provides five-minute simulated wind profiles for 126,000 sites across the continental United States for the period 2007-2013.[22]
- Solar PV profiles are from NREL's **Solar Integration National Dataset (SIND) Toolkit**, which provides 30-minute simulated profiles for both distributed and utility-scale solar PV installations for the period 2007-2013.[23]
- Solar thermal profiles are simulated using NREL's **System Advisor Model (SAM)**. The input files used in the simulation are from the same dataset used to generate the solar PV profiles in the SIND Toolkit.

This study uses a subset of the profiles available in these datasets in order to represent the renewable resources of the Common Case and the High Renewables Case. The profiles used in this study are selected from the original datasets based on geographic location and technology configuration. The geographic locations of wind and solar resources in the Common Case—shown in Figure 30—are based on information provided by WECC, supplemented with data obtained from the Energy Information Administration (EIA) for existing facilities. The additional profiles needed to reflect the incremental resources of the High Renewables Case are selected in each region from the remaining high-quality resource sites in each data set.

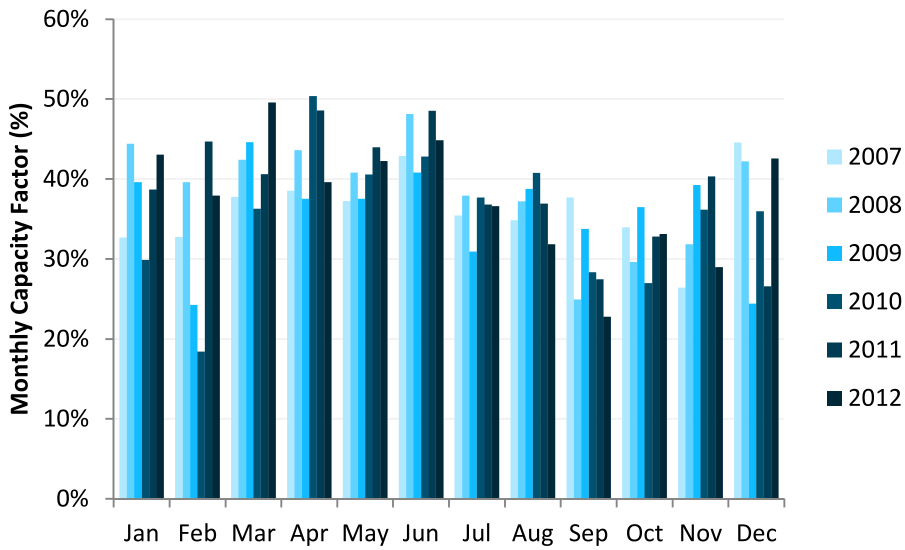
[22] While each dataset covers the historical period 2007-2013, this study uses the profiles from 2007-2012 because of the availability of time-synchronous load data
[23] As with the wind data, this study makes direct use of the solar profiles from 2007-2012.

Figure 30. Common Case wind & solar resources.



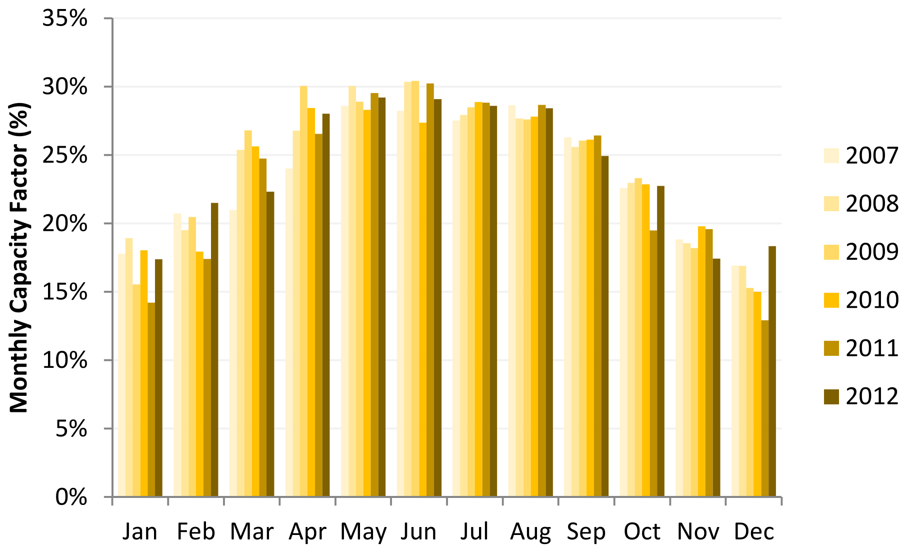
The use of six years of simulated renewable data allows the reliability assessment to capture the inter-annual variability in the expected output of renewable resources. This characteristic is particularly important for a robust representation of the output of wind resources, whose capacity factor can vary considerably from one year to the next. Figure 31, which shows the monthly capacity factors for the profiles used to represent the Common Case wind resources in the Northwest, illustrates this phenomenon.

Figure 31. Monthly average capacity factor for Common Case wind resources in the Northwest, 2007-2012 weather conditions.



Variability in output from one year to the next is less apparent for solar generation; Figure 32 shows the average monthly capacity factor for Common Case solar PV facilities in California. Nonetheless, the length of the simulated record available (2007-2012) helps to create a better distribution of possible solar production conditions and is also useful in characterizing the relationship between load and renewable output.

Figure 32. Monthly average capacity factor for Common Case solar PV resources in California, 2007-2012 weather conditions.



Forecasts for wind and solar production profiles for the day-ahead time frame are also developed by NREL. Wind power forecasts are an integral part of the WIND Toolkit dataset, which includes day-ahead, 4-hour-ahead, and hour-ahead wind power forecasts for each of the locations in the study. The day-ahead forecasts were developed using a more coarse numerical weather prediction run to simulate day-ahead forecasting accuracy. Smaller timescales blend truth data with the day-ahead forecasts to match forecasting accuracy observed at a large number of operational wind plants. More information can be found in Draxl et al. 2015. The day-ahead solar power forecasts were created using a similar statistical moment matching technique on the errors that was utilized in the WWSIS2.

3.1.3 DRAW METHODOLOGY

In order to make use of asynchronous datasets of load and renewable profiles, this study uses a stratified sampling methodology to create "draws"—twenty-four hour pairings of load and renewable outputs—for analysis. The sampling methodology, which pairs shapes based on season and type of day, is designed to preserve observed relationships between load and renewable output both across and within seasons to ensure that each draw reflects a plausible combination of load and renewables.

A four-step process is used to construct each draw:

+ Select a day from the historical record of load conditions at random. The same historical day is used for each region in order to preserve observed relationships between the regional loads.[24]
+ In each region, identify the "day-type" for the historical day drawn. For each month, the "day-type" is defined by the level of daily load relative to the entire historical record of daily load conditions for that month. This study uses fourteen day-types based on percentiles of daily load level, defined in Table 10.

[24] For the purposes of eliminating edge effects in the simulation results, the hourly profiles for the days immediately preceding and following the day of interest are included as inputs for the flexibility assessment such that each draw represents a 72-hour time strip; however, the resulting simulations for the first and last days are discarded when analyzing results, such that only the middle 24 hours are considered. The set-up of the production cost simulation is further discussed in Section 3.7.

Table 10. Day-type bins, defined by percentile of daily loads within each month.

Bin	Percentile	Load Level
1	<1%	Lowest
2	1-3%	Low
3	3-6%	Low
4	6-12%	Low
5	12-20%	Med-Low
6	20-30%	Med-Low
7	30-50%	Med
8	50-70%	Med
9	70-80%	Med-High
10	80-88%	Med-High
11	88-94%	High
12	94-97%	High
13	97-99%	High
14	>99%	Highest

+ In each region, select a daily wind profile at random from the appropriate month and day-type bin. Daily profiles are selected independently in each region to match the corresponding day-type conditions.
+ In each region, select a daily solar profile at random from the appropriate month and day-type bin. As with wind, profiles are selected independently in each region to match the corresponding day-type conditions.

The result of this process is a synthetic day-long record of load, wind, and solar profiles across the entirety of the study footprint that captures the key relationships among these variables in each region. The four steps of this process are illustrated in Figure 33.

Figure 33. Illustration of draw methodology.

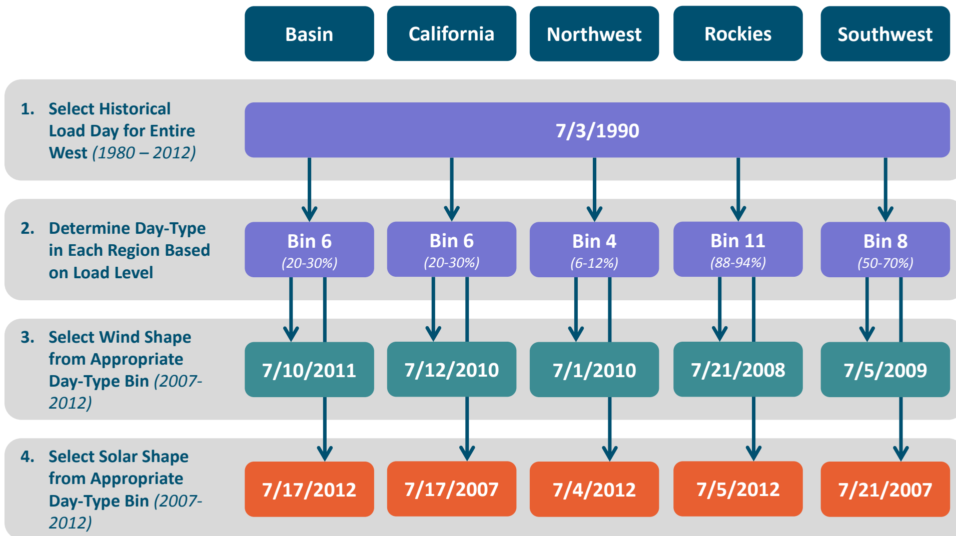
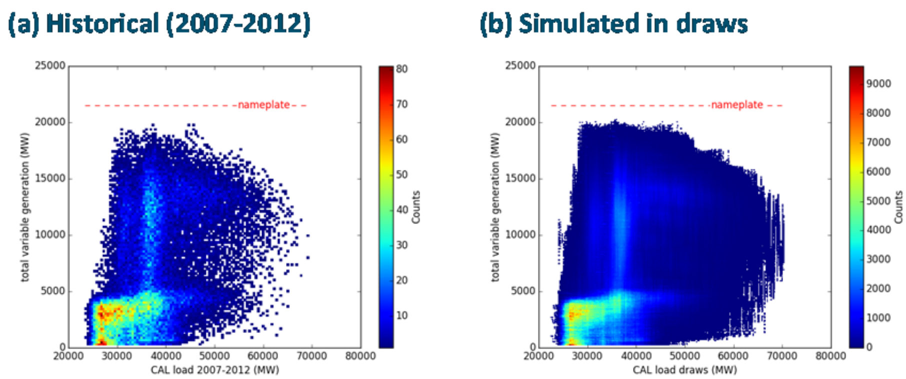


Figure 34 provides a visual comparison of the results of the methodology used to create the draws with time-synchronous load and renewable data from 2007 to 2012 for California in the Common Case. The method used to create the draws generally captures the appropriate frequency of different pairings of load and renewable output but provides a more complete distribution of possible conditions—particularly during the high load periods that are critical for reliability modeling. Additional comparative figures between time-synchronous and sampled load and renewable data are included in Section 7.2.1.

Figure 34. Comparison of load and renewable samples in historical data and draws, Common Case, California[25]



3.2 Conventional Generators

With the exception of hydroelectric resources (discussed in Section 3.2.2), all non-wind/solar generation is modeled based on the assumptions included in the 2024 Common Case. The TEPPC Common Case includes both existing resources as well as planned additions—investments in new generation resources needed to meet policy objectives, local reliability constraints, or portfolio needs—both of which are included in the analysis. However, the "gap" units—plants that were added by TEPPC modelers to the Common Case in each region to meet an assumed regional planning reserve requirement target—are removed from this analysis. Key inputs and assumptions used in the resource adequacy and flexibility assessments are described in subsequent sections.

[25] "Total variable generation" in this figure includes all wind, utility-scale and distributed solar PV, and solar thermal resources.

3.2.1 RESOURCE ADEQUACY ASSESSMENT

In the resource adequacy assessment, each generator is assumed to be capable of providing its full rated capacity to meet load, subject to its availability given an assumed forced outage rate. For modeling purposes, this category includes traditional dispatchable resources—gas, coal, nuclear—as well as storage, demand response, and baseload renewable resources. Each resource's contribution to reliability depends on two key input parameters: (1) the maximum available output of the resource, and (2) its expected forced outage rate. Both of these parameters are defined for each resource in the TEPPC Common Case, from which this analysis draws directly the assumptions of which generation resources are assumed online.

The maximum capacity available for each resource is defined for each month of the year separately; the ratings of different resources change from one month to the next as a result of seasonal trends in temperature, which affects the maximum output of thermal units. Table 11 and Table 12 show the breakdown of available capacity in each region based on rated capacity in January and July, respectively. The impact of seasonal trends on resource availability is perhaps most evident for gas-fired generators: for example, the capability of the gas fleet in the Southwest is reduced by 5% from 28,462 MW in the winter to 27,164 MW in the summer due to the effects of temperature on output. In regions whose load peaks during the summer, this effect results in a need for additional capacity to meet a given reliability target.

Table 11. January capacity ratings, conventional resources by generator type (MW).

Type	Basin	California	Northwest	Rockies	Southwest
Nuclear	—	2,300	1,145	—	3,937
Coal	6,974	2,121	3,001	6,455	7,807
Gas	5,079	37,625	9,095	6,953	28,462
Biomass	85	1,265	706	4	37
Geothermal	911	2,244	—	—	—
Storage	—	1,285	—	—	—
DR	1,035	2,268	222	525	759
Other	96	521	84	150	61
Total	**14,180**	**49,628**	**14,253**	**14,087**	**41,063**

Table 12. July capacity ratings, conventional resources by generator type (MW).

Type	Basin	California	Northwest	Rockies	Southwest
Nuclear	—	2,300	1,130	—	3,937
Coal	6,964	2,121	2,999	6,433	7,803
Gas	4,851	37,117	8,471	6,504	27,164
Biomass	82	1,309	679	4	36
Geothermal	910	2,218	—	—	—
Storage	—	1,285	—	—	—
DR	1,035	2,268	222	525	759
Other	96	513	84	130	51
Total	**13,938**	**49,131**	**13,585**	**13,596**	**39,749**

The second important input needed to characterize these units is their forced outage rates. The fact that the forced outage rate does not appear in the calculation of a system's planning reserve margin belies its importance to the measure of system reliability. In fact, the probability of forced outages during the system peak is one of the drivers of the need to maintain a planning reserve margin, and the two are directly linked: LOLP analysis will indicate that a system with a generation fleet that has a high risk of forced outages will need to hold a higher planning reserve margin than an otherwise equivalent system with lower forced outage rates in order to meet the same standard of reliability.

This study relies on the forced outage rates assumed in the TEPPC Common Case, originally derived from NERC's Generating Availability Data System (GADS). The average forced outage rate for each type of generator in each region is shown in Table 13. The forced outage rates in the TEPPC Common Case generally vary from 3-5% depending on technology type and size. Compared to assumptions used in other LOLP studies, these outage rates appear to be relatively low; should WECC continue to investigate using LOLP analysis to measure system reliability, additional scrutiny of these assumptions may be warranted given their importance in this type of analysis. This is particularly true in the event that higher renewable penetrations may result in increased ramping and cycling among gas and coal generators, which has been linked to increased probabilities of forced outage in a number of studies.[26]

[26] Citation to EGS study

Table 13. Average forced outage rates by generator type (%).

Type	Basin	California	Northwest	Rockies	Southwest
Nuclear	—	3.1%	3.1%	—	3.1%
Coal	4.7%	3.9%	4.4%	4.9%	4.6%
Gas	3.2%	3.2%	3.3%	3.2%	3.1%
Biomass	3.1%	3.1%	3.1%	3.1%	3.1%
Geothermal	3.1%	3.1%	—	—	3.1%
Storage	—	3.0%	—	—	—
DR	0.0%	0.0%	0.0%	0.0%	0.0%
Other	4.0%	4.5%	3.1%	9.0%	1.5%

3.2.2 FLEXIBILITY ASSESSMENT

The production cost modeling of the flexibility assessment requires additional parameters that characterize how each unit can operate and at what cost. In addition to the maximum capacity, a number of additional inputs are specified for each plant, including:

+ Minimum stable level (MW);
+ Heat rate (Btu/kWh);
+ Variable O&M ($/MWh);
+ Maximum ramp up and ramp down rates (MW/min);
+ Minimum up and down times (hrs);
+ Maintenance rates (%);
+ Start cost ($); and
+ Start fuel requirements (MMBtu).

Each of these input assumptions impacts the flexibility of each individual unit, either as a constraint on its dispatch or as a cost that affects its position in the merit order. This analysis relies predominantly on data developed through the TEPPC 2024 Common Case, which provides unit-specific information for each of these fields. Average characteristics for the major categories of coal and gas generators are summarized in Table 14. Detailed unit-specific assumptions can be obtained from WECC's 2024 Common Case database.

Table 14. Average operating characteristics for thermal generators.

Characteristic	Units	Coal	Gas CCGT	Gas CT	Gas ST
Maximum Output	MW	314	304	56	75
Minimum Output	MW	128	161	25	17
Heat Rate	Btu/kWh	10,089	7,396	10,369	11,294
Max Ramp Up	MW/hr	290	255	224	115
Max Ramp Down	MW/hr	290	253	224	115
Min Up Time	hr	165	8	3	12
Min Down Time	hr	48	4	2	8

This study does deviate from the TEPPC dataset's assumptions regarding operating parameters for nuclear generators. The Common Case includes three nuclear facilities: Diablo Canyon Power Plant (California), Palo Verde (Southwest), and Columbia Generating Station (Northwest). In this study, each nuclear unit is modeled using assumptions that reflect current planning assumptions—to the extent they exist—in each region. Diablo Canyon and Palo Verde are treated as must-run generation with Pmin equal to Pmax, and therefore have no flexibility to cycle output. Columbia Generating Station is also treated as must-run with no ramping flexibility, but can be committed and operated at any range between full capacity and a minimum level of 70% of its rated capacity. Maintenance schedules for plant refueling are based on historical patterns of maintenance for these plants.

3.3 Hydroelectric Constraints

Unlike most resources in the Western Interconnection, hydroelectric plants are generally use-limited, constrained in their operations not only by the technical limits of the generators themselves but by the underlying hydrological conditions. The availability of water to provide for generation, which varies from season to season and year to year, acts as a constraint on both the contribution of hydro resources to reliability and their flexibility in operations. In order to capture the unique and variable characteristics of each region's hydroelectric fleet, this study incorporates the historical range of hydrological conditions experienced from 1970-2008. The ability of the hydroelectric system in each region to contribute to system reliability, as well as the flexibility it affords the system in operations, depends directly on the associated hydrological conditions. This section describes:

+ How monthly budgets are developed for incorporation into the draws;
+ How the peaking capability of the hydro system is represented in Phase 1 as a function of monthly hydro conditions; and
+ How the operations of the hydro system are modeled in Phase 2 based on the monthly energy budget.

3.3.1 MONTHLY HYDRO BUDGETS

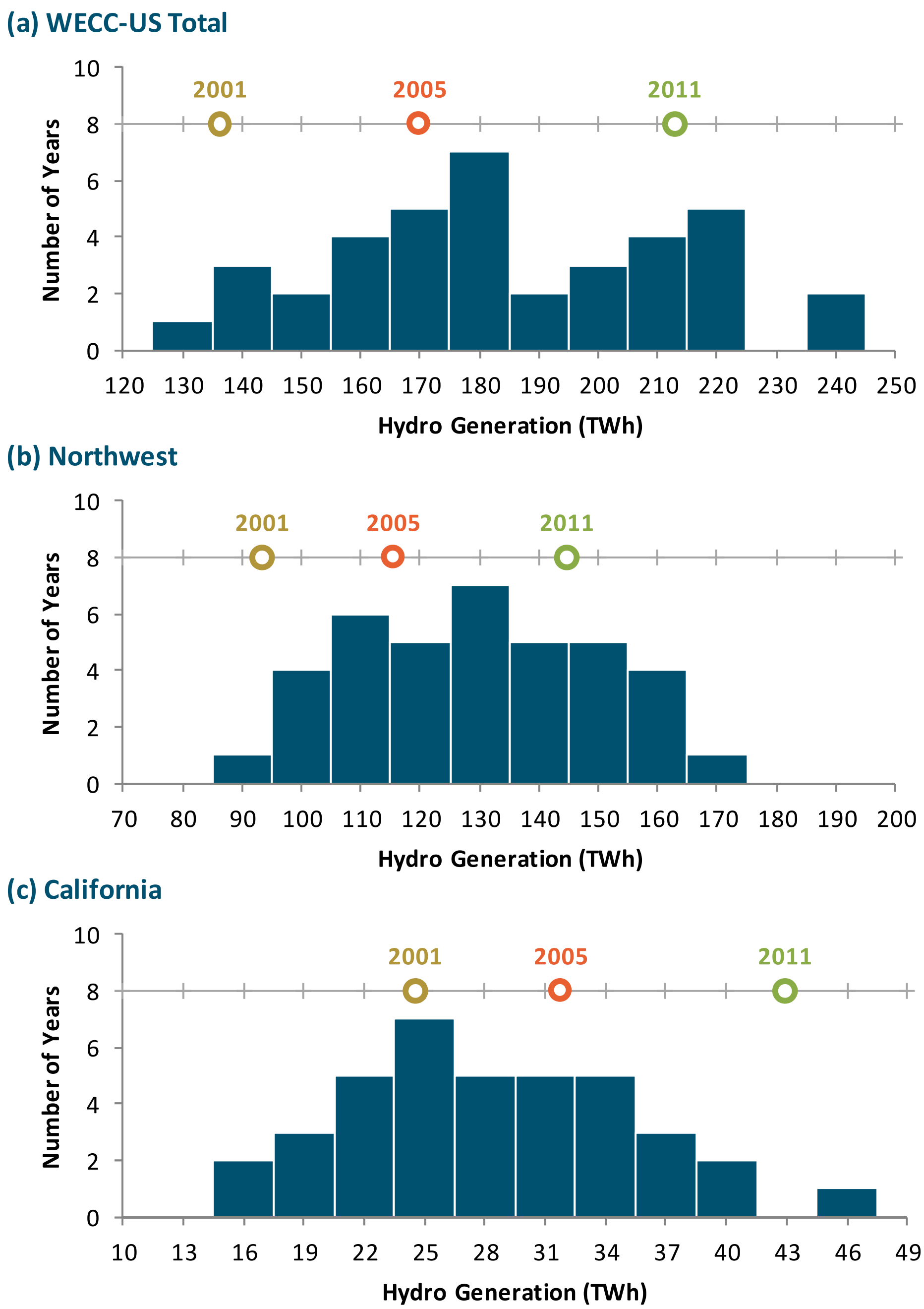
Monthly energy budgets for the fleet of hydroelectric generators are derived from two sources in this study:

+ **Energy Information Administration (EIA) (1970-2012).** EIA Form 906 was used to calculate historical monthly hydro output for all hydro generators in the California, Southwest, Rocky Mountain, and Basin regions.

- **Northwest Power and Conservation Council (1928-2008).** The NWPCC provided simulated output from the Northwest hydro fleet based on current operating procedures and monthly hydrological conditions across a long historical record.

From each of these datasets, this study uses the hydroelectric data from 1970-2008, the full extent of the chronological overlap between the two. Figure 35 shows examples of the distributions of annual energy budgets for the study footprint as a whole as well as for the Northwest and California regions over this historical period. The specific annual budgets for 2001, 2005, and 2011, commonly used to represent dry, average, and wet hydro years, are shown for comparison purposes.

Figure 35. Distribution of annual energy budgets.



3.3.2 SUSTAINED PEAKING CAPABILITY

The RECAP model treats time steps independently for ease of computation, which makes it necessary to create approximations for energy- or use-limited resources. Hydro is the most significant of these and presents a substantial modeling challenge when it comes to peaking capability for reliability.

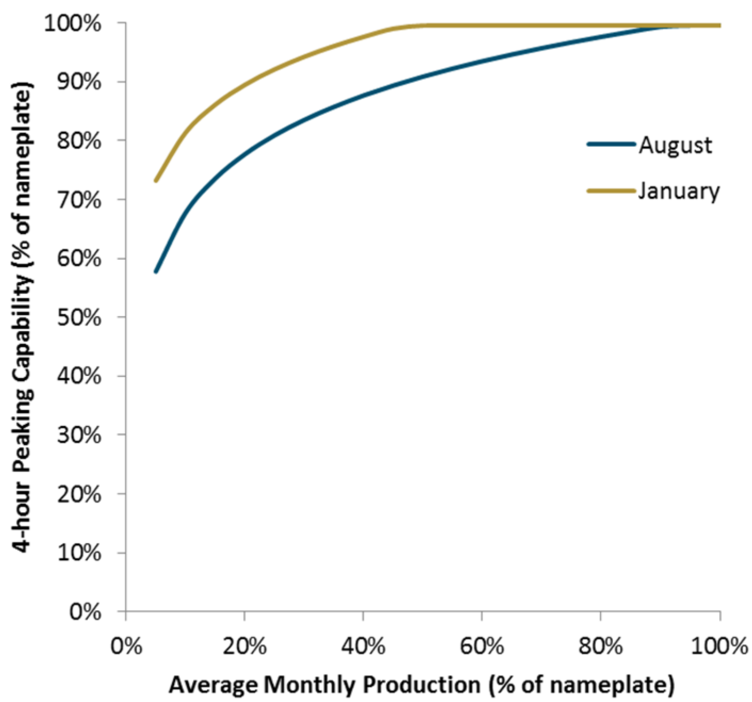
In the Northwest, where the majority of generating capacity is hydroelectric, the contribution of the hydro fleet to reliability has historically been modeled based on its "sustained peaking capability": the ability of the fleet to sustain a certain level of output across durations of two, four, and ten hours through different seasons and under different hydrologic conditions. These constraints are derived by the NWPCC through a hydrologic model of the operations of the major river systems and dams in the Pacific Northwest.

For this study, hydro peaking capability within RECAP was constrained using relationships between the monthly hydro budget for a region and the sustained 4-hour peaking capability of its hydro generators.[27] The 4-hour peaking capability is used because in a power system with adequate reliability, resource adequacy shortages rarely last longer than 4 hours—either native load drops below available capacity or additional resource can be brought online. The average outage duration in RECAP is 2 hours when the loss of load frequency is equal to one event per ten years. Because of this, 4-hour sustained peaking was judged to be appropriately conservative, without being over-constraining. The

[27] Because RECAP treats each time step independently, a full implementation of the sustained hydro peaking capability constraints is not possible. Therefore, a simplified application of the sustained peaking constraints was used.

relationship between hydrologic conditions and peaking capability is specific to month and developed using the sustained peaking constraints provided by the NWPCC, as shown in Figure 36.

Figure 36. Example hydro peaking constraints used in RECAP analysis



In the absence of comparable data for other regions, this study assumes that the relationship between monthly energy budgets and 4-hour sustained peaking capability is constant across regions, applying the normalized curves to the hydro fleet in each region. In benchmarking exercises with other regions where other detailed studies of hydro resource adequacy has been done, principally California, the sustained peaking functions resulted in an effective load carrying capability for hydro that agreed closely with current planning assumptions

(10,878 hydro ELCC vs. 10,928 hydro dependable capacity from the California Energy Commission).[28] Because of this agreement, and due to the lack of detailed information for other regions, the generalized sustained peaking relationships were judged appropriate for this study.

3.3.3 OPERATIONAL CONSTRAINTS

Accurately modeling the operations of hydroelectric generators is a perpetual challenge in production cost modeling. The operations of many hydroelectric generators in the Western Interconnection are governed not only by electric system conditions but by a host of additional factors, including flood control, navigation, and irrigation. These factors, coupled with the inherent challenge in modeling use-limited resources, make capturing the full physical capability of the hydro system in a production cost model a difficult proposition.

Rather than relying on the physical ratings and capabilities of the hydroelectric generators that make up each region's fleet, this study uses a model of hydro operations that derives constraints and assumptions largely from analysis of the actual historical operations of each region's hydro fleet. In deriving constraints from actual historical conditions, the methods used herein attempt to replicate the flexibility that has historically been demonstrated by each region's hydro system, such that even if all constraints cannot be represented within the model, the simulated dispatch of the hydro fleet does not imply a radical change from what has occurred historically.

[28] California Energy Commission, *Summer 2012 Electricity Supply and Demand Outlook*. Available at: http://www.energy.ca.gov/2012publications/CEC-200-2012-003/CEC-200-2012-003.pdf.

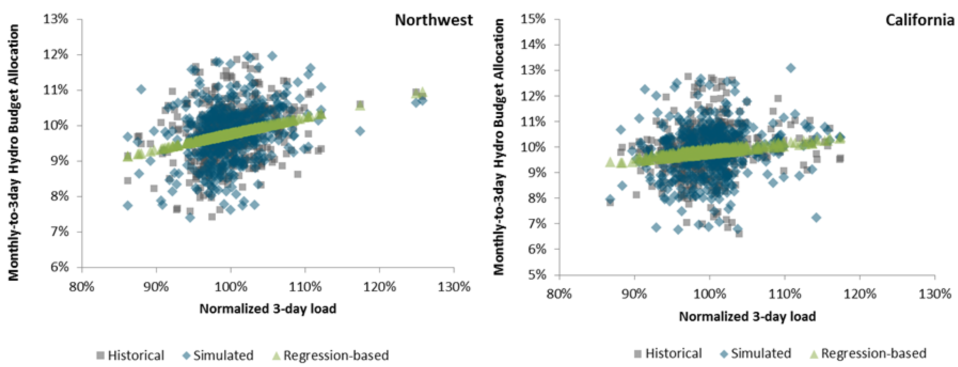
To model hydro operations in the production cost model, several constraints are derived based on the monthly energy budget sampled for each draw. First, from the monthly energy budget, a three-day energy budget consistent with the time horizon of the draw is derived. In addition, constraints on the minimum and maximum generating capacity, as well as hour-to-hour and multi-hour ramps, are imposed upon the dispensation of the energy. The derivation of these constraints is discussed in the subsequent sections.

3.3.3.1 Allocation of Monthly Hydro Budgets to 3-Day Draws

The monthly hydro energy budget sampled for each draw is translated to a three-day energy budget based on a regression analysis of historical daily hydro generation and historical daily load conditions. The historical data exhibit strong correlation between load and daily hydro allocation in some months, but the relationship is not consistent across seasons—an indicator of the fact that other factors limit the utilization of hydro resources in different seasons.

To account for all drivers, linear regression is used to establish a relationship between the historical three-day hydro budget (as a percentage of the monthly budget) and the historical three-day load (normalized to the historical average load for the respective month) in order to quantify the relationship between hydro budget and load. A distribution of the residuals is created, and random perturbations are drawn from this distribution to simulate drivers other than load level. The historical data, along with the linear regression and the simulated hydro budgets, are shown in Figure 37.

Figure 37. Historical, regression-based, and simulated hydro 3-day energy budgets as a function of 3-day net load.

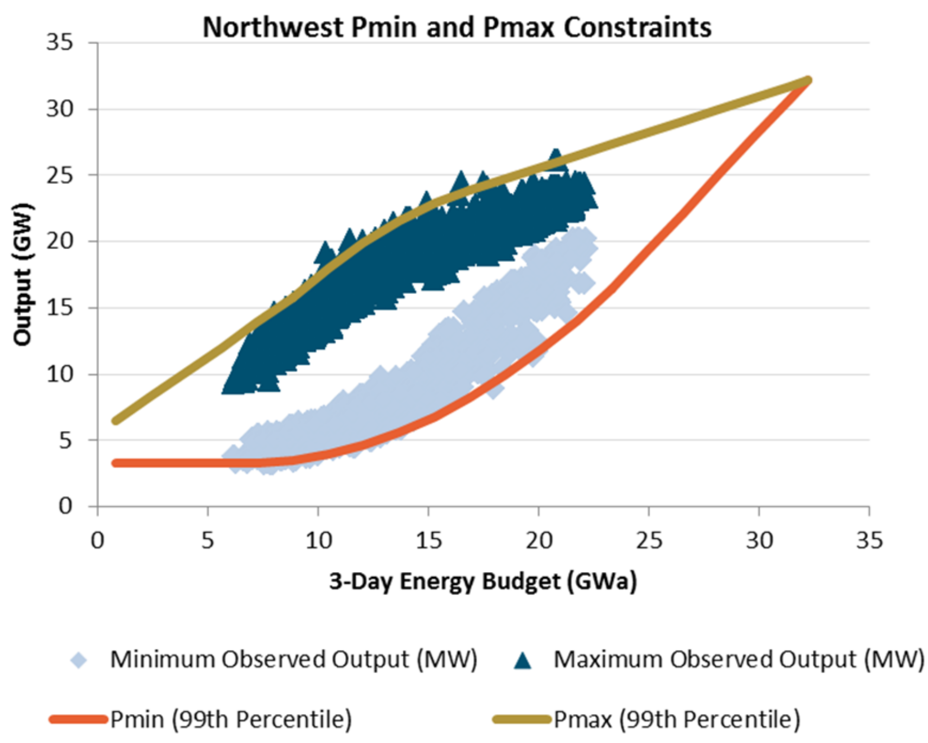


Rather than using solely simulated load to translate a monthly energy budget to three days, this study uses simulated net load (load net of solar and wind production) along with the observed historical relationship presented above. This modeling choice is intended to reflect the fact that under increasing penetrations of renewable generation, operators will increasingly respond to both load and renewable conditions in scheduling hydro resources.

3.3.3.2 Hydro Operational Constraints

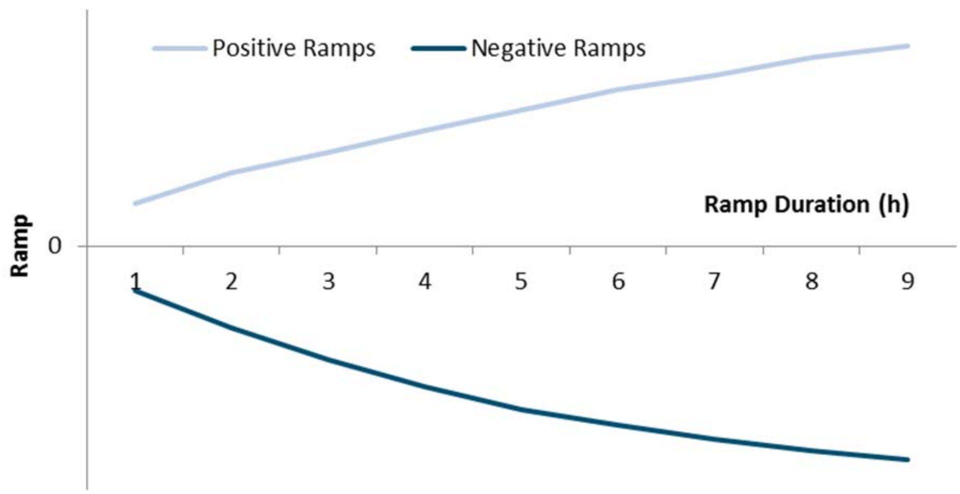
The constraints on minimum output (Pmin), maximum output (Pmax), and ramping capability are developed based on plant-level hourly historical output data from TEPPC for 2001, 2005, 2010, and 2011. Data are aggregated to the regional level. The Pmin and Pmax constraints vary as a function of the energy budget. The goal behind this approach is to account for the impact of water availability conditions on the operational flexibility of hydro plants. The Pmin and Pmax functions are derived from the observed relationship between minimum (or maximum) hourly output and the average hourly output level during 3-day periods in the four years of TEPPC historical data. We fit a curve through the 99th percentile of data points for both the minimum and maximum output level to represent the limit on hydro capability.

Figure 38. Development of regional minimum (Pmin) and maximum (Pmax) output constraints.



In addition to constraints on minimum and maximum output level, we also impose a limit on the ability of hydro generators to change output from hour to hour. Historical hourly hydro operations data is used to derive a ramping envelope for each region such as that illustrated in Figure 39, representing the 99th percentile of observed upward and downward monotonic ramps across multiple discrete durations. In the production cost modeling, this envelope is used to constrain the ramping capability of each region's hydro fleet for ramps of durations between one and four hours.

Figure 39. Multi-hour ramping constraints in REFLEX.



Subject to the energy budget, minimum output, maximum output, and ramp rate constraints described above, REFLEX optimally dispatches the available hydro generation within the drawn operating days.

3.4 Interregional Power Exchange

3.4.1 RESOURCE ADEQUACY ASSESSMENT

In the reliability assessment, each region is modeled independently from its neighbors. While a West-wide reliability model may provide interesting information, the regional scope of the reliability assessment is chosen in order to better reflect the geographic scale at which resource planning occurs and reliability-driven investment decisions are made. Consequently, in each region, an assumption on the degree to which imports may contribute to system reliability is needed.

The availability of imports and their contribution to the reliability of a region depends on a number of factors, including the physical limits of the transmission systems, a region's long-term contracts or ownership of remote resources, the balance between loads and resources in different parts of the West, and the underlying economics of power markets. There is not a single consensus approach used to determine the potential contribution of imports to meeting a region's reliability requirements; rather, a number of approaches have been used in an attempt to quantify the availability of imports:

+ **Physical limits of the transmission system.** The physical limits of transmission lines that connect an electric system to its neighbors provide an upper-bound estimate of the possible contribution of imports to an electric system's reliability. This approach has been used in California—historically a major net importer from both the Northwest and Southwest during its summer peak periods—by both the California Energy Commission (CEC) and the California Public Utilities Commission (CPUC).
+ **Ownership of remote resources.** A number of major generating resources in the West are either owned or contracted by utilities: the output from the Palo Verde nuclear facility, Hoover Dam, and a number of large coal generators. In its evaluation of resource adequacy in each region of the Western Interconnection during the development of the Common Case, TEPPC relies on this information to characterize the contribution of imports (and exports) to the planning reserve margin in each area.

+ **Assessment of surplus generation capacity in neighboring systems.** A resource planner may attempt to evaluate the future availability of surplus generation capacity on neighboring systems. For example, in its resource adequacy assessment, the Northwest Power and Conservation Council (NWPCC) examined the balance of loads and resources in California to inform its decision to assume the availability of 2,500 MW of imports during the winter period due to California's relatively low winter loads and the consequent availability of capacity needed to meet its own summer peak.[29]

+ **Observation of historical import patterns.** Historical data on the dynamics of imports to an electric system provides another useful point of reference. The California ISO, in its *2014 Summer Loads & Resources Assessment*,[30] establishes a range of import availability of 8,500 – 11,000 MW that is, in part, informed by its historical operating experience.

This study uses a combination of these approaches, deferring to regional planning efforts for this important assumption when possible. For both California and the Northwest, the question of how much to rely on imports for reliability has been asked and answered in a number of regional planning forums, and this study uses the general results from these exercises to inform its analysis. In the other three regions, no such regional-level planning forum exists; for these regions, the availability of imports (and, in parallel, the obligation to export) is determined on the basis of remote ownership of

[29] Northwest Power and Conservation Council, *Pacific Northwest Power Supply Adequacy Assessment for 2019*. Available at: http://www.nwcouncil.org/media/7084800/2014-04.pdf.
[30] California Independent System Operator, *2014 Summer Loads & Resources Assessment*. Available at: http://www.caiso.com/Documents/2014SummerAssessment.pdf.

generation resources. The contribution of imports to reliability in each region, as well as the information used to derive each assumption, is shown in Table 15.

Table 15. Assumptions used to quantify imports & exports in resource adequacy assessment

Region	Net Capacity (MW)	Description
Basin	445	**Remote resource ownership (imports):** • Craig 1 & 2: 169 MW • Hayden 1 & 2: 87 MW • Intermountain Power Project: 450 MW **Remote resource ownership (exports):** • North Valmy 1 & 2: 261 MW
California	11,768	**Regional planning assumption:** • *CEC Summer Assessment*[31]: 13,118 MW **Adjustment for Intermountain Power Project:** • Modeled in LADWP; 1,350 MW deducted from import capability (CA ownership share of 1,800 MW plant)
Northwest	2,500	**Regional planning assumption:** • NWPCC[32]: 2,500 MW • Available in winter only (0 MW in summer)
Rocky Mountains	-602	**Remote resource ownership (exports):** • Craig 1 & 2: 423 MW • Hayden 1 & 2: 179 MW
Southwest	-1,737	**Remote resource ownership (imports):** • Craig 1 & 2: 254 MW • North Valmy: 261 MW **Remote resource ownership (exports):** • Hoover: 1,265 MW • Palo Verde 1, 2 & 3: 1,078 MW

3.4.2 FLEXIBILITY ASSESSMENT

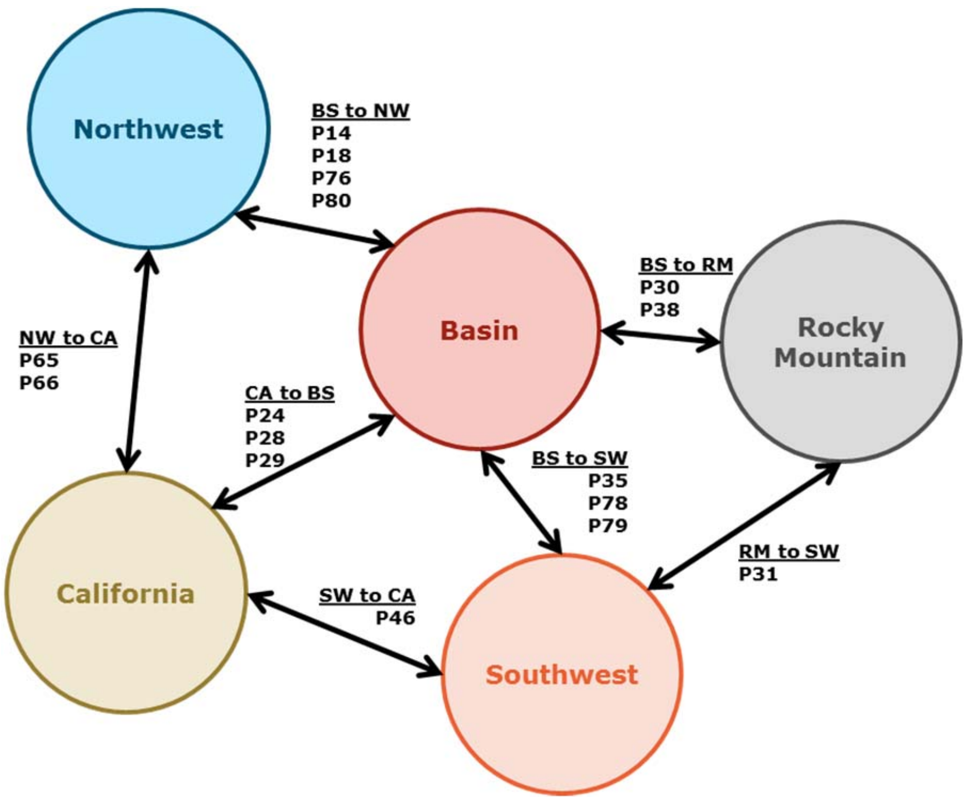
The flexibility assessment uses a zonal transport model with five regions to characterize the transmission system of the Western Interconnection. Power

[31] California Energy Commission, *Summer 2012 Electricity Supply and Demand Outlook*. Available at: http://www.energy.ca.gov/2012publications/CEC-200-2012-003/CEC-200-2012-003.pdf.

[32] Northwest Power and Conservation Council, *Pacific Northwest Power Supply Adequacy Assessment for 2019*. Available at: http://www.nwcouncil.org/media/7084800/2014-04.pdf.

flows among the five regions are limited based on the WECC paths that connect them; no internal transmission constraints are modeled within each of the regions. Figure 40 shows the topology (and corresponding interregional paths) used in this study.

Figure 40. Regional transmission topology used in production cost analysis.



The zonal transmission topology used in this study represents a simplification to the transmission network of the Western Interconnection made for modeling expedience; as a consequence, there are certain dynamics that this study will not capture. For example, no internal transmission constraints are enforced within any of the zones, so potential congestion within any single zone (e.g.

between Northern and Southern California on Path 26) would not be captured. Additional work may be necessary to understand the degree to which intraregional transmission infrastructure impacts operational flexibility.

The limits on flow between regions are established based on the WECC paths that connect them. In this study, the Reference Grid constrains flows over these interfaces so that they behave similarly to historically observed flow patterns. Flow constraints in the Reference Grid take two forms: limits on the hourly flow volume across each interface; and limits on the hour-to-hour ramps over each interface. Both the hourly flow constraints and the ramping constraints are derived from historical flow data, using the 0.1^{st} percentile and the 99.9^{th} percentile of observations to build the constraints. The sources and years of available flow data used for this analysis are described in Table 16.

