



Capturing the Impact of Storage and Other Flexible Technologies on Electric System Planning

Elaine Hale, Brady Stoll, and Trieu Mai
National Renewable Energy Laboratory

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

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

Technical Report
NREL/TP-6A20-65726
May 2016

Contract No. DE-AC36-08GO28308



Capturing the Impact of Storage and Other Flexible Technologies on Electric System Planning

Elaine Hale, Brady Stoll, and Trieu Mai
National Renewable Energy Laboratory

Prepared under Task Nos. SA12.0381, SA15.1015,
and SA15.0910

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

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

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

Technical Report
NREL/TP-6A20-65726
May 2016

Contract No. DE-AC36-08GO28308

NOTICE

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

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

Available electronically at SciTech Connect <http://www.osti.gov/scitech>

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

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
OSTI <http://www.osti.gov>
Phone: 865.576.8401
Fax: 865.576.5728
Email: reports@osti.gov

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

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312
NTIS <http://www.ntis.gov>
Phone: 800.553.6847 or 703.605.6000
Fax: 703.605.6900
Email: orders@ntis.gov

Cover Photos by Dennis Schroeder: (left to right) NREL 26173, NREL 18302, NREL 19758, NREL 29642, NREL 19795.

NREL prints on paper that contains recycled content.

Acknowledgments

We thank Paul Denholm, Jennie Jorgenson, Dave Mooney, and Daniel Steinberg (National Renewable Energy Laboratory), Mark Dyson (Rocky Mountain Institute), Nils Johnson (International Institute for Applied Systems Analysis), Hassan Hijazi (Australian National University / Commonwealth Scientific and Industrial Research Organisation), Kerry Cheung, Paul Donohoo-Vallett, and Ookie Ma (U.S. Department of Energy) for their thoughtful reviews, comments, and suggestions. We acknowledge Bethany Frew (National Renewable Energy Laboratory) for her supporting analysis. We also thank Steve Capanna and Ookie Ma (U.S. Department of Energy) for supporting this work. This research was funded by the U.S. Department of Energy under contract number DE-AC36-08GO28308. Any errors or omissions are the sole responsibility of the authors.

Executive Summary

Power systems of the future are likely to require additional flexibility due to the operating characteristics of many clean energy technologies, particularly those relying on renewable energy sources. This subject has been well studied from an operational perspective, but it has been more difficult to incorporate into capacity expansion models (CEMs) that study investment decisions on the decadal scale. There are two primary reasons for this. First, the necessary input data, including cost and resource projections, for flexibility options like demand response and storage are not widely available and are highly uncertain. Second, it is computationally difficult to simultaneously represent both investment and operational decisions in detail, with the latter being necessary to adequately value system flexibility in realistic systems.

The primary purpose of this report is to present new capabilities that were developed for a particular CEM, NREL's Resource Planning Model (RPM), to better reflect the impact of variable wind and solar generation on system operations and resource adequacy, and, complementarily, to model energy-constrained flexibility resources. The additional variable generation modeling capabilities enable a more-accurate representation of the need for additional flexibility in systems with high penetrations of variable generation and, as a result, better reflect the value of technologies that can provide flexibility. We demonstrate the use of the new capabilities by examining the role of two broad technological sources of flexibility in power systems—utility-scale storage and interruptible load—and we model how they might contribute to a range of grid services for operations and planning. A wide range of different energy storage technologies is modeled, including multiple types of battery technologies. We also model interruptible load, a type of demand response that can be used to reduce demand, usually during peak hours, within pre-determined capacity and duration limits.

The new capabilities use load duration curves from hourly data spanning a full year, thus enabling RPM to capture the impact of variable generation and the potential value of storage and interruptible load during infrequent, but important periods and hours within the year. The techniques enable RPM to accurately calculate the need for grid services and simultaneously consider, in its investment decision, the potential for storage and interruptible load to provide value to the system in multiple distinct ways: firm capacity for resource adequacy, operating reserve provision, energy arbitrage, and reductions in renewable curtailment. The ability of storage and interruptible load to lower system costs by providing these services, as well as the capital and implementation costs and other characteristics of the flexible technologies, are factored into RPM's economic investment decisions. These new techniques adapted for RPM can also be implemented in other CEMs.

Using these capabilities, we analyze deployment of storage and interruptible load in the Western Interconnection through 2030. Two distinct sets of underlying assumptions are modeled: baseline conditions that achieve about 32% renewable energy penetration by 2030 and higher renewable energy assumptions that lead to penetrations reaching about 40% in 2030. The scenario analysis is designed primarily to demonstrate new model capabilities, but it also includes initial estimates of the potential cost reductions and preferred storage and interruptible load configurations needed to achieve market deployment in the western United States.

Key findings from our analysis include:

- Capacity expansion models can be formulated—in a computationally tractable form—to reflect the need for system flexibility and the potential for storage and interruptible load to fulfill those needs.
- Under baseline scenarios of the Western Interconnection through 2030, interruptible load is found to be competitive and is deployed even when the annual costs approach the annualized capital costs of natural gas combustion turbines (\$83 per kilowatt-year [kW-yr]).
- Interruptible load is deployed primarily to meet growing planning reserve requirements, thereby reducing the need for future natural gas combustion turbine capacity that is otherwise used for peaking needs, and also shifting renewable capacity from capacity-constrained regions to those where the resource is more favorable. While resource adequacy needs are the principal drivers of interruptible load deployment, we find that in some regions, interruptible load is also deployed to meet spinning contingency reserve requirements.
- Under scenarios with greater renewable energy, the higher amounts of renewable capacity reduces the need for new thermal or other firm capacity resources, resulting in less natural gas and interruptible load deployment. However, even under these conditions, interruptible load at the highest assumed annual cost levels that also remain below combustion turbine annualized capital costs are adopted in certain regions by 2030. Additional plant retirements beyond those assumed, possibly driven by greater renewable energy penetration or policies that support renewable adoption, might drive greater demand for firm capacity services than in the higher renewable scenarios modeled.
- Under scenarios that assume storage cost and performance parameters developed from the DOE/EPRI 2013 Electricity Storage Handbook, new storage is not deployed in the Western Interconnection in the 2015–2030 timeframe. However, scenarios in which capital costs for battery storage technologies are 50% lower than the 2013 reported values do result in some modeled deployment of storage. It should be noted that RPM is unable to fully capture some of the values streams for storage technologies, including sub-hourly flexibility and potential intra-zonal congestion benefits.
- If technology evolution leads to reductions in storage costs significantly below 2013 estimated values (down to about \$100/kWh), more-significant storage capacity is developed by the model, including over 8 gigawatts (GW) of new storage across the Western Interconnection by 2030.
- The value of storage to the system varies greatly by region based on the needs of the regions, with value stacking being an important aspect of these installations. While firm capacity needs are found to be an important driver for new storage development, the ability of storage to contribute to other system needs (e.g., operating reserves) or lower system costs (e.g., energy-shifting) can also drive new storage deployment. This is clearly observed in situations when the least-expensive storage option (on a capacity basis) is passed over for a different option that offers greater overall value.
 - Storage deployment is strongest when firm capacity or reserves are needed by the system, and costs are competitive with alternative resources, including natural gas combustion turbines.

- Storage enables increased renewable penetration in regions with a high need for flexibility reserves, which grow with renewable penetrations. Additionally, energy shifting and curtailment reduction increase the value of energy storage in regions, such as California, with already high penetrations of variable generation.

These findings are based on an initial modeling analysis using new methods. Additional research could help improve these methods and the data inputs used in RPM. Furthermore, the analysis does not consider the need for sub-hourly services or the associated value opportunities for flexibility options that could be used during these shorter timescales. In addition, while the methods use 8,760 hourly data, they rely on data from a single year (2006). A more robust assessment could include multiple years of data. Moreover, new energy policies that might be enacted in the future would change scenario results and alter the quantitative findings reported above. Finally, like many other similar models, RPM has certain limitations, including its system-wide approach that may not accurately reflect current market or institutional flexibility (or inflexibility) or the costs and value of flexibility from the perspective of developers, utilities, system operators, regulators, or other power sector participants. Nonetheless, our research demonstrates a new capability to begin to incorporate considerations of flexibility in large-scale capacity expansion models and provides initial estimates on the deployment potential for interruptible load and storage in the western United States.