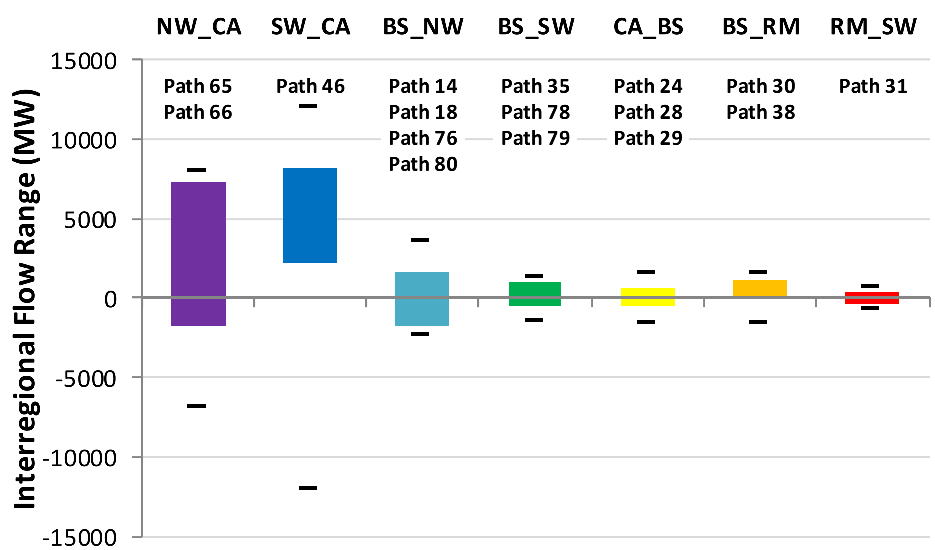
Table 16. Data sources for derivation of historically-based flow limits over interfaces.

Interface	Data Source	Years of Data
NW_CA	BPA	2007-2013
SW_CA	WECC Path Flows	2008-2012
BS_NW	WECC Path Flows	2010-2012
BS_SW	WECC Path Flows	2008-2012
CA_BS	WECC Path Flows	2010-2012
BS_RM	WECC Path Flows	2010-2012
RM_SW	WECC Path Flows	2008-2012

The resulting flow and ramping constraints are shown for each interface in Figure 41 and Figure 42, respectively.

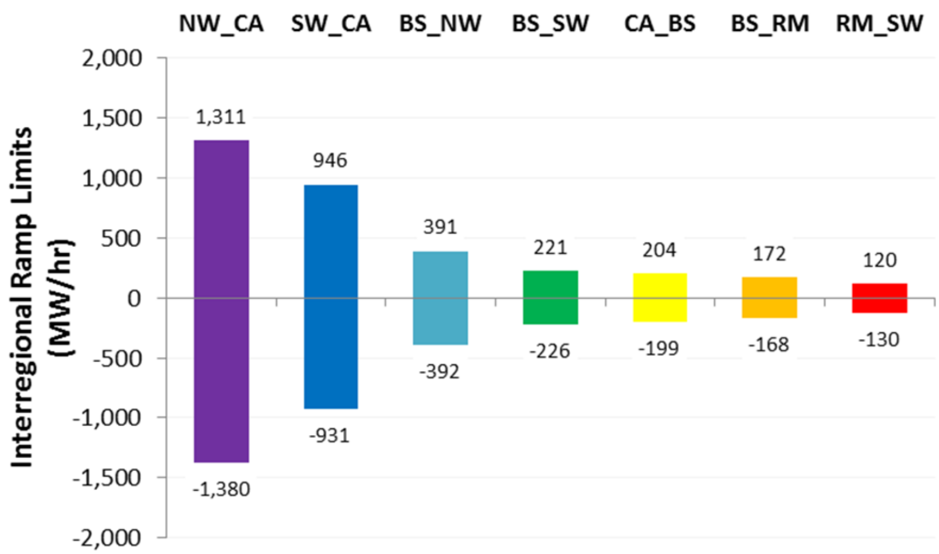
This analysis also considers a relaxation of the Reference Grid assumptions, in which improved regional coordination allows more complete utilization of the transmission connecting each of the five regions, in term of both flow volumes and manageable ramps across the interfaces. In this Regional Coordination case, the flow constraint for each interface is equal to the sum of the path ratings comprising the interface and the historically-based ramping constraints are lifted. The path rating-based constraints are shown in the black hashes in Figure 41.

Figure 41. Historical and physical limits on intertie flows.



Colored bars show historical range; black hashes show combined path ratings

Figure 42. Constraints on the ramping limits across each interregional interface.



3.5 Costs of Flexibility Violations

Within the framework used by REFLEX, power system inflexibility will translate into curtailment of renewable energy and/or unserved energy. As described in Section 2.4.2, flexibility challenges driven by renewables can be managed in two ways: (1) renewable energy delivery can be prioritized, which may result in upward shortages on the system (i.e., unserved energy); or (2) meeting load can be prioritized over delivering renewable energy, which may require renewable curtailment to accommodate the inflexibility of the conventional resource fleet. Both operating practices have a cost, and this analysis prioritizes meeting load. In other words, this analysis assumes the system operator curtails renewables before involuntarily shedding load as a consequence of flexibility challenges.

In production cost simulations, operator actions are prioritized by applying a $/MWh cost (“penalty price”) to unserved energy and renewable curtailment. When the penalty price for unserved energy is higher than the penalty price for renewable curtailment, then the simulation will choose to curtail renewables before shedding load because the marginal cost of that action is lower. The values selected for these penalty prices are discussed further below.

3.5.1 UNSERVED ENERGY

The economic cost of unserved energy is commonly referred to as the value of lost load (VOLL). From the customer perspective, VOLL measures a customer’s willingness to pay for reliable service in each hour. A body of research devoted to this question has found that VOLL varies significantly by a number of factors, including customer class, timing of the outage and its duration, but generally falls between $2,000/MWh to $300,000/MWh, with small commercial and industrial customers at the higher end of the range and residential customers at the lower end of the range.[33]

In production cost models, loss of load events are observed at the system level, so the penalty price for unserved energy is set near the average customer VOLL. For this study, unserved energy was penalized at $50,000/MWh across all scenarios. This penalty price is high enough to steer the model away from choosing to shed load unless it is a last resort.

[33] Sullivan, M. J., M. Mercurio, M., J. Schellenberg, and M.A Freeman, “Estimated Value of Service Reliability for Electric Utility Customers in the United States,” Ernest Orlando Lawrence Berkeley National Laboratory, LBNL-2132E, 2009, available at: **https://emp.lbl.gov/sites/all/files/REPORT%20lbnl-2132e.pdf**.

3.5.2 RENEWABLE CURTAILMENT

The cost of renewable curtailment has been of interest in recent years as renewable penetrations have increased and policies signal for continued development over the coming years. While there is less literature to draw from, there is also more clarity on the economic cost of renewable curtailment driven by renewables policies. Again, the penalty price depends on the perspective and the jurisdiction.

In jurisdictions with a mandatory renewable energy target, the direct consequence of renewable curtailment is the need to procure additional renewable energy to comply with the target. The cost of renewable curtailment in these jurisdictions is the "replacement cost" of the curtailed renewable energy or the "overbuild cost" to ensure compliance (these costs are equivalent, but here we use the term "replacement cost"). The replacement cost can be approximated for use in a production cost model by considering the cost of incremental renewable build, the expected curtailment levels of the incremental build, and the avoided cost associated with the incremental build.[34,35]

---

[34] Note that in a production cost model where the installed capacities are fixed, the replacement energy is not physically modeled. If the renewable portfolio was designed assuming no curtailment, then the system will not be reach the intended target once curtailment is accounted for. The cost applied to renewable curtailment in these simulations is an attempt to express both the additional fixed costs associated with renewable overbuild and the variable cost savings of the additional renewable energy into a variable cost term in the optimization. In this way, the approach represents an approximation of the economic impact of renewable curtailment on the system. A more accurate approach would be to endogenize the renewable build in the optimization so that overbuild and curtailment are explicitly linked and fully consistent, but this approach increases the computational complexity of the problem substantially.

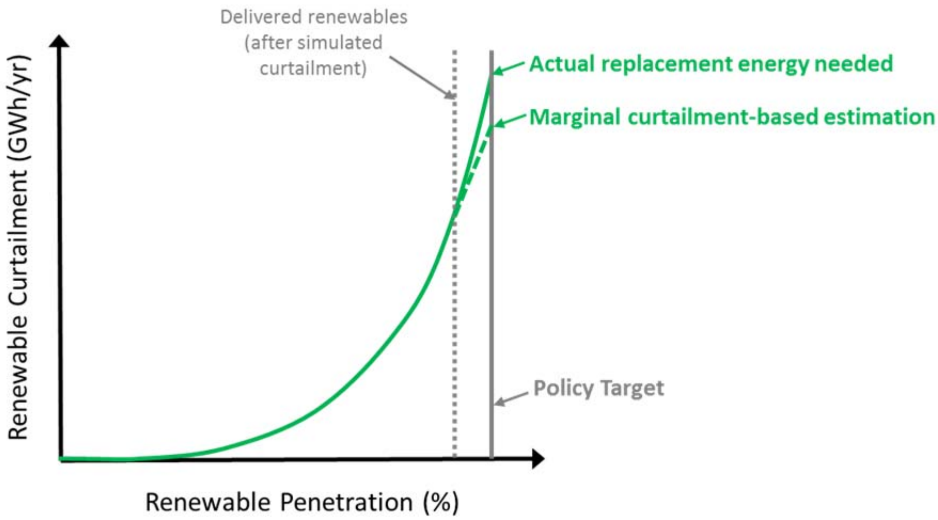
[35] In addition, there may be out-of-pocket costs such as increased O&M due to frequent curtailment and loss of tax benefits such as Production Tax Credits (PTC). These costs should also be incorporated into the curtailment penalty.

When using the replacement cost framework for penalizing renewable curtailment, the penalty can be approximated from the procurement cost of the renewable energy in \$/MWh (assuming no curtailment), the energy value of the incremental delivered renewable energy, and the marginal curtailment of the incremental renewables:

$$[penalty, \$/\mathrm{MWh}] = \frac{[procurement\ cost, \$/\mathrm{MWh}]}{1 - [marginal\ curtailment, \%]} - [energy\ value, \$/\mathrm{MWh}]$$

The marginal curtailment is defined as the fraction of the incremental renewable build that is assumed to also be curtailed. The marginal curtailment is calculated by comparing the incremental renewable output availability over time to the curtailment already occurring on the system in those hours – any incremental renewable energy available in curtailment hours is assumed to be curtailed. In general, the marginal curtailment calculated in this manner represents an underestimate of the curtailment experienced by the incremental resources, as the incremental resources themselves may increase the frequency of curtailment events. This is illustrated conceptually in Figure 43, which shows how a production cost case that was built to physical compliance with renewable energy targets (neglecting curtailment) can be used to calculate a lower bound on the replacement energy needed for actual compliance.

Figure 43. Marginal curtailment-based estimation of replacement energy need due to renewable curtailment observed in a production cost analysis.



This approach yields values that depend on the costs of renewable resources available in each region in the year of the study and on the conditions in the system in that year, but estimates for the mid-2020s to 2030 with various combinations of wind and solar resources tend to fall between $80/MWh and $120/MWh. For this study, a base assumption of $100/MWh was selected to be consistent with the replacement cost framework as a way of understanding the long-run cost of renewable curtailment.

There are alternative approaches to costing renewable curtailment in production cost modeling and these are explored through sensitivities in this analysis. At the low end, the cost of curtailment reflects the opportunity cost for a utility that does not face a renewable energy target compliance obligation and faces only the loss of a production tax credit and any additional O&M during curtailment hours. This study assumes a low curtailment price of $30/MWh as an estimate of this out-of-pocket cost. On the high end, a price of $300/MWh serves as an upper bound on the cost of economic curtailment; this is the current assumption used by the CAISO in its flexibility modeling for California's Long-Term Procurement Plan,[36] so it represents the cost of curtailment in a specific regulatory context. These two sensitivities are discussed in Section 5.3.1.

3.6 Reserves

In the flexibility analysis, operating reserves are held in each hour to account for contingencies, subhourly fluctuations in loads and resource availability, and forecast errors. These requirements and the resources that are capable of contributing to the requirements are summarized below.

3.6.1 CONTINGENCY AND REGULATION RESERVES

Spinning and regulation reserves are both modeled in PLEXOS as a headroom requirement on committed units within each region in each hour. Non-spinning contingency reserves are not modeled. The spinning reserve component is modeled as 3% of load in each region in each hour. Regulation reserves are modeled as 1% of the load in each region in the Common Case. In the High Renewables Case, regulation reserves are increased to 1.5% of load to account for the increase in minutely variability of net load due to higher renewable penetrations. This assumed increase is of the same order of magnitude as found

[36] https://www.caiso.com/Documents/Presentation_2014LTPPSystemFlexibilityStudy_SHcall.pdf

in a number of studies that have examined the impact of higher renewable penetrations on the need for regulation reserves.[37] Only dispatchable technologies are allowed to contribute upward capability to spinning and regulation reserves. Some resources are able to contribute portions of their upward capability to reserve requirements in multiple regions based on contractual arrangements reflected in the TEPPC Common Case (see Table 17).

Table 17. Reserve allocation shares for units serving multiple regions.

Units	Basin	California	Northwest	Rockies	Southwest
Colstrip_3 & 4	10%	0%	90%	0%	0%
Craig_1 & 2	19%	0%	0%	52%	29%
Hayden_1	25%	0%	0%	76%	0%
Hayden_2	13%	0%	0%	37%	50%
Hoover Dam	0%	61%	0%	0%	39%
Intermountain1 & 2	9%	91%	0%	0%	0%
Jim_Bridger_1, 2, 3, & 4	33%	0%	67%	0%	0%
NorthValmy2	50%	0%	0%	0%	50%
Parker_Dam_1, 2, 3, & 4	0%	50%	0%	0%	50%
San_Juan_4	7%	39%	0%	0%	54%

3.6.2 FLEXIBILITY RESERVES

Flexibility (or load following) reserves represent capacity that must be reserved in each hour to accommodate both forecast errors and subhourly fluctuations in the load and availability of renewables. Because renewable resource output is both harder to forecast and more variable than load, as renewable penetration

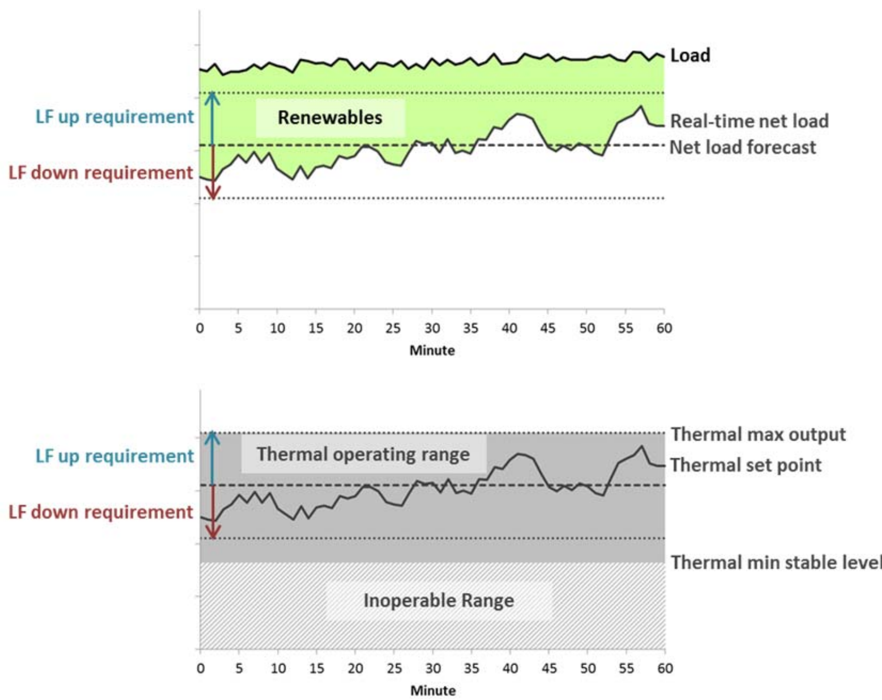
[37] Ela, E., M. Milligan, B. Kirby, "Operating Reserves and Variable Generation," National Renewable Energy Laboratory, Technical Report NREL/TP-5500-51978, August 2011. <http://www.nrel.gov/docs/fy11osti/51978.pdf>

increases, the need for flexibility reserves generally also increases. Whereas in most studies, this requirement is fixed, this study determines the appropriate level of reserve provision in each hour of the simulation endogenously based on the tradeoffs between the costs of meeting the flexibility reserve requirement and the economic consequences of a flexibility reserve shortfall. This section describes the theoretical basis for the approach taken in this study as well as the derivation of inputs used in the modeling process.

3.6.2.1 Theory

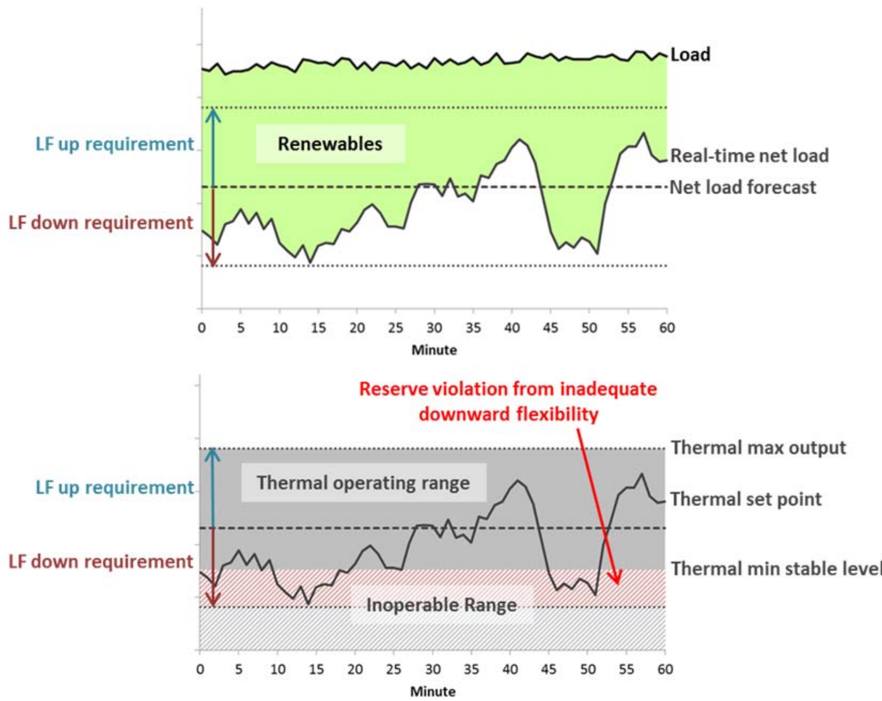
In conventional production cost modeling, flexibility reserve requirements are fixed on an hourly scale; these requirements must be met by reserving capacity from conventional flexible resources. At modest penetrations of renewable generation, the conventional fleet may be capable of operating over a range large enough to meet both upward and downward flexibility reserve requirements, as shown in Figure 44. At this level, these requirements have some impact on the optimal unit commitment and dispatch, but do not present a major challenge for operations, as the thermal fleet is capable of meeting the net load while also providing adequate operating range for both the upward and downward load following reserves.

Figure 44. Net load, load following (LF) reserve requirements, and thermal operating range for moderate renewable penetration during an hour that is not flexibility constrained.



At higher renewable penetrations, the need for flexibility reserves increases; at the same time, net load decreases. Accordingly, the set points of the conventional resources must decrease to accept the additional renewable energy while the necessary operating range to meet the reserve requirements increases in size. In this circumstance, meeting both the upward and downward flexibility reserve requirements may indicate the need to operate thermal units below their minimum stable levels, and so a flexibility reserve violation is encountered; this phenomenon is illustrated in Figure 45. In this hour, the thermal fleet is not capable of meeting the net load while also providing adequate operating range for both the upward and downward load following reserves. As a consequence, the system observes a downward load following violation in this hour.

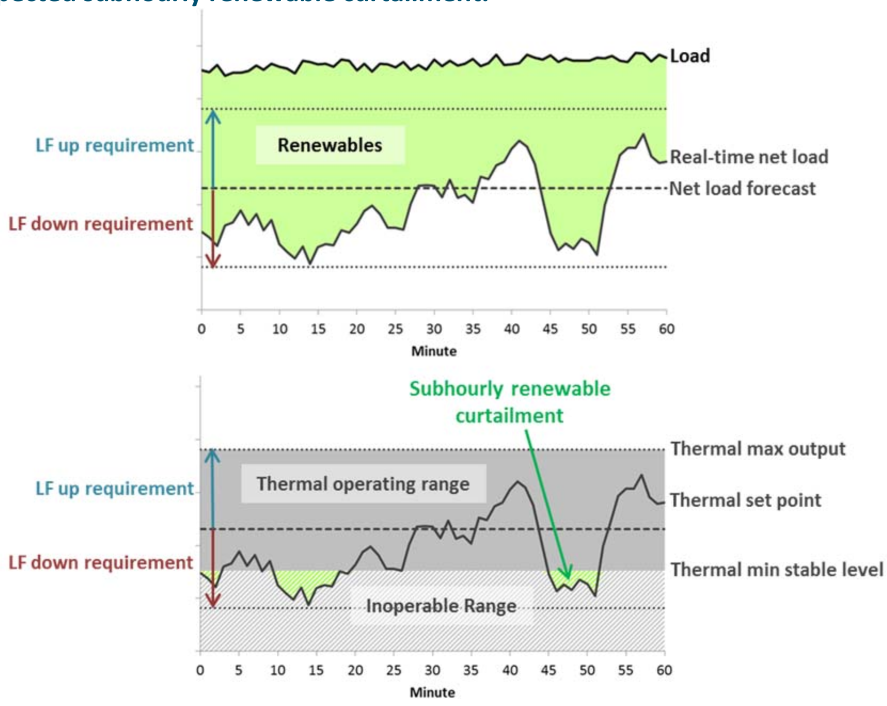
Figure 45. Net load, load following (LF) reserve requirements, and thermal operating range for high renewable penetration during an hour that is flexibility constrained.



In operations, the consequence of a failure to reserve adequate flexible capacity can vary. If nothing is done to correct it, the violation is experienced by the entire Interconnection through an increase in the Balancing Authority's Area Control Error (ACE). If the violation is small or if it serves to bring the interconnected system closer to its frequency target of 60 Hz, this may not have any negative consequences for the Balancing Authority. If the violation is large, occurs frequently, or moves the Interconnection away from 60 Hz, the BA may be subject to fines for violation of Control Performance Standards.

REFLEX does not capture the intricacies of CPS compliance; rather, it imposes economic penalties as if the BAA had prevented the violations from occurring. The BAA is assumed to do this either through real-time load shedding (in the case of upward violations) or real-time renewable curtailment (in the case of downward violations). This provides a signal that is commensurate with the potential economic consequences of insufficient flexible capacity. This is illustrated in Figure 46.

Figure 46. Manifestation of a downward reserve shortage as additional expected subhourly renewable curtailment.



This approach is valid as a modeling method even if the system operator does not have the ability to curtail renewables in real time, because it provides an appropriate economic penalty. However, if the system operator does have that capability, this approach captures the operator's choice to hold less capacity for downward flexibility reserves with the expectation that renewable generation will be curtailed on a subhourly timescale. In effect, the renewables themselves are providing the downward flexibility reserves.

By linking reserve shortfalls to their economic consequences (curtailment or unserved energy), the simulation determines the appropriate level of reserves to hold as part of the cost minimization problem. Including the economic consequences of reserve shortfalls in the objective function also allows the simulation to hold lower levels of flexibility reserves when it is economic to do so. One example of such a circumstance is during extended periods (e.g. multiple hours) of curtailment, when holding downward flexibility reserves would exacerbate curtailment (due to the commitment of incremental thermal generation, which then must run at its minimum generation levels) while a downward reserve shortfall would result in a limited amount of real-time curtailment.

3.6.2.2 Implementation

The inputs needed to model flexibility reserves endogenously comprise two components: (1) a "baseline" flexibility reserve requirement for each region and each hour that reflects the statistical variability and uncertainty of load, wind, and solar at that time; and (2) a normalized "surface" that links a shortfall in reserves to subhourly unserved energy or renewable curtailment. These two inputs are combined to yield a function for each region and each hour that links the actual provision of flexibility reserves to some expected quantity of subhourly unserved energy or curtailment.

The first of these inputs, the "baseline" flexibility reserve requirement, is derived using the standard methods developed in the Eastern Wind Integration and Transmission Study[38] to calculate load following reserves that can accommodate 95% of 5-minute real-time deviations from the hourly forecasted net load. This level of flexibility reserves is assumed to be adequate to avoid all subhourly unserved energy and renewable curtailment, subject to NERC standards. That is, it is assumed that deviations beyond this level are allowed under CPS and BAAL and can therefore be ignored in the flexibility assessment.

The second input is a normalized "surface" that links reserve shortfalls to their subhourly consequences, allowing the model to carry fewer load following reserves than the "requirement" described above, when conditions warrant. This surface is developed through 5-min simulation of a wide range of net load forecasts and conditions that might be encountered. In each hour interval within this set of system states, deviations of the 5-min actual net load from the net load forecast are compared to various potential load following reserve policies (i.e. quantities, in MW, of reserves that could be carried for a given hour). For each load following reserve policy, the simulation determines how much renewable curtailment (or unserved energy) would have been experienced within the hour. When the several hours of simulation are

[38] EnerNex Corporation, "Eastern Wind Integration and Transmission Study," National Renewable Energy Laboratory, Subcontract Report NREL/SR-5500-47078, February 2011.
<http://www.nrel.gov/docs/fy11osti/47078.pdf>

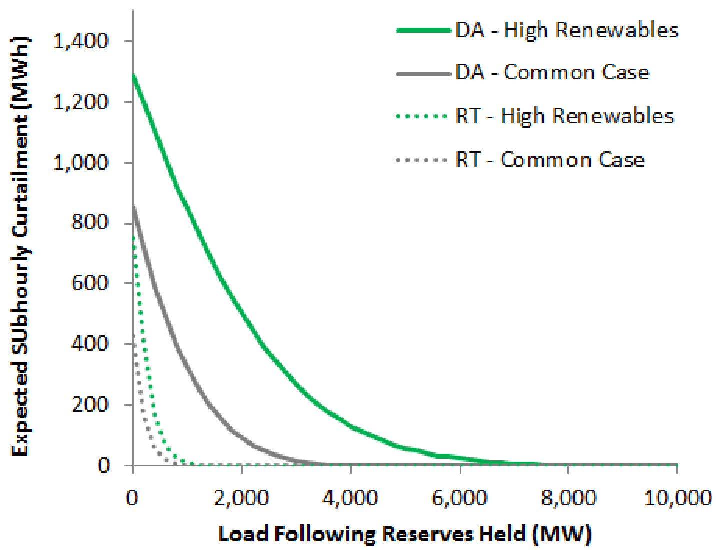
considered in aggregate, a parameterization can be developed that relates the reserve provision to the expected subhourly renewable curtailment in each hour as a function of the load, wind, and solar forecast for that hour.

When these two components are combined, the result is a function that links the reserve provision in each hour with the subhourly consequences. Examples of these functions for downward flexibility reserves in a single region and a single hour are shown for California in Figure 47.[39] Along each curve, the point along the y-axis represents the amount of subhourly curtailment that would be expected if no flexibility reserves are held; the point where the curve intercepts the x-axis represents the "baseline" reserve requirement; between these two points, increasing quantities of flexibility reserves result in decreasing expected quantities of subhourly curtailment.

Surfaces are developed for two timeframes in which unit commitment decisions are made: day-ahead (DA) and hour-ahead (or real-time, RT). In each case, the surfaces inform the model's decision about the quantity of load following reserves to include (both upward and downward) in unit commitment decisions.

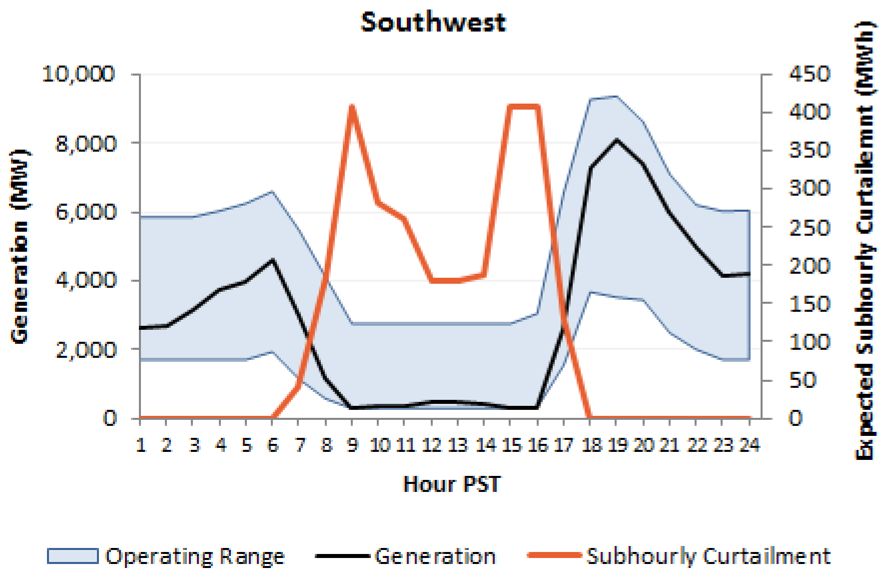
[39] In this figure, the DA requirements reflect both day-ahead forecast error and subhourly variability, whereas the RT curves reflect only subhourly variability.

Figure 47. Subhourly curtailment parameterization for HE08 in an example California draw for both the High Renewables Case and the Common Case



In each hour of a REFLEX simulation, the optimization uses these functions to consider the renewable curtailment (and unserved energy) that is expected to occur within the hour—along with their costs—when determining the thermal unit commitment and hence the downward reserve provisions among thermal units. The asymmetry of penalties on renewable curtailment ($100/MWh) and unserved energy ($50,000/MWh) typically yield unit commitment solutions that meet the conventional upward reserve requirement in all (or nearly all) hours, but that experience downward reserve shortages in high renewable output hours, resulting in expected subhourly curtailment. As a result, conventional resource set points sit very close to the minimum stable level during curtailment hours.

Figure 48. Operating range of conventional dispatchable generators in the Southwest on an example April day and the consequence of inadequate downward flexibility (subhourly curtailment).



This behavior is illustrated for an example April day in the Southwest in Figure 48. During nighttime hours on this day, the dispatchable generators sit at set points well within their full operating range in order to provide adequate upward and downward load following reserves. During daylight hours on this day, dispatchable generators are shut down to accommodate high output from solar resources. This reduces the operating range of the dispatchable units, making it harder to provide both upward and downward reserves. Because inadequate upward reserve shortages are substantially more costly than downward reserve shortages, the system prioritizes meeting upward reserve requirements over providing downward reserves, causing the dispatchable units to sit at low set points relative to the operating range in these hours. Note that the system does not choose to commit additional units to provide additional downward load following reserves, as this would require increased renewable curtailment to accommodate the minimum stable levels of the additional units.

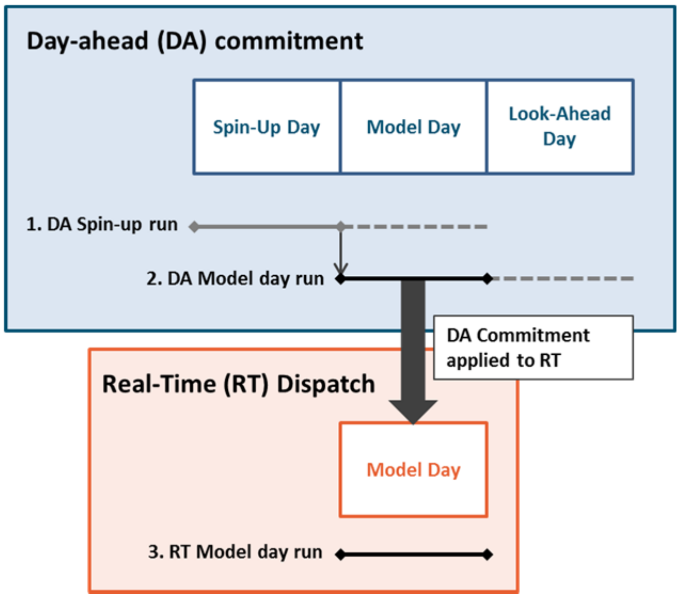
3.7 Production Cost Model Configuration

The draw methodology described above requires configuration of the PLEXOS model to simulate independent days in the flexibility analysis. The modeling of individual days in production cost introduces several challenges, including:

+ The application of constraints that span multiple hours, like ramp rate limits, across the beginning and end of the model day
+ The ability to ensure that long-start units begin and end the model day with appropriate commitment states; and
+ The allocation of a hydropower energy budget to the model day without information about the preceding day

The methods utilized in this study aim to address these challenges, but it is acknowledged that further investigation is needed into the effects of the adopted framework on the behavior of the system. This is discussed further in Section 7.2. To both enforce constraints that span multiple hours across the beginning and end of the model day and to determine the beginning commitment state of long-start units, PLEXOS considers a full three day period, where the drawn day is surrounded by the preceding and following days of hourly load, wind, and solar conditions. Operations for each three-day draw are simulated through a series of simulations, which are summarized in Figure 49 and described below. Only the results of the Model Day—the second day of the simulation—are presented in this report.

Figure 49. PLEXOS model configuration adapted for day draw methodology. Solid lines are optimization periods with 1-hour time steps;



The day-ahead commitment schedules on the model day are determined through a two-step optimization that makes use of day-ahead load, wind, and solar forecasts: first PLEXOS solves for the optimal commitment in each hour of the spin-up day, which takes into account the conditions on the model day as a "look-ahead" period with a four-hour resolution; next PLEXOS solves for the optimal commitment in each hour of the model day, given the hourly commitment in the spin-up day and taking into account the conditions on the look-ahead day with a four-hour resolution. The commitment states are then brought into the real-time simulation. The real time simulation simultaneously optimizes over the full 24-hour period of the model day given the actual hourly load, wind, and solar. All units are allowed to re-dispatch in real time, but only quick-start units (listed in Table 18) are allowed to adjust their commitment.

Table 18. Thermal unit commitment, dispatch, and reserve assumptions.

Thermal Technology	Final Commitment Stage	Final Redispatch Stage	Provides Spinning Reserves	Provides Flexibility Reserves
CCWhole-NatGas-Aero	DA	RT	Yes	Yes
CCWhole-NatGas-Industrial	DA	RT	Yes	Yes
CCWhole-NatGas-SingleShaft	DA	RT	Yes	Yes
CCWhole-SynGas	DA	RT	No	No
CrossCompoundWhole-Coal	DA	RT	No	No
ST-Coal	DA	RT	Yes	No
ST-NatGas	DA	RT	Yes	Yes
ST-Nuclear	Always Committed	RT	No	No
ST-Other	DA	RT	Yes	Yes
ST-OtherGas	DA	RT	Yes	Yes
ST-WasteHeat	DA	RT	No	No
CT-NatGas-Aero	RT	RT	Yes	Yes
CT-NatGas-Industrial	RT	RT	Yes	Yes
CT-OilDistillate	RT	RT	Yes	Yes
CT-OtherGas	RT	RT	Yes	Yes
CT-SynGas	RT	RT	No	No
ICE-NatGas	RT	RT	Yes	Yes
ICE-OilDistillate	RT	RT	Yes	Yes

4 Resource Adequacy Assessment Results

4.1 Regional Reliability Results

4.1.1 RELIABILITY STATISTICS

The results of the stochastic reliability assessment for the 2024 Common Case, summarized in Table 19, indicate that each region meets the study's assumed threshold for reliability of LOLF < 0.1. A small probability for loss of load events is identified in the Basin and Rockies regions; however, the size of these risks is not large enough to necessitate the addition of incremental capacity. Consequently, this modeling effort identifies no need for additional capacity beyond the resources of the Common Case to meet traditional reliability thresholds.

Table 19. Reliability statistics in each region, Common Case.

Reliability Metric[40]	Basin	California	Northwest	Rockies	Southwest
Loss of Load Frequency	0.02	—	—	0.04	—
Loss of Load Expectation	0.04	—	—	0.09	—
Expected Unserved Energy	10.6	—	—	21.9	—

[40] No operating reserves are assumed, consistent with the original conception of the "1-in-10 standard."

These results do not imply the adequacy of *today's* generation fleet to meet loads reliably in 2024. In addition to a substantial buildout of renewable resources to meet state policy goals, the Common Case also includes a number of conventional generator additions based on plans submitted to TEPPC by its members as well as based on information gathered by stakeholders in the TEPPC process. These additional resources represent planned investments without which the system might not meet the requisite reliability thresholds evaluated in this study.

The addition of incremental renewable generation to meet the targets of the High Renewables Case results in a system in each region that is more reliable than the Common Case. In the High Renewables Case, each region meets the LOLF threshold of 0.1 without the need to add additional capacity; this is not surprising as the Common Case already exceeded this standard. In fact, no reliability events are observed in the High Renewables Case due to the surplus of capacity present in the case.

4.1.2 PLANNING RESERVE MARGINS

While the primary function of the RECAP model is to compute these reliability statistics for an electric system, it also provides a snapshot of the more traditional planning reserve margin. The resulting planning reserve margin in the Common Case for each region is summarized in Table 20.

Table 20. Planning reserve margins in each region, Common Case (MW).

Type	Basin	California	Northwest	Rockies	Southwest
Conventional Generators	14,182	49,779	14,314	14,093	41,065
Hydro	1,915	10,878	26,209	1,489	3,568
Wind & Solar	1,002	7,615	854	845	2,749
Imports	445	11,768	2,500	-603	-1,737
Total Supply	17,544	80,040	43,877	15,824	45,645
1-in-2 Peak Demand	15,013	64,007	33,196	13,286	34,574
Reserve Margin (%)	*17%*	*25%*	*32%*	*19%*	*32%*

Results for the High Renewables Case are shown in Table 21. The additional renewables lead to large PRMs across the Western Interconnection.

Table 21. Planning reserve margins in each region, High Renewables Case (MW).

Type	Basin	California	Northwest	Rockies	Southwest
Conventional Generators	14,182	49,779	14,314	14,093	41,065
Hydro	1,915	10,878	26,209	1,489	3,568
Wind & Solar	2,983	10,855	2533	2438	5,439
Imports	445	11,768	2,500	-603	-1,737
Total Supply	19,525	83,280	45,556	17,417	48,335
1-in-2 Peak Demand	15,013	64,007	33,196	13,286	34,574
Reserve Margin (%)	30%	30%	37%	31%	40%

The calculated planning reserve margin is highly dependent on the conventions used in its evaluation, which are outlined in Table 22. In some cases, these conventions differ from the accounting method historically used by TEPPC to ensure resource adequacy in the development of the Common Case—most notably in the use of ELCC to measure the contribution of variable resources and hydro to the planning reserve margin.

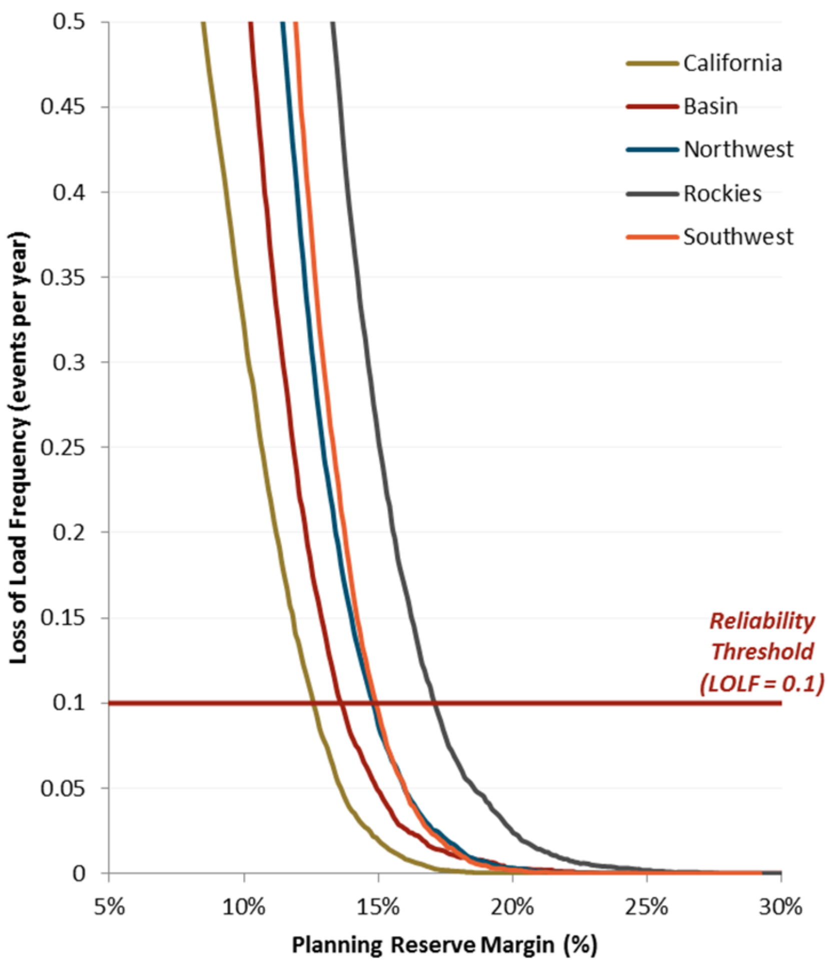
Table 22. Conventions used to count capacity towards the "planning reserve margin"

Category	RECAP Convention
Dispatchable Generation • *Biomass* • *Geothermal* • *Gas* • *Coal* • *Nuclear* • *DR* • *Storage*	100% of maximum rated capacity
Hydro • *Conventional hydro* • *Small hydro*	Effective load carrying capability
Variable Renewable Generation • *Solar PV* • *Solar Thermal* • *Wind*	Effective load carrying capability
Imports/Exports • *Specified* • *Unspecified*	100% of capacity of remote contracted resources + regional planning assumption for unspecified imports

4.1.3 TARGET PLANNING RESERVE MARGINS

The RECAP model can be used to derive target planning reserve margins—the reserve margin needed to meet the reliability threshold of LOLF = 0.1—given the characteristics of its loads and resources. Figure 50 shows the relationship between the loss of load frequency and the planning reserve margin in each region. While each region's curve is unique, the general functional form is the same: increasing the reserve margin of an electric system causes a decrease in the expected LOLF; this decrease is nonlinear and shows diminishing returns as more capacity is added to the system. For each region, the point on this curve at which the LOLF is equal to 0.1 represents the planning reserve margin needed to meet that reliability standard.[41] Based on this analysis, each region's target reserve margin is shown in Table 23.

Figure 50. Loss of load frequency as a function of planning reserve margin



[41] It is important to note that while this study interprets the "1-in-10" standard to reflect one loss of load event in ten years, there is not uniform agreement on what standard should be used. The "1-in-10" standard itself has several interpretations, each of which would imply a different target planning reserve margin.

Table 23. Target planning reserve margins for each region

Type	Basin	California	Northwest	Rockies	Southwest
Target Reserve Margin (%)	14%	13%	15%	17%	15%

The difference in target planning reserve margin standards between regions is expected and is primarily due to the difference in region size, composition of the generation fleet, and contingency size. First, a larger region will have greater load and resource diversity; load diversity dampens extreme peak loads while resource diversity reduces the likelihood of a critical forced outage occurring simultaneously. Resource type is important due to its outage frequency, repair rates, and or seasonal patterns in dependable capacity. Finally, a system with large contingencies, such as one with a large portion of load met with only a few generators, will tend to need a higher planning reserve margin because generator failure in just 1-2 locations may be enough to cause a loss of load event.

4.2 Renewable Effective Load Carrying Capability

One of the distinguishing features of the RECAP model is its ability to produce estimates of the effective load carrying capability (ELCC) of variable generation, a measure of its contribution to reliability relative to a perfectly reliable generation resource. This metric provides a more analytically robust measure the capacity value of variable resources than the rules of thumb commonly used today. ELCC is therefore a useful metric for resource planning in several respects: (1) it can be used to measure capacity contributions from variable renewable resources in the traditional planning reserve margin framework;[42] and (2) is provides a useful measure of the value provided by prospective renewable investments with which procurement decisions can be more effectively made to mitigate costs to ratepayers.

4.2.1 TECHNOLOGY-SPECIFIC ELCC CURVES

Marginal ELCC curves for wind and solar technologies in each region are plotted in Figure 51. Each curve is derived assuming that only that variable resource is present on the system; such curves inherently do not account for the benefits of resource diversity, which can result in greater ELCCs for variable resources (see Section 4.2.2). These results suggest a number of observations on the nature of ELCC for renewable resources across the Western Interconnection:

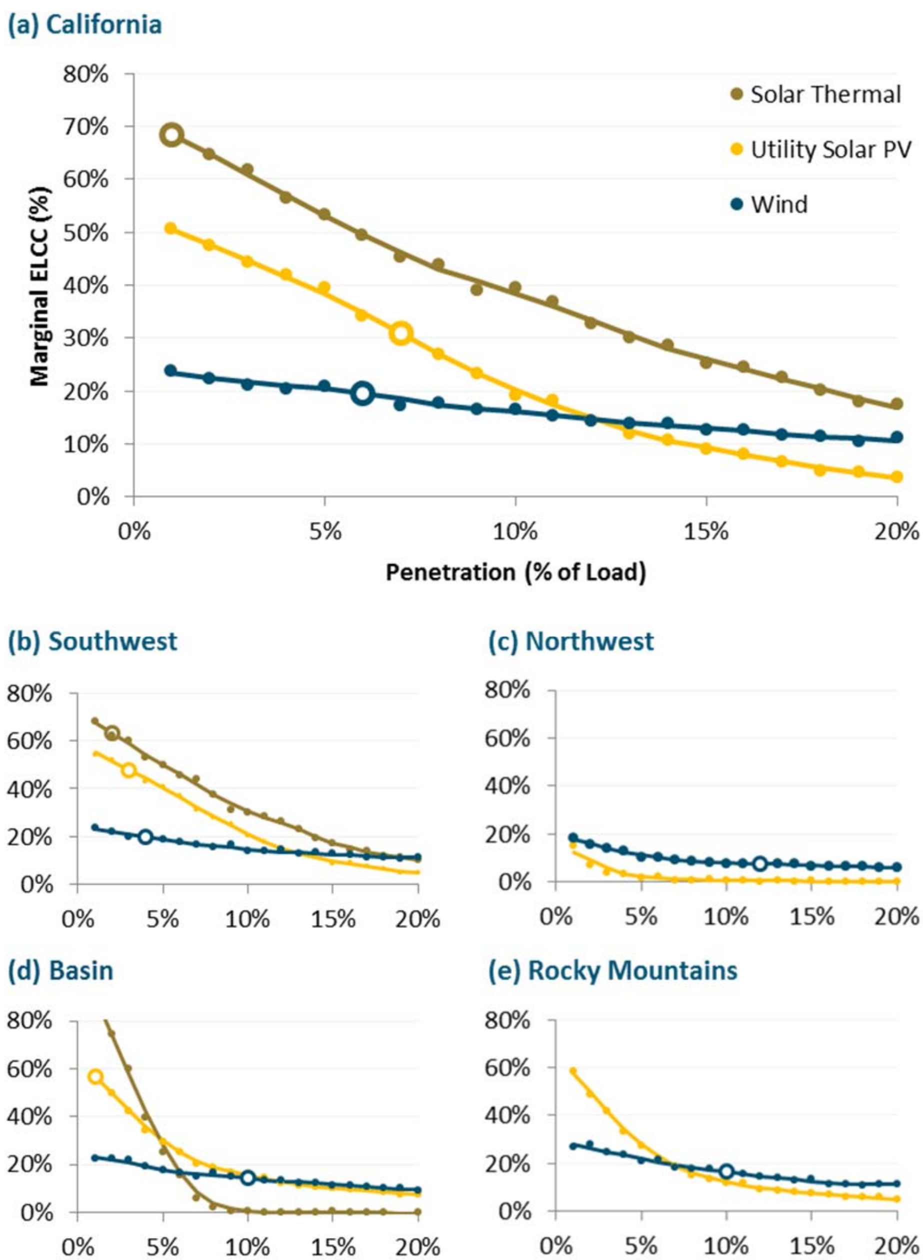
+ In all regions, wind and solar ELCC values exhibit diminishing returns to scale. As penetrations increase, such variable resources have a reduced benefit to system reliability as the net peak demand shifts away from hours in which production is concentrated.
+ In most regions, the marginal ELCC of solar PV at low penetrations is relatively high (50-60%) and aligns with common rules of thumb used to attribute capacity credits to solar in many planning exercises. This reflects the coincidence of solar generation with peak load conditions in most of the regions. The notable exception to this observation is the

[42] In California, the CPUC is currently working to incorporate ELCC values for wind and solar resources in its Resource Adequacy proceeding.

Northwest, where the timing of the peak demand during the evening in the winter results in very low ELCC values for solar resources.

+ The marginal ELCC of solar PV decreases rapidly in most regions with increasing penetration; this is a result of the effect of the shifting of the net load peak towards the evening when solar production drops. At high penetrations, traditional rules of thumb based on the coincidence of solar production with peak load become highly inaccurate in the value they attribute to solar resources.
+ Marginal ELCC for solar thermal resources (calculated only for California the Southwest, and the Basin, which each have a small penetration of solar thermal resources in the Common Case) exhibits the same general trend as solar PV; however, the marginal ELCC values are slightly higher due to the fact that a number of these plants are assumed to have thermal storage, which allows them to sustain higher levels of output through peak periods in the late afternoon and early evening.
+ At low penetrations, marginal ELCC values for wind range from 15-30% of nameplate capacity; as penetration increases, the marginal ELCC declines, though at a rate slower than the decline for solar technologies.

Figure 51. Marginal ELCC for wind & solar resources by region

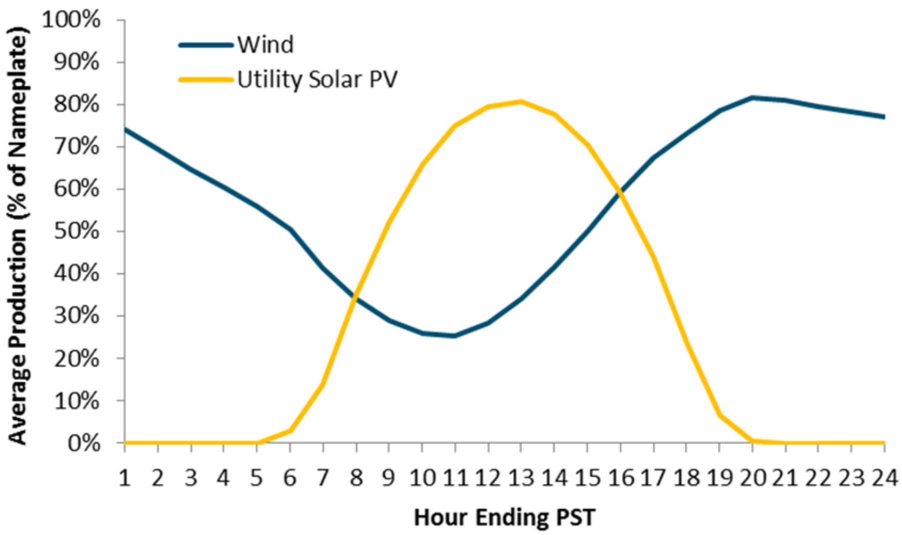


4.2.2 PORTFOLIO EFFECTS & BENEFITS OF DIVERSITY

The technology specific ELCC curves shown in section 4.2.2 are useful for illustrating diminishing returns from a single resource type for resource adequacy, but they do not give an accurate picture of the total resource adequacy value of a portfolio of variable generation. The portfolio ELCC value differs from the sum of the individual resource ELCC values due to two factors: portfolio effects—positive or negative resource complementarity—and diversity—reduction in production variability.

To illustrate the impact of complementary production profiles, take Figure 52, which shows average wind and solar production in July. As increasing amounts of solar are added to the system, the net peak will shift later in the evening; this effect is reflected in the declining marginal ELCC curves. However, this shift in the timing of the net peak load also increases the contribution of wind, which has higher expected production later in the evening, to system reliability. The net result is that the wind and solar resources complement one another and result in a greater ELCC together than the sum of the two separately.

Figure 52. Average wind and solar in California for July



The magnitude of this effect can be quantified using Figure 53, which shows the combined ELCC of wind and solar resources for a variety of portfolio combinations of the two in California. In the 2024 Common Case, California has approximately 10% solar PV and 6% wind penetration by energy.[43] Referencing Figure 53, this total resource mix yields an ELCC of 6,924 MW; however, when examined separately, a 6% wind penetration and a 10% solar penetration yield ELCC values of 1,213 MW and 4,916 MW, respectively. The sum of 1,213 MW and 4,916 MW gives 6,129 MW, short of the total portfolio value by 795 MW. The additional 795 MW (shown in Figure 54) is the portfolio effect and is 12% of the total portfolio ELCC value. Resources with a high degree of correlation, such

[43] For purposes of illustrative simplicity, the small penetration of solar thermal in California is ignored in this example.

as solar thermal and solar PV, show negative portfolio effects because they both tend to shift net peak load in the same ways.

Figure 53. Effective load carrying capability in megawatts as a function of wind and solar penetration in California

Solar Penetration (% Annual Load)	Wind Penetration (% Annual Load) 0%	2%	4%	6%	8%	10%	12%	14%	16%	18%	20%
0%	0	432	822	1,213	1,539	1,849	2,127	2,388	2,622	2,839	3,041
2%	1,346	1,803	2,235	2,638	2,983	3,314	3,628	3,908	4,160	4,402	4,621
4%	2,526	3,050	3,526	3,962	4,349	4,696	5,032	5,332	5,622	5,868	6,105
6%	3,533	4,109	4,648	5,141	5,560	5,952	6,319	6,648	6,959	7,245	7,498
8%	4,331	4,994	5,576	6,114	6,613	7,048	7,436	7,803	8,131	8,440	8,725
10%	4,916	5,664	6,324	6,924	7,448	7,939	8,378	8,769	9,142	9,482	9,775
12%	5,362	6,128	6,853	7,511	8,115	8,648	9,136	9,569	9,966	10,331	10,648
14%	5,670	6,500	7,255	7,958	8,608	9,198	9,743	10,228	10,648	11,051	11,412
16%	5,903	6,758	7,573	8,308	8,991	9,627	10,200	10,715	11,184	11,599	11,988
18%	6,060	6,938	7,784	8,570	9,288	9,958	10,566	11,108	11,598	12,064	12,462
20%	6,172	7,081	7,952	8,768	9,503	10,208	10,844	11,421	11,935	12,414	12,861

Figure 54. ELCC diversity benefits between wind and solar PV resources

Solar Penetration (% Annual Load)	Wind Penetration (% Annual Load) 0%	2%	4%	6%	8%	10%	12%	14%	16%	18%	20%
0%	0	0	0	0	0	0	0	0	0	0	0
2%	0	+25	+67	+79	+98	+119	+155	+174	+192	+217	+234
4%	0	+92	+178	+223	+284	+321	+379	+418	+474	+503	+538
6%	0	+144	+293	+395	+488	+570	+659	+727	+804	+873	+924
8%	0	+231	+423	+570	+743	+868	+978	+1,084	+1,178	+1,270	+1,353
10%	0	+316	+586	+795	+993	+1,174	+1,335	+1,465	+1,604	+1,727	+1,818
12%	0	+334	+669	+936	+1,214	+1,437	+1,647	+1,819	+1,982	+2,130	+2,245
14%	0	+398	+763	+1,075	+1,399	+1,679	+1,946	+2,170	+2,356	+2,542	+2,701
16%	0	+423	+848	+1,192	+1,549	+1,875	+2,170	+2,424	+2,659	+2,857	+3,044
18%	0	+446	+902	+1,297	+1,689	+2,049	+2,379	+2,660	+2,916	+3,165	+3,361
20%	0	+477	+958	+1,383	+1,792	+2,187	+2,545	+2,861	+3,141	+3,403	+3,648

Diversity within a given resource type also has value because a reduction in variability on peak, for a given expected value, leads to a higher ELCC. This means that the ELCC of wind or solar resources will tend to be higher when the size of the area examined is increased, provided a consistent resource quality.

4.3 Future Application of RECAP

The analysis of reliability in this study is intended as a precursor to the flexibility assessment and as a demonstration of a modeling framework; it is not a substitute for a rigorous planning reserve margin study in each of the regions of the Western Interconnection, including an additional examination of regional reliability standards. Nonetheless, the RECAP model provides a flexible platform from which more detailed reliability analysis of individual balancing authorities, regions, or the Western Interconnection as a whole is possible. In order to apply the RECAP model in the context of its continued use in reliability analysis at WECC, additional refinement to inputs and assumptions will improve the model's characterization of reliability in the area of focus.

+ **Develop renewable profiles that reflect expected output patterns from actual plants.** In this study, renewable profiles are based on data sets whose assumptions underlying resource performance do not necessarily align with observations of existing plants. For example, in a number of cases, the power curve used to derive profiles in the WIND Toolkit yields a higher estimate of plant capacity factors than are observed in existing wind plants at the same geographic location. Developing a single data set of renewable profiles that provides a reasonable representation of both existing and future resources will continue to present a challenge to resource planners, yet as efforts to develop additional data on wind and solar performance continue to evolve, ensuring the best possible representation of performance is crucial to the study of reliability.

+ **Review forced outage rate assumptions.** The forced outage rates of generation resources—a characteristic oft-overlooked in the PRM framework of reliability—have a first-order impact on the results of LOLP analysis. NERC's GADS provides a reasonably comprehensive dataset of observed historical patterns of forced outages, but studies should also consider whether forced outage rates may change in the future, especially due to increased cycling and ramping requirements at higher renewable penetrations.
+ **Determine appropriate assumptions for contribution of imports to reliability.** As discussed in Section 3.3.3, one notable challenge in reliability planning is the determination of how much to rely on imports for reliability, a determination that typically requires some discretion on the part of the resource planner. Depending on the footprint of the area of study, the strength of its transmission links to neighboring areas, the historical utilization of those interties during peak periods, and the availability of generation resources in neighboring areas may all provide some basis for this assumption.
+ **Conduct detailed assessment of sustained peaking capability of hydroelectric fleet.** This study incorporates a portion of the detail from the NWPCC's evaluation of the sustained peaking capability of the hydroelectric fleet in the Pacific Northwest; however, comparable efforts to measure the possible contribution of hydro resources under a variety of hydro conditions have not been undertaken in such detail in other regions of the West. In the absence of such measurements, this study generalizes the constraints used in resource planning in the Northwest and applies them to the other regions of the West in order to approximate limitations in the output of hydro resources across a variety of conditions. To the extent that hydro resources account for a

non-trivial share of the generating capacity in the footprint of the area of interest (e.g. California in this study), a more detailed representation of the potential range of peaking capabilities of hydro resources will enhance the reliability assessment.

5 Flexibility Assessment Results

5.1 Need for Flexibility

The addition of variable renewable generation to the electric system increases the need for operational flexibility to ensure that operators can continuously balance load with generation to provide reliable service. The ability of the system to provide the needed flexibility depends on three primary factors:

+ **Operating range:** The ability to operate to serve a wide range of potential net load conditions.
+ **Ramping capability:** The ability to change quickly from one net load condition to another.
+ **Forecast accuracy:** The ability to adjust system commitment and dispatch to accommodate forecast errors in net load.

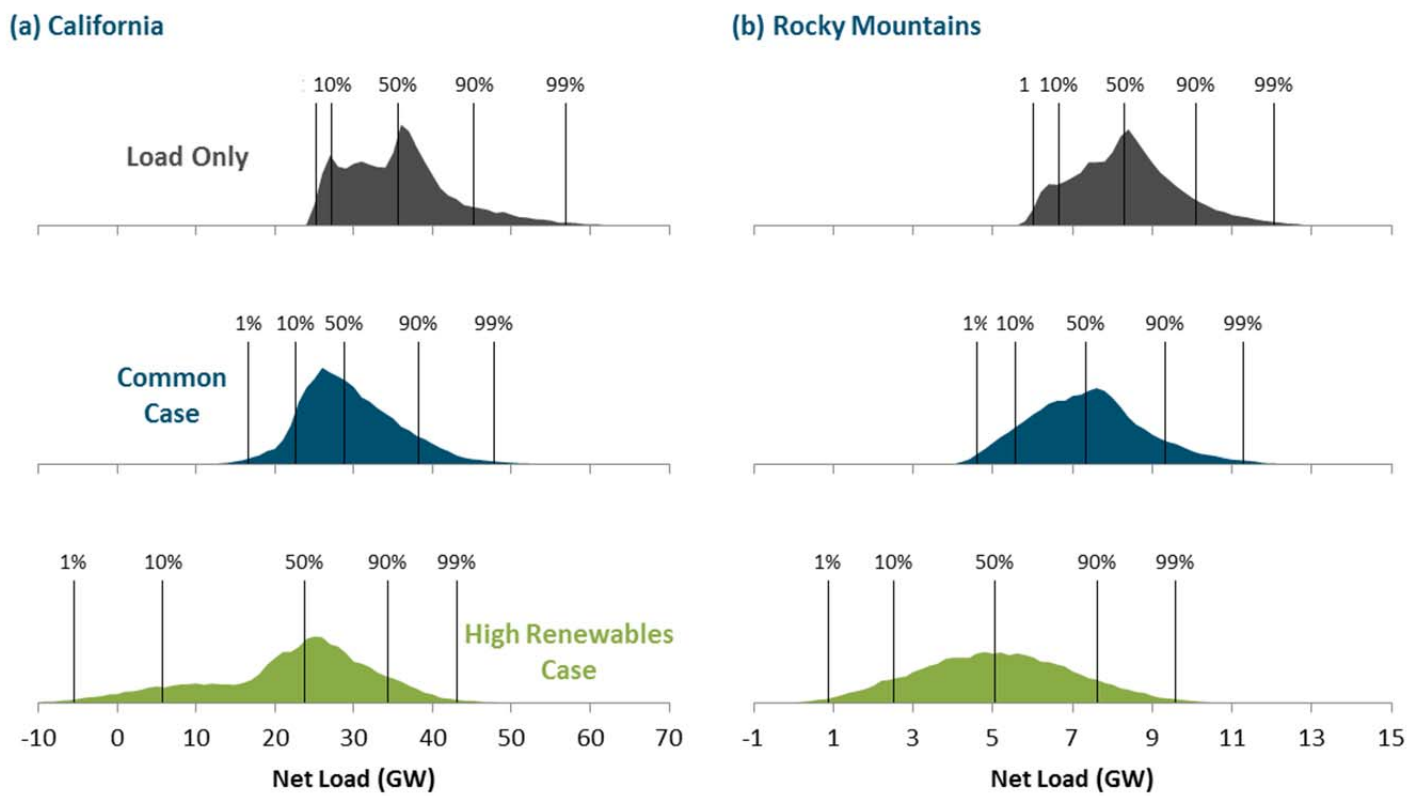
None of these three considerations are new to system operators, who have historically dispatched resources to meet a wide range of load conditions while managing uncertainties in load. Rather, the addition of renewable generation to the system magnifies the need for flexibility to respond to each of these phenomena.

5.1.1 OPERATING RANGE

Because the output of variable renewable resources is concentrated in time periods throughout the day and year that do not correlate with load, higher penetrations of these resources require a system to operate across a wider range of potential net load conditions. Examples of the distributions of load and net load for California and the Rocky Mountains[44] are shown in Figure 55. In both regions, the distribution of load conditions has a long tail on the upper end, indicating the low probability for extreme high load conditions around which electric systems have traditionally been planned to serve load reliably. As renewable generation is added, several impacts are notable. First, the width of the distribution increases. Second, particularly in regions with high solar penetration such as California, a tail on the lower end of the distribution appears at high penetrations of renewable generation; this is due primarily to the concentration of solar generation during the middle of the day, which, during off-peak seasons, results in very low net load conditions.

[44] Examples are shown for these two regions because of their differences: California is the largest region analyzed in this study and relies predominantly on solar PV in the High Renewables Case, whereas the Rocky Mountain region is the smallest of the five and relies mainly on wind.

Figure 55. Distributions of hourly load and net load, Common Case & High Renewables Case



5.1.2 RAMPING CAPABILITY

A second impact of increased penetrations of variable renewable generation is the increase in magnitude of net load ramps, which occur across a multitude of time scales. Within the hour, the subhourly variability and intermittency of renewable generation requires operators to hold increased flexibility reserves; on a longer time scale, diurnal ramps in renewable production contribute to the need to meet larger multi-hour ramps in net load. Table 24 and

Table 25 summarize the extreme three-hour ramps in load and net load for each region.

Table 24. 99th percentile, three-hour upward net load ramp (MW).

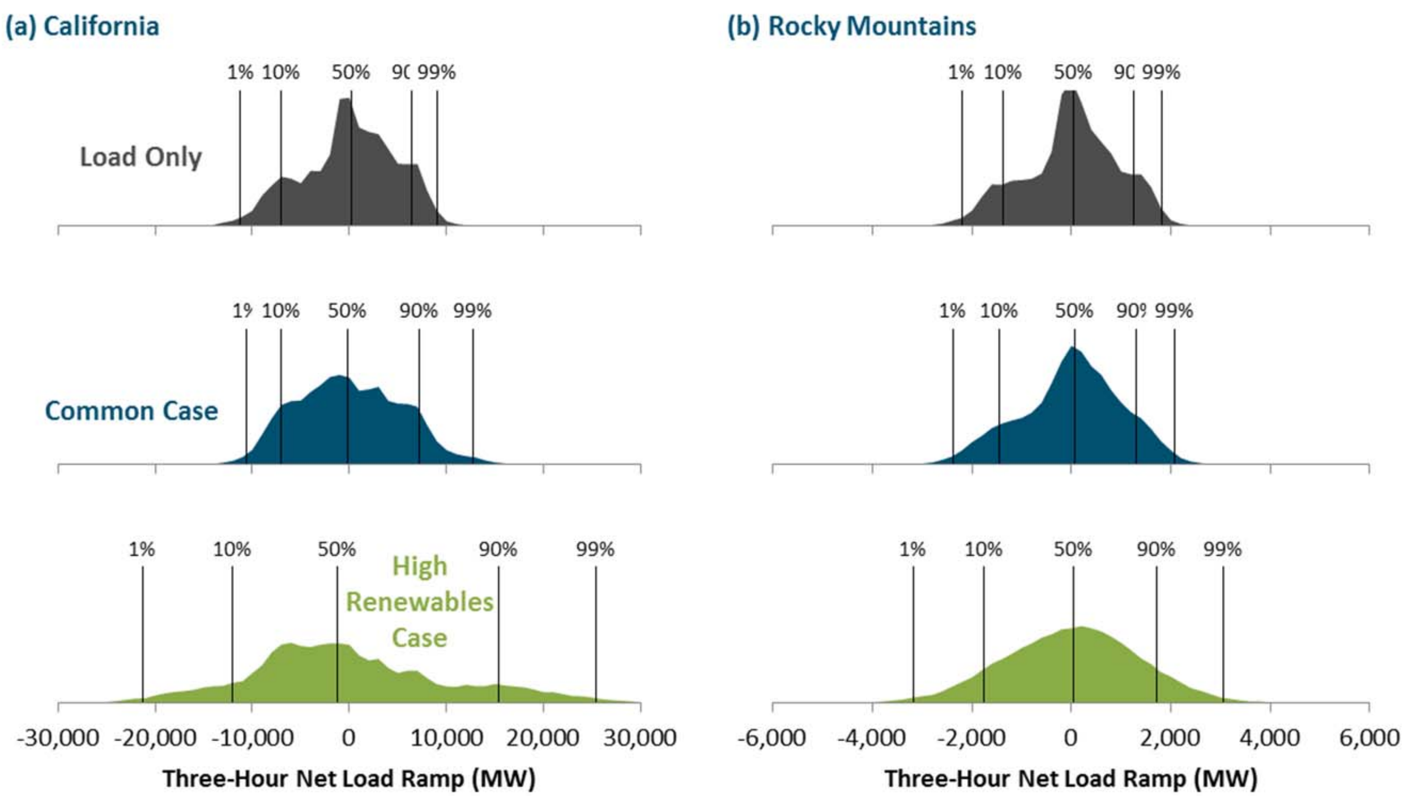
Type	Basin	California	Northwest	Rockies	Southwest
Load Only	+2,227	+9,003	+6,164	+1,824	+5,395
Common Case	+2,218	+12,715	+6,429	+2,079	+4,556
High Renewables Case	+3,570	+25,338	+6,707	+3,064	+13,748

Table 25. 99th percentile, three hour downward net load ramp (MW).

Type	Basin	California	Northwest	Rockies	Southwest
Load Only	-2,616	-10,941	-4,848	-2,195	-5,720
Common Case	-2,558	-10,609	-5,082	-2,377	-5,506
High Renewables Case	-3,180	-21,290	-5,624	-3,181	-13,045

Figure 56 shows the full distribution of three-hour ramps in load and net load for each renewable portfolio in California and the Rocky Mountains.

Figure 56. Distributions of three-hour net load ramps, Common Case & High Renewables Case



It should be noted that these ramping distributions are estimated *before* taking into consideration the ability of the system to manage them; that is, the net load ramps that must be met by conventional generators in production cost modeling are reduced in some hours because renewable output is curtailed either due to oversupply or insufficient ramping capability.

5.1.3 FORECAST ACCURACY

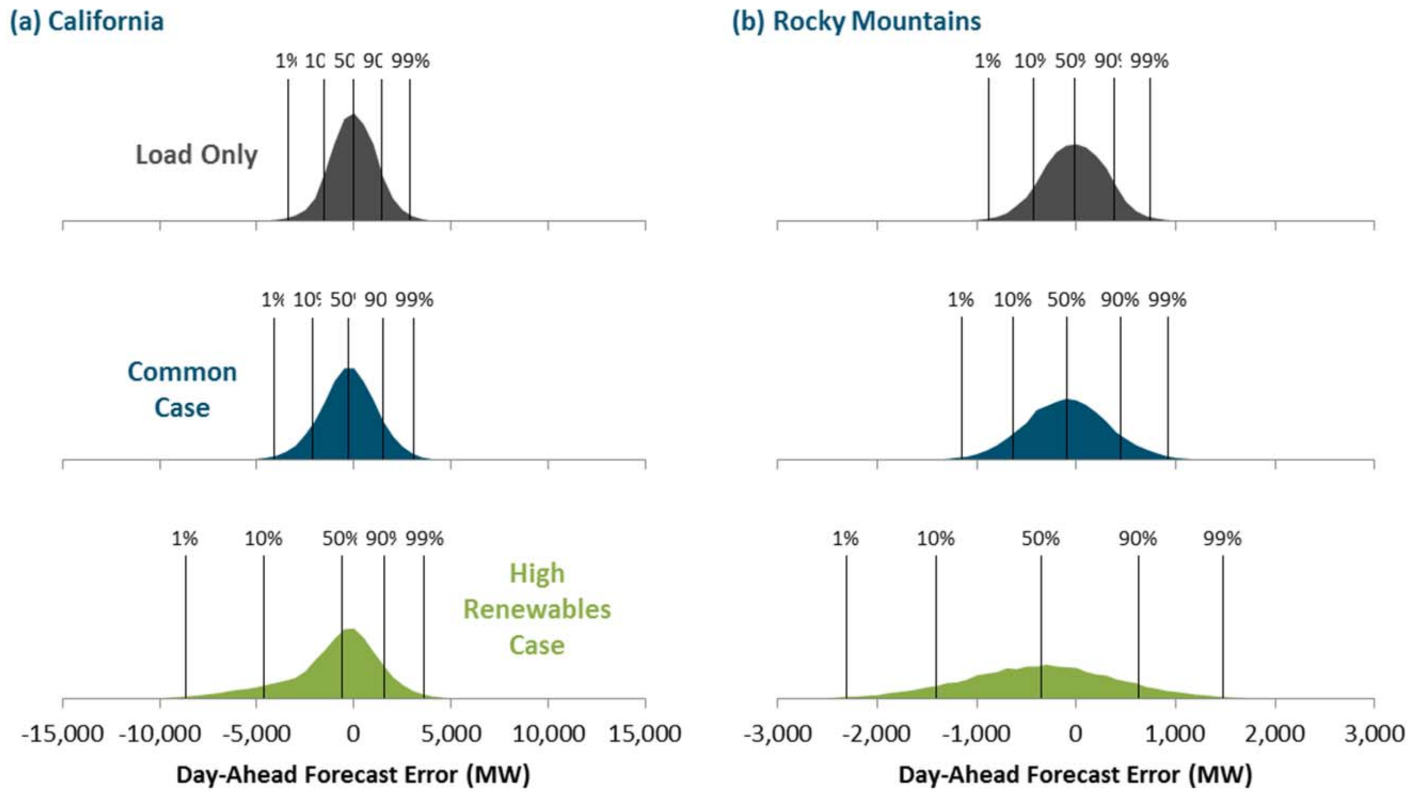
The third impact of renewables that introduces the need for additional flexibility is the increase in the net load forecast uncertainty due to the forecast errors associated with wind and solar generation. Table 26 shows the mean absolute error (MAE) for load as well as for net load in the Common Case and the High Renewables Case. In each region, the impact of adding uncertain renewable resources is significant: day-ahead net load forecast errors are increased by 105-162% in the Common Case and by 217-317% in the High Renewables Case.

Table 26. Summary of mean absolute error for load and net load (MW).

Type	Basin	California	Northwest	Rockies	Southwest
Load Only	243	968	619	261	566
Common Case	637	2,265	1,745	568	1,161
High Renewables Case	875	3,372	2,574	1,090	1,794

Figure 57 shows profiles for the distributions of load and net load forecasts for California and the Rocky Mountains, respectively. For the reasons discussed above, the distributions of net load forecasts under the High Renewables Case are considerably broader in both regions, and the frequency of large under- and over-forecasts of net load is considerably higher than for load alone.

Figure 57. Distributions of day-ahead net load forecast error, California and Rocky Mountains.



5.2 Reference Grid Results

5.2.1 OVERVIEW OF FINDINGS

This study uses the term ‘Reference Grid’ to define a set of assumptions that define the underlying flexibility of the Western Interconnection’s electric system; these assumptions are used to model both the Common Case and High Renewables Case portfolios. Like all of the scenarios considered within this study, the Reference Grid case is not intended to be predictive of the future, but rather to provide indicative information on the potential outcomes under a pre-defined set of assumptions. The key assumptions that constitute the Reference Grid scenario are:

+ Flows and ramps across interregional interties are limited to historical ranges;
+ Regional hydroelectric fleet operations are limited to historically observed range;
+ Renewable curtailment is allowed at a penalty cost of $100/MWh;
+ Flexibility reserve provisions are determined endogenously as described in Section 3.6.2; and
+ The non-renewable, non-hydro generation fleet is modeled based on characteristics in TEPPC’s 2024 Common Case.

The resulting simulated operations of the 2024 Common Case renewable portfolio are summarized in Table 27. Across the footprint of the study, renewable generation accounts for approximately 20% of regional load; the remainder is mostly served by a combination of gas (26%), coal (26%), hydro (21%), and nuclear (7%) generation.

Table 27. Generation mix in each region, Common Case (GWh)

Type	Basin	California	Northwest	Rockies	Southwest
Nuclear	—	18,587	9,494	—	32,000
Coal	53,534	17,940	23,872	53,439	67,752
Gas	9,809	128,769	23,970	6,623	48,992
Hydro	10,446	26,417	124,083	4,059	10,239
Pumped Storage	—	1,899	—	650	36
Storage	—	993	—	—	—
Renewables	18,335	90,120	30,493	8,485	15,800
Net Imports	-12,353	41,914	-18,530	216	-12,007
Total	**79,772**	**326,639**	**193,383**	**73,471**	**162,812**

Table 28 shows the generation mix for each region in the Western Interconnection under the High Renewables Case. Across the study footprint, renewable generation accounts for just below 40% of regional loads. The addition of significant quantities of renewable generation results in predictable reductions in the output from other generation resources: gas (39% reduction across study footprint) and coal generation (33% reduction) are displaced by the addition of incremental renewables, resulting in reductions in operating costs and reduced emissions across the footprint of the study.

Table 28. Generation mix in each region, High Renewables Case (GWh)

Type	Basin	California	Northwest	Rockies	Southwest
Nuclear	—	18,882	9,259	—	32,281
Coal	37,335	11,178	15,487	34,685	45,708
Gas	4,272	89,827	6,263	3,369	30,120
Hydro	10,578	26,504	124,140	4,093	10,278
Pumped Storage	—	5,522	—	1,181	301
Storage	—	2,462	—	—	—
Renewables	31,859	158,328	58,908	29,084	64,092
Net Imports	-4,336	34,560	-17,833	1,418	-15,179
Total	**79,709**	**347,264**	**196,225**	**73,830**	**167,601**

Challenges to system flexibility are manifest in the simulation through either unserved energy or renewable curtailment. In the 2024 Common Case, both of these indicators of flexibility challenges are observed exceedingly rarely. As expected based on the resource adequacy analysis from Phase 1, unserved energy is not observed in any of the Common Case draws. Renewable curtailment is also a rare phenomenon; the total amount of curtailment observed across all regions amounts to less than 0.1% of the total available renewable generation. These metrics are promising indicators of the flexibility of the generation fleet to accommodate renewable penetrations consistent with the Common Case: the non-renewable fleet is able to serve load across all conditions examined while allowing almost all of the available renewable generation to be delivered.

At higher penetrations of renewable generation, the system begins to encounter flexibility challenges. Notably, no unserved energy is observed in the High Renewables Case; rather, constraints on flexibility are realized through renewable curtailment, which is experienced in increasing frequencies and magnitudes relative to the Common Case. Renewable curtailment is experienced on two different time horizons through the REFLEX modeling framework:

+ At an **hourly** time scale, renewable curtailment is scheduled to maintain a balance between the real-time hourly net load and the supply of disptachable resources.
+ At a **subhourly** time scale, renewable curtailment is realized when the system is short on downward flexibility reserves, implying some expectation of curtailment in order to follow the five-minute variability of the net load signal.

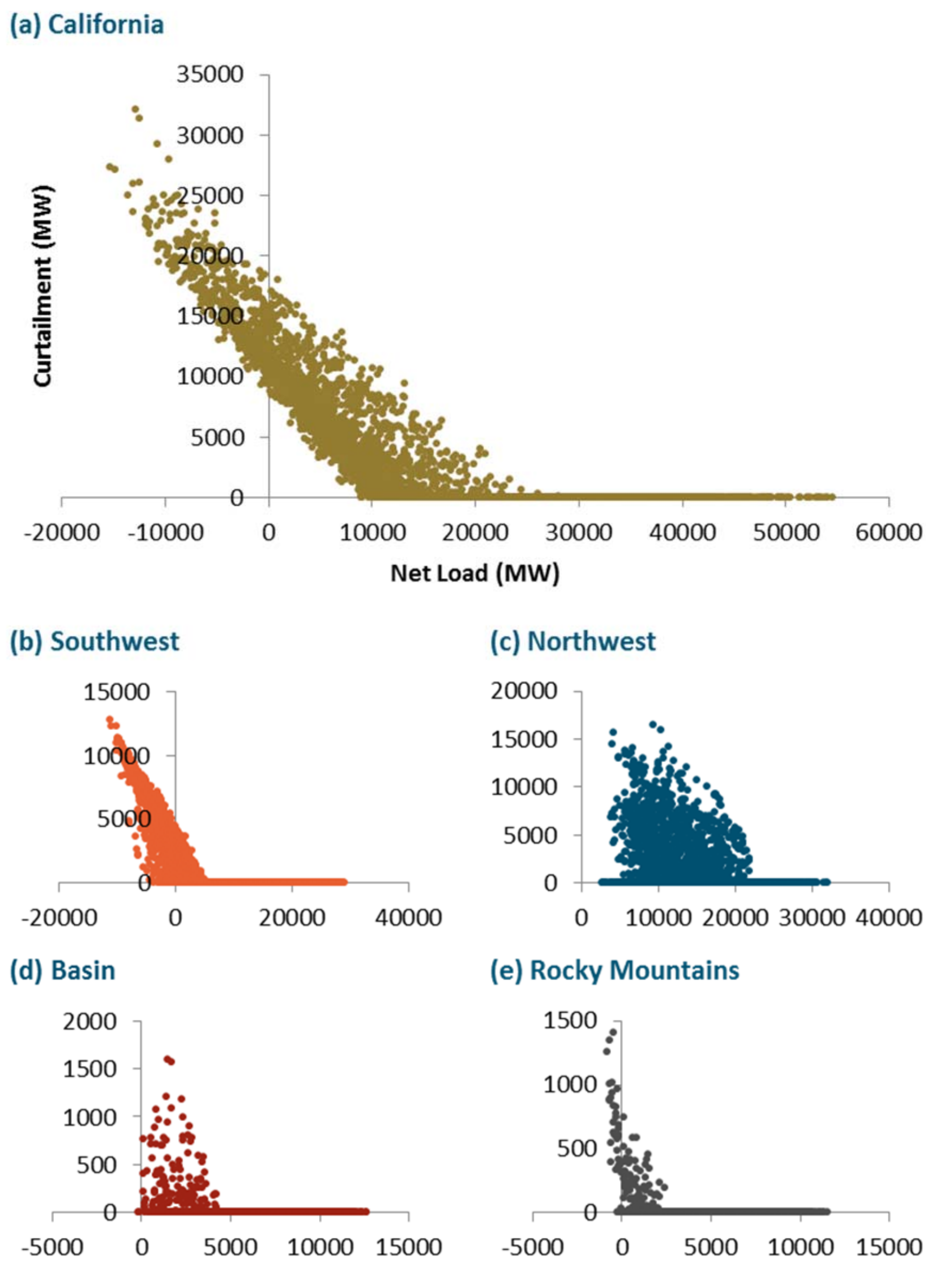
These two categories of curtailment are manifest to differing extents in each of the regions, as summarized in Table 29.

Table 29. Summary of renewable curtailment observed in the High Renewables Case.

	Type	Basin	California	Northwest	Rockies	Southwest
Total Energy Curtailed (GWh)	Hourly	34	13,686	3,586	29	4,076
	Subhourly	96	64	5	136	589
	Total	**130**	**13,749**	**3,591**	**165**	**4,664**
Percent of Renewable Gen (%)	Hourly	0.1%	8.6%	5.6%	0.1%	6.4%
	Subhourly	0.30%	0.04%	0.01%	0.47%	0.92%
	Total	**0.4%**	**8.7%**	**5.6%**	**0.6%**	**7.3%**

Renewable curtailment is seen most frequently during low net periods, which arise from a combination of low load and high renewable output. Figure 58 compares each region's net load for each hour across all of the simulated draws with the amount of renewable curtailment observed in that hour. While the shape of each scatter plot differs for each region, a clear trend emerges: renewable curtailment is confined to periods of low net load. This observation is indicative of the major cause of renewable curtailment under high penetrations: “oversupply,” when the capacity of a system to use (and/or export) renewable generation is exceeded by the amount available, operators must curtail the excess in order to maintain the balance between loads and resources. This is certainly not the only reason to curtail renewable generation in operations, but at high penetrations of renewables, it becomes the dominant cause.

Figure 58. Renewable curtailment as a function of net load in each region.



** Note different scales in each region*

Flexibility challenges may also be encountered due to net load ramps that are larger than the gross load ramps to which system operators are accustomed. The magnitudes of the net load ramps encountered in each region are discussed in Section 5.1.2. The operational strategies for managing these large net load ramps vary by region. Figure 59 and Figure 60 show the upward and downward (respectively) ramping by technology utilized to manage the ten largest three-hour net load ramps observed in each region. A variety of resources are used to help meet ramping needs in each region, including a large contribution from coal in the Basin and Rocky Mountain regions and to a lesser extent in the Southwest. Ramping in the Northwest is largely handled through dispatch of the hydropower fleet. In California, the Northwest, and the Southwest, where significant curtailment is observed, renewable curtailment helps to reduce the burden placed on conventional generators to respond to the largest net load ramps. In this way, renewable curtailment is not only a manifestation of overgeneration conditions, but may also capture the impact of flexibility constraints on the conventional generator fleet. This is investigated further in Section 5.4.3.

Figure 59. The average contribution by resource type to meeting the ten largest 3-hr net load upward ramps in each region.

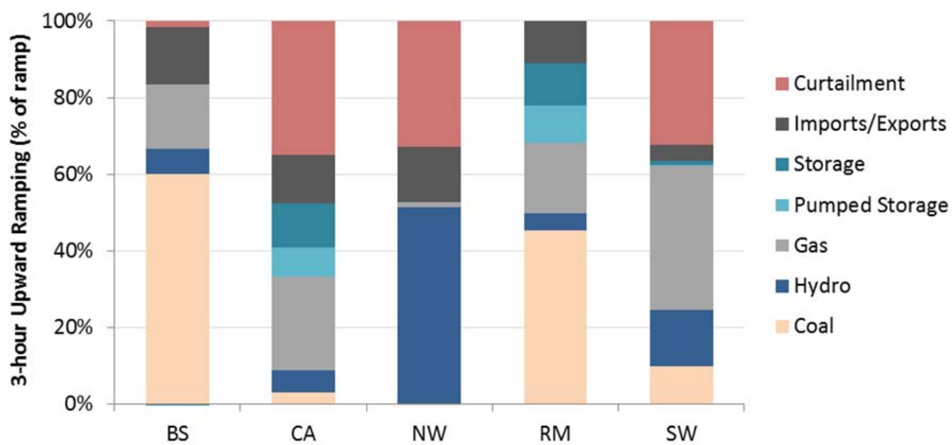
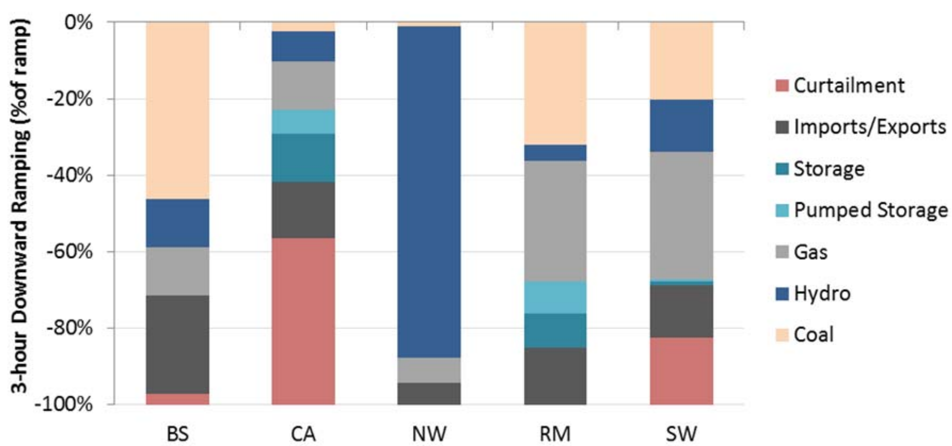


Figure 60. The average contribution by resource type to meeting the ten largest 3-hr net load downward ramps in each region.



Finally, net load forecast errors may contribute to renewable integration challenges. In this analysis, day-ahead forecasts of the net load are used to establish the commitment of long-start units, which limits the ability of the system to utilize these resources when mitigating large forecast errors. The forecast error mitigation strategies are summarized for each region in Figure 61 and Figure 62. In Figure 61, the average redispatch from day-ahead to real-time is shown over the ten hours with the largest day-ahead net load over-forecasts, as a fraction of the total day-ahead forecast error. The same is shown for day-ahead under-forecasts in Figure 62.

Figure 61. The average redispatch by resource type in response to the ten largest day-ahead net load over-forecasts, expressed as a percent of the total day-ahead net load forecast error.

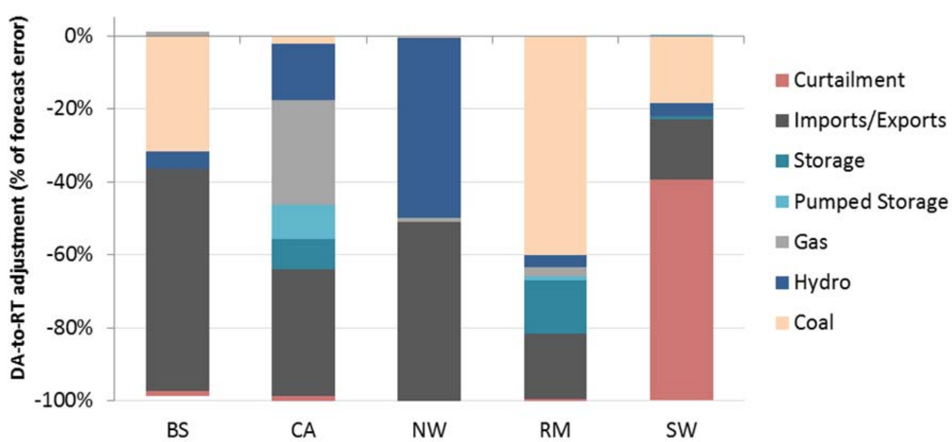
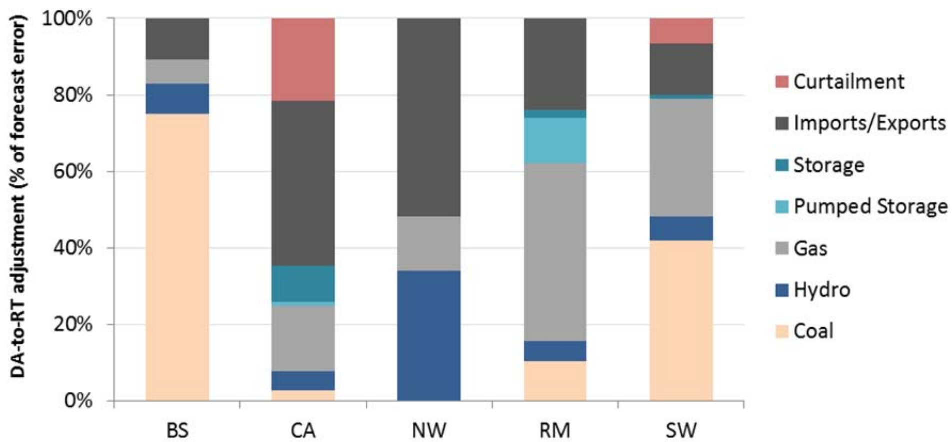


Figure 62. The average redispatch by resource type in response to the ten largest day-ahead net load under-forecasts, expressed as a percent of the total day-ahead net load forecast error.