Table of Contents

1	Introduction	1
2	The Resource Planning Model	3
2.1	Spatial Structure and Initial Conditions	4
2.2	Modeling Investment Decisions	7
2.3	Modeling System Operations and Flexibility	11
2.3.1	Co-Optimization of Operations and Investments	13
2.3.2	Net Load Duration Curve-Based Flexibility Adjustments	14
3	Modeling Flexibility Investments	19
3.1	Interruptible Load	19
3.2	Storage	21
4	Scenario Framework	23
4.1	Interruptible Load Scenarios	24
4.2	Storage Scenarios	26
4.3	Overall Framework	28
5	Results	28
5.1	Interruptible Load	31
5.1.1	General Trends in the Least Constrained Scenarios	31
5.1.2	Discerning the Primary Grid Service for which Interruptible Load is Enabled	37
5.1.3	Summary of Price Point Findings	41
5.2	Utility-Scale Storage	45
5.2.1	General Trends in Storage Deployment	45
5.2.2	Description of Storage Operations	51
6	Discussion	60
	References	62
	Appendix A: Net Load Duration Curve Modeling for Variable Generation	67
A.1	Capacity Value	68
A.2	Renewable Curtailment	70
	Appendix B: Net Load Duration Curve Modeling for Flexibility Investments	75
B.1	Interruptible Load Methodology	75
B.2	Storage Methodology	76
	Appendix C: Detailed Equation Listings	79
C.1	Interruptible Load Model	79
C.2	Storage Model	80

List of Figures

Figure 1. Combined nodal and zonal spatial structure of the Resource Planning Model with a Sacramento Municipal Utility District (SMUD) focus region.....	6
Figure 2. Comparison of 8,760 hourly and RPM 96-hour sampled load duration curves (LDCs) for the 2010 Western Interconnection (left) and net load duration curves (NLDCs) for an illustrative 40% VG scenario (right).....	12
Figure 3. NERC sub-regions in the western U.S. used for planning reserve regions in RPM (NERC 2015a).....	13
Figure 4. Conceptual depiction of the NLDC-based capacity value estimation method.....	15
Figure 5. Marginal and existing capacity value from utility-scale solar and wind by year from an example of an RPM scenario that reaches 40% VG penetration by 2030.....	16
Figure 6. Conceptual depiction of the NLDC-based minimum curtailment estimation method.....	17
Figure 7. Existing and marginal curtailment fractions for variable resources installed in RPM for each solve year for an example of a scenario with 40% VG penetration by 2030.....	18
Figure 8. Base scenario system-wide capacity expansion over time.....	29
Figure 9. Difference in capacity between the High RE and Base scenarios.....	30
Figure 10. Hourly dispatch for the High RE scenario.....	30
Figure 11. Capacity differences for the 150h scenario relative to the Base scenario.....	32
Figure 12. Deployment of interruptible load for the 150h scenario by year and NERC sub-region.....	33
Figure 13. Capacity differences by year and NERC sub-region for the 150h scenario.....	34
Figure 14. Capacity differences for the High RE 150h scenario relative to the High RE scenario.....	36
Figure 15. Deployment of interruptible load for the High RE 150h scenario by year and NERC sub-region.....	37
Figure 16. Deployment fraction differences versus the logarithm of capacity value ratios, Base-Interruptible Load scenarios, 2020.....	40
Figure 17. Deployment fraction differences versus the logarithm of capacity value ratios (left) and the logarithm of annual energy capacity ratios (right) in the High RE-Interruptible Load scenarios, 2030.....	41
Figure 18. Differences in installed capacity in 2030 for the four Base-Storage Cost scenarios relative to the Base scenario.....	46
Figure 19. Difference in capacity installed for each of the Base-Storage Cost scenarios compared to the Base scenario by NERC sub-region in 2030.....	47
Figure 20. Storage capacity in 2030 by technology sub-class and NERC sub-region for the Base-Storage Cost scenarios.....	48
Figure 21. Differences in installed capacity in 2030 for the four High RE-Storage Cost scenarios relative to the High RE scenario.....	49
Figure 22. Difference in capacity installed compared to the high renewables scenario for each of the storage scenarios and NERC sub-region in 2030.....	49
Figure 23. Storage capacity in 2030 by technology sub-class and NERC sub-region for the High RE-Storage Cost scenarios.....	50
Figure 24. Yearly capacity differences at a system level for the 0p10 scenario for both the Base and High RE assumptions.....	51
Figure 25. Cumulative system-level storage capacity by technology sub-class for the Base-0p10 and High RE-0p10 scenarios.....	51
Figure 26. Differences in annual generation for the Base scenarios.....	52
Figure 27. Flexibility reserves provision by generator type, as a fraction of all flexibility reserves provided, for the entire Western Interconnection.....	53
Figure 28. Differences in Flexibility and Regulation reserves for the Wyoming reserve sharing group in 2030.....	54