Each region relies on a unique mix of resources to manage large forecast errors. Notably, all regions rely to some extent on deviations in imports and exports from day-ahead schedules to manage forecast errors. This finding highlights the fact that some degree of regional coordination (though not full coordination) is embedded within the Reference Grid. It is also consistent with a large body of research that suggests that forecast errors can be reduced by aggregating renewable resources over larger areas (i.e. when one region has over-forecasted its renewable output, another region may have under-forecasted its renewable output). In the Rocky Mountain region, where connections to the rest of the West are limited, over-forecasts are largely managed by dispatching coal resources down and under-forecasts are largely managed by dispatching natural gas resources up. With the exception of California, natural gas plant redispatch is relied upon more heavily for mitigating net load under-forecasts (in which it is dispatched upward relative to day-ahead schedules) than for mitigating over-forecasts (in which is it dispatched downward). This trend may be a reflection of the higher cost of dispatching gas resources and the limited flexibility of the coal fleet. Day-ahead schedules will tend to rely on coal to the extent possible to minimize costs, but coal resources cannot be committed in real-time. In the case of an under-forecast that exceeds the available headroom on the committed coal units, commitment of additional gas plants may be necessary to balance the load. In California and the Southwest, adjustments to renewable curtailment are also used to mitigate a portion of the forecast error in some of the hours with the largest forecast errors (curtailment is decreased in the case of a net load under-forecast and increased in the case of a net load over-forecast). This finding highlights the value of incorporating renewable curtailment into both day-ahead scheduling and real-time dispatch.

5.2.2 REGIONAL SUMMARIES

As the curtailment, ramping, and redispatch behavior suggest, the nature and magnitude of the challenges encountered vary from one region to the next, and depend on the characteristics of the region's load, the composition of its renewable portfolio, the composition of its non-renewable generation fleet, and the strength of its interties with neighboring regions. This section explores the nature of the integration challenges encountered in each region.

5.2.2.1 California

Table 30 summarizes the curtailment observed in the Common Case and High Renewables Case. Whereas in the Common Case, curtailment is rarely observed, it becomes a routine occurrence in the High Renewables Case, occurring in over

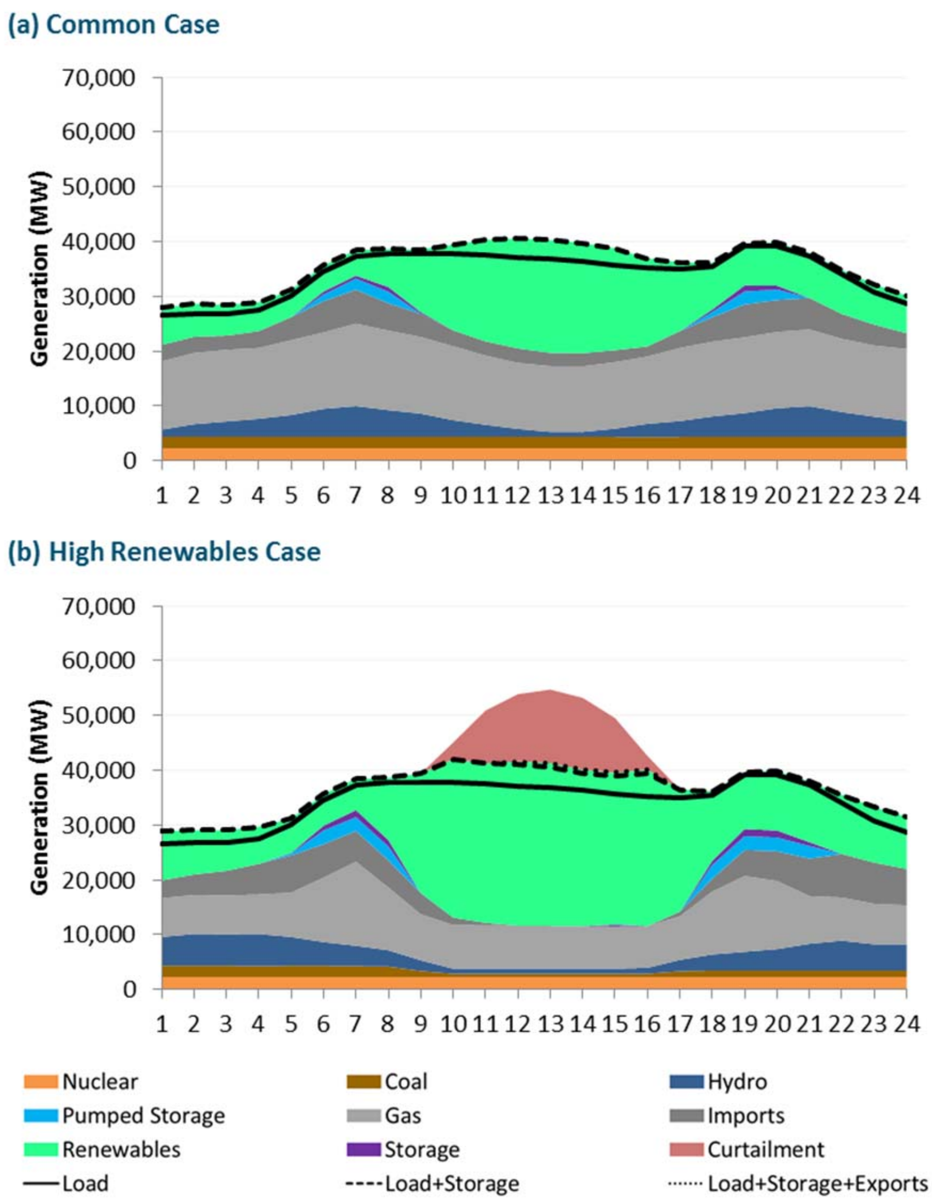
20% of the hours of the year and accounting for nearly 9% of the available renewable energy.

Table 30. Summary of curtailment observed in the Common & High Renewables Cases, California.

Type	Curtailment (% of renewable gen)		Frequency (% of hours)	
	Common	High Ren	Common	High Ren
Hourly	0.0%	8.6%	0.1%	21.2%
Subhourly	0.0%	0.0%	0.4%	25.9%
Total	0.0%	8.7%	0.5%	31.6%

Figure 63 summarizes the phenomenon that drives the need to curtail renewable generation, illustrating system operations for a typical spring day under both the Common Case and the High Renewables Case. In the Common Case, the strong diurnal output pattern of solar PV is apparent—renewable generation peaks at HE12—but is not so significant that the system cannot accommodate it; hydro, gas, and imports provide the necessary ramping capability to balance net load throughout the day. In the High Renewables Case, the output of renewables in the middle of the day increases significantly, such that even with all other resources operating at or close to minimum levels, the amount of renewable generation available exceeds what the system can accommodate. Without an alternative use for this surplus, this oversupply drives the need to curtail in order to achieve a balance between loads and resources.

Figure 63. Generation for average spring day, California.

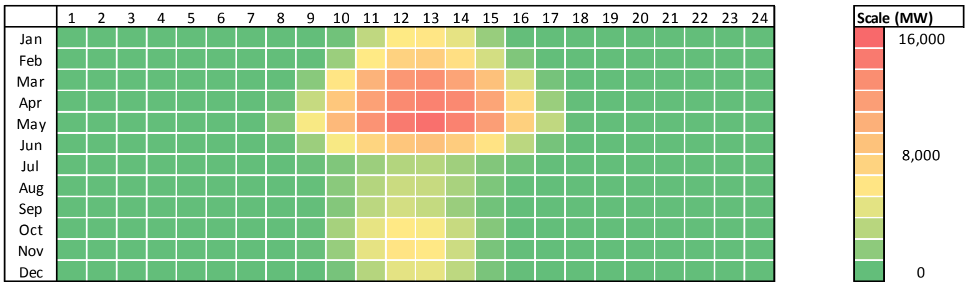


While Figure 63 summarizes operations for a single day, the dispatch exhibited here is indicative of regular diurnal patterns that are present in most days throughout the year at high renewable penetrations:

- Significant downward ramping capability is needed in the early morning hours to accommodate a net load ramp as solar output increases;
- Non-renewable generation resources reduce output to minimum levels during the middle of the day to maximize delivery of renewable generation;
- Any surplus daytime renewable generation that cannot be delivered to loads or exported during this period must be curtailed; and
- Significant upward ramping capability in the late afternoon to early evening hours is needed to meet a large upward net load ramp as solar production wanes.

Because the curtailment observed in California is driven primarily by the regular diurnal pattern of solar resources, it follows regular daily and seasonal patterns, as illustrated in Figure 64. Curtailment is generally observed in the daylight hours between HE09 and HE17 and is generally most significant between HE12 and HE 13. Seasonally, the spring-time months experience the largest amounts of curtailment, due to several factors: (1) relatively high-quality solar resources; (2) lower loads; and (3) higher hydroelectric system output due to spring runoff. In contrast, the summer months show the least curtailment, as California's summer peaking loads are large enough to absorb a significant share of the solar generation present during that season.

Figure 64. Seasonal and diurnal patterns of curtailment, High Renewables Case, California.



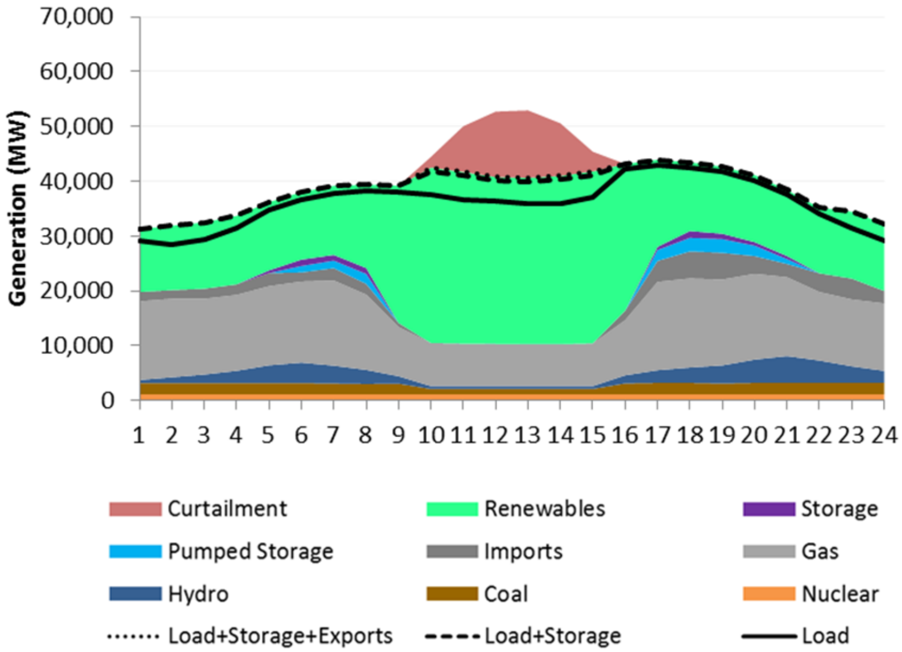
In this simulation, several assumed characteristics of the California system contribute to its limitations on downward flexibility, which exacerbate the amount of curtailment experienced. First, California's generation fleet includes a notable amount of "must-run" generation—generation whose flexibility to turn on and off in response to market signals is limited—including the Diablo Canyon Power Plant (DCPP) nuclear facility and a large fleet of non-dispatchable cogeneration resources (4,721 MW). Second, California is modeled with a minimum generation requirement that approximately 25% of loads must be met with qualifying thermal resources (generally dispatchable gas generation located in load pockets). Finally, the 'Reference Grid' assumptions limit California transmission flows to its historically observed range. Since California has historically been a major net importer of power from both the Northwest and the Southwest, its ability to export is modeled as limited.

While the predominant apparent challenge to the California fleet is the oversupply of renewables in the middle of the day, a secondary question is whether the fleet has adequate ramping capability to meet large diurnal net load ramps. The large net load ramps that occur in the early morning and late afternoon hours represent significant increases relative to historical experience.

In spite of this dramatic change, there is no indication of a shortage of ramping capability on the California system. The simulation results show these net load ramps being met primarily with a combination of gas resources, changes in net imports, and renewable curtailment.

Figure 65 presents an example of the California fleet operating to meet an extreme net load ramp—the largest three-hour net load ramp observed in the sample of draws analyzed. On this December day, from HE14 to HE17, the net load increases from -452 MW to 30,957 MW, an upward ramp of 31,409 MW over a three hour period. On this day, the available production from solar resources is high enough that renewable curtailment is required during mid-day to balance the load. The renewable curtailment softens the magnitude of the net load ramp by 9,622 MW; the remaining upward net load ramp is met by a combination of gas, imports, hydro & pumped storage.

Figure 65. Dispatch by technology type in California on the day with the largest upward net load ramp (a December day), High Renewables Case



This extreme day provides a useful illustration of the role that curtailment can play in meeting the system's needs for flexibility. While the primary cause of curtailment on this day is oversupply, the curtailment of renewable generation in the middle of the day softens the magnitude of the upward net load ramp as the sun sets dramatically, reducing its magnitude from over 31 GW to 21 GW across the three hours from HE14 to HE17. This type of anticipatory renewable curtailment can soften the magnitude of upward net load ramps and thereby reduce the need for upward ramping capability from traditional dispatchable resources; it is therefore a strategy that operators can employ to ensure the reliable service of loads even with significant hour-to-hour variability of renewable production.

To summarize, the main drivers of flexibility challenges in California in the High Renewables Case are:

- High solar oversupply;
- Must-run generation, including nuclear and cogeneration resources; and
- Minimum generation constraints to satisfy local reliability needs.

The impacts of these constraints are in part mitigated by:

- Renewable curtailment;
- Flexibility from the gas and hydro fleets;
- Energy storage resources, including existing pumped hydro and planned battery systems; and
- Exports to neighboring systems during hours of potential oversupply.

Even with these solutions in place, California faces significant flexibility challenges at 50% renewable penetration and may require additional solutions in order to meet renewable policy goals in the most cost effective manner. The impacts of additional solutions are discussed in Section 5.4.

5.2.2.2 Southwest

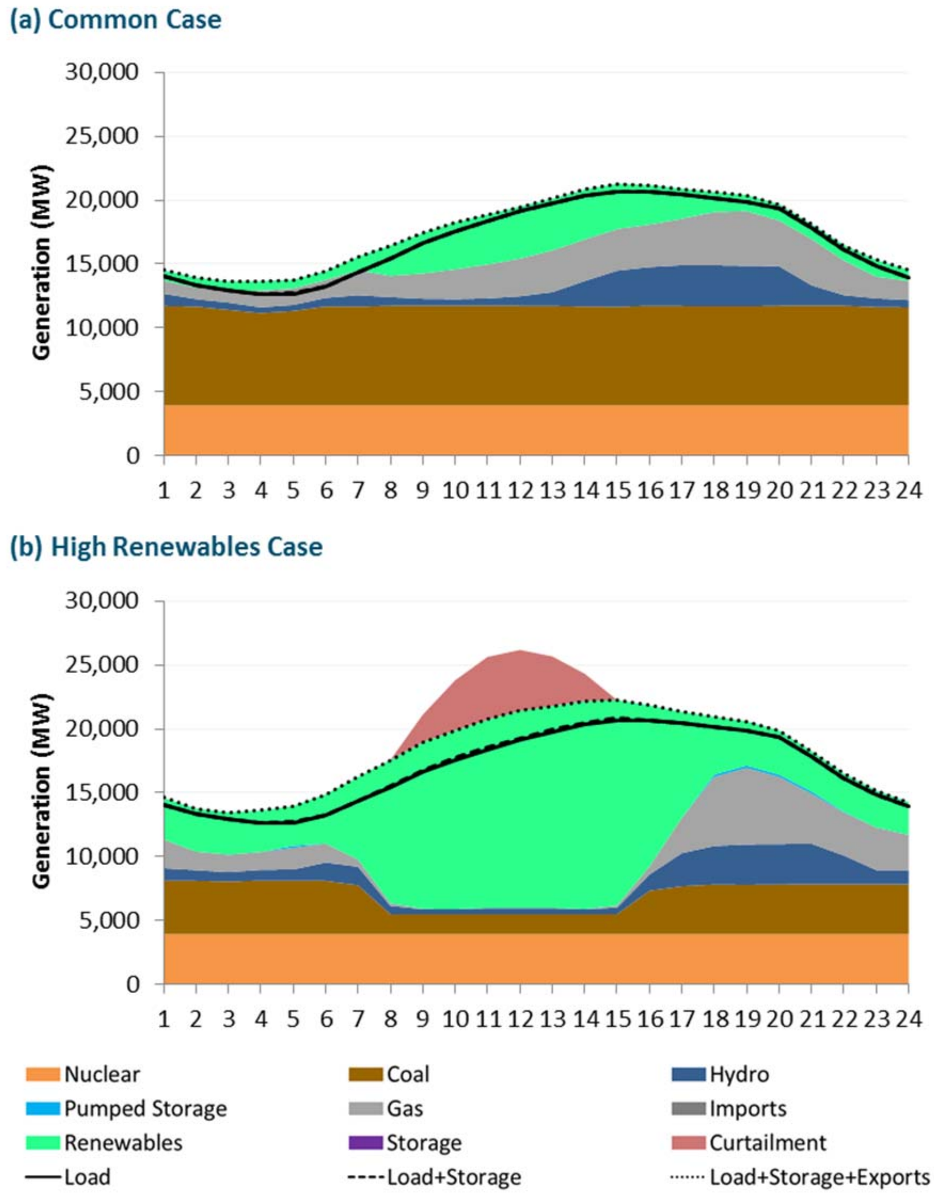
In the Southwest, renewable curtailment also plays a similarly significant role in operations in the High Renewables Case—summarized in Table 31—amounting to over 7% of available renewable energy and occurring at an hourly level in nearly 13% of hours.

Table 31. Summary of curtailment observed in the Common & High Renewables Cases, Southwest.

Type	Curtailment (% of renewable gen)		Frequency (% of hours)	
	Common	High Ren	Common	High Ren
Hourly	0.0%	6.4%	0.0%	12.9%
Subhourly	0.0%	0.9%	1.3%	39.0%
Total	0.0%	7.3%	1.3%	39.0%

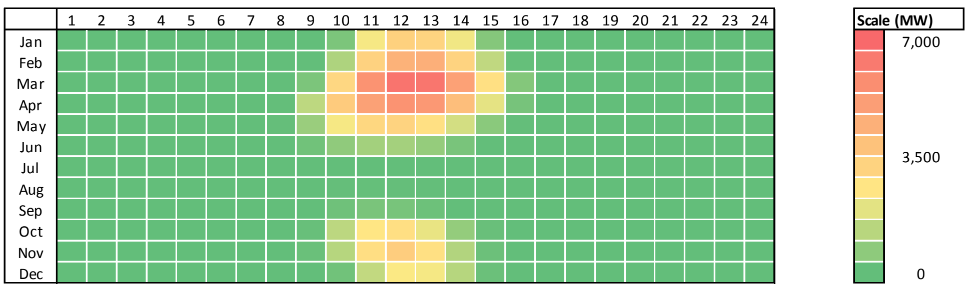
The nature of the oversupply challenge experienced in the Southwest, illustrated in Figure 66, is similar in many respects to that of California. Like California, the Southwest is endowed with high-quality, low cost solar PV resources whose output is concentrated in the middle of the day. Under current policy (the Common Case), the amount of renewables on the system is small enough that its impact on the net load shape is manageable (Figure 66a); at higher penetrations, extreme low net loads result in the need to curtail due to oversupply conditions (Figure 66b).

Figure 66. Generation for average spring day, Southwest.



Though the amount of curtailment observed in the Southwest is lower than in California, the seasonal and diurnal patterns of curtailment, shown in Figure 67, are strikingly similar: curtailment is a middle-of-the-day phenomenon whose volume is most concentrated in the low-load spring months and least concentrated in the peak load summer months.

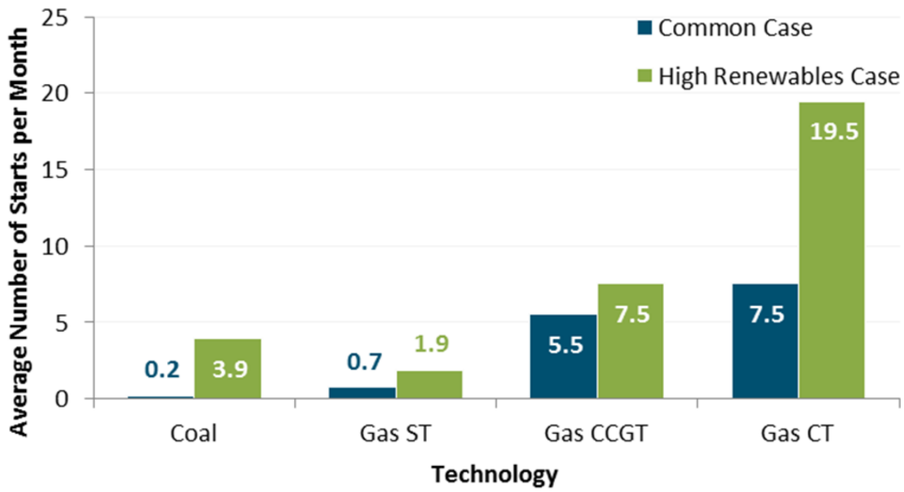
Figure 67. Seasonal and diurnal patterns of curtailment, High Renewables Case, Southwest.



Where the Southwest differs from California is in the composition of its non-renewable generation fleet, used to balance the net load. As shown in the dispatch plots of Figure 66, a significant share of the generation resources in the Southwest are nuclear and coal, which have traditionally operated in a baseload capacity (and continue to in the Common Case). The increase in solar PV penetration in the High Renewables Case exerts pressure on these resources to reduce output to accommodate solar PV.

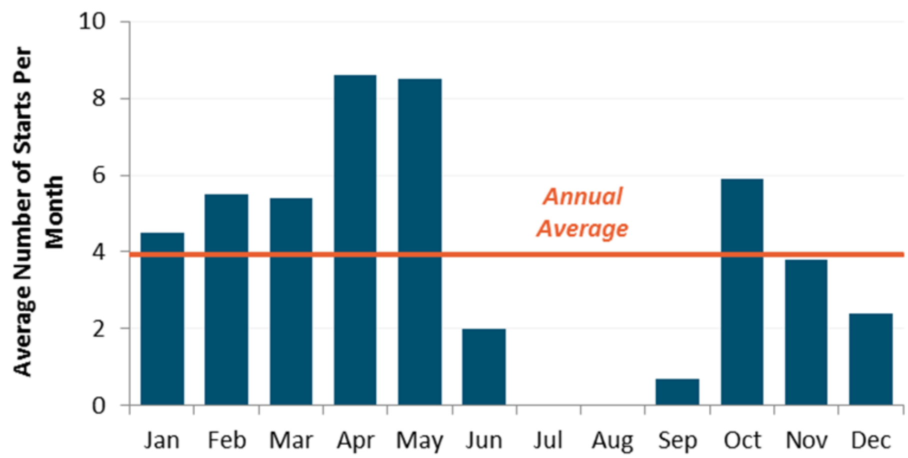
One of the notable impacts of the diurnal curtailment patterns observed in the Southwest is its impacts on the frequency of unit start-ups and shut-downs. In an effort to reduce the amount of thermal generating capacity online in the middle of the day, many units will turn off as the sun rises, only to start up once again to meet a steep net load ramp as the sun sets. This results in an increase in the frequency with which different types of generators cycle (i.e. fully start up and shut down) relative to the Common Case, as illustrated in Figure 68. The most notable changes are coal generators, which start up and shut down fewer than once per month in the Common Case but nearly four times per month in the High Renewables Case; and combustion turbines, whose frequency of starts and shutdowns increases from 7.5 to 19.5 times per month.

Figure 68. Average start-ups and shutdowns for thermal resources, Southwest



The changes implied in the operations of the coal fleet in the Southwest are significant. The implied number of starts in each month for the units in the Southwest coal fleet is shown in Figure 69. Particularly during the spring months of April and May when curtailment is most prevalent, coal units start an average of nearly nine times per month, or once every three to four days. The operations of the Southwest coal fleet with this amount of flexibility implies a significant increase in cycling relative to how coal plants have operated historically and may result in increased cycling costs; at the same time, this flexibility provides benefits by enabling delivery of incremental renewable generation. The value of operating coal with such flexibility—and, alternatively, the consequences of operating the fleet in a more traditional "baseload' capacity—is explored through sensitivity analysis presented in Section 0.

Figure 69. Average number of starts per month, Southwest coal units, High Renewables Case



The flexibility challenges identified for the Southwest region are similar to those identified for California. They include:

- High solar oversupply; and
- Must-run generation, like nuclear resources.

Mitigation solutions available to the system in the Reference Grid include:

- Renewable curtailment;
- Flexibility from the coal fleet (if coal units are allowed to cycle);
- Flexibility from the gas fleet;
- Exports to neighboring systems; and

+ Subhourly curtailment to avoid overcommitment of thermal units.

Despite these solutions, the Southwest has significant integration challenges with the renewable portfolio assumed in the High Renewables Case, even when frequent coal cycling is allowed to occur. These challenges are exacerbated when coal flexibility is further constrained (as is discussed in Section 5.3.1). In addition to pursuing more flexible coal operations, the Southwest may benefit from additional renewable integration solutions, like increased regional coordination and/or energy storage, as renewable penetrations increase in the region.

5.2.2.3 Northwest

Like the Southwest and California, the Northwest experiences limited renewable curtailment in the Common Case but a significant quantity in the High Renewables Case: renewable curtailment is observed in 10% of hours and accounts for nearly 6% of available renewable energy, as summarized in Table 32.

Table 32. Summary of curtailment observed in the Common & High Renewables Cases, Northwest.

Type	Curtailment (% of renewable gen)		Frequency (% of hours)	
	Common	High Ren	Common	High Ren
Hourly	0.1%	5.6%	0.1%	9.9%
Subhourly	0.0%	0.0%	0.0%	1.1%
Total	0.1%	5.6%	0.2%	11.0%

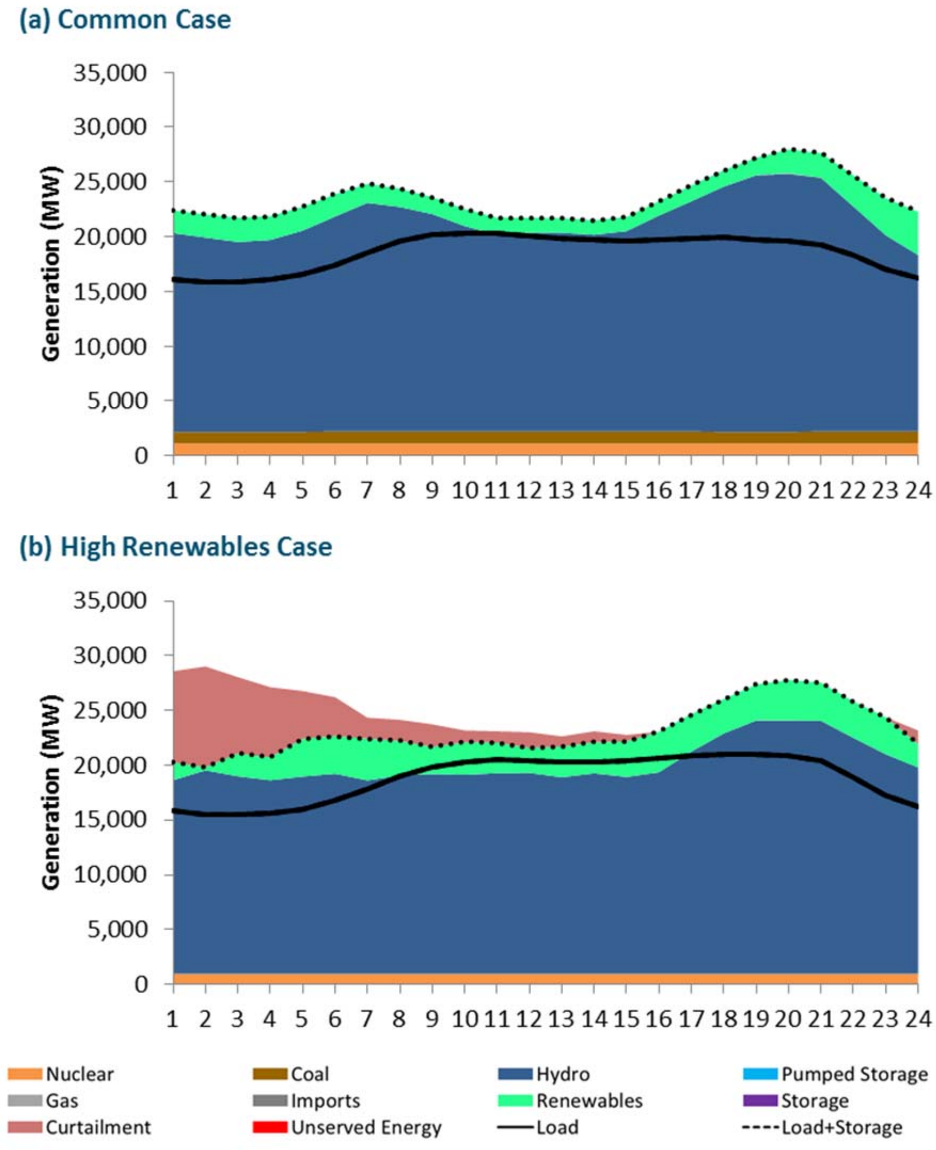
The nature of the oversupply challenge that leads to curtailment in the Northwest is unique among the regions of the Western Interconnection: whereas oversupply in California and the Southwest is driven by the concentration of solar PV in the middle of the day, oversupply in the Northwest is a result of the combination of high hydroelectric output and wind generation during the spring runoff months. In this respect, the oversupply challenge in the Northwest is not new—the Northwest has historically experienced periods where the capability of the hydro system has exceeded local loads, resulting in either off-system sales to California and other markets or the need to spill hydro—however, it is intensified under higher renewable penetrations.

At the same time, the Northwest experiences many fewer hours in which subhourly curtailment is significant; just 1.1% of hours under the High Renewables case, as compared to 39% of hours for the Southwest and 26% of hours for California, despite the fact that hourly curtailment is similar. This speaks to the vast capability of the region's hydroelectric system to meet subhourly renewable integration needs.

Figure 70 illustrates this challenge for a spring day under average hydroelectric conditions. In the Common Case, the combination of the hydro system output and renewable generation exceeds the load in the Northwest, but the region's export capability allows it to export between 2,000 and 10,000 MW over the course of the day. In the High Renewables Case, the increase in wind output—especially during off-peak hours—yields a surplus of generation that cannot be used locally or exported,[45] ultimately resulting in renewable curtailment.

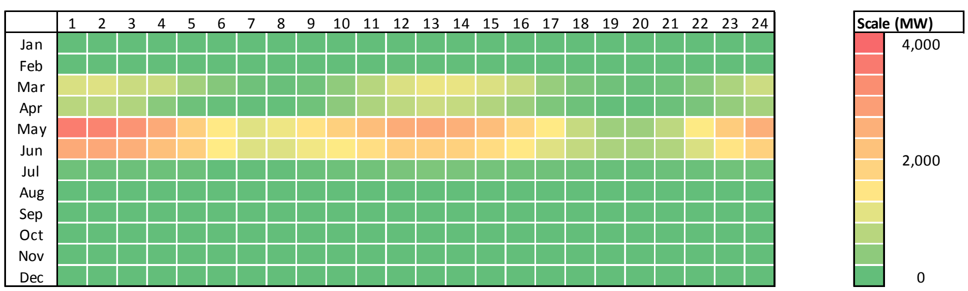
[45] With the addition of renewable generation in other regions of the Western Interconnection, demand for surplus exports from the Northwest is reduced as well, a secondary driver of the increase in curtailment.

Figure 70. Generation for average May day, Northwest



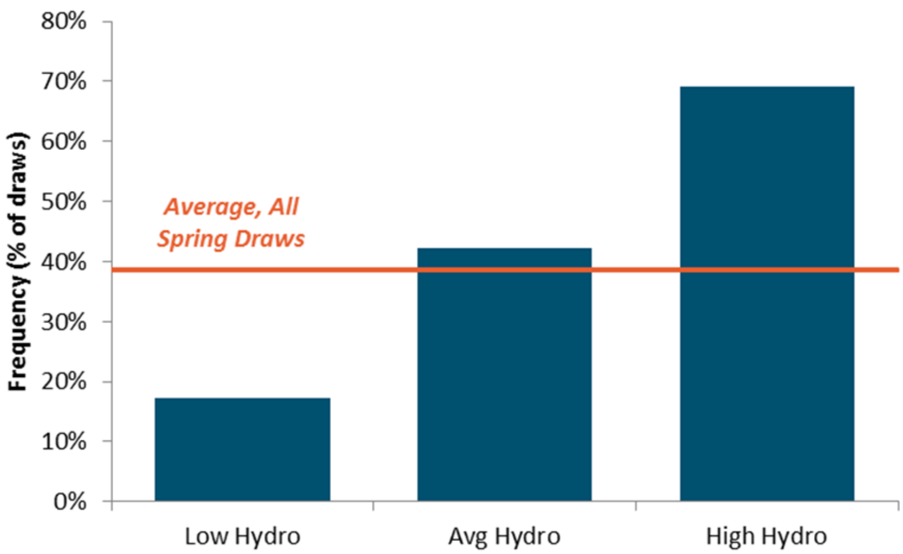
As the oversupply phenomenon in the Northwest has different underlying drivers from California and the Southwest, the seasonal and diurnal patterns of renewable curtailment observed in the Northwest are distinctive as well, as shown in Figure 71. Renewable curtailment is limited almost exclusively to the spring runoff period from March through June, and unlike in regions with high solar penetrations, is observed at all times of the day. The periods in which curtailment is most heavily concentrated are generally in the middle of the night (when loads are relatively low) and in the middle of the day (when the Northwest's ability to export its own surplus to other regions is limited by oversupply in neighboring regions).

Figure 71. Seasonal and diurnal patterns of curtailment, High Renewables Case, Northwest.



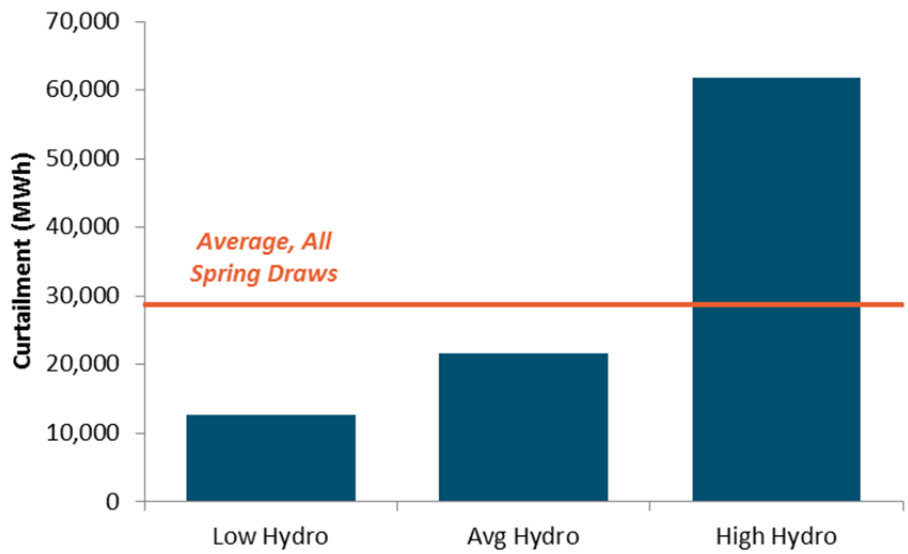
In the Western Interconnection, the Northwest region is unique in its reliance on hydroelectric generation to serve its load. The characteristics of the Northwest hydro system—its seasonal and year-to-year variability—have important implications for renewable integration in the Pacific Northwest. To examine the connection between hydro conditions and renewable curtailment in the Northwest, the draws of the spring months—March through June, when curtailment is most prevalent—are segmented according to their hydro conditions. The third of draws with the highest hydro energy in each month are classified as "wet"; the third with the lowest hydro energy in each month are "dry"; and the remaining middle third are "average." Figure 72 shows the frequency of curtailment under each of those conditions: whereas on average approximately 40% of days exhibited curtailment across all draws, curtailment was experienced in less than 20% of dry hydro draws while nearly 70% of wet hydro draws.

Figure 72. Frequency of curtailment under low, average, and high hydro conditions, Northwest, High Renewables Case (March – June)



The concentration of curtailment during relatively high hydro conditions is even clearer when the amount of curtailment is considered, as shown in Figure 73. The amount of curtailment experienced in high hydro conditions (approximately 60,000 MWh per day) is nearly three times larger than the amount that is experienced in average hydro conditions (approximately 20,000 MWh per day).

Figure 73. Average curtailment observed in spring under low, average, and high hydro conditions, Northwest, High Renewables Case (March – June)



The asymmetry of the relationship between hydro conditions and renewable curtailment has important implications for renewable integration in the Northwest. As shown in Figure 73, the amount of curtailment experienced across all conditions is considerably larger than the amount that would be expected under average conditions, skewed by the significant realization of curtailment that may occur during wet hydro years. This serves as an instructive example of the importance of considering the full range of conditions on a system in resource planning exercises and procurement decisions; failure to capture the breadth of hydro conditions in the Northwest could significantly understate the long-run expectation of curtailment.

To summarize, the operational behavior in the Northwest at higher renewable penetrations is unique from other regions in that it is dominated by the size and variability of the hydro resource. Long duration oversupply conditions arise due to the coincidence of high hydro and wind resource availability. However, the hydro resource also provides a high degree of hour-to-hour flexibility, which largely eliminates renewable integration challenges related to ramping and forecast error. Whether the Northwest can make such efficient use of the hydro fleet toward these ends, however, remains an open question, as this study assumes perfect coordination and no transmission constraints within the region. The Northwest also relies heavily on exports to neighboring systems during nighttime hours to alleviate or lessen oversupply, so the nature of the interactions between balancing areas (both within the Northwest and between the Northwest and other regions) will be critical in mitigating or exacerbating renewable integration challenges in the future.

5.2.2.4 Basin

While California, the Southwest, and the Northwest see considerable challenges in the High Renewables Case, the Basin region experiences fewer curtailment events. Curtailment observed in the Common Case and High Renewables Case in the Basin is summarized in Table 33. While some flexibility challenges are evidenced by the observation of small amounts of curtailment in the High Renewables Case, it remains a relatively rare phenomenon even at the 40% renewable penetration examined in this study.

However, it is notable that the system relies on subhourly curtailment of renewables as a subhourly balancing strategy in 41% of hours, even though curtailment occurs much less frequently. While the total magnitude of this subhourly curtailment is small, this is an important strategy in managing flexibility needs at the subhourly level.

Table 33. Summary of curtailment observed in the Common & High Renewables Cases, Basin.

Type	Curtailment (% of renewable gen)		Frequency (% of hours)	
	Common	High Ren	Common	High Ren
Hourly	0.0%	0.1%	0.0%	1.1%
Subhourly	0.0%	0.3%	4.1%	41.3%
Total	0.0%	0.4%	4.1%	41.3%

Figure 74 shows a typical spring day for the Basin in the Common Case and the High Renewables Case. In the Common Case, most load is served by coal generation that operates; renewable generation—predominantly geothermal and wind—produces at a relatively stable level throughout the day; and the region is a net exporter. In the High Renewables Case, a strong diurnal shape to renewable production appears with the addition of solar PV, causing midday cycling of coal resources. While the effects of an increased share of solar PV on diurnal operational patterns are evident, the effects are not so large as to require curtailment as is observed in the Southwest.

Much like the Southwest, the Basin region relies predominantly on coal and gas generation to balance net load throughout the year; in contrast, whereas the Southwest experiences significant curtailment at a 40% renewable penetration, very little is observed in the Basin at this same level. The primary factor that distinguishes these two regions is the diversity of the renewable portfolio: the Southwest's portfolio comprises predominantly solar (70%) and a smaller share of wind (30%), while the Basin's portfolio includes a mix of geothermal (25%), solar (30%), and wind (45%). The technological diversity inherent in this portfolio—particularly the limited penetration of solar PV resources—distributes renewable production more evenly across the hours of the year, which in turn means that the Basin is less likely to experience oversupply events that often lead to curtailment in regions whose reliance on solar PV is greater.

Figure 74. Generation for an average spring day, Basin

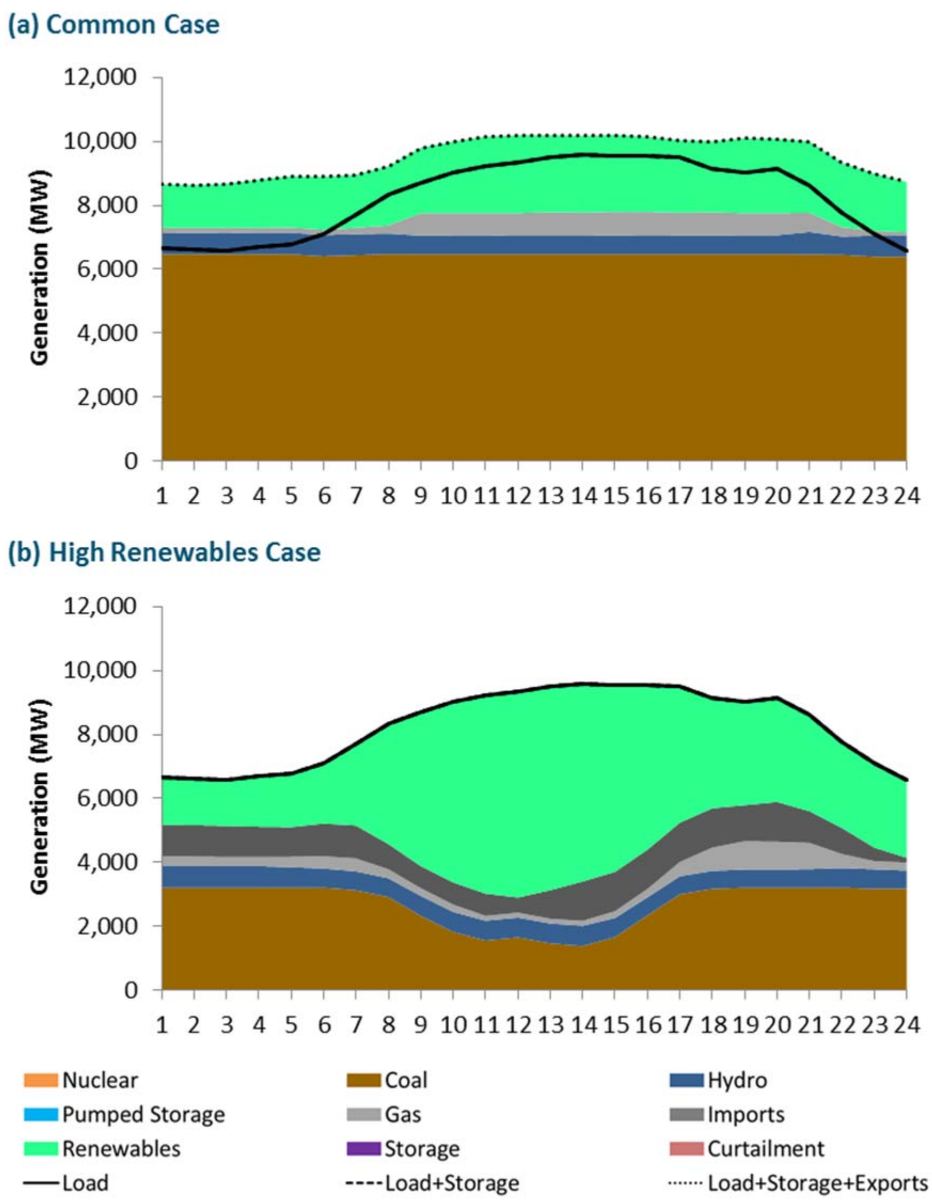
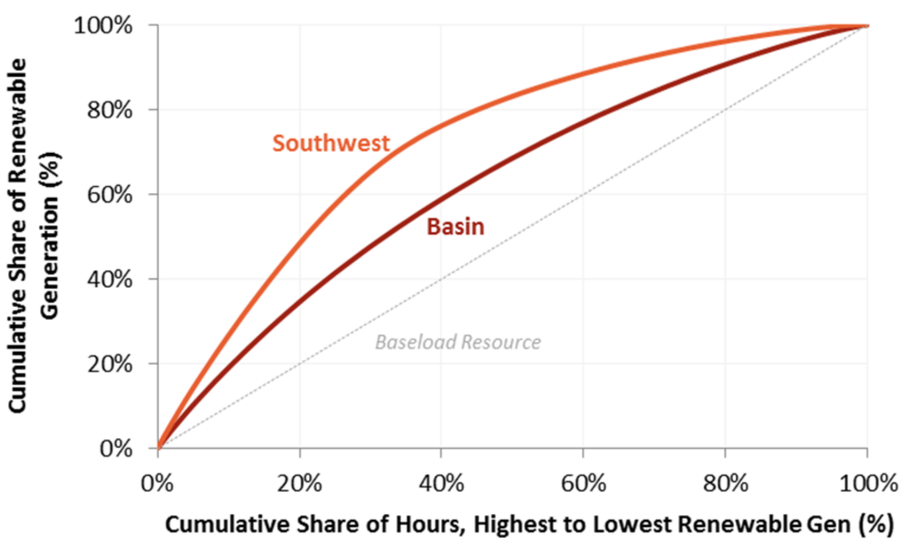


Figure 75 contrasts the portfolios of the Basin and Southwest region to highlight the effect of portfolio diversity. The effect of a higher penetration of solar PV—a resource whose production is physically limited to the daylight hours—is that renewable production is heavily concentrated in a limited number of hours. In the Southwest, the 20% of hours with the highest renewable production represent 48% of annual renewable generation; in the Basin, the 20% of hours with the highest renewable production represent only 35% of annual renewable generation.

Figure 75. Distribution of renewable generation throughout the year, Southwest and Basin, High Renewables Case



The contrast between the Basin and the Southwest—regions with similar reliance on a mix of coal and gas—serves to highlight the value of renewable portfolio diversity in achieving high penetration renewable goals. A lack of portfolio diversity creates a particularly significant challenge when the resource relied upon is solar PV, as opportunities for diversification are limited by the confinement of production to the daylight hours.

Just as the coal resources in the Southwest may be called upon to operate with more flexibility than historically under a high renewables future increased operational flexibility from the coal generators in the Basin may be needed to facilitate renewable integration. The observed frequency of unit start-ups and shut-downs (4.9 per unit per month) is similar to the Southwest. This observed flexibility is certainly one of the reasons that curtailment is observed so rarely in the High Renewables Case; the subject of coal flexibility and its implications for renewable integration are further explored through sensitivity analysis in Section5.3.1.

The renewable integration challenges (or lack thereof) identified in the Basin are largely a product of the modeling decisions made in this analysis. Three key integration strategies can be identified that enable the Basin region to integrate 40% renewables with limited curtailment:

- Renewable portfolio diversity;
- Flexibility from the coal fleet (if coal units are allowed to cycle);
- Subhourly curtailment to avoid overcommitment of thermal resources;
- Exports to neighboring regions; and
- To a lesser degree, flexibility from the gas and hydro fleets.

Both the portfolio diversity and the operational abilities of the coal fleet were selected to investigate a possible future in which some procurement and operational strategies are pursued to support renewable integration. However, the costs of these solutions as well as institutional barriers to implementation may cause additional challenges in any region, including the Basin. For example, the renewable portfolio diversity will depend on the relative costs of geothermal, wind, and solar resources and the coal operational flexibility will depend on both operator decisions and the costs of added wear and tear on coal plants. These types of economics questions will be central to questions around renewable integration in regions with coal resources and diverse renewable resource potential, like the Basin.

5.2.2.5 Rocky Mountains

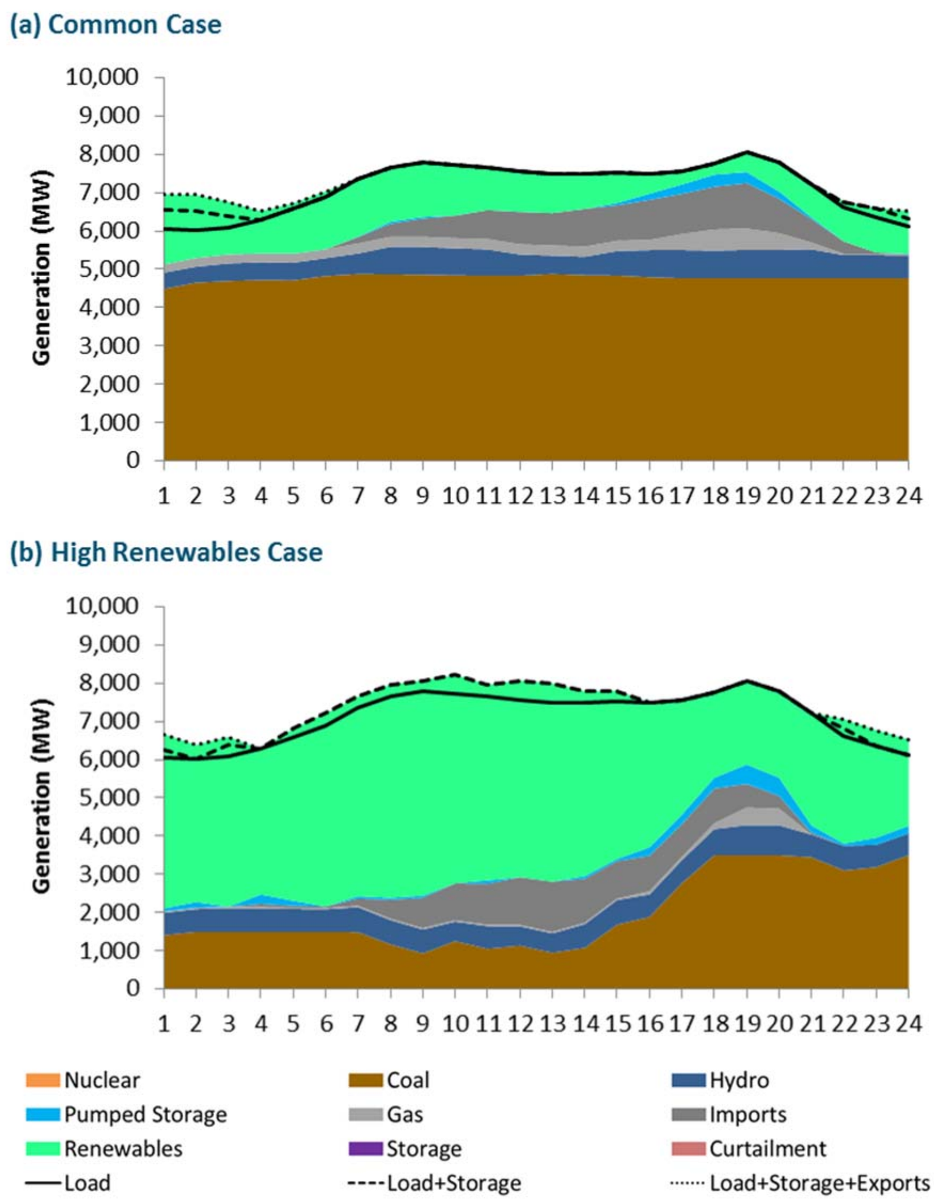
Like the Basin, renewable curtailment in the Rocky Mountain region is observed in very small quantities—even in the High Renewables Case. The region's curtailment statistics, shown in Table 34, highlight this fact. Also like the Basin region, subhourly curtailment is a key integration strategy, occurring in over half of hours under the High Renewable case despite very low hourly curtailment.

Table 34. Summary of curtailment observed in the Common & High Renewables Cases, Rocky Mountains.

Type	Curtailment (% of renewable gen)		Frequency (% of hours)	
	Common	High Ren	Common	High Ren
Hourly	0.0%	0.1%	0.0%	1.0%
Subhourly	0.1%	0.5%	19.9%	52.2%
Total	0.1%	0.6%	19.9%	52.2%

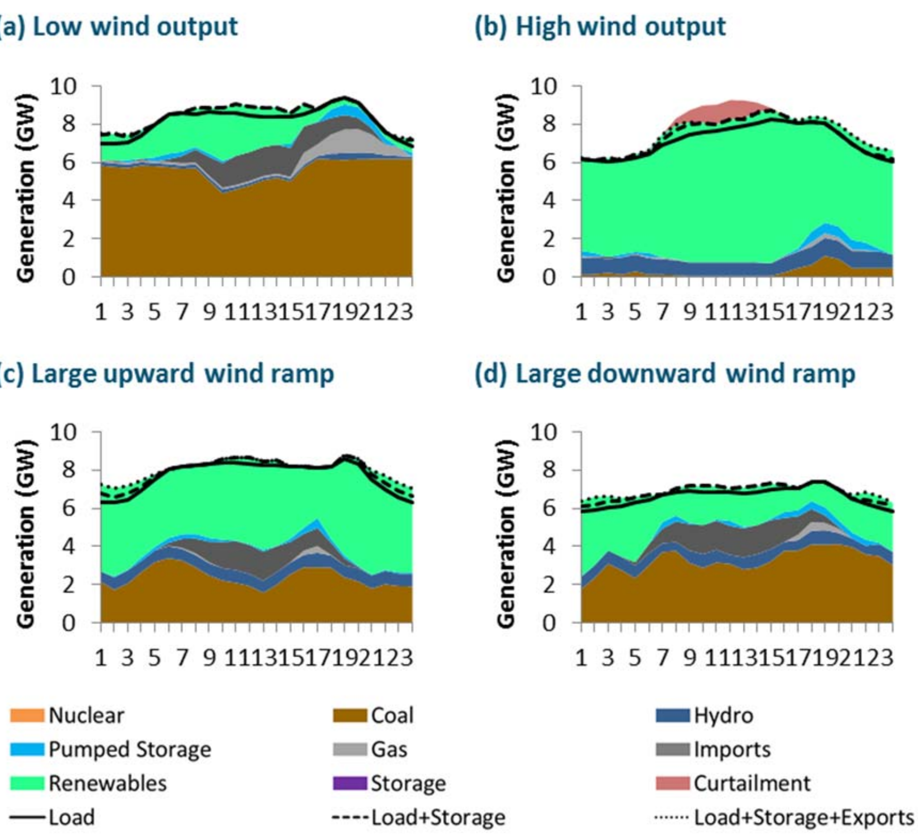
An indicative day of operations in the spring season is shown in Figure 76 for the Common Case and High Renewables Case. In the Common Case, output of renewable generation—predominantly wind—is limited; a large portion of the region's coal fleet is committed and runs at full capacity throughout the day; and the hourly net load signal is balanced with a combination of gas and hydro resources and scheduled imports. In the High Renewables Case, the increased output of the renewable fleet—still predominantly wind—results in significantly fewer coal resources committed and online through the simulation. Those coal resources that are online operate more flexibly in this scenario than in the Common Case and contribute to meeting the large upward ramp in net load observed between HE13 and HE19.

Figure 76. Generation for an average spring day, Rocky Mountains



Like in other regions with high penetrations of wind generation, choosing a single day as indicative of how the system operates—even during a specific season—is challenging, as the output profile for wind generation varies considerably from one day to the next. Whereas regions such as California and the Southwest experience relatively regular diurnal patterns due to high solar penetrations, day-to-day operations in the Rocky Mountains appear considerably different with changes in renewable output within individual seasons. Figure 77 presents four different day "types" for the Rockies in the spring: (1) low wind output, (2) high wind output, (3) large upward wind ramp, and (4) large downward wind ramp.

Figure 77. Variations in net load conditions observed during spring season, Rocky Mountains, High Renewables Case.



As evidenced by Figure 77b, the Rocky Mountains does experience some days where extreme low net load conditions present flexibility challenges to the system. Four strategies are used to accommodate this extremely high renewable penetration on this day:

+ Thermal units are cycled, with some coal units ramping down to their minimum generation levels and other coal units as well as gas units shutting down for the daylight hours in which solar output stresses the system.
+ Pumped hydro storage units pump during daylight hours to store excess renewable energy and discharge during early morning and evening hours to help balance the net load.
+ Excess wind power is exported to other regions during early morning and late evening hours; exports are notably not utilized during daylight hours, as other regions are also in curtailment conditions during those hours.
+ Excess renewable power is curtailed during midday to avoid imbalance (overgeneration) on the system.

Some coal units in the Rocky Mountain region are kept on throughout the day, even during curtailment hours, in order to provide upward reserves in case of forecast error or subhourly imbalance.

While the amount of curtailment shown in the Rocky Mountain region is low when compared with other regions, it nonetheless plays an important role in the region's day-to-day operations under a high renewables future—in particular, subhourly curtailment is observed in 52% of hours in the High Renewables Case. Unlike regions that experience large curtailment events in hourly dispatch that reflect an oversupply of energy, the subhourly curtailment observed in the Rocky Mountains reflects the use of renewables to meet the system's need for downward flexibility reserves: in the simulation, renewables are actively dispatched within the hour to meet the five-minute variable signal of net load. While this represents a relatively small amount of energy (0.5% of renewable generation annually), it provides an important benefit to the system by allowing the system to avoid committing thermal resources to meet the needs for downward flexibility reserves. The role of subhourly curtailment and the benefits that enabling this degree of renewable participation are further explored in Section 5.3.3.

To summarize, the Rocky Mountain region faces integration challenges in the High Renewables Case related largely to the predominance of wind in the renewable portfolio, which carries with it large forecast errors and high subhourly variability. However, the Rocky Mountain region is able to mitigate much of the renewable integration challenges through:

+ Flexibility from the coal fleet (if coal units are allowed to cycle);
+ Flexibility from pumped storage, gas and hydro resources; and
+ Subhourly renewable curtailment to avoid overcommitment of thermal resources.

5.2.3 COMMON OBSERVATIONS

While the nature of operations and the constraints on flexibility encountered vary from one region to the next, a number of observations and findings of this analysis are cross-cutting:

+ **The comparatively low penetrations of renewable generation observed in the Common Case do not place severe stress on system flexibility.** Coal plants across the West continue to run in a relatively baseload manner; gas and hydro facilities are used to meet net load ramps; and nearly all available renewable generation is delivered to load without any reliability events observed. Because the constraints on flexibility are observed rarely in the Common Case, the remainder of this study's investigation into the flexibility of the Western systems focuses on the High Renewables Case.
+ **Curtailment serves as the relief valve for operational challenges under the High Renewables Case.** Curtailment is primarily observed during periods of oversupply, but its role in enabling high penetrations of renewable generation is more nuanced: in addition to allowing each system to maintain a balance between loads and generation during periods of oversupply, it softens the magnitudes of net load ramps and provides operators with a flexible means of responding to forecast error when a system's dispatchable resources are limited in flexibility. Subhourly renewable curtailment reduces or eliminates the reliance on conventional resources for providing downward flexibility within the hour, allowing the system to operate at more efficient set points and with less total curtailment.
+ **In the coal-heavy regions of the Western Interconnection, growing penetrations of renewable generation will exert pressure on the coal fleet to operate more flexibly than it has historically.** Whether this is technically possible—and whether operators choose to adjust operations as shown in these simulations—remains to be seen; the consequences of not operating coal with the amount of flexibility assumed in these cases and instead continuing to operate coal in a baseload capacity is explored in sensitivity analysis in Section 0.

+ **No region experiences a shortage of upward or downward ramping capability in the simulation.** Much of the original motivation for examining operational flexibility centered around the question of whether the thermal fleet in various parts of the West would be capable of meeting extreme upward net load ramps (e.g. could California's gas fleet meet the upward ramp required of it by the duck chart?). Across the West, no shortage of ramping capability is experienced in either the Common or the High Renewables Cases. In many instances, the magnitude of net load ramps is softened by renewable curtailment in low net load conditions.

The subsequent sections of this investigation explore sensitivity analyses of the High Renewables Case (Section 5.3) and potential strategies to mitigate flexibility challenges across the Western Interconnection (Section 5.4).

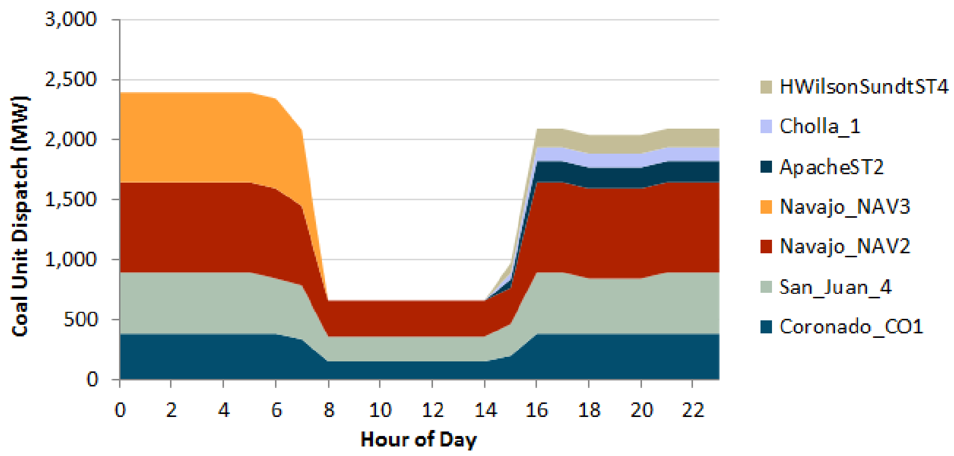
5.3 Sensitivity Analysis of High Renewables Case

Three sensitivities were tested in order to understand the impacts of specific cost and operating assumptions made in this analysis. The sensitivities focus on: (1) the ability of coal units to start and stop within a model day; (2) the cost penalty placed on renewable curtailment; and (3) the treatment of downward load following reserve requirements and subhourly curtailment.

5.3.1 COAL FLEXIBILITY

As a base assumption, coal units throughout the West are allowed to start up and shut down within each day to help balance the net load. Because the simulation horizon for each draw is a three-day period, minimum up time and minimum down time constraints that span the beginning or the end of the three-day period are not enforced in the simulation. As has been discussed in prior sections, this relaxation results in a high degree of flexibility from the coal fleet in each region under Reference Grid assumptions. An example of the coal cycling behavior observed due to this relaxation is shown in Figure 78.

Figure 78. Cycling behavior of coal units in the Southwest, High Renewables Case.



On this April day, three of the coal units in the Southwest (Coronago_CO1, San_Juan_4, and Navajo_NAV2) are committed for the full day and ramp down to their minimum stable levels during daylight hours to accommodate solar power. Four additional coal units are cycled throughout the day to help meet nighttime loads while avoiding daytime renewable curtailment. One unit (Navajo_NAV3) begins the day committed in order to help meet early morning loads. This unit shuts down at 8am to avoid renewable curtailment, but its minimum down time restricts it from turning back on at night. Instead, three other units (ApacheST2, Cholla_1, and HWilsonSundtST4) are turned on at 3pm to help meeting evening loads. While this alternating cycling behavior is observed on the lowest net load days in high solar regions, coal units spend most days in the simulation either committed on or off for the entirety of the day. The coal dispatch is summarized across all the draws for the Reference Grid High Renewables simulation in Table 35.

Table 35. Summary statistics for coal fleet in the High Renewables Case under Reference Grid assumptions.

Attribute	Basin	California	Northwest	Rockies	Southwest
Total Capacity (MW)	7,570	2,044	3,279	6,960	6,808
Capacity Factor (%)	56%	62%	54%	57%	77%
Starts (Avg # per unit-yr)	**59.3**	**26.9**	**61.1**	**53.0**	**47.2**

The high renewable penetration notably depresses average coal capacity factors due both to the cycling behavior described above and a greater number of days with enough renewable energy to allow the coal units to shut off entirely. On average, coal units across the West cycle about 25-60 times per year, with the most cycling occurring in spring months – coal units in the Southwest average between 8 and 9 cycles per month in the months of April and May and coal units in the Northwest average between 6 and 9 cycles per month in January through March.[46]

[46] The day draw methodology prevents direct calculation of the number of starts that occur across sequential days. The average number of starts presented in this report combines starts observed within simulated days and the expected number of starts occurring between days assuming that the modeled days within each month are randomly ordered. Given this assumption, the expected number of inter-day starts is a function of the number of days a unit is on in a month and the number of total days in the month. To the extent that high net load days or low net load days are clustered within months due to multi-day weather patterns, the system may experience fewer starts on average than are reported here.

To put the unit starts into context, the typical minimum up and down times for coal units in the Common Case are 168 and 48 hours, respectively. If coal units were to cycle at this maximum rate, they would experience approximately 3.4 cycles per month. The cycling observed under base assumptions is in alignment with this figure is some regions and some months, but as is shown in Table 36, average cycling in most months exceeds this frequency and in some months more than doubles this frequency.

Table 36. Average number of starts (average number per unit per month) for each month of the year.

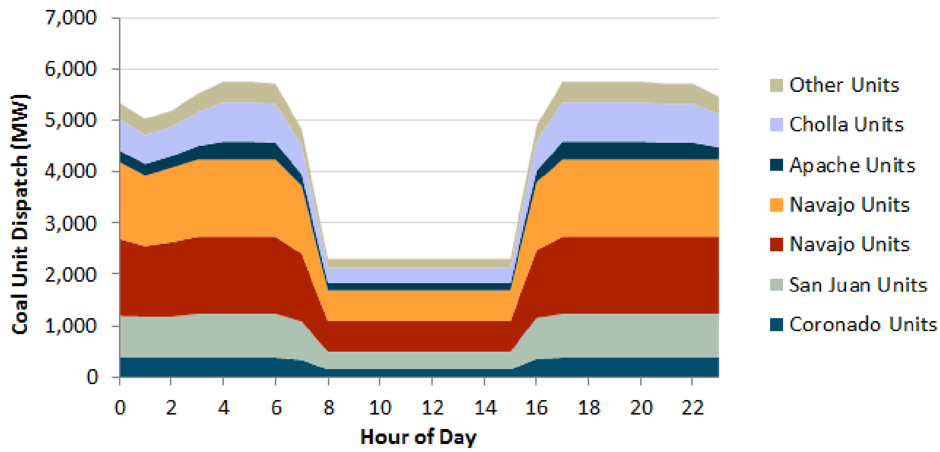
Scenario	Basin	California	Northwest	Rockies	Southwest
January	5.6	3.0	7.8	5.9	4.5
February	5.0	1.9	6.0	4.8	5.5
March	7.4	2.8	9.0	3.9	5.4
April	7.3	1.8	7.7	3.2	8.6
May	7.1	0.7	2.0	4.3	8.5
June	6.2	1.7	4.0	4.1	2.0
July	2.9	3.0	8.0	3.4	0.0
August	2.5	3.0	4.5	4.5	0.0
September	2.3	2.3	2.5	4.5	0.7
October	3.8	2.6	1.2	4.1	5.9
November	3.8	1.7	3.0	5.4	3.8
December	5.5	2.5	5.5	4.7	2.4

The frequency of cycling observed in the High Renewables case is much higher than is experienced today and would fail to meet the minimum up and down time constraints for coal units in the Common Case database if they were enforced across days. However, coal flexibility and the technical and economic feasibility of increased cycling of coal plants remain subjects of ongoing investigation. The findings in this analysis support the need for additional investigation in this area.

In order to investigate the impact of the high levels of coal flexibility that are assumed in the Reference Grid, an additional sensitivity on coal unit startup and shutdown costs was performed. In this sensitivity, coal unit startup and shutdown costs were set to $1 billion/MWh to prevent coal units from changing their commitment within each three-day period, bringing the coal commitment behavior into closer alignment with the minimum up and down time constraints.

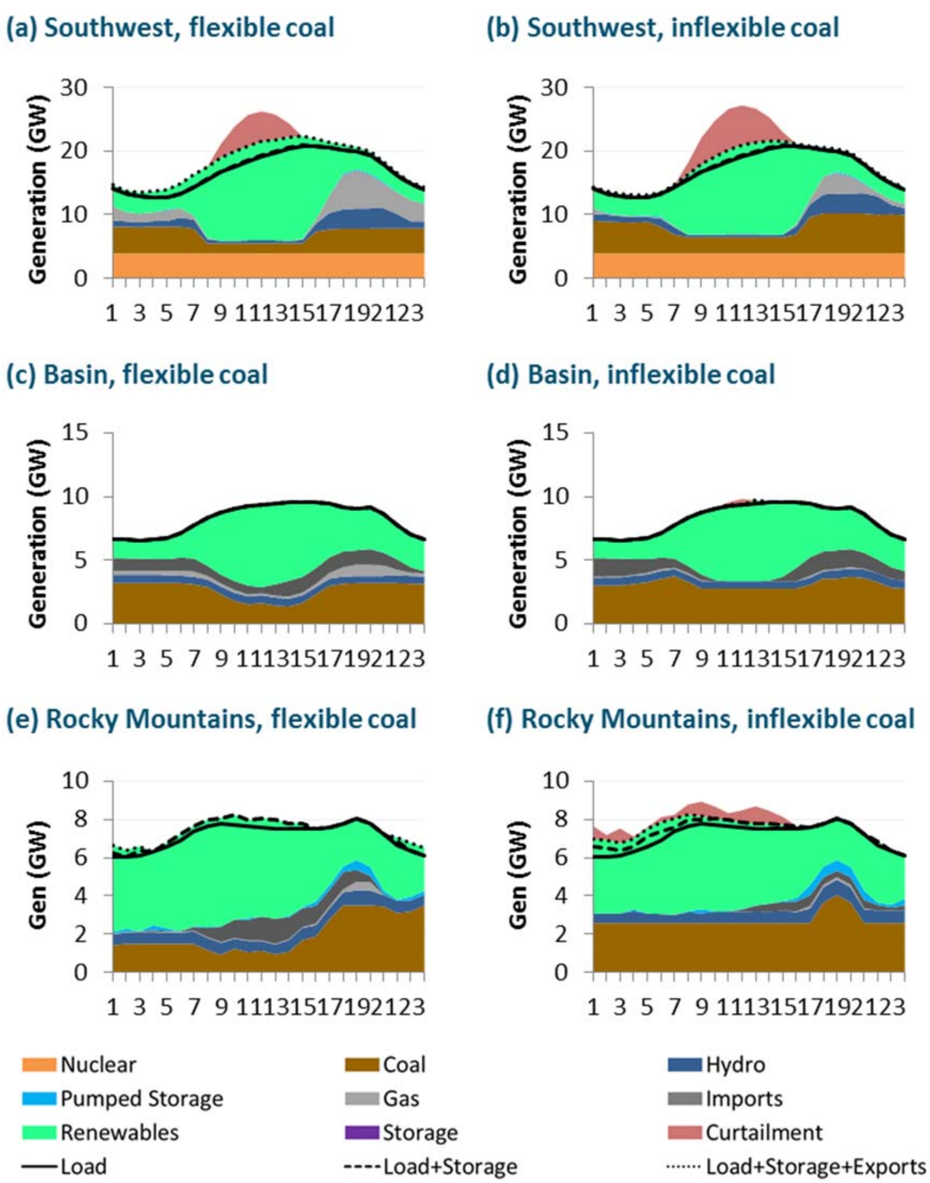
The coal dispatch for the same April day shown in Figure 78 is shown for the coal sensitivity simulation in Figure 79. Note that all committed coal units remain online throughout the day as a result of the high startup and shutdown costs, resulting in higher utilization. Coal dispatch on this day with the Reference Grid assumptions ranges from a daytime minimum of 656 MW to a nighttime maximum of 2,391 MW. Without the ability to turn units on and off throughout the three-day simulation period, enough coal units must be committed to meet the largest net load hour during the full three day period. As a result, the coal dispatch on the day of interest increases to 5,882MW at night and 2,353MW during the day.

Figure 79. Coal dispatch on the April day in the Southwest in the limited coal flexibility sensitivity.



In addition to operations in the Southwest, this sensitivity has a particularly notable impact on the Basin and the Rockies, where a sizable portion of the installed thermal fleet is made up of coal generators. Figure 80 illustrates the major impacts of limitations on coal flexibility with an example day for each of these three regions. In each region, without the ability to shut down or start up coal units within the day, coal generators operate at higher levels—closer to the baseload role that they have filled in the past. The greater dispatch of coal resources is balanced by two effects: (1) a reduction in the reliance on natural gas generators; and (2) an increase in renewable curtailment during low net load periods.

Figure 80. Example spring days under flexible and inflexible coal assumptions



The impacts of coal flexibility on the annual coal dispatch and renewable curtailment are summarized in Table 37 and Table 38, respectively. The increased coal dispatch in the Inflexible Coal sensitivity also results in an additional 22.7 MMtCO2/yr of greenhouse gas emissions relative to the flexible coal assumptions embedded in the Reference Grid.

Table 37. Impact of limited flexibility on coal capacity factors

Scenario	Basin	California	Northwest	Rockies	Southwest
Flexible Coal	56%	62%	54%	57%	77%
Inflexible Coal	65%	80%	68%	64%	87%
Difference	**+9%**	**+18%**	**+15%**	**+7%**	**+10%**

Table 38. Impact of inflexible coal on renewable curtailment.

Scenario	Basin	California	Northwest	Rockies	Southwest
Flexible Coal	0.4%	8.7%	5.6%	0.6%	7.3%
Inflexible Coal	2.2%	9.9%	9.2%	3.0%	10.1%
Difference	**+1.8%**	**+1.2%**	**+3.6%**	**+2.4%**	**+2.8%**

A notable secondary impact of the reduction in coal flexibility is the resulting increase in curtailment in California and the Northwest, both regions in which coal makes up a relatively small share of generation. The mechanism for this impact is a reduction in the willingness of neighboring coal-heavy regions to accept exports, as, without the capability to back down coal generation beyond a certain level, they have no use for the surplus power themselves. This illustrates one of the key uncertainties that faces California decision-makers today trying to understand the size of potential export markets for a solar surplus: to what extent will operators in other regions be willing to cycle coal generators to accept low cost surplus exports from California?

5.3.2 SCHEDULED CURTAILMENT & CURTAILMENT PENALTIES

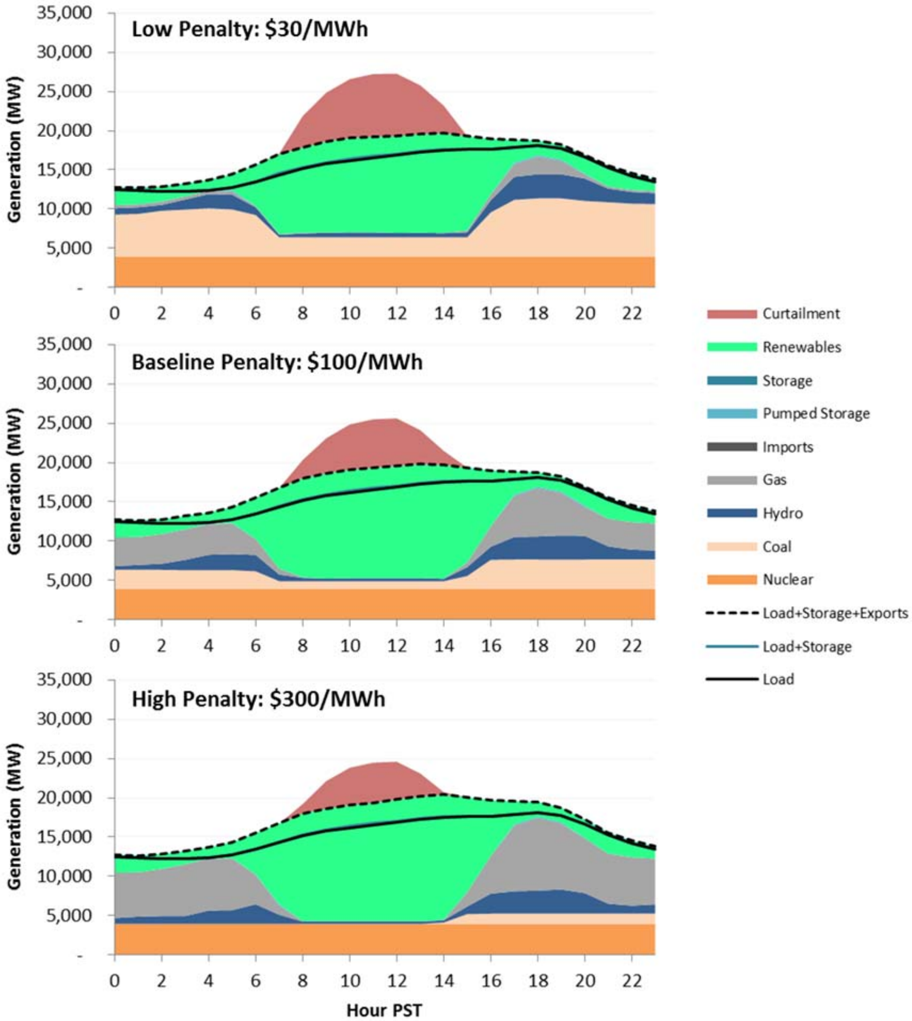
While the coal flexibility sensitivity examines the impact of potential operating constraints on the system, economic considerations also impact the performance of the system under higher renewable penetrations. One critical economic assumption in this analysis is the penalty price placed on renewable curtailment. In this modeling framework, renewable curtailment is given a penalty price that is intended to capture the value to the system of delivering renewable energy. This penalty price plays into an economic decision around the best way to operate the conventional resource fleet to balance the system while utilizing available renewable energy. If, for example, renewable curtailment is relatively inexpensive, then the least-cost dispatch may include some renewable curtailment to make room on the system for inexpensive and relatively inflexible thermal units to meet demand. If renewable curtailment is more costly, it may instead be worth dispatching more expensive and more flexible resources around the renewables in order to avoid curtailment in hours of high renewable output.

The range of costs that may be applied to renewable curtailment are discussed in detail in Section 3.5.2. As a base assumption, this study assumes that the variable cost applied to curtailment reflects the cost of procuring additional renewable energy to ensure compliance with a renewable energy target despite the observed curtailment. This is approximated to be $100/MWh. In addition to the base assumption, two sensitivities were run to examine the impact of the renewable curtailment penalty on operations and on the volume of renewable curtailment experienced by the system. On the low end, a penalty of $30/MWh is used to represent an estimate of the out-of-pocket costs that may be incurred by a utility during curtailment, representing compensation for lost tax credits (e.g. the production tax credit) as well as any additional O&M incurred during curtailment. On the high end, a price of $300/MWh represents an extreme high price for economic curtailment; this is the current assumption used by the CAISO in its flexibility modeling for California's Long-Term Procurement Plan[47].

The most pronounced impact from adjusting the curtailment penalty is observed in the Southwest region. The dispatch on an April day in the Southwest is shown for the base curtailment cost ($100/MWh) and the two sensitivities ($30/MWh and $300/MWh) in Figure 81.

[47] https://www.caiso.com/Documents/Presentation_2014LTPPSystemFlexibilityStudy_SHcall.pdf

Figure 81. Impact of renewable curtailment penalty price on dispatch in the Southwest region on an April day



On this day, the renewable curtailment penalty directly determines whether the net load is met predominantly with coal versus gas. When the renewable curtailment price is low, it is more economic to keep coal units on and curtail excess renewable energy during the middle of the day. However, as the curtailment penalty increases, the economics shift in favor of the more flexible natural gas plants, which can shut down in the morning to accommodate the solar power and start up in the evening to meet the evening net load peak. At the base assumption of $100/MWh, the net load is met with a mix of coal and natural gas resources on this day, but at $300/MWh, nearly all the coal is uneconomic on this day.

The impact of the renewable curtailment price on the observed renewable curtailment is summarized across all the regions in Table 39 and the impact on the coal capacity factor is summarized in Table 40. In both the Basin and Rocky Mountain regions, the renewable curtailment price has little impact on both the observed curtailment and the coal dispatch, largely because the absolute amount of renewable curtailment in the case is very low in these regions. In the Northwest, while significant curtailment is observed, the impact of the curtailment penalty is also relatively small because curtailment is driven largely by an oversupply of hydro and renewable energy relative to load, rather than inflexibility related to economic thermal dispatch. California does experience some coal to gas switching as a result of increasing the renewable curtailment penalty, but the impact on curtailment is quite small because of the relatively small size of the coal fleet and because a region-wide minimum generation constraint often binds in curtailment hours, preventing further unloading of thermal resources.

Table 39. Renewable curtailment under alternative curtailment penalty prices.

Scenario	Basin	California	Northwest	Rockies	Southwest
Low Penalty ($30/MWh)	0.4%	9.0%	5.8%	0.6%	9.2%
Base Penalty ($100/MWh)	**0.4%**	**8.7%**	**5.6%**	**0.6%**	**7.3%**
High Penalty ($300/MWh)	0.3%	8.3%	5.3%	0.5%	5.1%

Table 40. Coal capacity factors under alternative curtailment penalty prices.

Scenario	Basin	California	Northwest	Rockies	Southwest
Low Penalty ($30/MWh)	57%	69%	55%	57%	89%
Base Penalty ($100/MWh)	56%	62%	54%	57%	77%
High Penalty ($300/MWh)	56%	33%	54%	57%	46%

In the Southwest, where there is significant renewable curtailment and a tradeoff between the cost and flexibility of the coal and gas fleets in the region, the dispatch is highly sensitive to the renewable curtailment price, as is suggested by the example day shown in Figure 81. Across the range of curtailment prices in this set of sensitivities, the renewable curtailment in the Southwest ranges from 5.1% to 9.2% and the average coal capacity factor ranges from 46% to 89%. This suggests that renewable integration policies in the Southwest might not only drive renewable curtailment, but may also have considerable implications for coal generation and greenhouse gas emissions (see Table 41) in the region. This phenomenon would likely also be obersved in the Basin and Rockies regions under higher levels of renewable penetration.

Table 41. Greenhouse gas implications of the sensitivity of the coal dispatch to the penalty applied to renewable curtailment.

Renewable Curtailment Penalty	Total Footprint GHG Emissions (MMtCO2/yr)	Difference Relative to Base Penalty (MMtCO2/yr)
Low Penalty ($30/MWh)	227	+7
Base Penalty ($100/MWh)	220	-
High Penalty ($300/MWh)	205	-15

The economic tradeoff between coal dispatch and renewable curtailment identified for the Southwest may also be relevant to specific balancing areas within the Western Interconnection that have a high reliance on coal and gas resources and face increasing renewable penetrations. In addition, the sensitivity to curtailment prices may be impacted by the flexibility of the coal fleet described in Section 5.3.1. This analysis stops short of testing the interactive effects of the renewable curtailment price and the cost of cycling coal plants, but this constitutes an important area of further research for many balancing areas in the West.

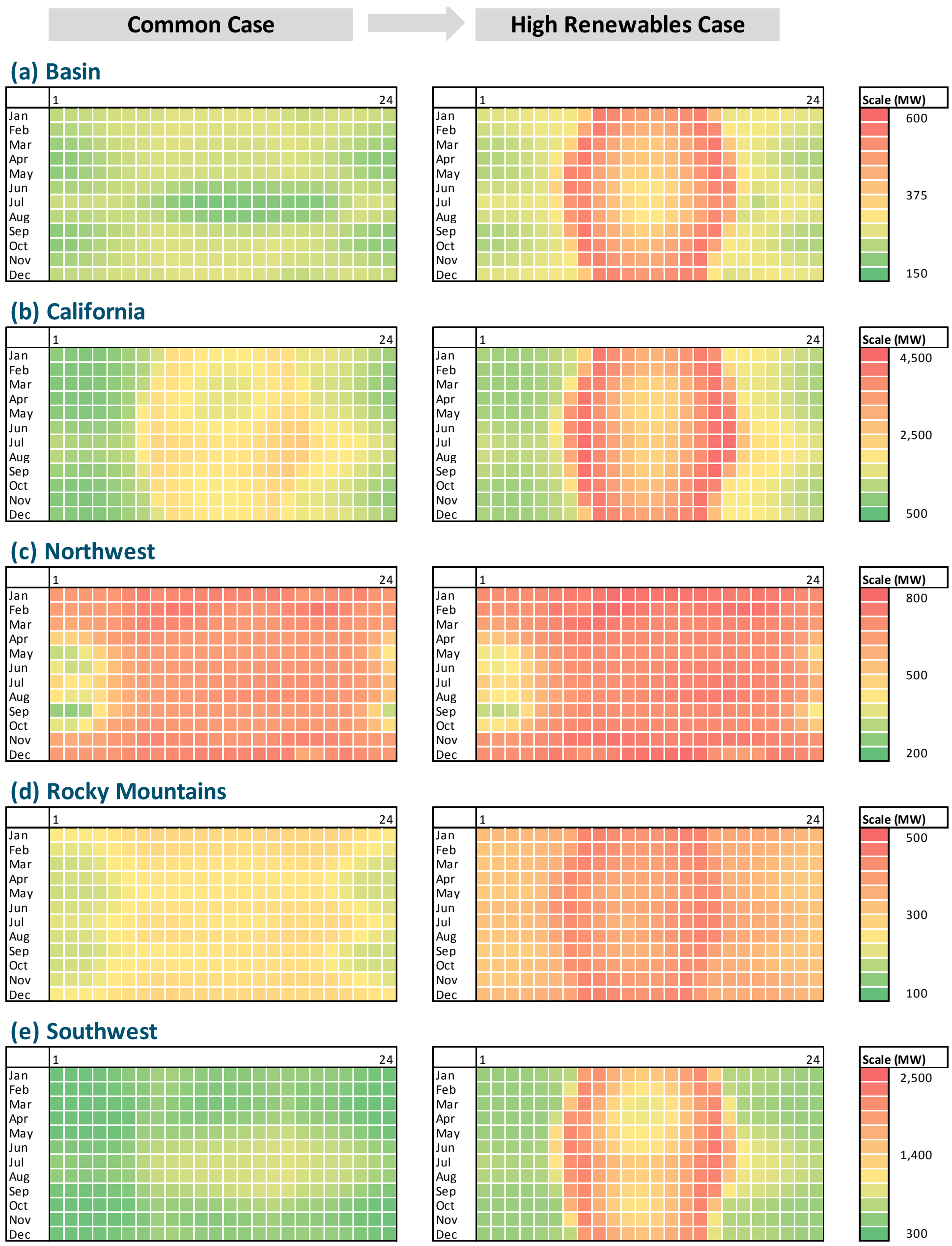
5.3.3 FLEXIBILITY RESERVES & SUBHOURLY CURTAILMENT

Related to the economics of renewable curtailment discussed above is the ability of renewable resources to respond (or curtail) on a real-time basis to provide additional flexibility to the system. Section 3.6.2 discusses the approach used in this study to simulate the provision of flexibility reserves under Reference Grid assumptions. The approach assumes that renewable resources in all regions can be curtailed within the hour to avoid real-time overgeneration when it is not feasible or too costly to meet downward flexibility reserve requirements with conventional resources. This approach is also applied to upward flexibility reserves—subhourly unserved energy may be experienced in the case of an upward flexibility reserve shortage. Because unserved energy is penalized at $50,000/MWh, upward reserve violations are rarely experienced.

Subhourly renewable curtailment, however, which is penalized at $100/MWh, is often the least expensive approach to providing adequate downward flexibility within the hour when the system is constrained in the downward direction. This section describes both the impact of the increased renewable penetration on flexibility reserve requirements between the Common Case and the High Renewables case and the behavior of the system in response to these reserve requirements and the additional constraints placed on the system by higher renewable penetrations.

The average flexibility reserve requirements in the real-time stage are shown on a month-hour basis for each region and for both the Common Case and the High Renewables case in Figure 82. As is shown, the increased renewable build in the High Renewables Case generally increases flexibility reserve requirements relative to the Common Case. Notably, regions with significant solar build have a distinct seasonal and diurnal pattern, in which the flexibility reserve requirements peak during sunrise and sunset hours. In these time periods, solar resources experience significant ramps. A portion of these ramps can be accommodated by hour-to-hour ramping of committed conventional resources as well as starting up or shutting down units at the bottom or top of each hour. However, additional flexibility is required within each hour to balance the net load in real-time. This subhourly ramping capability is set aside in the model through the provision of the flexibility reserves.

Figure 82. Average real-time flexibility reserve requirements by month-hr in the Common Case and High Renewables Case.



While reserve requirements in these hours can be quite large (the maximum requirement in California exceeds 4,500 MW), the response needed from conventional resources within these hours is largely predictable, so grid operators can plan for these time periods when setting generator schedules in advance of the operating day. Additional wind resources also increase the flexibility reserve requirements, but these impacts have less seasonal and diurnal predictability. As a result, the impact of increased subhourly flexibility needs in wind-heavy regions like the Northwest are fairly diffuse across seasons and time of day and operators will face the added challenge of evaluating flexibility reserve requirements on an ongoing basis as new load and meteorological forecasts change.

In addition to the flexibility reserve requirements, this study also examines the response of the system to these requirements and the extent to which the system experiences subhourly unserved energy or curtailment as an economic result of holding inadequate reserves. To illustrate this, the available reserves for two example operating days in the Basin region are shown in Figure 83 and Figure 84. On the August day, thermal resources are committed in order to provide both upward and downward reserves throughout the day, avoiding subhourly renewable curtailment and unserved energy (not shown). In contrast, on the April day in Figure 84, thermal units ramp down to minimum stable levels during daylight hours due to low net load conditions. The resulting downward flexibility reserve shortage causes between 100 and 120 MWh of expected subhourly renewable curtailment in each of the hours with the reserve shortage.

Figure 83. Operating range of units providing flexibility reserves in the Basin region on August day.

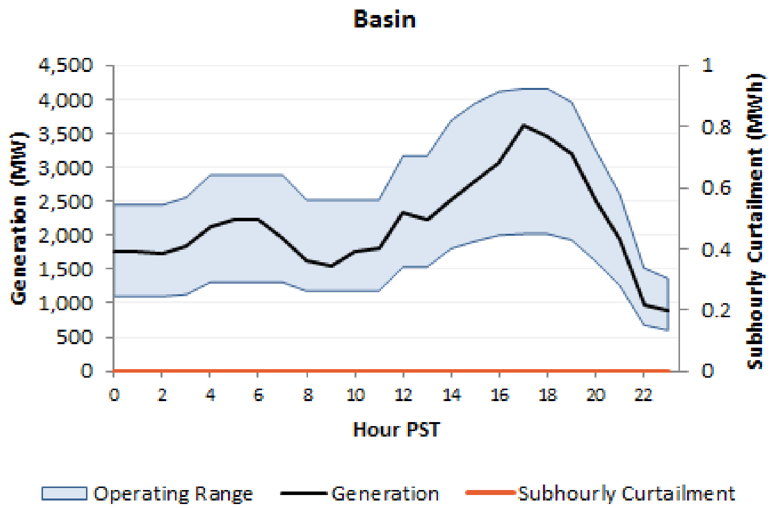
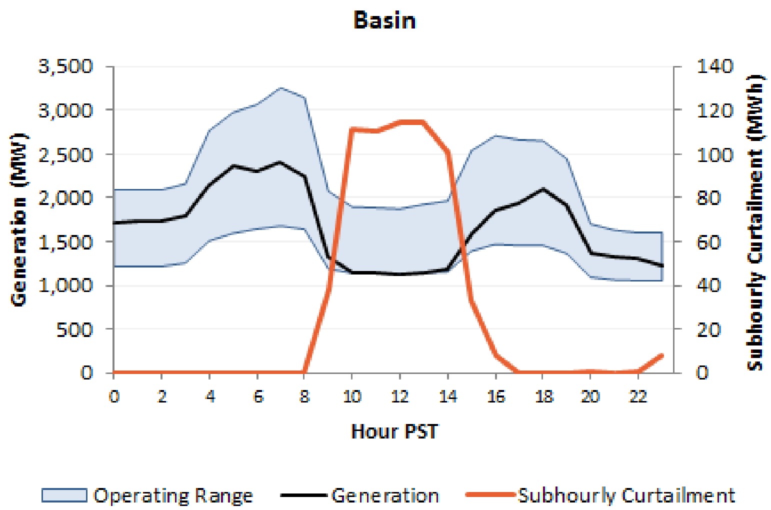


Figure 84. Operating range of units providing flexibility reserves in the Basin region on April day. Low net load conditions push thermal units to their minimal stable levels in hours 09 through 14, resulting in subhourly renewable curtailment.



The subhourly curtailment in these hours can also be understood as the renewables themselves providing downward reserve services through real-time dispatch. In some time periods—specifically those with high renewable output—this is the most economic approach to managing subhourly flexibility needs. The alternative in these time periods would be to operate conventional units at higher set points in order to allow them to ramp down within the hour, exacerbating the hourly curtailment challenge. By allowing real-time dispatch of renewables, some curtailment is experienced within the hour, but hourly curtailment related to downward reserve and thermal constraints can be avoided. In higher net load hours, these units tend to sit at higher more efficient set points and the downward reserve provisions can be met at zero or low cost without participation from renewables. This is shown in the nighttime hours in Figure 84.

The same trends that are observed on the example days in the Basin region are observed throughout the West in the High Renewables case, with the exception of the Northwest, where the large hydro fleet provides adequate upward and downward flexibility to avoid flexibility reserve shortages in almost all hours. The reserve requirements and reserve provisions met with conventional resources in each region are summarized in Table 42. In all regions but the Northwest, real-time dispatchability of renewables is relied upon to meet downward reserve requirements in one quarter to one half of all hours.

Table 42. Average upward and downward flexibility reserve requirements, provisions, and frequency of shortages in the High Renewables case.

Metric	Basin	California	Northwest	Rockies	Southwest
Average Reserve Requirement (MW)	354	2,308	670	365	1,108
Average Upward Reserve Provision (% of Requirement)	100%	100%	109%	100%	100%
% of Hours with Upward Reserve Shortage	0.033%	0.066%	0%	1.1%	0.93%
Average Downward Reserve Provision (% of Requirement)	75%	87%	101%	76%	65%
% of Hours with Downward Reserve Shortage	41%	26%	1.1%	52%	39%

The seasonal and diurnal trends in subhourly renewable curtailment are shown in Figure 85. In most regions, real-time dispatch of renewables to provide downward flexibility within the hour is relied upon predominantly during daytime hours when the region experiences hourly curtailment, though the Rocky Mountain region also experiences nighttime subhourly curtailment due to high wind conditions. For comparison, the subhourly curtailment is also shown for the Common Case, though the Rocky Mountain region is the only region with non-negligible subhourly curtailment in the Common Case. Several factors influence the levels of observed subhourly curtailment.

High renewable penetration and the variability of the renewable resource output drives the need for flexibility reserves, but other factors influence whether the system is capable of meeting those reserves with conventional generation. For example, these simulations assume that while coal resources can be redispatched within the day away from day-ahead schedules, they cannot be dispatched in real-time away from their hourly schedules. Flexibility reserves in real-time must therefore be provides by gas resources, hydro resources, or subhourly curtailment.

This largely explains why the Basin and Rocky Mountain regions, which experience very little curtailment on the hourly level, experience similar levels of subhourly curtailment to California, where renewable oversupply challenges are much greater but natural gas, hydro, and storage resources provide significant subhourly flexibility. The very high levels of subhourly curtailment in the Southwest show the combined impact of a system that frequently faces oversupply conditions and has limited subhourly flexibility due to the coal dispatch.

Figure 85. Subhourly curtailment experienced as a result of flexibility reserve shortages in the High Renewables Case.

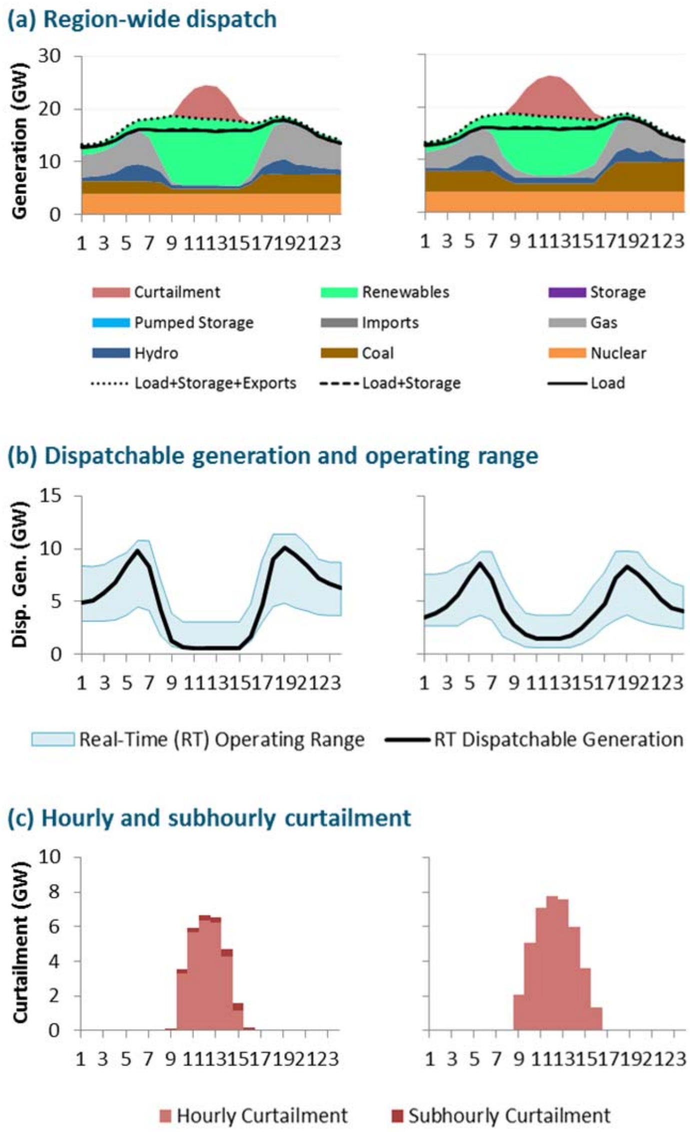
The economic tradeoff in the Reference Grid simulations between providing downward flexibility reserves with conventional resources versus subhourly renewable curtailment makes two key assumptions about the operation of the grid:

1. That the communications and controls are available for renewable resources to be dispatched downward in real-time to balance the load; and

2. That the grid operator has sufficient confidence in the real-time response from renewable generators to unload thermal generation.

Some, but not all, jurisdictions in the West already have the ability to curtail renewable resources in real-time and to incorporate this anticipated curtailment into decisions around thermal schedules. In order to quantify the value of these types of operational strategies at higher renewable penetrations, a sensitivity in which downward flexibility reserve violations were penalized as at the same cost as upward flexibility reserve violations was tested. The increased penalty on downward reserve violations ensures that conventional resources are used when possible to provide downward flexibility within the hour, rather than allowing renewable curtailment in real-time. The impact of this more restrictive reserve provision strategy is shown for an example March day in the Southwest in Figure 86.

Figure 86. The impact of increasing the penalty on subhourly curtailment, shown for an example March day in the Southwest (left panel illustrates the operations when subhourly curtailment is penalized at the cost of curtailment ($100/MWh) and the right panel shows the operations when subhourly curtailment in penalized at $50,000/MWh).



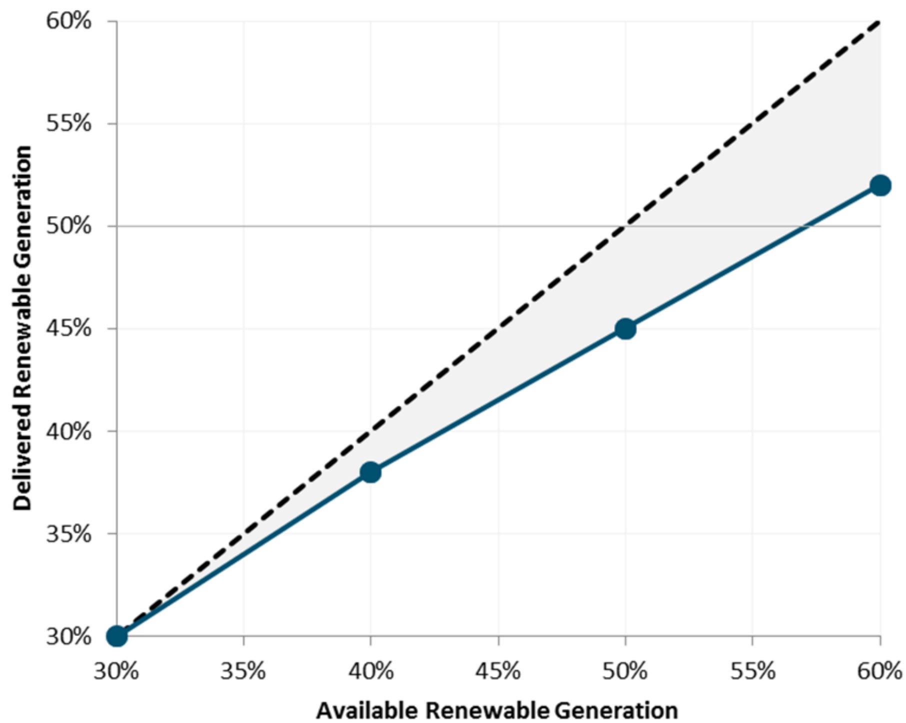
As the example day illustrates, disallowing (or highly penalizing) real-time renewable curtailment exacerbates the curtailment challenge for systems with high penetrations of renewables. In the middle of the day, when renewable curtailment occurs on this March day, generators can be seen in Figure 86b operating at the bottom of the operating range when subhourly curtailment is penalized at $100/MWh, opting to experience some subhourly curtailment to avoid even larger amounts of curtailment at the hourly level. When real-time curtailment is not available to the system as a relatively low cost strategy for managing subhourly fluctuations, dispatchable generator set points increase to provide subhourly flexibility, further constraining the system and increasing hourly renewable curtailment (Figure 86c).

By allowing real-time renewable curtailment, both total curtailment and total cost can be reduced. On this day, the Southwest region experiences 29.2 GWh of curtailment (including both hourly and subhourly) when renewable curtailment is used to help mitigate real-time flexibility challenges and 40.6 GWh of curtailment when renewable resources are not able to provide real-time balancing services. Allowing real-time curtailment also results in $487k of production cost savings on this day by allowing more efficient utilization of the thermal fleet. This sensitivity highlights the value of allowing renewable resources to participate in real-time markets or to be re-dispatched by the operator in real-time from both the renewables integration and operating cost perspectives.

5.4 Additional Enabling Strategies for Renewable Integration

The critical role of curtailment as an operational strategy to ensure reliability and efficient operations in high penetration renewable scenarios has direct implications for policies aimed at increasing the penetration of renewables. The foregone production from renewable resources that occurs when operators choose to curtail creates a challenge to meeting policy goals such as RPS targets or greenhouse gas intensity limits: while the renewable fleet may be built with the ability to produce a certain amount of generation over the course of the year, the amount that can be delivered to the electric system (in the absence of solutions) is reduced due to the impacts of curtailment. As illustrated in Figure 87, the size of this challenge increases nonlinearly; that is, as the concentration of intermittent resources becomes increasingly mismatched with the instantaneous demand for electricity, the share of renewable generation that must be curtailed to preserve reliability also increases.

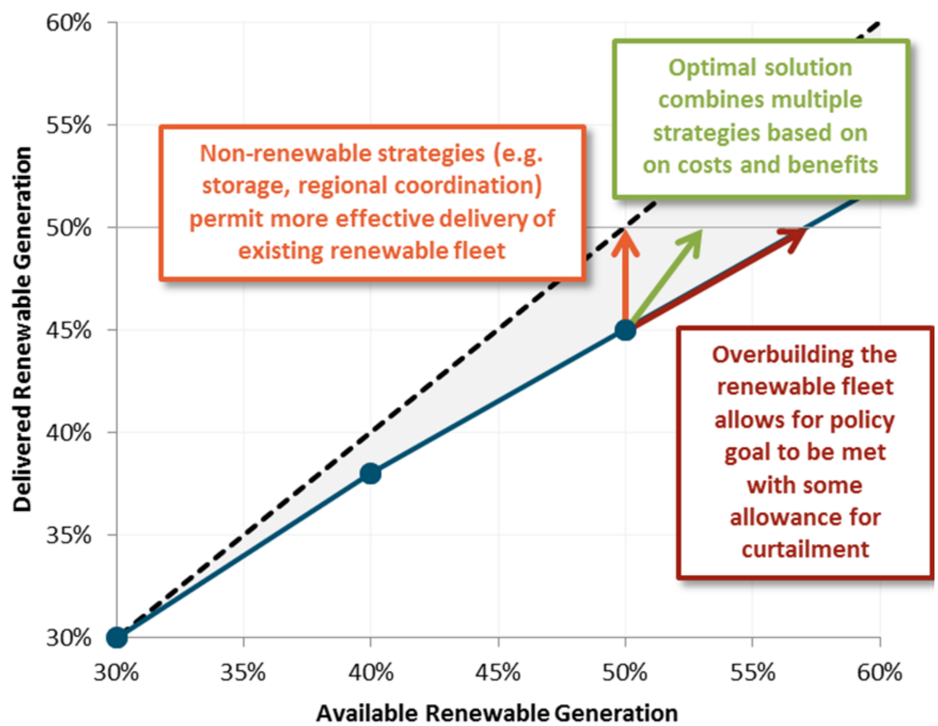
Figure 87. Impact of curtailment on achievement of targeted policy goal.



Overcoming this challenge to meet policy goals and continue along a pathway towards the decarbonization of the electric sector can be achieved through a variety of renewable integration strategies, illustrated conceptually in Figure 88. The simplest strategy to overcome the effect of curtailment is to overbuild the renewable fleet such that it has the capability to produce more, on an annual basis, than is required to meet policy goals, effectively establishing an allowance for renewable curtailment while still delivering the requisite amount of generation to the electric system (illustrated by the red arrow). By applying a cost penalty to renewable curtailment that is intended to represent the "replacement cost" for renewable generation, this study implicitly treats overbuilding as the "default" solution for renewable integration against which the value provided by alternative strategies may be measured.

Alternative strategies to overbuilding the renewable portfolio allow for the achievement of policy goals by avoiding curtailment, allowing for the delivery of additional renewable generation without the need to overbuild the fleet (shown by the orange arrow). Such strategies generally either entail increasing the downward flexibility of the existing generation fleet to accommodate more renewable production or finding an alternative market for generation that would otherwise be curtailed.

Figure 88. Illustration of strategies to overcome integration challenges.



This study examines three specific strategies for renewable integration intended to reduce or avoid curtailment in day-to-day operations: (1) increased regional coordination; (2) investments in energy storage; and (3) adding flexible generation to the thermal fleet. The measures considered in this study are not intended to represent a comprehensive set of the potential strategies or investments that a utility might consider, but do provide a useful indication of the types of characteristics that relieve challenges encountered in renewable integration. In this study, the potential value of each of these strategies is evaluated; however, this study does not attempt to determine which combination may be cost-effective or optimal. Determining an "optimized" combination of integration strategies (illustrated by the green arrow) is a question of economics, and depends not only upon the value that each can provide but upon the cost to implement the individual strategies as well.

It is also worth noting that the flexibility solutions investigated in this study are considered after significant levels of coordination have already been achieved within each region, allowing each region to take full advantage of the load and resource diversity within its boundaries. In regions with more Balkanized operations, the solutions investigated herein will tend to have increased value. These results are therefore not intended to offer a conclusive assessment of the benefits of flexible resources in the West, but are instead meant to highlight trends, identify interactive effects, and help develop intuition that can be used for future investigation into flexibility needs in specific systems.

5.4.1 INCREASED REGIONAL COORDINATION

Aside from the deregulated power markets of the California ISO and the Alberta Electric System Operator, the Western Interconnection operates through bilateral contracts among a mix of vertically integrated utilities, public power utilities, and federal power marketing authorities. In this environment, interregional power exchange in the West is largely guided by longstanding contractual arrangements and well-established seasonal patterns. While other markets exist, such as those at Palo Verde and Mid-C, transactions in each of these settings are generally bilateral in nature. Because of this structure, the Western Interconnection does not function as a single integrated single market.

With increasing penetrations of renewable generation, the balance between supply and demand in each region will change, and economic signals will adjust; however, there is no guarantee that operators will respond efficiently or immediately to those signals in the absence of a single integrated and well-functioning market. This creates the potential for an institutional barrier for renewable integration—the technical capability of the Western system as a whole to absorb renewable generation may be greater than allowed by the existence of current market arrangements and institutional conventions.

Overcoming this institutional barrier is one of the strategies for renewable integration in this study, and it presents a large potential value by harvesting the diversity of loads and renewable resources in the Western Interconnection. By relaxing limits on interregional power exchange from historically observed flows to the physical limits of interregional transmission paths, this study examines the benefits of increased regional coordination.

Figure 90 compares the amount of curtailment observed in each region in the High Renewables Case for both the historical intertie limits (a Reference Grid assumption) and the physical intertie limits. The relaxation of this constraint—and the consequent implied emphasis upon regional coordination in operations—provides significant value through its reduction in renewable curtailment, which drops from 6.4% across the study footprint under historical limits to 3.0% under physical limits. Figure 89 provides detailed region-by-region curtailment impacts.

Figure 89. Impact of intertie limits on renewable curtailment.

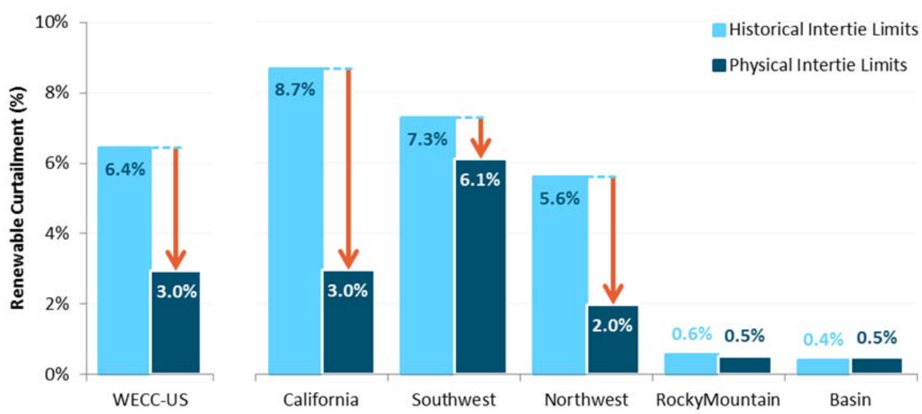


Table 43. Impact of increased regional coordination upon renewable curtailment

Scenario	Type	Basin	California	Northwest	Rockies	Southwest
Historical Intertie Limits	Scheduled	0.1%	8.6%	6.1%	0.1%	6.4%
	Subhourly	0.3%	0.0%	0.0%	0.5%	0.9%
	Total	0.4%	8.6%	6.1%	0.1%	7.3%
Physical Intertie Limits	Scheduled	0.1%	3.0%	1.5%	0.0%	5.1%
	Subhourly	0.4%	0.0%	0.1%	0.5%	0.9%
	Total	0.5%	3.1%	1.6%	0.5%	6.0%
Difference	**Scheduled**	**-0.0%**	**-5.6%**	**-4.6%**	**-0.1%**	**-1.3%**
	Subhourly	**+0.1%**	**-0.0%**	**+0.1%**	**-0.1%**	**-0.0%**
	Total	**+0.1%**	**-5.6%**	**-4.5%**	**-0.0%**	**-1.3%**

The reduction in curtailment results from the fact that surplus generation that would have otherwise been curtailed can find an alternative market, displacing primarily fossil-fueled generators in other regions. However, the depth of those alternative markets is not infinite, and each region's appetite for low-cost renewable generation from its neighbors depends on its own resource balance. Consequently, regional coordination reduces, but does not eliminate, renewable curtailment; during periods when multiple regions experience curtailment simultaneously, finding a market for surplus power may be difficult.

Figure 90 shows the impact of this change on the seasonality of curtailment within each region. Reductions in curtailment outside of the spring months and during nighttime hours are notable; each of these periods corresponds to a situation in which at least one region's ability to import allows for a substantial reduction in curtailment. However, curtailment remains prevalent during the spring months—especially in the middle of the day—due to the coincident surplus of generation across multiple regions.

Figure 90. Impact of increased regional coordination on curtailment in each region.

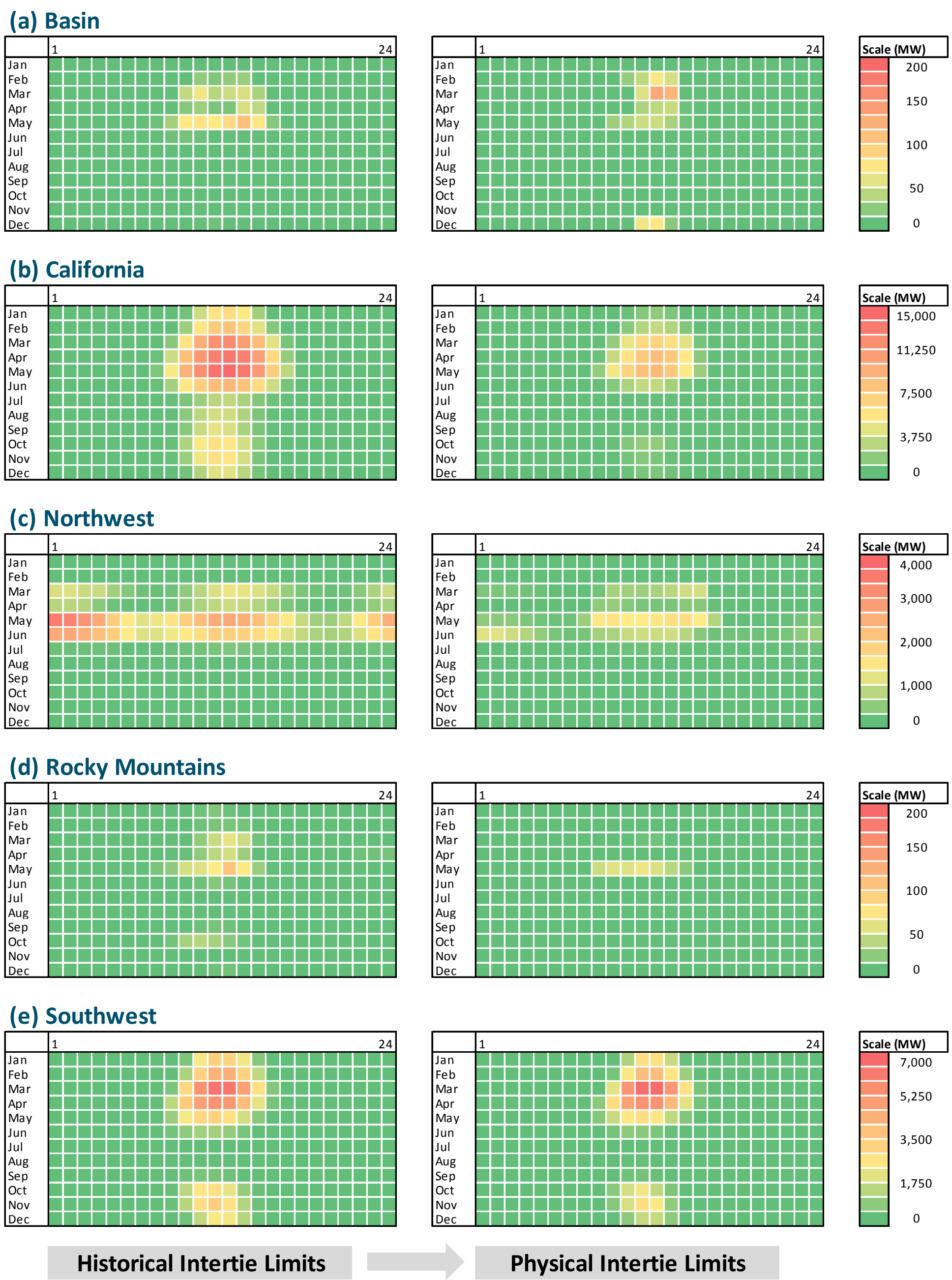
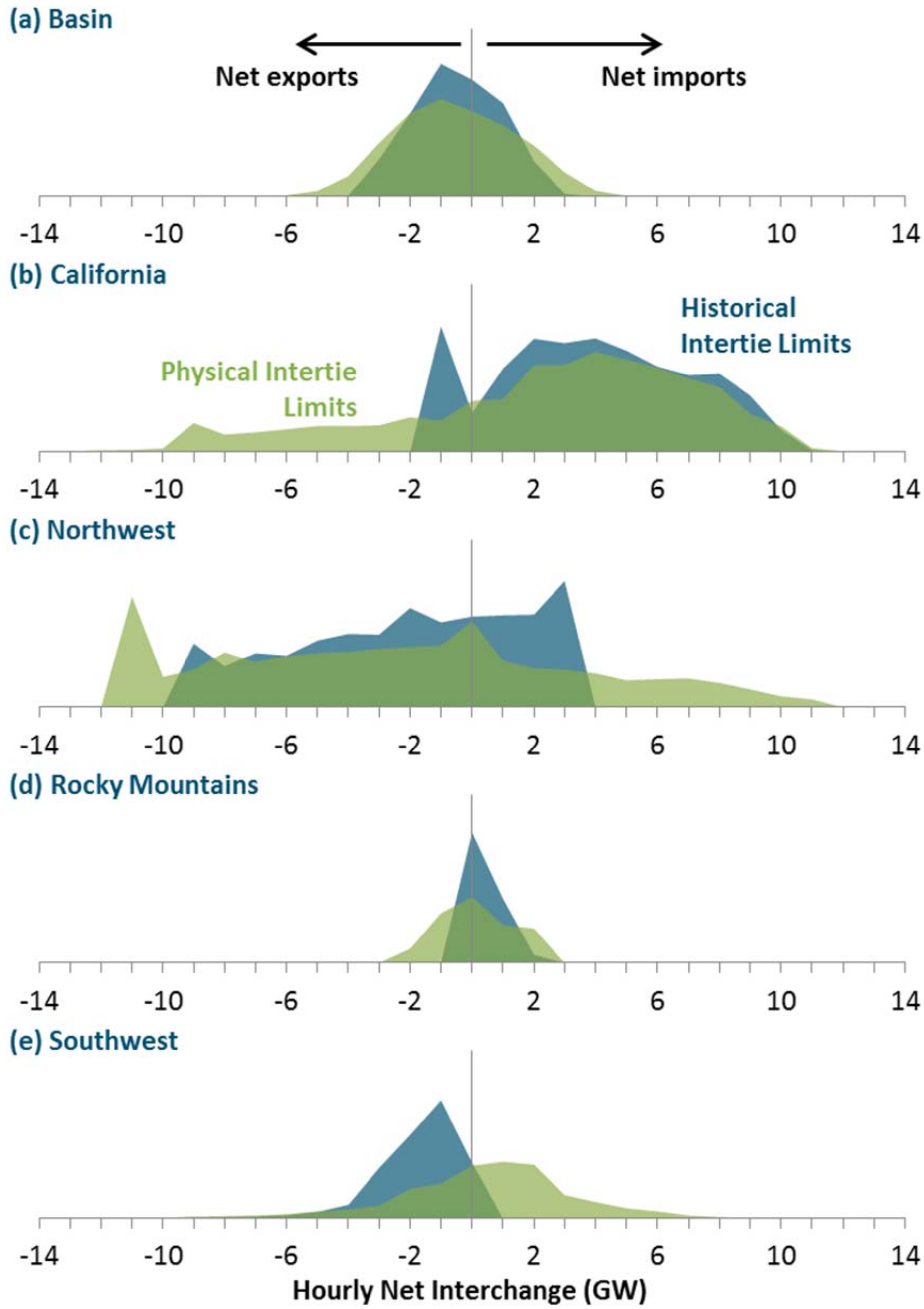


Figure 91. Distributions of net interchange by region.



The increase in regional coordination that is implied by the relaxation of intertie limits to physical constraints is also interesting for what it implies about patterns in interregional transmission flow, which could change dramatically from what has been observed historically. Figure 91 illustrates this phenomenon, showing the distributions of flow of net interchange in the High Renewables Case under the two intertie scenarios.

A number of key impacts of relaxing intertie constraints are captured in Figure 91, specifically:

+ When allowed to export generation up to physical path limits, California becomes a major net exporter in certain hours.
+ One of the large destination markets for California's exports is the Northwest, which becomes a major net importer during certain periods of the year (as indicated by the long tail on the right side of the distribution).
+ Expanded export capability also benefits the Northwest, as it is able to access additional markets for its oversupply in the spring runoff period. One of the important destination markets for this surplus generation in the spring is the Southwest, as generation is wheeled through California and across Path 46.
+ On an aggregate annual basis, the Southwest becomes a net importer of generation. While Path 46 has historically never flowed from west to east, the High Renewables Case suggests a potential transformation with respect to how this path is used. With the relaxation of intertie constraints, the Southwest benefits from exports from California as well as the Northwest (via California) during periods when it is not experiencing its own oversupply.

The regional coordination sensitivity also provides information about utilization of specific interties when flows are allowed up to the full path ratings. Table 44 summarizes the percent of hours in which path rating constraints bind in each direction in both the Common Case and the High Renewables case.

Table 44. Percent of hours in which the intertie constraints bind in the Common Case and High Renewables Case when intertie flows are allowed up to full path ratings.

Intertie Definition		Minimum Limit % of Hours Binding			Maximum Limit % of Hours Binding		
From Region	To Region	Limit (MW)	Common Case	High Renewables Case	Limit (MW)	Common Case	High Renewables Case
Basin	Northwest	-3,406	1%	11%	4,637	0%	1%
Basin	Rockies	-1,530	0%	2%	1,530	1%	12%
Basin	Southwest	-1,480	0%	10%	1,465	15%	21%
California	Basin	-1,550	39%	21%	1,760	0%	9%
Northwest	California	-6,775	0%	7%	8,020	9%	16%
Rockies	Southwest	-690	4%	22%	690	19%	25%
Southwest	California	-11,200	0%	0%	11,200	0%	0%

In general, increasing the renewable penetration increases the reliance on transmission to balance loads and resources across the West, as indicated by an increase the percent of hours in which paths are full utilized. This trend corroborates both the findings in this study regarding the value of regional coordination as well as the common finding in the literature that transmission to integrate renewables over larger geographical footprints reduces bulk integration challenges.

5.4.2 INVESTMENT IN ENERGY STORAGE

While institutional renewable integration solutions like improved regional coordination show promise in alleviating renewable integration challenges, there is considerable interest in the value of physical solutions as well – in particular flexible assets like energy storage. Investments in energy storage facilitate renewable integration by increasing both the amount of upward and downward flexibility in an electric system. Over timescales of hours, energy storage resources can store renewable energy in periods of potential overgeneration and discharge the energy in later hours to offset thermal dispatch. At shorter timescales, energy storage can provide reserves to help reduce the inefficiencies and constraints associated with providing these services with thermal resources. To test the value of energy storage for integrating renewables, the High Renewables Case was run with additional energy storage resources in three different configurations, with 2-hr, 6-hr, and 12-hr energy storage devices, summarized in Table 45. In each scenario, energy storage resources are built in the three regions with the largest curtailment: California, the Southwest, and the Northwest. The storage build in each region was guided by the magnitude of the curtailment challenge in each region under Reference Grid assumptions.

The energy storage resources in these scenarios were modeled as a single energy storage system in each region with a maximum state of charge based on the device duration (i.e. a 2-hr device has a maximum state of charge large enough to discharge at the maximum discharge rate for 2 hours). The storage systems provide flexibility reserves in discharging mode, based on the full discharging operating range, but not in charging mode.[48] A $17.65/MWh variable charge was applied to the energy storage dispatch in order to penalize storage losses at the same rate as renewable curtailment.[49] This charge prevents the model from using energy storage resources to intentionally burn energy in order to avoid renewable curtailment and the $100/MWh penalty associated with it.

Table 45. Energy storage resources added to each Storage Scenario.

Storage Scenario	Storage Build	Maximum Duration	Round-trip efficiency
2-hr storage	California: 4,000 MW; Southwest: 1,000 MW; Northwest: 1,000 MW	2 hours in all regions	85%
6-hr storage	California: 4,000 MW; Southwest: 1,000 MW; Northwest: 1,000 MW	6 hours in all regions	85%
12-hr storage	California: 4,000 MW; Southwest: 1,000 MW; Northwest: 1,000 MW	12 hours in all regions	85%

The impact of the energy storage resources on curtailment in each region is summarized in Table 46. While the energy storage resources in each storage scenario reduce renewable curtailment in both California and the Southwest, the Northwest sees a slight increase in renewable curtailment with the modeled energy storage build. The results suggest that the renewable integration value

[48] This assumption represents a compromise between the reserve capabilities of battery systems, which can utilize the full range from maximum charging rate to maximum discharging rate at subhourly time scales, and pumped storage systems, which can provide reserves over the discharging range when discharging and potentially the charging range when charging if variable speed pumps are installed.

[49] The variable charge on energy storage dispatch also acts as a hurdle rate for gas-on-gas or coal-on-gas arbitrage, so that the energy storage behavior specifically reflects the renewable integration benefits, but may not fully capture the economic benefits of energy storage resources operating on the system.

of the energy storage resources is driven largely by the ability to store excess renewable energy during midday and to discharge this energy to meet the evening peak in solar-dominated regions. An example of the storage dispatch in California on an example day (Figure 92) illustrates this dynamic. Notably, in the Northwest, where overgeneration is driven more by daily hydro energy constraints than by diurnal mismatches between load and renewable availability, the daily energy storage resources do not alleviate renewable curtailment. While it was not specifically investigated herein, this suggests that managing imbalances in the Northwest may require longer duration energy storage.[50]

The ability of the energy storage devices to avoid renewable curtailment is also highly dependent on the device duration (i.e. maximum state of charge) in the short- to medium-duration regime, as is illustrated in Table 46. In both California and the Southwest, the curtailment avoided by the 6-hr storage device is approximately double the curtailment avoided by the 2-hr device because curtailment events in those regions tend to last for several hours in the middle of the day. Despite the significant added value of building a 6-hr device versus a 2-hr device in these simulations, very little additional value is found for a 12-hr storage device over a 6-hr device, as the likelihood of encountering a 12-hr curtailment event in a solar-dominated system is exceedingly low. The diminishing value associated with increasing the duration of a fixed discharge capacity device suggests that there is an economically optimal device duration for a given system—this optimum occurs where the value associated with

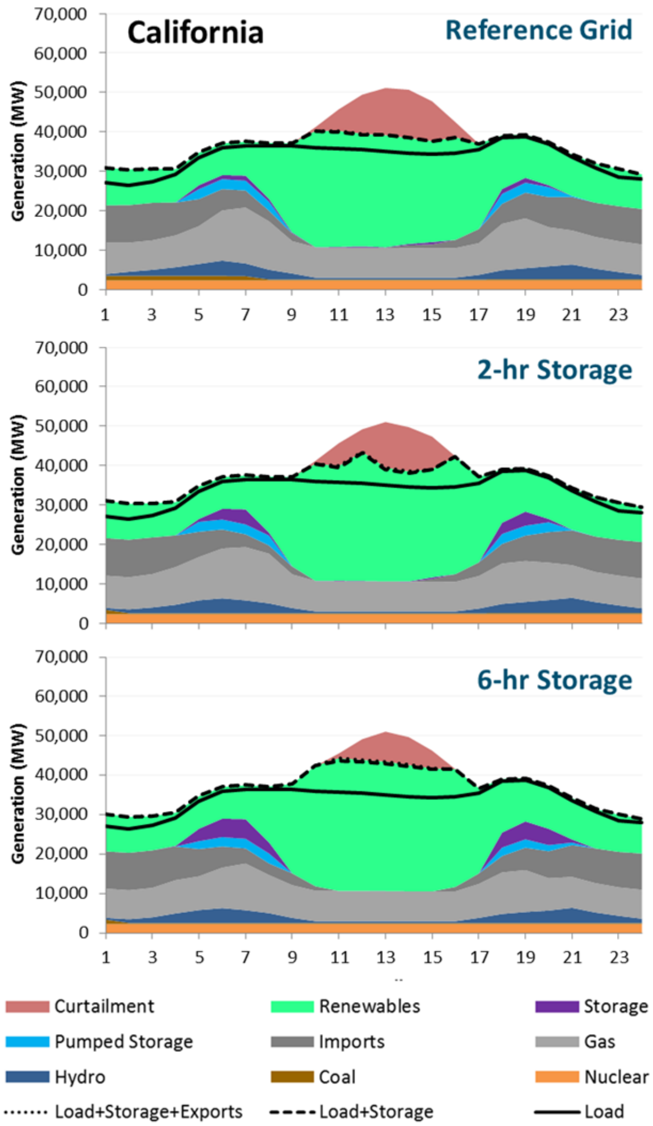
[50] Note that because British Columbia has been excluded from this analysis, the Northwest is not capable of utilizing the long duration hydro storage capability of its northern neighbor in the simulation.

increasing the maximum state of charge offsets its associated cost. Identification of this optimal storage size is beyond the scope of this study, but the economic trade-offs between maximum discharge capacity, maximum state of charge, avoided curtailment, and avoided fuel burn should be considered in economic evaluations of energy storage resources for renewable integration.

Table 46. Impact of energy storage on renewable curtailment (as a % of available renewables) in each storage scenario.

Scenario	Type	Basin	California	Northwest	Rockies	Southwest
Reference Grid	Scheduled	0.1%	8.6%	5.6%	0.1%	6.4%
	Subhourly	0.3%	0.0%	0.0%	0.5%	0.9%
	Total	0.4%	8.7%	5.6%	0.6%	7.3%
2-hr Storage	Scheduled	0.1%	7.1%	5.7%	0.1%	5.3%
	Subhourly	0.3%	0.0%	0.0%	0.5%	0.8%
	Total	0.4%	7.2%	5.7%	0.6%	6.2%
	Difference	**-0.0%**	**-1.5%**	**+0.1%**	**+0.0%**	**-1.1%**
6-hr Storage	Scheduled	0.1%	5.8%	5.9%	0.1%	4.2%
	Subhourly	0.3%	0.0%	0.0%	0.5%	0.9%
	Total	0.4%	5.8%	5.9%	0.6%	5.1%
	Difference	**-0.0%**	**-2.9%**	**+0.3%**	**+0.0%**	**-2.2%**
12-hr Storage	Scheduled	0.1%	5.7%	5.8%	0.1%	4.2%
	Subhourly	0.3%	0.0%	0.0%	0.5%	0.9%
	Total	0.4%	5.7%	5.8%	0.6%	5.1%
	Difference	**0.0%**	**-2.9%**	**+0.2%**	**0.0%**	**-2.2%**

Figure 92. Impact of 2-hr and 6-hr energy storage on the dispatch on a March day.

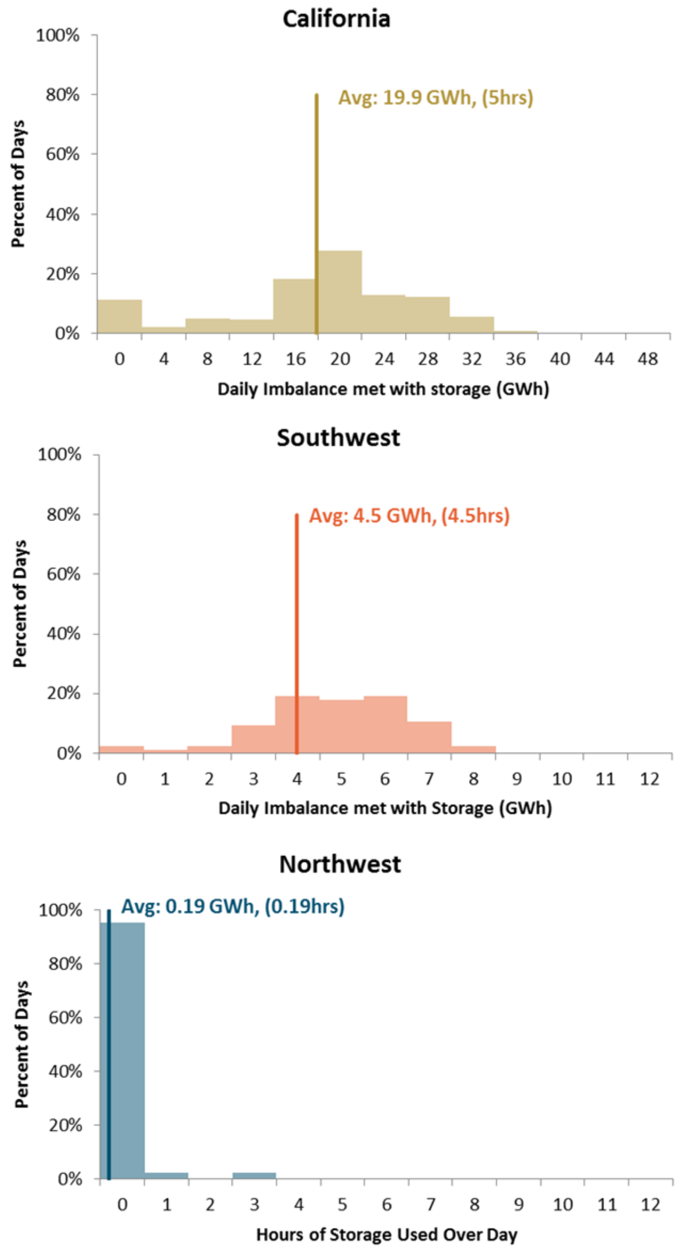


To gain additional insight into the demand for energy storage capability in the three high curtailment regions, the storage dispatch in the 12-hr Storage scenario was used to characterize the distributions of the daily imbalances met with the long duration device. These distributions are shown in Figure 93. In both California and the Southwest, the majority of days utilize less than 9 hours of energy storage capability, with an average utilization of 5.0 hours in California and 4.5 hours in the Southwest. In both California and the Southwest, the energy storage utilization is driven by the size and duration of midday solar oversupply events and the duration of the evening peak. In contrast, the energy storage resource in the Northwest has little to no utilization in nearly all modeled days, resulting in an average utilization of 0.19 hours. The limited utilization of energy storage in the Northwest further indicates that imbalances in the Northwest are driven by daily constraints rather than hour-to-hour fluctuations, which can largely be managed by the hydro resource. Note that while the region-wide challenges in the Northwest suggest a limited role for daily energy storage, energy storage resources operated within specific balancing areas may have heavier utilization depending on local resources and operating constraints.

In general, the storage utilization distributions support the finding that energy storage resources may have declining marginal benefits in solar-dominated regions as the maximum state of charge increases, particularly above 6-8 hours of discharging capability. The storage utilization distributions show that there are very few days in which more than 8 hours of discharging capability would be utilized even if it were available. However, two important caveats apply to these observations:

- Since the reserve contribution of energy storage resources was limited to the discharging range, this analysis potentially underestimates the benefits provided by shorter duration battery storage devices that can provide reserves over the full range from maximum charging to maximum discharging. If fully utilized, this enhanced reserve capability may allow some systems to unload conventional resources that would otherwise be needed for providing reserves, making additional room on the system to accommodate renewable energy. These benefits may not show up as increased storage utilization for managing hourly imbalance, but may reduce curtailment nonetheless; and
- While this study has investigated energy storage systems with independent discharge capabilities and maximum states of charge, real energy storage resources may have fewer options with respect to duration than have been explored here. For example, the duration of a pumped storage facility will depend on the reservoir topography among other factors and the duration of a battery system may be chemically constrained. The observations that a 6-hr storage device has significant benefits over a 2-hr device and that the marginal benefits of going from a 6-hr to a 12-hr device are small do not imply that all 6-hr energy storage devices will be more cost effective than all 2-hr and 12-hr energy storage devices. The costs and benefits of energy storage options should be evaluated for specific systems with full accounting of the specific energy storage resource costs and capabilities to make this determination on a case-by-case basis.

Figure 93. Distribution of energy storage utilization across draws in each region. 1 hour of storage utilization indicates that the stored energy over the course of the day was capable of being discharged at the maximum discharge rate for 1 hour.



5.4.2.1 Sensitivity: Energy Storage with Full Regional Coordination

One of the important dynamics to understand is how potential renewable integration strategies may interact with one another. As each individual strategy will exhibit diminishing returns with scale, so too will the benefits of combining strategies be smaller than the sum of their values when examined independently. To explore this dynamic, the 12-hr energy storage case discussed above was combined with the assumptions underlying the increased regional coordination strategy to illustrate the interactions between these two. The impacts of the energy storage resources on renewable curtailment are summarized in Table 47 for both the Reference Grid assumptions (Historical Intertie Limits) and the scenario with enhanced regional coordination (Physical Intertie Limits).[51]

[51] Note that in the 12-hr energy storage scenario with Physical Intertie Limits, the flexibility reserve provisions were not modeled for the energy storage devices, so the total curtailment may be overestimated, but the relative impact of excluding reserves is anticipated to be very small.

Table 47. Impact of energy storage resources on renewable curtailment by region under the Reference Grid assumptions (Historical Intertie Limits) and with more enhanced regional coordination (Physical Intertie Limits).

Intertie Limits	Storage	Basin	California	Northwest	Rockies	Southwest	WECC-wide
Historical	Reference	0.4%	8.7%	5.6%	0.6%	7.3%	**6.4%**
	12-hr Storage	0.4%	5.7%	5.8%	0.6%	5.1%	**4.7%**
	Difference	*0.0%*	*-2.9%*	*+0.2%*	*0.0%*	*-2.2%*	***-1.7%***
Physical	Reference	0.5%	3.0%	2.0%	0.5%	6.1%	**3.0%**
	12-hr Storage	0.4%	1.4%	1.5%	0.5%	3.6%	**1.7%**
	Difference	*-0.1%*	*-1.6%*	*-0.5%*	*0.0%*	*-2.5%*	***-1.3%***

In general, relaxing the constraints on the interties lessens the impact of energy storage on renewable curtailment across the West, largely because the system experiences curtailment conditions less frequently when it can more fully utilize the interties. In California, for example, the 4,000 MW of 12-hr energy storage avoids 2.9% curtailment when historically-based intertie constraints are imposed and only 1.6% curtailment when these constraints are lifted. These findings support the expectation that the presence of renewable integration solutions may reduce the value proposition of additional solutions if the renewable penetration (and hence the size of the market for integration solutions) remains fixed. By this logic, the optimal amount of energy storage is expected to be larger in a Balkanized system than a more fully aggregated system that makes use of diverse loads and resources over larger geographic areas. While this tradeoff has been explored in this study for the five large regions across the Western Interconnection, a growing body of research has found that this tradeoff tends to hold generally – that smaller balancing areas face larger renewable integration challenges than larger balancing areas. A direct implication for this study is that regions without significant curtailment or energy storage benefits identified herein may still benefit from integration solutions like energy storage if coordination within the region is not adequate to make full use of the internal resource diversity and flexibility.

A secondary and seemingly contradictory observation can be made in the Northwest, where increased regional coordination appears to enhance the effectiveness of energy storage resources. While curtailment increases slightly in the Northwest by adding storage resources under the Reference Grid assumptions, the energy storage resources are found to reduce curtailment in the Northwest when intertie limits are relaxed to the full path ratings. As is described in Section 5.4.1, relaxation of the intertie flow constraints allows the Northwest to take advantage of the nighttime market for zero marginal cost energy in solar-dominated regions. With increased nighttime exports, the curtailment remaining in the Northwest region tends to occur during daytime hours, leading to a diurnal imbalance in the Northwest that is driven by solar generation in other parts of the West. Daily energy storage resources, which cannot avoid a day-long energy imbalance, but can help mitigate daytime curtailment, are therefore more utilized in the Northwest under the enhanced regional coordination assumptions. This regional benefit, however, is offset by a reduction in the avoided curtailment in the rest of the Western Interconnection, so that the total impact of regional coordination on the value of energy storage for renewable integration remains negative.

5.4.3 INVESTMENTS IN GAS FLEXIBILITY

Flexible natural gas resources may also provide benefits to systems with increased renewable penetrations. Improved flexibility in the gas fleet may be pursued through a variety of avenues, including:

- Investments in new flexible gas generation resources. Many jurisdictions in the West are increasingly considering procurement of more flexible gas resources, including aeroderivative combustion turbines, reciprocating engines, and flexible combined cycle units. These resources may be considered for helping to meet a traditional planning reserve margin as load growth and coal retirements lead to anticipated capacity shortages around the West, while providing additional flexibility benefits. They may also be considered for procurement in excess of the traditional capacity need in order to supplement a system with additional flexibility. These new units can provide additional flexibility through low minimum stable levels, high maximum ramping rates, short minimum up and minimum down time requirements, and/or quick start capability.
- Retrofits or refurbishments of existing units. For systems that do not face capacity shortages, but do require additional flexibility, retrofits or refurbishments may be considered for existing units to enhance their operating flexibility. In particular, these updates may aim to reduce minimum stable levels, increase maximum ramping rates, and shorten start times.

This study examines the impacts of increased gas flexibility by testing the system with new flexible gas resources built on top of the existing fleet. While a wide range of characteristics differentiate the available flexible gas technologies described above, this study focuses specifically on the benefits of flexible combined cycle units, which combine relatively low minimum stable levels and high ramp rates with low heat rates to potentially displace both inflexible but relatively inexpensive gas plants and flexible but high heat rate plants. The operating parameters for the new flexible combined cycle units are listed in Table 48 and are juxtaposed against the fleet-wide average operating parameters for combined cycle units and combustion turbines in the Common Case. Incremental build of even more flexible combustion turbines and reciprocating engines was not considered largely because these resources are intended to manage the relatively large subhourly fluctuations that arise in smaller systems with renewables, which are largely mitigated in this study by aggregation of loads and resources across each region. Exclusion from this analysis therefore does not preclude substantial renewable integration value for these resources under specific circumstances.

Table 48. Generator specifications for the new flexible combined cycle (CCGT) units added in the Gas Flexibility Analysis, shown against the fleet-wide average unit specifications for CCGTs and combustion turbines (CTs) in the Reference Grid.

Parameter	Existing CCGT Fleet	Existing CT Fleet	New Flexible CCGT
Maximum output (MW/unit)	375	57	500
Minimum stable level (% of maximum output)	51%	41%	30%
Maximum ramp rate (% of Pmax per min)	0.9%	4.9%	1.7%
Minimum up time (hrs)	8.0	3.1	1
Minimum down time (hrs)	4.5	2.4	1
Heat rate at Pmin (kBtu/MWh)	8,117	13,152	8,000
Heat rate at Pmax (kBtu/MWh)	7,374	10,248	7,000

In order to make a direct comparison with the energy storage scenarios, the same capacity additions were applied to each region in both the energy storage and the flexible gas scenarios (4,000 MW in California, 1,000 MW in the Northwest, and 1,000 MW in the Southwest). The impact of the new flexible gas resources on renewable curtailment are summarized in Table 49. In all regions, the impact of the flexible gas resources on curtailment is less than 1%, or within the margin of error. This finding suggests that the curtailment observed in this study is not being driven by limited flexibility in the gas fleet.

Table 49. Impact of new flexible gas resources on renewable curtailment in each region (as % of available renewables).

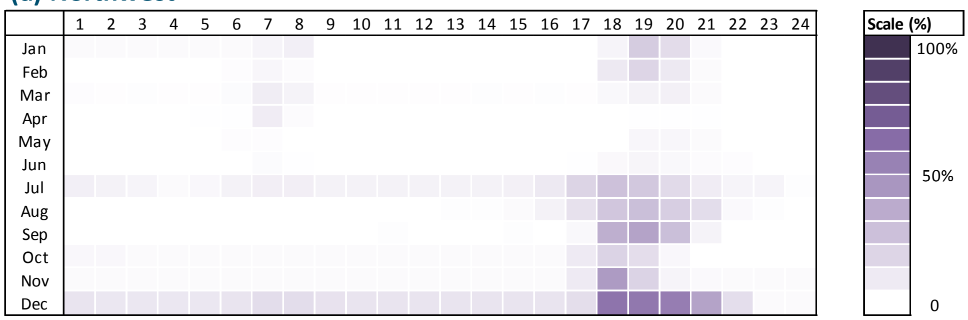
Scenario	Type	Basin	California	Northwest	Rockies	Southwest
Reference Grid	Scheduled	0.11%	8.65%	5.59%	0.10%	6.36%
	Subhourly	0.30%	0.04%	0.01%	0.47%	0.92%
	Total	0.41%	8.69%	5.60%	0.57%	7.28%
Flexible Gas	Scheduled	0.10%	8.64%	5.55%	0.10%	6.39%
	Subhourly	0.30%	0.04%	0.01%	0.48%	0.92%
	Total	0.40%	8.68%	5.56%	0.57%	7.31%
Difference	**Scheduled**	**-0.01%**	**-0.01%**	**-0.04%**	**0.00%**	**+0.02%**
	Subhourly	**0.00%**	**0.00%**	**0.00%**	**+0.01%**	**0.00%**
	Total	**-0.01%**	**-0.01%**	**-0.04%**	**0.00%**	**+0.02%**

The seasonal and diurnal dispatch patterns for the flexible CCGTs (Figure 94) suggest these units are used largely to meet the relatively high net loads during the shoulder hours in the solar-dominated California and Southwest regions. In the Northwest, these units have much lighter utilization, likely due to the vast flexibility of the hydro fleet.

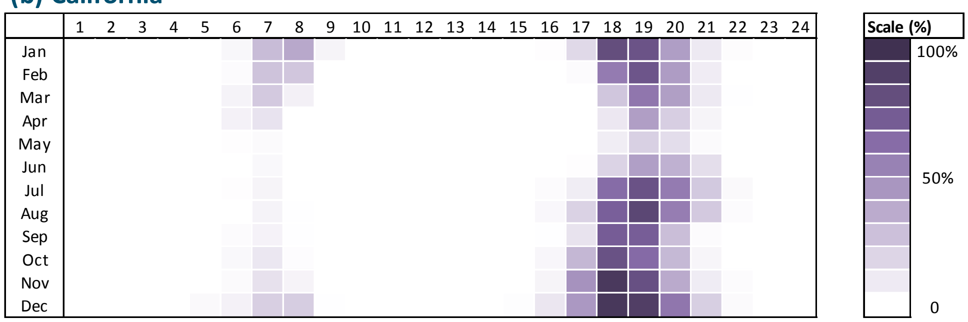
Despite the fairly frequent utilization of the new flexible CCGT units in California and the Southwest in the Flexible Gas scenario, there is not an appreciable impact of these units on either cost or emissions across the study area. Similar to the curtailment impact, the change in both cost and emissions between the High Renewables Case and the Flexible Gas Sensitivity are within the margin of error.

Figure 94. Average capacity factor by month-hour for new flexible CCGT units in the Flexible Gas scenario.

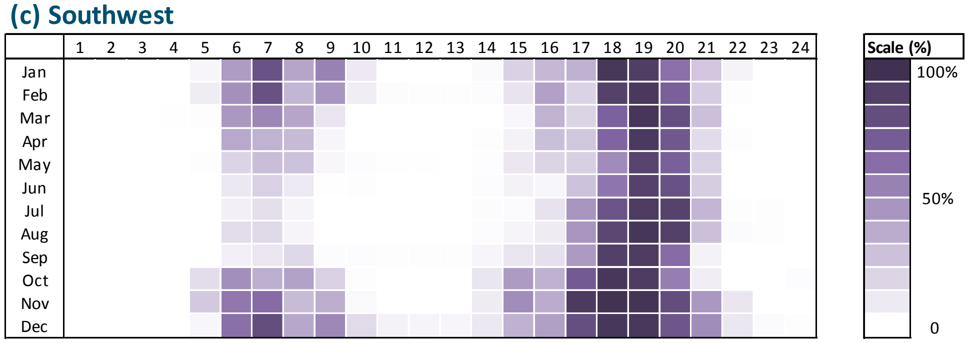
(a) Northwest



(b) California



(c) Southwest



6 Conclusions

6.1 Summary of Technical Findings

The technical findings from this analysis largely reinforce conclusions reached by prior studies of high renewable penetrations; namely:

- Operating a system reliably at high penetrations of renewable generation is technically feasible;
- Renewable curtailment plays a key role in operating electric systems at high renewable penetrations;
- Regional coordination offers a low-hanging fruit as an enabling strategy for renewable integration; and
- Measures that increase an electric system's capability to serve loads during low net load conditions have the greatest potential to ease integration challenges.

While the findings are shared with other technical studies, this work provides enhanced detail on the nature of many of these conclusions to illuminate challenges that emerge as electric systems move to higher penetrations.

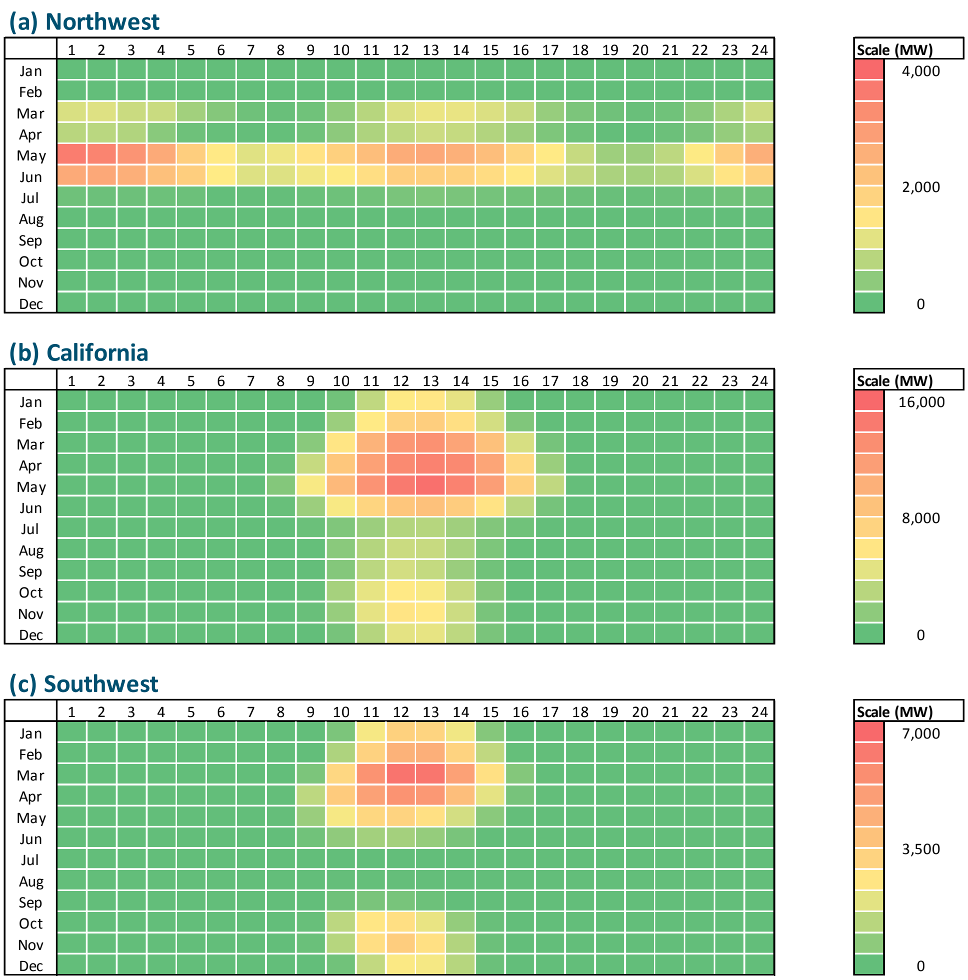
Achieving current policy goals requires modest adjustments in how systems operate on a day-to-day basis. In most jurisdictions, the Common Case renewable portfolio causes a modest change in the shape of net loads. Resources that have traditionally operated in a baseload capacity at high capacity factors (coal and nuclear) continue to do so; gas and hydro resources provide most of the flexibility needed to follow the net load signal from one hour to the next; and nearly all renewable generation is delivered to the system. These low levels of renewable penetration do not challenge the flexibility of the system significantly or frequently, and existing institutions and practices appear adequate to integrate these levels of renewables on to the system.

Integrating high penetrations of renewable generation while serving load reliably presents new challenges for operations, but it is technically feasible. The addition of large quantities of renewable generation throughout the region in the High Renewables Case imposes new challenges on system operators, in particular: (1) systems operate to meet an expanded range of net load conditions; (2) systems must operate to meet much larger hour-to-hour ramps in net load; and (3) systems must adjust flexibly to accommodate the forecast uncertainty of wind and solar resources. Across all scenarios and sensitivities of the High Renewables Case examined throughout the study, each region's system shows itself capable of accomplishing these three tasks without experiencing an increase in the risk of reliability events relative to the Common Case.

Renewable curtailment, while negligible at low renewable penetrations, plays a key role in system operations at high penetrations. One of the premises of the model developed for this study is that renewable generation is dispatchable, and this is one of the key tools that allow the region's electric system to respond to the new demands placed upon it. Curtailment is observed in all regions under the High Renewables Case across all scenarios. Its role in facilitating operations at high renewable portfolios is multifaceted: (1) it helps mitigate oversupply events when renewable production exceeds the capability of a system to absorb it; (2) it helps soften upward and downward ramps in net load that must be met by traditional dispatchable resources; and (3) it provides operators with another mechanism to adjust operations between day-ahead and real-time scheduling processes; and (4) it offers a substitute for holding downward flexibility reserves, allowing for the commitment of fewer thermal resources to serve loads. The prevalence of curtailment across the cases considered is indicative of systems that would encounter challenges balancing generation without load in its absence.

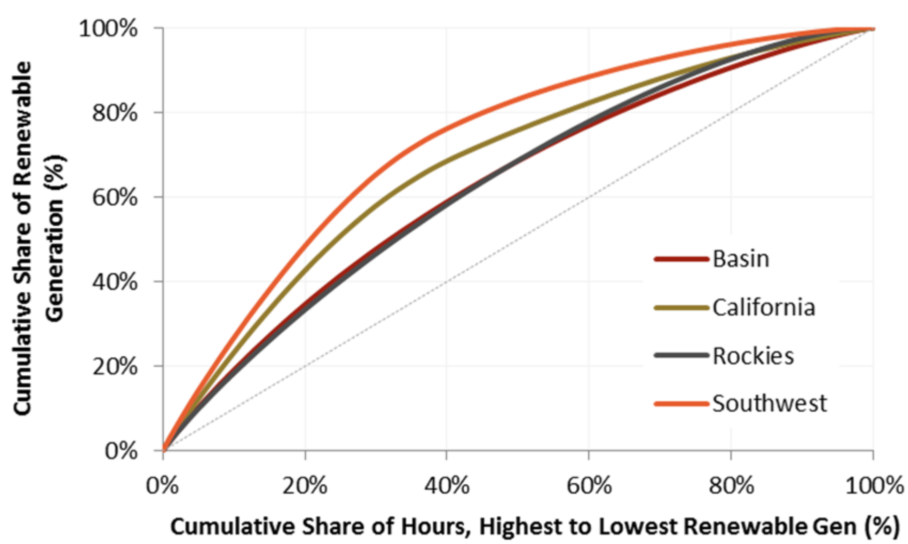
The nature of integration challenges experienced in the High Renewables Case vary from one region to the next and depend on the seasonality of load, the composition of the renewable portfolio, and the characteristics of non-renewable generators. In the Northwest, renewable curtailment is observed predominantly in the spring months and results from the coincidence of the spring hydro runoff with relatively low loads and periods of high wind production. This result is highlighted in Figure 95a.California and the Southwest (Figure 95b and c, respectively), whose portfolios rely predominantly on solar PV, experience frequent curtailment in the middle of the day throughout the year due to the concentration of solar output during these periods.

Figure 95. Heat maps for average renewable curtailment, High Renewables Case.



In the Basin and Rocky Mountains, curtailment is observed relatively infrequently. What distinguishes these regions from California and the Southwest—all of which rely predominantly on thermal generation to meet net load—is the composition of their respective renewable portfolios. In the Basin, the renewable portfolio combines significant quantities of geothermal, solar PV, and wind resources to yield a technologically diverse set of renewable resources whose output is distributed across much of the year. The Rocky Mountain portfolio comprises primarily wind resources, whose geographic dispersion similarly distributes generation more uniformly throughout the year. The effect of this relative uniformity of production limits the size and frequency of major oversupply events such as those observed in the solar-heavy California and Southwest regions. The diversity inherent in the portfolios of the Rocky Mountains and the Basin is compared with the California and Southwest portfolios in Figure 97.

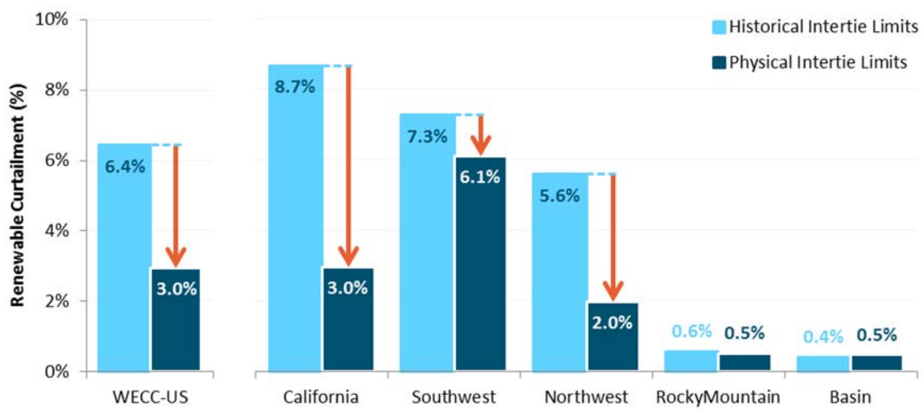
Figure 96. Distribution of renewable generation throughout the year for the Basin, California, Rockies, and Southwest regions.



The flexibility with which coal plants can be operated in the future is another key factor with a significant impact on renewable integration challenges. Three regions—the Basin, the Rocky Mountains, and the Southwest—currently rely on a large fleet of coal generators that have historically operated in a baseload capacity to serve load throughout much of the year. As the penetration of renewables in these regions increase, market signals will exert pressure on the coal fleet to operate more flexibly than it has historically. The degree to which this is technically and economically feasible will dictate the magnitude of the integration challenges each region faces.

Improving regional coordination offers a low-hanging fruit among integration strategies. There are many options for improving regional coordination, which include measures as simple as improving existing scheduling processes and as comprehensive as balancing authority consolidation. Without attempting to model any specific measure to foster regional coordination, this study provides two useful bookend scenarios to illustrate the value of harvesting the full diversity of loads and generation resources across multiple regions in the Western Interconnection: (1) a first in which interregional flows are limited to the historically observed range; and (2) a second in which limits on flows are relaxed up to the full physical rating of the existing transmission paths. The difference between these scenarios is stark: renewable curtailment is reduced from 6.4% to 3.0% through improved regional coordination. Figure 97 illustrates each region’s benefits achieved through reductions in renewable curtailment.

Figure 97. Regional reductions in renewable curtailment achieved through improved regional coordination.



While enhanced regional coordination provides an obvious benefit to regions for whom it reduces curtailment, it can also produce substantial value to those entities on the other side of export transactions. Specifically, by purchasing surplus generation from neighboring regions or balancing authorities that would otherwise curtail that generation, a region may reduce its own cost of serving load by reducing its own fuel purchases and operations and maintenance costs. Thus, regional coordination creates mutual beneficiaries: “sellers” reduce the cost to comply with clean energy goals, and “buyers” receive low-cost power to displace traditional generation resources.

Measures that expand an electric system’s ability to meet very low net load conditions ease renewable integration challenges. Existing resource adequacy rules and the enforcement of planning reserve margins in most jurisdictions in the West ensure that electric systems are capable of meeting demands during the highest load periods of the year; however, many jurisdictions do not have existing processes defined to ensure systems are capable of dispatching to meet low net loads. A number of such measures are explored in this study, including (but not limited to):

+ **Subhourly renewable curtailment.** Enabling subhourly curtailment increases the downward flexibility of an electric system as a whole by eliminating the necessity to hold certain quantities of downward flexibility reserves with generation resources operating above their minimum stable level.
+ **Investment in energy storage.** Particularly when coupled with a renewable portfolio with high penetrations of solar generation, whose regular diurnal output is well suited for shifting to off-peak periods, energy storage can be used to mitigate oversupply and to provide ramping capability during periods of extreme net load ramps.
+ **Ensuring sufficient downward flexibility of the thermal fleet.** The sensitivity analysis of coal flexibility serves to highlight the potential value of downward flexibility from thermal resources. Existing coal generators may be capable of operating in new ways to accommodate renewable generation, but may incur additional costs and experience degradation doing so. In the event that coal resources cannot operate with the degree of flexibility envisioned in this study—or to the extent that additional coal plants retire, requiring replacement capacity—procurement of new flexible gas resources could provide renewable integration benefits in these regions.

Those measures explored within this study are not intended to present a comprehensive palette of integration strategies; many other measures on both the supply and demand side could play a role in facilitating renewable integration, including demand response, flexible loads, vehicle electrification, TOU rates, etc.

6.2 Implications for Flexibility Planning

This study was envisioned not only as a means to characterize flexibility challenges for the Western fleet under high penetrations, but to identify best practices for these types of analyses as well as areas where additional exploration is necessary. Continued refinement of modeling techniques and constructs explored herein will serve to highlight renewable integration challenges and tradeoffs. A number of the topics identified here are explored in further depth in Section 7, Technical Lessons Learned.

One of the key factors that distinguishes this study from traditional production cost analysis of high renewable penetrations is the use of Monte Carlo production cost modeling to capture conditions that span multiple years. Typically, production cost analyses are used to examine a single calendar year; however, choosing a year that is “typical” or “representative” with respect to all conditions (load, wind, solar, hydro) is a difficult task, and a single typical year may fail to capture outlier events that occur at the tails of the distribution but which strain a system’s flexibility. The Monte Carlo day sampling used in this study has the advantage of allowing it to make use of much larger historical records of load, wind, solar, and hydro data.

Of course, this does raise a question of how many days of analysis are needed to characterize a result within a tolerable range of confidence, as there is an obvious tradeoff between the computational resources that can be brought to bear upon a problem and the accuracy with which it should be characterized. In this study, the five hundred draws that have been selected randomly according to the stratified sampling methodology appear to provide a reasonably representative characterization of the key metrics that are the focus of this study. This topic is explored in Section 7.1.

The technical results of this study also serve to highlight the importance of a number of key assumptions on the resulting renewable integration challenges identified. These topics reflect areas where, in the future, flexibility planners must carefully consider assumptions in their analysis, as impacts on results will be notable; in addition, these topics merit further investigation to enhance the understanding of the nature of renewable integration challenges.

- **How much should an entity rely on the market to help resolve flexibility challenges?** The contrast between scenarios with historical and physical limits on interregional power exchange highlights the significance of assumptions regarding each region's ability to import and export power. While many utilities have experience assessing the amount of imports they may rely upon during peak periods, in the future, utilities will be faced with the new challenge of determining the potential size of export markets.
- **How much flexibility can be provided by coal generators?** Historically, most coal generators in the West have operated at relatively high capacity factors, running at or near full output during much of the year. At high renewable penetrations, market signals will exert pressure on coal plants to operate more flexibly, but there are technical, economic, and institutional factors that may limit their abilities to do so.

In addition to these two key drivers of flexibility challenges highlighted by the results discussed in this report, several other important factors are further discussed in Section 7.3:

+ **What level of minimum thermal generation is necessary within a system for frequency response and voltage support?** In this study, a minimum generation constraint is applied only to the California fleet. This assumption is not intended to suggest that such constraints might not exist on other systems, but is merely a reflection of a lack of available information on their operating constraints, particularly as low net load conditions exert pressure on dispatchable fleets to reduce output as low as possible. Further work to identify and characterize such constraints will help to provide an enhanced view of potential integration challenges outside of California.
+ **How much flexibility can the hydro fleet be relied upon to provide?** This study conservatively limits the ramping capability of the hydroelectric fleet in each region to the range across which it has historically operated. Particularly in regions that rely heavily on hydroelectric generation for load service, this assumption may understate the flexibility of the existing system, and planners should seek to understand how constraints on the flexibility of hydro resources will impact renewable integration challenges.
+ **To what extent can neighboring entities be relied upon to provide hour-to-hour ramping services to help balance net load?** In this study, much of the reduction in curtailment that results from relaxing constraints on the interties results from the increased volume of interchange made possible by a wider range of limits on the interties. A more nuanced question that planners will need to confront is the degree to which ramping across the interties will be possible given the

bilateral conventions of power exchange in the Western Interconnection.

This study's regional approach provides a useful perspective on broad patterns that may characterize a region's operations as an integrated whole, and in doing so has made a number of simplifying assumptions regarding the transmission network and the degree of coordination that exists within each region with respect to operational practice. It remains true that procurement decisions and operations remain under the jurisdiction of individual constituent utilities and balancing authorities. Because of the simplifications and assumptions made in this work, it should not necessarily be assumed that the findings that apply to a single region in this work would equally apply to individual entities within it.

6.3 Policy Implications

The technical findings and conclusions reached through this study have a number of implications that are relevant for regulators and policymakers seeking to enable higher penetrations of renewable generation on the system and to ease the associated challenges.

Adjusting institutional practices and conventions to enable routine economic curtailment is a fundamental necessity to achieving high penetrations. Renewable curtailment serves as the relief valve that allows a system to operate reliably in spite of the increased demand for flexibility imposed by renewable generation. Ensuring that curtailment is available and can be used efficiently in day-to-day operations requires a number of steps:

+ **Market structures and scheduling processes must be organized to allow participation of renewable generators.** Within organized markets, this means ensuring that utilities can submit bids into the market on behalf of renewable generators that reflect the "replacement cost" of that resources as well as ensuring that renewable plants are not excessively penalized for deviations from their schedules due to forecast errors. In environments in which vertically integrated utilities or another type of scheduling coordinator is responsible for determining system dispatch, that operator must begin to consider the role of renewable curtailment in dispatch decisions and establish a "strike price" for curtailment that matches its renewable replacement cost.

+ **Contracts between utilities and renewable facilities must be structured to allow for economic curtailment.** Historically, many power purchase agreements have been set up to pay renewables for the generation that they produce and have included provisions limiting curtailment under the premise that limiting risk and ensuring an adequate revenue stream to the project are necessary to secure reasonable financing. In the future, contract structures must evolve to allow for economic curtailment while still providing a reasonably certain revenue stream to the developer. **Compensated curtailment,** under which developers are paid a PPA price both for generation that is delivered to the system as well for estimated generation that is curtailed, would be one means of achieving this goal. This model has a number of advantages: (1) it provides the developer with a reasonably certain cash flow that is not subject to market risk; (2) it places the risk of curtailment on the electric utility, who, as manager of an entire generation portfolio, is in a better position to evaluate and manage that risk; and (3) it aligns the resource's marginal cost to the utility with the resource's actual marginal cost to produce, which in turn encourages the utility's efficient use of the resource in the market or scheduling processes.

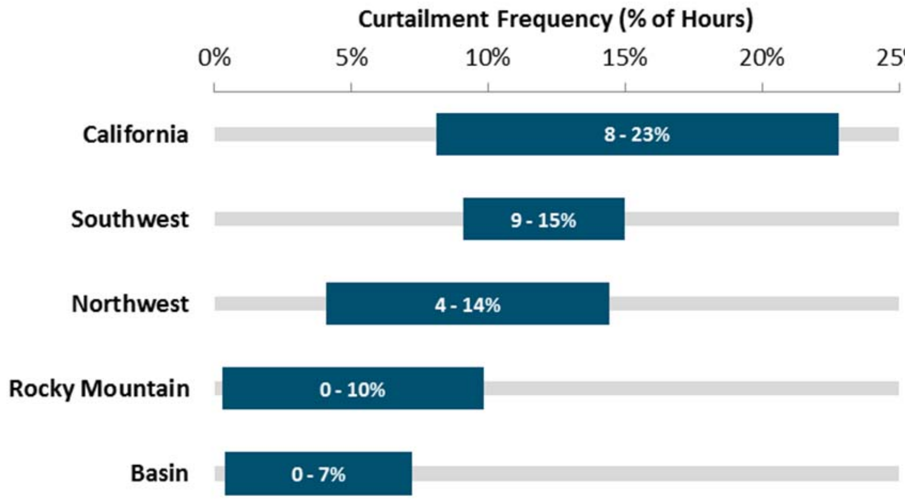
One step beyond ensuring that renewables can be curtailed at an hourly level is allowing participation in subhourly scheduling processes. Allowing renewable participation in scheduling and dispatch processes at a subhourly level can provide significant benefits in systems operating under high penetrations of renewable generation, as it allows the operator to avoid holding downward flexibility reserves with thermal resources. This is especially beneficial in systems that experience frequent curtailment at an hourly level, as it allows that system to operate its thermal resources at lower output levels, reducing the overall magnitude of renewable curtailment.

Another key step to enabling reliable and efficient operations under high penetrations is ensuring operators fully understand the conditions and circumstances under which renewable curtailment is necessary or desirable. In some instances—namely, in oversupply conditions—the need to curtail is relatively intuitive; however, in other instances, the important role of curtailment may not be so obvious. For example, an operator faced with a choice between keeping a specific coal unit online and curtailing renewables or decommitting that coal unit to allow additional renewable generation should make that decision with knowledge of the confidence in the net load forecast as well as an understanding of the consequences of possible forecast errors. Similarly, an operator anticipating a large upward net load ramp may decide to curtail renewable generation prospectively to spread the ramp across a longer duration if the ramp rates of conventional dispatchable units are limited. Additional work is necessary to identify such operating practices and conditions in which renewable curtailment may be necessary outside of oversupply conditions to ensure reliable service.

The role of operating reserves at avoiding unserved energy under unexpected upward ramping events must also be considered. Resources under governor response or Automated Generation Control (AGC) respond quickly to small deviations in net load. Contingency reserves (spinning and supplemental or “non-spin” reserves) are used to manage large disturbances such as the sudden loss of a generator or transmission line. Additional categories of reserve products—for instance, “load following” or “flexibility” reserves—have been contemplated at higher renewable penetration, but have not yet been formalized. How these reserves are deployed will impact the magnitude of challenges encountered at higher renewable penetrations.

With the prevalence of renewable curtailment, wholesale market signals in the future will transform radically from historical patterns. Figure 98 shows the observed frequency of hourly curtailment in each region across all scenarios examined under the High Renewables Case. In California, the region with the most curtailment, curtailment is observed between 8 and 23% of the hours of the year, implying that market prices could be negative for up to a quarter of the year. Even in regions where curtailment is limited (the Basin and Rocky Mountains), curtailment is observed up to 10% of the hours of the year. These markets appear very different from the historical paradigm under which the wholesale market has generally closely followed the avoided costs of thermal generation in the Western Interconnection.

Figure 98. Observed range of hourly curtailment frequency across all scenarios in each region.



With the frequent occurrence of curtailment at high penetrations, the consequences of extended periods of negative pricing must be examined and understood. Historically, the centralized markets and bilateral exchanges of the Western Interconnection have, for the most part, followed the variable costs of producing power—most often the costs of fuel and O&M for coal and gas plants. In a future in which renewable curtailment becomes routine, forcing utilities to compete to deliver renewable generation to the loads to comply with RPS targets, the dynamics of wholesale markets will change dramatically. How the dynamics of negative pricing ultimately play out remains a major uncertainty; nonetheless, with frequent low or negative prices in a high renewables future, utilities, other market participants, and regulators will be confronted by a host of new questions:

- How should generators that provide other services to the system during periods of low negative prices be compensated?
- How can the proper signal for investment in generation resources be provided as frequent negative prices further erode margins in energy markets?
- Do negative prices create new issues for loads, who, rather than paying for power from the wholesale market during periods of curtailment, would be paid to consume?
- At what point does the prevalence of negative prices lead to new policy mechanisms other than production quotas to promote the development of new renewable energy?
- How should retail tariffs for electricity service be designed with consideration for wholesale market signals?

These and other questions will require consideration as penetrations of renewables continue to increase.

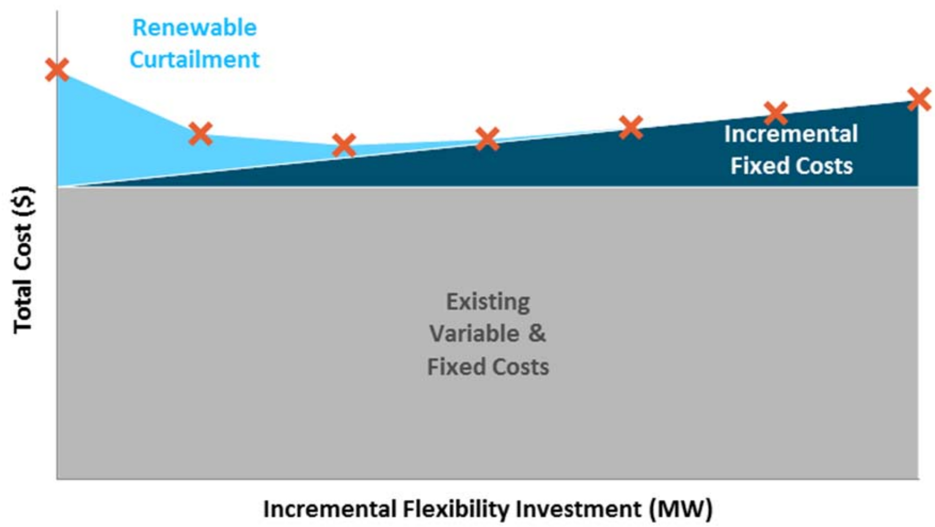
While renewable curtailment is identified as the predominant challenge in operations at high renewable penetrations, its magnitude can be mitigated through efficient coordination of operations throughout the Western Interconnection. Today's balkanized operations may act as an institutional barrier to efficient renewable integration; by allowing full utilization of the natural diversity of loads and resources throughout the Western Interconnection, regional coordination offers a low-hanging fruit to mitigating integration challenges. A number of studies have identified the significant operational benefits that can be achieved through balancing authority consolidation, a conclusion that is supported by the reduction in renewable curtailment at high penetrations identified in this study.

Many supply- and demand-side solutions merit further investigation to understand their possible roles in a high renewable penetration electric system. This study examines a select few of the multitude of possible supply- and demand-side portfolio measures available to utilities to illustrate how different attributes do (or do not) provide value to electric systems at high penetrations of renewable generation. The solutions examined within this study illustrate how different "types" of flexibility impact a system to differing degrees: whereas storage effectively mitigates renewable curtailment through its ability to charge during periods of surplus, fast-ramping flexible gas resources have a comparatively limited impact on operations, displacing less efficient gas generation resources but effecting minimal changes in curtailment.

The ability of renewable curtailment to serve as an "avoided cost" of flexibility points to an economic decision-making framework through which entities in the Western Interconnection can evaluate potential investments in flexibility and ultimately rationalize procurement decisions. As the need for operational flexibility has grown, a number of efforts have explored whether additional planning standards—analogous to those used for resource adequacy today—are necessary to ensure that when the operating day comes, the generation fleet is sufficiently flexible to do serve load reliably. As this study demonstrates, so long as (1) the generation fleet is capable of meeting extreme peak demands, and (2) the operator can use curtailment as a relief valve for flexibility constraints, the operator can preferentially dispatch the system to avoid unserved energy. Thus, the consequence of a non-renewable fleet whose flexibility is inadequate to balance net load is renewable curtailment, whose implied cost is orders of magnitude smaller than the cost of unserved energy. In this respect, the determination of flexibility adequacy is entirely different from resource adequacy: for resource adequacy, conservative planning standards are justified on the basis of ensuring that costly outages are experienced exceedingly rarely; for flexibility adequacy, the appropriate amount of flexibility for a generation system is instead an economic balance between the costs of "inadequacy" (renewable curtailment) and the costs of procuring additional flexibility.

Because renewable curtailment serves as an "avoided cost" of flexibility, the question of "flexibility adequacy" is economic, rather than technical. Renewable curtailment imposes a cost upon ratepayers, reflected in this study by the idea of the "replacement cost," and, to the extent it can be reduced through investments in flexibility, its reduction provides benefits to ratepayers. At the same time, designing and investing in an electric system that is capable of delivering all renewable generation to loads at high penetrations is, itself, cost-prohibitive. Between these two extremes is a point at which the costs of some new investments or programs that provide flexibility may be justified by the curtailment they avoid, but the cost of further investments would exceed the benefits. This idea is illustrated in Figure 99, which shows the tradeoff between the costs of renewable curtailment with the costs of a possible theoretical measure undertaken to avoid it.

Figure 99. Illustration of an economic framework for flexibility investment.



While not performed in the context of this study, this type of economic assessment of flexibility solutions to support renewables integration will depend on rigorous modeling of system operations combined with accurate representation of the costs and non-operational benefits of various solutions. The specific types of investments to enable renewable integration that are found necessary will vary from one jurisdiction to the next, but the overarching framework through which those necessary investments are identified may be consistent. Implementation of such an economic framework for decision-making for flexibility will foster the transition to high renewable penetration, enabling the achievement of policy goals and decarbonization while mitigating the ultimate impacts of those changes to the quality and cost of service received by ratepayers.

7 Technical Lessons Learned

One goal of this analysis was to glean insights into the nature of flexibility modeling in systems with high penetrations of renewables. In any production cost modeling analysis, many decisions must be made regarding the configuration of the model and the specific constraints applied to the system. Here we discuss some of the specific decisions made in this analysis, implications of these modeling decisions, and potential alternatives.

7.1 Convergence Behavior of Draws

One of the guiding principles of the REFLEX approach to flexibility analysis is that such a study should capture the full range of potential conditions for load, wind, solar, and hydro. In order to enhance subsequent efforts, this study seeks to inform the question of how many such randomly sampled draws are necessary to characterize these distributions fully in the context of ensuring accurate modeling results. While approximately 500 draws were analyzed across all the scenarios considered in this study, the team conducted further analysis on the 'Reference Grid' scenarios for both the Common Case and the High Renewables Case, simulating an additional 1,000 draws in order to provide a rich record of draws to inform this question.

The question of whether enough draws have been considered will depend on what result the modeling effort seeks to characterize. In the instance of loss-of-load-probability modeling (used in Phase 1 of this study), thousands of years of simulations are needed to achieve convergence of key reliability indicators (e.g. loss of load frequency, expected unserved energy), as they are observed exceedingly rarely. In contrast, total annual production cost—a key output commonly extracted from production simulation models—converges much more quickly, and should not require such extensive sampling to characterize with some degree of confidence. While the analysis yields many interesting results and metrics, curtailment is the primary metric used to quantify flexibility challenges in each region.

For the convergence analysis conducted here, we simulated more than 1,500 draws for the 'Reference Grid' scenario for the Common Case and the High Renewables Case. After optimizing operations for each day, we calculated the updated expected value of a range of model outputs. Figure 95 shows an example of the convergence behavior of curtailment as percentage of all renewables for the California, Southwest, and Northwest regions, updated after each draw (the Basin and Rocky Mountain regions are excluded as curtailment frequency in those regions is rare). In all three regions, this metric remains unstable for several hundred draws before eventually steadying when about a thousand days have been drawn. WECC-wide production cost, shown in Figure 101, appears to stabilize after only a couple of hundred draws.

Figure 100. Mean percent of renewable curtailment in the California, Southwest, and Northwest regions as a function of the number of draws ('Reference Grid' High Renewables Case).

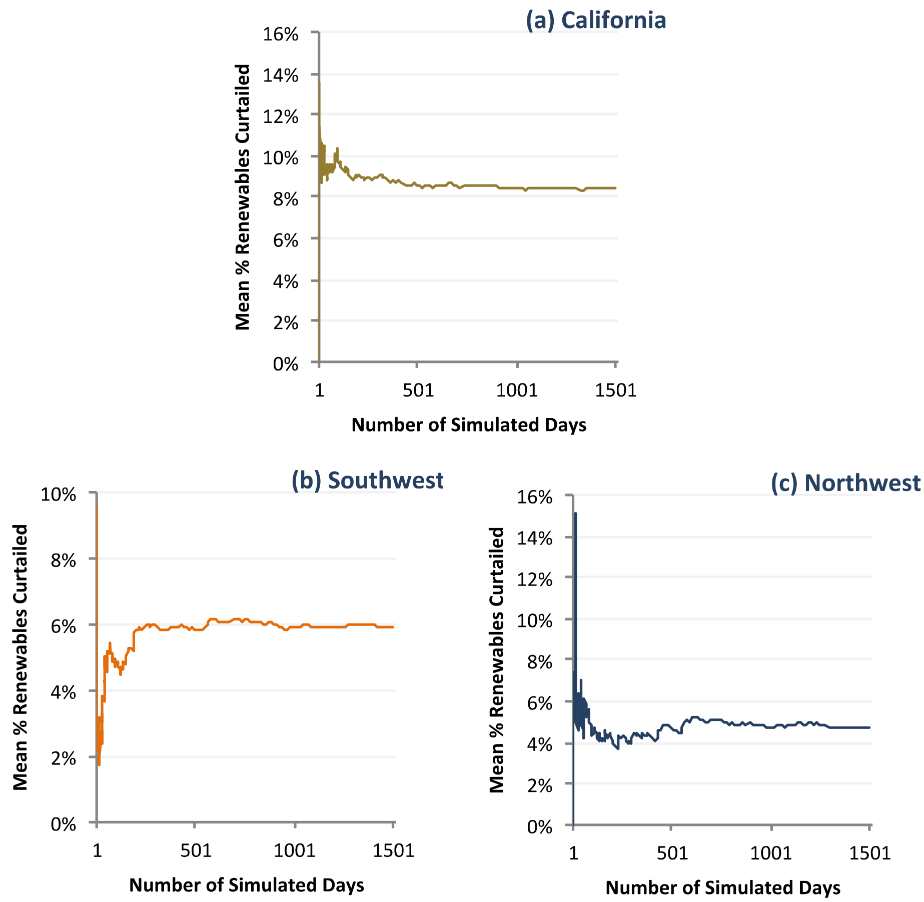
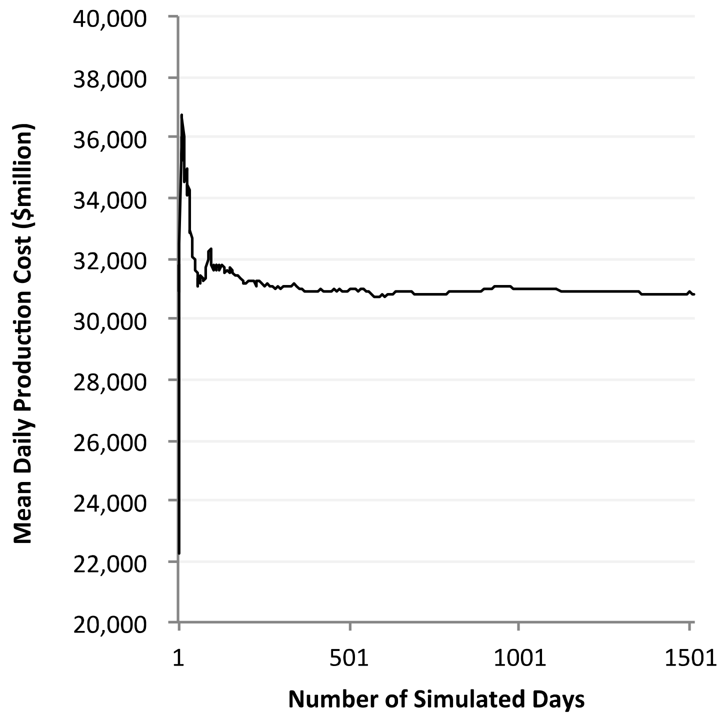


Figure 101. Mean WECC-wide production cost as a function of the number of draws ('Reference Grid' High Renewables Case).



To compare the level of convergence across different metrics and across regions, we calculated the relative standard error (the standard error divided by the sample mean) of several model outputs for each region, including the total daily production cost (fuel and variable O&M), the total daily curtailment, and the total daily upward and downward subhourly imbalances. The results of this analysis are shown in Figure 102 and Figure 103 for the High Renewables Case and Common Case respectively.

Of the four metrics analyzed, daily production cost converges fastest. Daily production cost by region was generally estimated to within relative standard error of 1 percent or less in the 1,500 draws simulated here in both the Common Case (Figure 102b) and the High Renewables Case (Figure 103b); in the Northwest, the final relative standard error for daily production cost after 1,500 draws was higher than in the rest of the regions, but less than 2 percent. The somewhat slower convergence observed in the Northwest is likely due to the variability in daily production cost introduced by different hydro conditions (hydro has zero operational costs), both within a given year and between years. The sample size needed for convergence of the production cost metric stays approximately the same between the Common Case and High Renewables Case. The relative standard error observed after 1,500 draws is on the order of half a percentage point higher in the High Renewables Case.

The scheduled hourly curtailment and subhourly imbalances metrics take longer to converge than production cost. After 1,500 draws, scheduled curtailment in the High Renewables Case converges to a relative standard error of 2 percent in California and 3 percent in the Southwest, the two regions with the most frequent occurrence of curtailment (Figure 103). On the other hand, as scheduled curtailment remains rare in the Basin and Rocky Mountain regions even in the High Renewables Case, the relative standard errors are 19 percent and 15 percent respectively after 1,500 draws. Curtailment in the Northwest is dependent on inter-annual variation in hydro and wind conditions. The final relative standard error in the Northwest is 9 percent, indicating the need for more than 1,500 draws to fully converge.

Downward subhourly imbalances converge faster than scheduled curtailment in the California, Southwest, Basin, and Rocky Mountain regions. In these regions, subhourly curtailment is frequently used in the High Renewables Case to provide economic downward flexibility, so this metric converges to a relative standard error of around 2 percent in the 1,500 draws modeled. In the Northwest, the flexibility of the hydro fleet in real time precludes frequent downward subhourly violations, and convergence of this metric is slower (Figure 103d). Similarly, upward violations are rare in all regions due to the high penalty on unserved energy, so achieving a high level of convergence requires an increase in sample size (Figure 103c).

The results in this study are based on approximately 500 draws and the convergence analysis presented here helps to inform the confidence in the conclusions regarding system flexibility needs and solutions. Of particular interest is ability to assess the level of curtailment, the main indicator of system flexibility challenges explored in this study.

A sample size of 500 draws appears sufficient to evaluate curtailment levels with a high degree of confidence in regions with high penetration levels of solar. The diurnal periodicity of solar output—in combination with the diurnal load pattern—results in a predictable curtailment pattern. Solar PV generation is concentrated in the middle of the day throughout the year. In these regions, curtailment is therefore frequent and routine, and can be characterized relatively well in 500 draws. In California, the relative standard error is 5% after 500 draws; in the Southwest, it is 6%. Additional draws are beneficial to increasing confidence in the curtailment levels in these regions, but the marginal benefit is relatively small.

In the wind- and hydro-heavy Northwest region, the inter-annual variability in wind output and hydro availability results in considerably lower confidence in the curtailment results. The relative standard error in a sample size of 500 draws is more than 15% (and remains at 9% even after 1500 draws). Additional draws are required to fully capture the underlying distribution of wind, hydro, and load conditions in the Northwest. The draw methodology offers clear advantages over using a single year of data for systems with large inter-annual variability in resource availability.

Figure 102. Convergence behavior or key model outputs by region in the 'Reference Grid' Common Case.

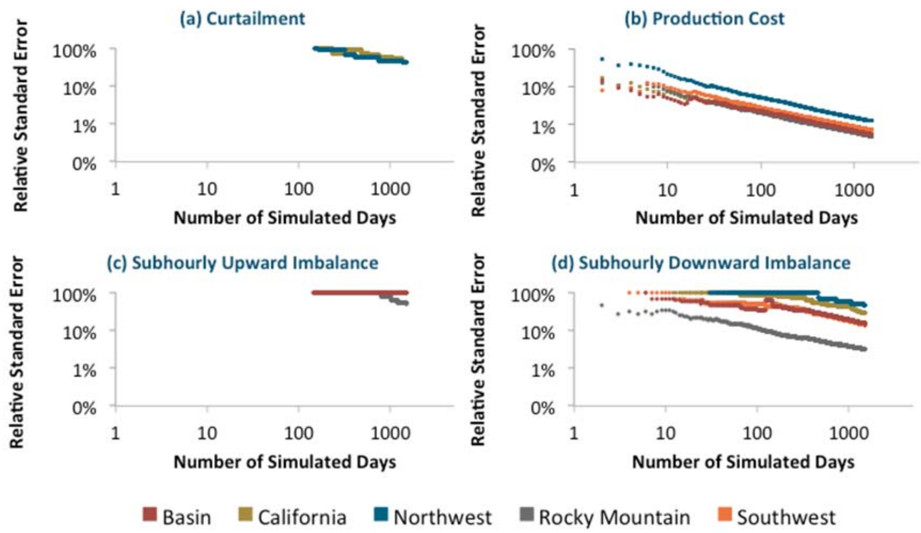
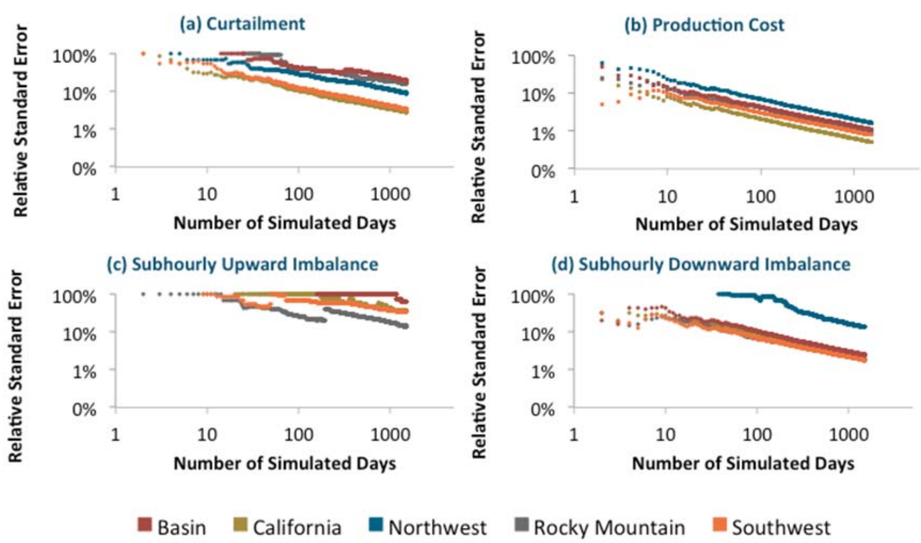


Figure 103. Convergence behavior or key model outputs by region in the 'Reference Grid' High Renewables Case.



7.2 Three-Day Draw Methodology

As with prior applications of the REFLEX model, this study relies on the analysis of a large number of three-day "draws," each of which is intended to represent a plausible combination of load, wind, solar, and hydro conditions. The first and third day are included in the simulation for the purposes of eliminating edge effects but are discarded when compiling results, leaving the middle day from each draw as a twenty-four hour unit for analysis. Through Monte Carlo sampling of conditions for each simulation, this approach is intended to provide a reasonably robust distribution of conditions intended to mimic the long-run distribution.

The approach of using draws of snapshot days used in REFLEX reflects the marriage of analytical techniques used in loss-of-load-probability modeling to capture robust distributions with the technical rigor of production cost modeling. Because load conditions and the output of wind and solar resources vary hour to hour, season to season, and year to year, representing each with multiple years of hourly profiles is helpful to ensure that a robust distribution of possible conditions have been captured. Historically, however, public multi-year time-synchronous data sets for load, wind, and solar have not been available. The stratified sampling method used in this study was designed to produce a much larger set of possible conditions on the system than have been measured simultaneously in the historical record. While this approach has the benefit of allowing for investigation of a broader set of system conditions, it also introduces some modeling challenges related to the treatment of correlations between variables and accurately capturing operational phenomena that occur over time scales that are longer than the three day period. These are discussed in this section.

7.2.1 RELATIONSHIP BETWEEN LOAD, WIND AND SOLAR

In order to construct draws for analysis, this study uses a stratified sampling methodology that matches historical load, wind, and solar profiles according to season, day-type, and load level. The sampling approach used in this study is intended to preserve key relationships and correlations among these variables. Preservation of the appropriate correlations will depend on how the load, wind, and solar data are binned, as described in Section 3.1.3. In general, using smaller bins improves the correlations between variables but decreases the number of possible unique draws and at some point potentially over-states the importance of specific historical days. Using larger bins allows for a richer set of sampled days, but may not accurately capture important correlations. The distributions shown in Figure 104 through Figure 108 compare the actual historical realized load & renewable conditions with those simulated by the draw methodology.

Figure 104. Frequency of load & renewable pairings for Common Case portfolio, Basin region.

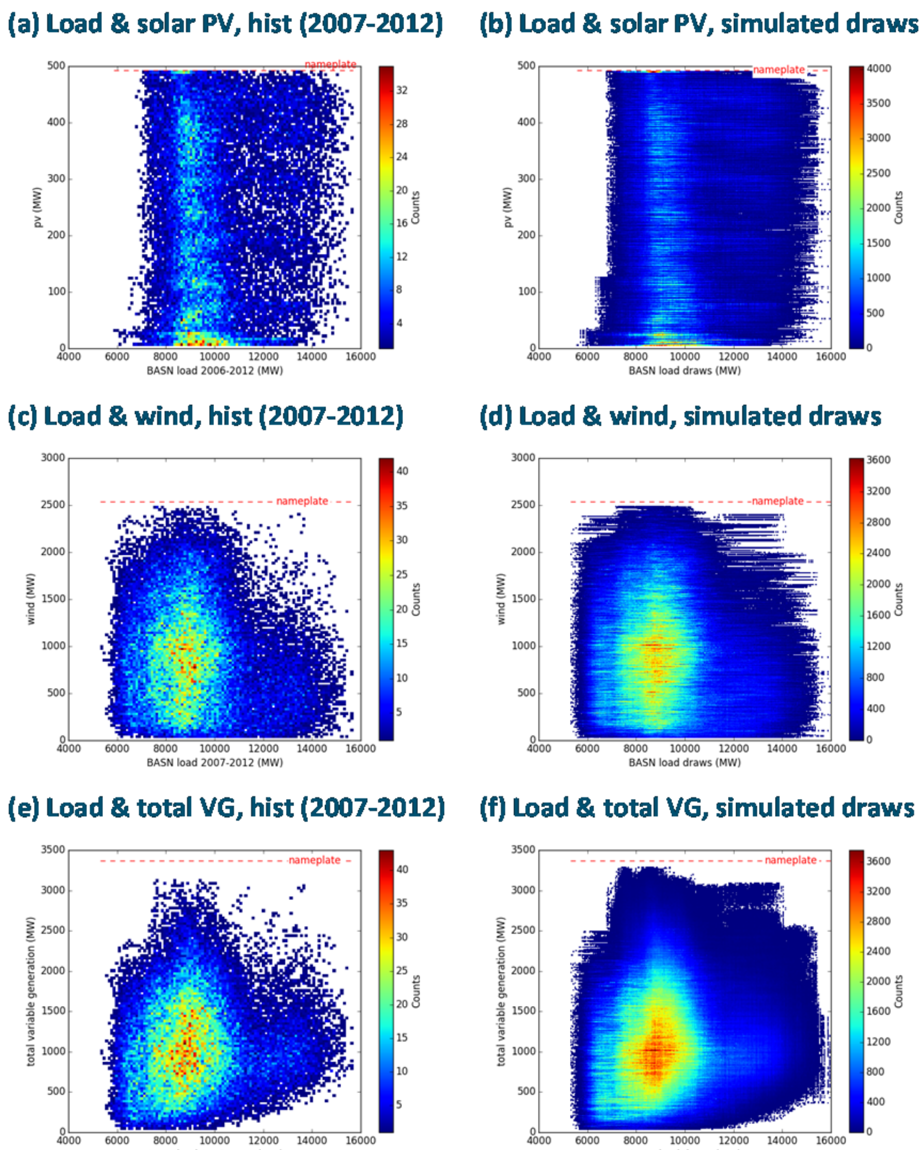


Figure 105. Frequency of load & renewable pairings for Common Case portfolio, California region.

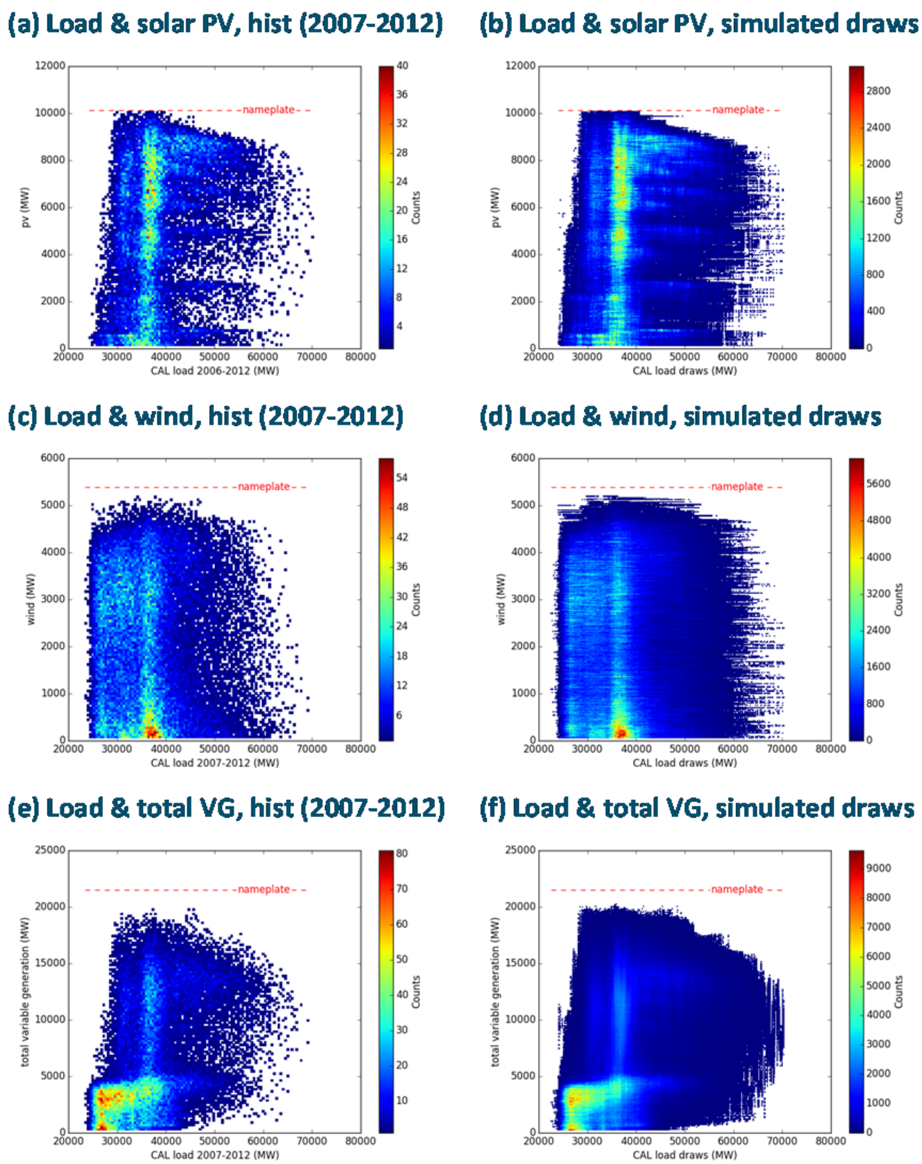


Figure 106. Frequency of load & renewable pairings for Common Case portfolio, Northwest region.

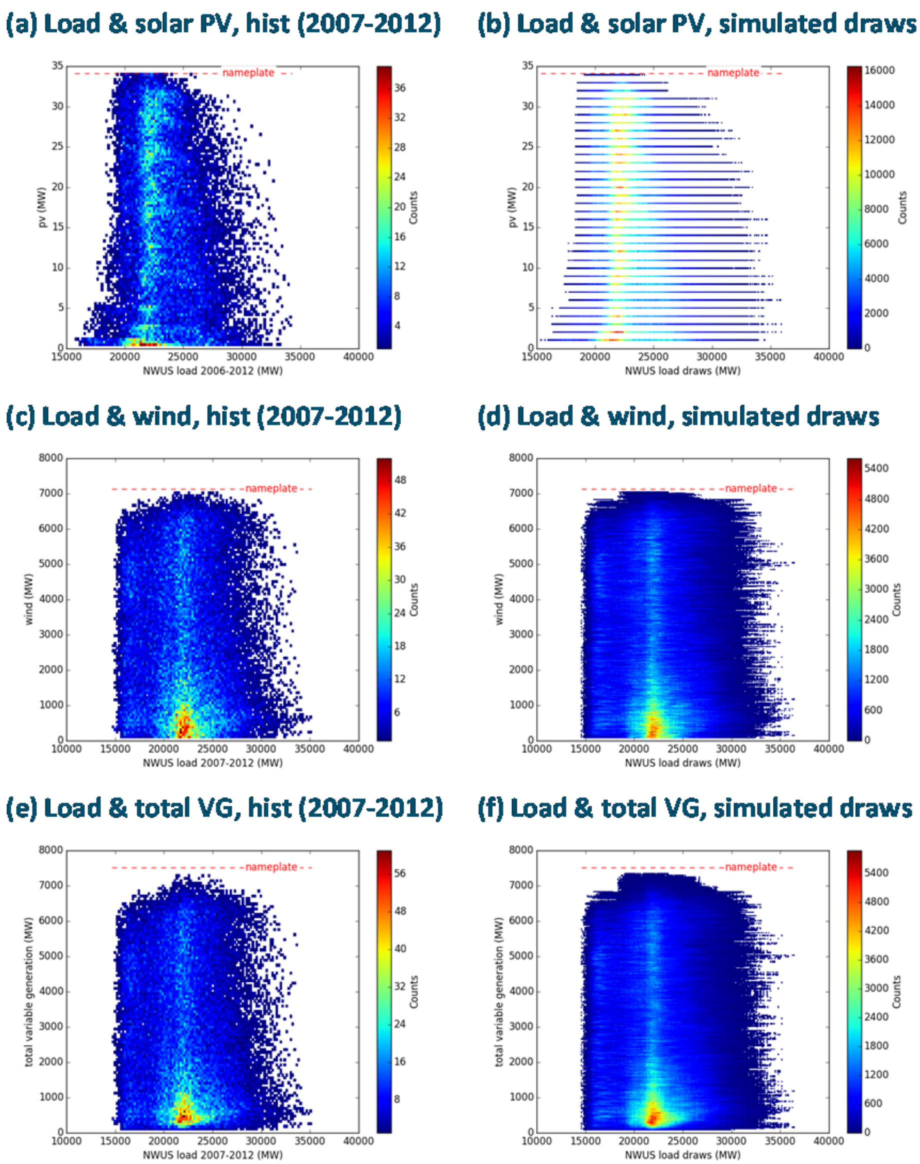


Figure 107. Frequency of load & renewable pairings for Common Case portfolio, Rockies region.

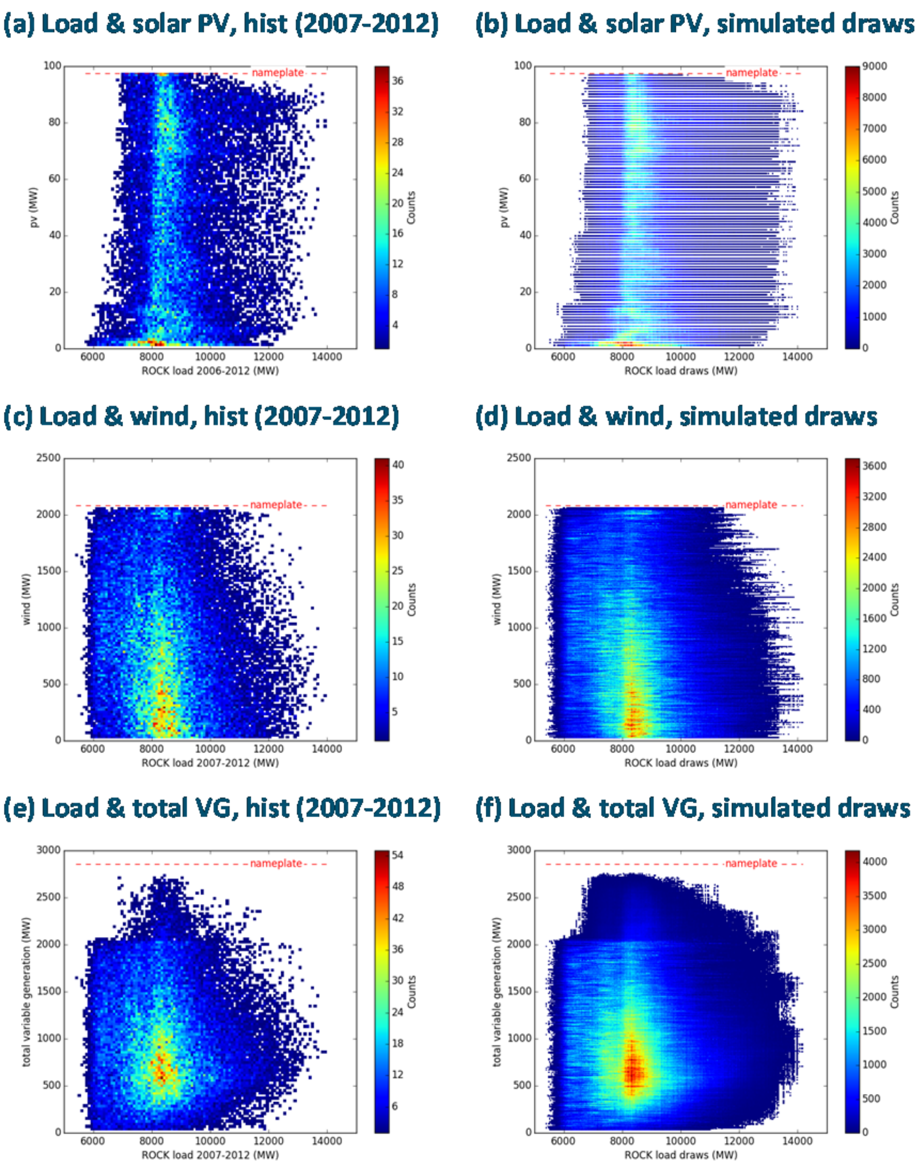
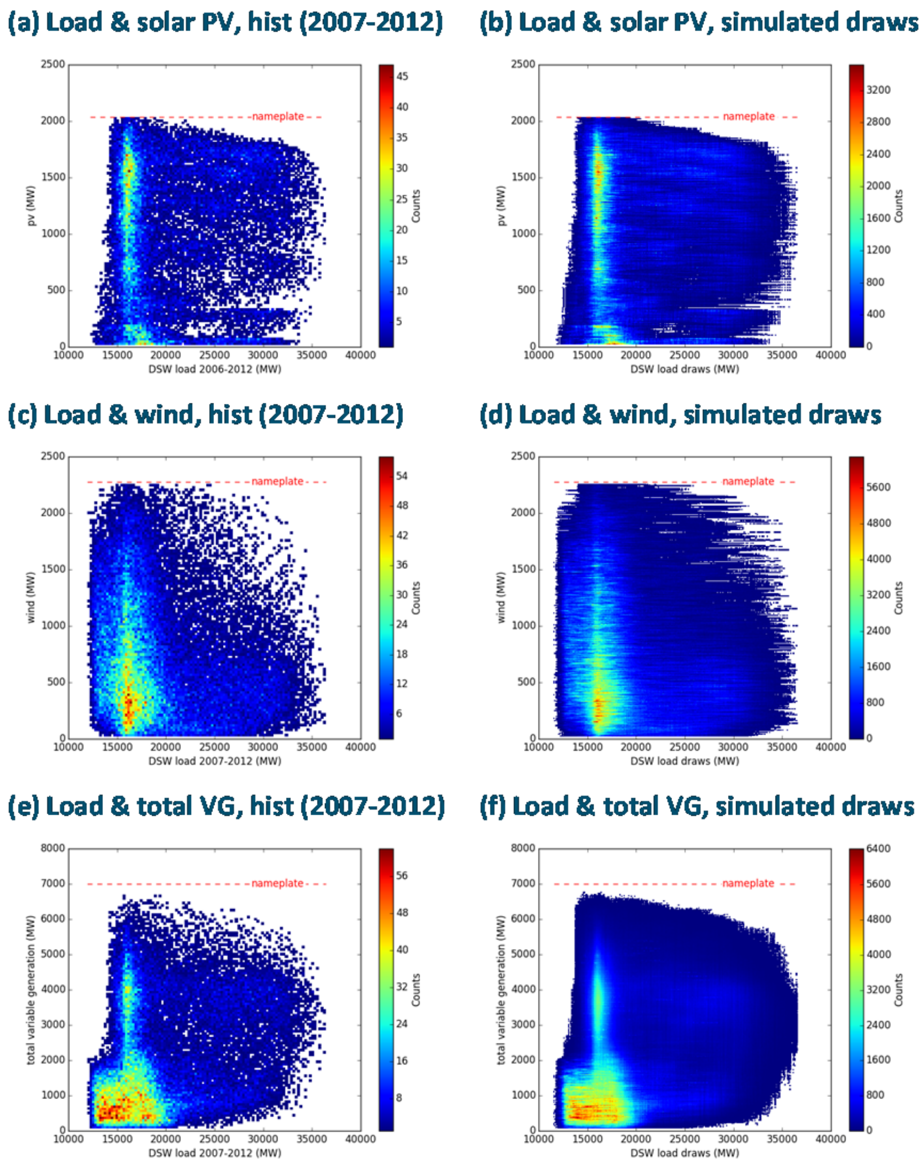


Figure 108. Frequency of load & renewable pairings for Common Case portfolio, Southwest region.



7.2.2 USE-LIMITED & INFLEXIBLE RESOURCES

The approach of analyzing a single operating day was originally conceptualized in the context of analyzing high penetrations of renewable generation on the California system, whose non-renewable generation fleet is composed predominantly of relatively flexible gas generators with smaller contributions from hydroelectric and nuclear resources. In this system, the treatment of individual draws as independent from one another was justified by the fact that gas resources can be turned on and off relatively quickly and have few constraints that impact their operation from one day to the next.

In other regions of the Western Interconnection, other non-renewable resources (nuclear, hydro, and coal) play a larger role in serving load. To the extent possible, this study captured as much fidelity as possible on the modeling of these units, but the application of the three-day window required simplifications in the modeling of hydro and coal generators.

7.2.2.1 Hydro

Modeling the operations of the hydroelectric fleet of generators in the Western Interconnection presents a challenge in any production cost exercise: while various hydro systems are somewhat responsive to wholesale market conditions and provide some flexibility to the electric system, they are also constrained in their ability to operate by a host of other environmental, political, and social factors. Accordingly, most models of the hydro system used in production cost analysis of the Western Interconnection are developed with the intent of mimicking the observed patterns of behavior of the hydro system in the historical record rather than explicitly modeling all of the constraints faced by hydro operators.

This study is similar in this regard—constraints for the hydro system's ability to generate and to ramp are derived through analysis of historical patterns of output as described in Section 3.3. As each draw comprises a period of three days, a corresponding energy budget to match the three-day period is developed for each draw based on hydro conditions for the corresponding month drawn from the historical record of 1970-2008. Due to the configuration of the optimization across each three day period in PLEXOS, it was necessary to allocate the three-day energy budget to each day. This was done by simply giving an equal portion of the three day budget to each day.

What this method fails to capture is the flexibility inherent in the hydro system to shift water between days in order to accommodate anticipated changes in net load conditions over time. Hydro operators are not forced to expend a fixed hydro energy budget each day, as this study assumes; rather, depending on the specific plant or system, they have some latitude to determine the amount to generate on any given day in response to the electric system conditions. The limited time horizon examined in each three-day draw does not allow for this type of inter-day flexibility to be considered directly by the optimization and the allocation of energy budgets within the three-day period further constrains the hydro inter-day operations.

As an alternative approach, improved characterization of hydro operational flexibility may be achieved by optimizing over longer periods – several days or weeks, for example. Configuring a production cost model in this way has two primary impacts. First, it may dramatically increase the problem size for each optimization step. Second, it overestimates the ability of the system to optimally plan for phenomena that will be experienced in the relatively distant future. If optimizing over a week, for example, the allocation of hydro energy to the first day will be based in part on the operational needs of the seventh day, which in real operations may be highly uncertain at higher renewable penetrations. Some modeling methods also involve a longer term (e.g. a full month) optimization to allocate hydro energy to shorter windows (e.g. days) in order to approximate the operator’s ability to move hydro dispatch across days without performing a full optimization of detailed dispatch across the longer step. Some methodologies also involve the ability to carry hydro energy imbalances from one optimization step to the next. This can occur over long optimization steps, like days or weeks, or over short time scales, like hours. With these types of configurations, there is a risk that a small amount of drift in the cumulative hydro energy could influence economic comparisons between cases.

It is often useful to test multiple configurations for both runtime performance and reasonableness of results. In this analysis both the day draw approach and the configuration options in PLEXOS limited the types of methodologies that could be tested for the hydro energy budget allocation. In future studies, the tradeoffs of the various approaches to modeling hydro systems should be considered based on the specific application.

7.2.2.2 Coal

The three-day simulation window also presents a challenge for inflexible baseload units with extended start and shutdown periods. For nuclear units, this is avoided by assuming that nuclear plants are must-run: the few nuclear units in the Western Interconnection rarely, if ever, cycle on or off in response to market conditions, as the cost of such cycling would likely be prohibitive. Coal generators present a challenge insofar as they offer more flexibility than nuclear units to adjust output or to cycle on and off but do not offer the same amount of flexibility as gas resources.

In the 2024 Common Case, to capture the limited flexibility of the coal fleet, TEPPC imposes minimum up and down times on coal generators to prevent frequent cycling on and off. Most plants are generically assumed to have a minimum up time of 168 hours (1 week) and a minimum down time of 48 hours (2 days). Start costs are also applied to discourage frequent unit cycling. The day draw methodology as it is implemented in PLEXOS limits the ability to impose these constraints and to account for the start costs in the objective function because each optimization does not "see" hours before and after the optimization step. For example, a unit may begin the three-day simulation in the on state without ever incurring the cost of being turned on and it may turn off in any hour throughout the three-day period because the optimization assumes that it has been on indefinitely. As a result, this study identified coordinated cycling behavior in the High Renewables Case in which some coal plants started the day on, only to shut down to accommodate mid-day solar generation, and other units would start up at sunset to meet the evening net load peak (see Section 5.3.1). The frequency of cycling implied by this behavior exceeds the maximum cycling frequency associated with the minimum up and down times in the Common Case database.

More accurate representation of the cycling limits of coal units could be achieved by optimizing over a longer period, like a week, though the boundary effect described above may still be encountered unless sequential weeks are linked in the optimization – so that the operations in each week are constrained based on the commitment and dispatch decisions of the prior week. Regardless of the length of the optimization step, the draw methodology, in which sequential draws are completely independent, does not allow for this linking of sequential periods. This limitation suggests that more accurate representation of coal flexibility may be achieved by performing a traditional sequential production cost modeling exercise. The remaining challenge in such an exercise would be to accurately reflect the probability distributions of variables that have significant inter-annual variability, like hydro conditions.

In solar-dominated systems, one approach to imposing constraints on coal cycling while also using the day draw methodology is to impose a periodic boundary condition on all inter-timestep constraints on the day of interest. This ensures that the generators behave on that day as if they had to meet the same demands in the prior day and will have to meet the same demands in the next day. While this assumption may be a reasonable approximation for solar-dominated systems, it may impose unrealistic constraints on wind-dominated systems, which may experience very different operational states at the beginning and end of a given day.

7.2.3 POSSIBLE ALTERNATIVE APPROACHES

The simplifications required in the modeling of coal and hydro generation for short time horizons suggest that a longer window for simulation may be appropriate to capture both the limitations and availability of flexibility from these units. Applying the stratified sampling methodology across a longer time period presents challenges in appropriately capturing correlations and reduces, but does not eliminate the boundary effects that impact coal cycling.

Refined methods are made possible by the fact that NREL's latest wind and solar PV libraries (the WIND and SIND datasets, respectively) each span the same historical period of 2007-2013—a period for which historical load profiles are also publicly available through WECC. The existence of data sets for these three variables that (1) coincide with one another, and (2) span a large number of years of historical conditions creates new possibilities for modeling techniques that could be used in this type of analysis.

With these datasets it may be possible to perform a single sequential production cost modeling exercise over the full seven year period with optimization steps long enough to improve the characterization of hydro and coal fleet flexibility and a historical record long enough to improve the characterization of key probability distributions over a single year analysis. Because of its high degree of inter-annual variability, it may be most challenging to reflect the long run distribution of hydro availability with only seven years of data. Two hybrid approaches may be used to address this concern: 1) time-synchronized load, wind, and solar data may be used to characterize hourly conditions but hydro availability could be drawn from the longer historical record; or 2) the operational and cost results across the seven years of days or weeks could be weighted to best match underlying distributions of all key variables including hydro availability. Both of these approaches introduce new modeling challenges and the best approach for a given application will likely be a unique compromise between operational fidelity and statistical accuracy.

Computational limitations may also inhibit a full analysis of seven years of operating conditions. In this case, it may be desirable to select a subset of the days or weeks for which to model operations. In this type of scheme, days or weeks could be directly drawn from the seven years of data without the need to synthetically match profiles from different days as was performed in this study. These draws could be repeated until convergence is achieved on the key operational metrics of interest. This type of approach could also incorporate smart sampling and weighting to reduce the number of draws required to represent the long run distributions of conditions. One drawback of this and any other sampling approach is that the operations will still encounter the boundary effects described above, which may influence coal cycling and pose challenges to simulating longer duration energy storage.

Each of these proposed refinements is made possible by the expanded availability of high-quality renewable generation data, and each offers two improvements upon the synthetic day-draw method used in this study: (1) historically accurate, weather-matched load, wind, and solar profiles; and (2) the ability to model longer time horizons than a single day. For these improvements, each method merits further investigation as the discipline of operational flexibility analysis continues to evolve.

7.3 Flexibility-Related Constraints

In addition to the technical constraints imposed on generating units, this analysis also incorporates some constraints related to flexibility that have not been imposed in prior renewable integration analyses and some constraints that are increasingly being relied upon in production cost modeling exercises and have large impacts on operational flexibility.

7.3.1 MINIMUM GENERATION CONSTRAINTS

Minimum generation constraints are imposed in some balancing areas in the California region. These constraints originally reflected CAISO market constraints intended to limit the hourly imports into SCE and SDG&E in case of an extreme underfrequency and intertie separation event.[52] These market constraints were eventually lifted, but the CAISO has continued to use an evolving set of minimum generation constraints in its production cost modeling exercises for California's Long Term Procurement Plan (LTPP). In this analysis, the constraints are imposed based on the assumptions in the 2024 Common Case, which reflect the 2014 LTPP minimum generation constraints. These require 25% of the generation in the SCE, SDG&E, and LADWP balancing areas to be met with local generation from a subset of generators. Since specific balancing areas are not modeled in this study, these constraints are approximated based on a fixed share of the total California load in each hour,

[52] http://www.caiso.com/Documents/TechnicalBulletin-ImportLimitDefinitionandManagementinSupport-Under-FrequencyLoadShedding_UFLS.pdf

calculated as 25% times the average load share of each balancing area with a minimum generation constraint.

These constraints have a significant impact on the oversupply challenges identified in the California region by ensuring that gas generation stays on in the middle of the day when low net load conditions drive renewable curtailment. Minimum generation constraints were found to bind between 45% and 60% of all hours (depending on the balancing area) in the High Renewables Case and in 100% of all curtailment hours in California, suggesting that the operational challenges identified for California may be highly sensitive to these constraint definitions. Given the critical importance of minimum generation constraints to the solar oversupply storyline in California, future flexibility analyses should take care to incorporate the most recent knowledge on this topic into production cost modeling efforts.

The role of the minimum generation constraints in California also suggest that similar types of constraints may bind in other regions if they were to be imposed. Currently no other regions in the study area incorporate minimum generation constraints, which potentially relieves the other systems from exacerbated oversupply conditions in this study. At present, it is unclear if similar constraints do not exist for other regions because those regions are technically capable of operating at lower thermal dispatch levels or if the less aggressive renewable policy goals in those regions have not yet necessitated investigation into the technical challenges that minimum generations constraints are intended to reflect. This is an important area for future investigation not only in California, but across the West.

7.3.2 HYDRO MULTIHOUR RAMPING CONSTRAINTS

As has been discussed, the modeling of hydro resources in production cost modeling is challenging due to the complicated interacting constraints placed on hydro systems due to technical, environmental, and social factors. For this reason, hydro resource constraints are often derived from historical dispatch data in order to best reflect the aggregated effects of these factors without modeling them explicitly. In addition to constraining the minimum output, maximum output, and daily energy generated from hydro resources, this study also applies ramping constraints on hydro resources based on historical ramping behavior. These constraints are imposed over multiple hours in order to limit the ability of the hydro fleet to sustain a ramp, much like the hydro fleet has limited sustained peaking capability.

To test the sensitivity of the results to these multihour hydro ramping constraints, both the Common Case and the High Renewables Case were run in sensitivities with the multihour hydro ramps lifted. The resulting renewable curtailment, which is summarized in Table 50, suggests that inclusion of the multihour ramping constraints on hydro does not have a significant impact on the results of this study. These sensitivities corroborate the finding that ramping capability is not a significant driver of renewable integration challenges on the system.

Table 50. Annual renewable curtailment in the Common Case and the High Renewables Case with and without mulithour ramping constraints on hydro output in each region.

Scenario	Hydro Ramping Constraints	Basin	California	Northwest	Rockies	Southwest	WECC-US
Common Case	Imposed	0.01%	0.00%	0.06%	0.13%	0.00%	**0.02%**
	Lifted	0.01%	0.00%	0.06%	0.12%	0.00%	**0.02%**
High Renewables Case	Imposed	0.4%	8.7%	5.6%	0.6%	7.3%	**6.4%**
	Lifted	0.4%	8.5%	5.9%	0.6%	7.3%	**6.4%**

7.3.3 INTERTIE RAMPING CONSTRAINTS

Single hour ramping constraints on intertie were also imposed as a base assumption in this study. These constraints served to ensure that interactions between regions largely reflected the level of flexibility that has been historically available from bilateral trades. Other studies have incorporated hurdle rates to represent friction between balancing areas in an attempt to capture similar behavior. The best way to reflect institutional inefficiencies in trades between balancing areas in a Balkanized system like the Western Interconnection remains an active area of investigation and discussion. The ramping constraints in this study were largely motivated by the desire to incorporate a flexibility planning assumption for imports and exports similar to the types of import assumptions used in resource adequacy studies. For example, CAISO incorporates some level of imports that can be counted upon when determining the system-wide capacity need. Similarly, by constraining ramps over interties based on historical behavior, this study sought to approximate some amount of flexibility that could be counted upon by a given region from neighboring balancing areas when conducting flexibility analyses.

Given the ongoing and rich discussion about balancing area interactions in production cost models, an additional sensitivity was run to isolate the impact of imposing the ramping constraints from the impact of the combined ramping and intertie flow constraints that are reflected in the Reference Grid assumptions. As shown in Table 51, imposing the historically-based intertie ramp limits affects the distribution of curtailment across the five regions, but has a relatively small impact on the total curtailment across the study area relative to the impact of imposing the historically-based flow limits. Similar to the hydro multihour ramping constraint sensitivity, this result is generally consistent with the findings that the value of regional coordination is primarily driven by the ability to find other markets across the West for excess renewable generation and that additional ramping capability has a minimal impact on the curtailment challenge.

Table 51. Renewable curtailment in the High Renewables Case under various intertie constraints.

Intertie Limits	Intertie Ramp Limits	Basin	California	Northwest	Rockies	Southwest	WECC-US
Physical	None	0.5%	3.0%	2.0%	0.5%	6.1%	**3.0%**
Physical	Historical	0.7%	3.5%	0.8%	0.7%	7.1%	**3.2%**
Historical	Historical	0.4%	8.7%	5.6%	0.6%	7.3%	**6.4%**