Figure 29. Differences in annual generation for the High RE scenarios, showing all High RE-Storage Cost scenarios versus the High RE scenario in 2030 (left) and differences in High RE 0.10 costs and High RE annual generation by year (right)	54
Figure 30. Reserves provision by generator type, as a fraction of all reserves provided, for the entire Western Interconnection, showing all High RE-Storage Cost scenarios in 2030 (left) and the High RE-0p10 scenario annual provision by year (right)	55
Figure 31. Differences in Flexibility reserves for the Colorado and California-South reserve sharing groups in 2030, under High RE assumptions	56
Figure 32. Dispatch of the 2h NaCl-Ni energy storage technology in SCE in the Base-0p10 scenario	57
Figure 33. System dispatch for SCE in the Base-0p10 scenario	58
Figure 34. Dispatch of the 1h Li-ion energy storage technology in NEVP in the High RE-0p10 scenario	58
Figure 35. System dispatch for NEVP in the High RE-0p10 scenario	59
Figure A1. Sample load duration curve (LDC), net load duration curve (NLDC), and incremental load duration curve (ILDC)	67
Figure A2. Example of capacity value calculation using the load duration curve methodology	68
Figure A3. Marginal capacity values of utility-scale solar and wind for regions in which the model built new capacity and did not build new capacity for each of the model solve years	69
Figure A5. Example of load duration curve, net load duration curve, and incremental load duration curve and the effective minimum generation level used to calculate curtailment	70
Figure A6. Effective minimum generation regression for each NERC region based on the penetration of variable generation sources	72
Figure A8. Hourly curtailment for WECC	73
Figure A9. Marginal curtailment fractions of utility-scale solar and wind for regions in which the model built new capacity and did not build new capacity for each of the model solve years	74
Figure B1. Marginal capacity values of interruptible load resources for resources deployed and not deployed for each of the model solve years	76
Figure B2. Example of load duration curve impacts from storage technologies on the capacity value and curtailment	77
Figure B3. Example of storage impacts on the net load duration curve, and the curtailment fallout	77
Figure B4. Example of load duration curve from an incremental amount of storage and the impact new storage technologies can have on curtailment	78

List of Tables

Table 1. Start Year (2010), Prescribed, and Retiring Capacity in NREL’s Resource Planning Model.....	7
Table 2. Technology Cost and Performance Assumptions for New Generation Capacity	10
Table 3. Full Factorial Design of Interruptible Load Technology Sub-Classes.....	25
Table 4. Interruptible Load Scenario Framework.....	26
Table 5. Storage Technology Sub-Classes Modeled in RPM.....	27
Table 6. Scenario Framework for Storage Technologies.....	27
Table 7. Summary of Scenario Framework	28
Table 8. Interruptible Load Resource Available in each NERC Sub-Region.....	33
Table 9. Planning Reserve Constraint Summary for the 150h Scenario.....	35
Table 10. Minimum Prices for Grid Services from Interruptible Load as a Function of Annual Cost and Annual Energy Capacity	38
Table 11. Interruptible Load Deployment and Resource by Year for the 150h Base Scenario	41
Table 12. Price Points Achieved in the Base 150h Scenario	43
Table 13. Price Points Achieved in the High RE 150h Scenario.....	44
Table 14. Variable Generation Curtailment in 2030.....	56

1 Introduction

Historical and anticipated growth in variable renewable electricity has led to increasing emphasis on power system flexibility. Flexibility is a measure of a system’s ability to change supply or demand as needed to accommodate variability and uncertainty at different time scales. Power system planners have always considered flexibility to accommodate load variability and uncertainty as well as unexpected generation and transmission outages. Increasing generation from variable renewable sources such as wind and solar add to this existing need. In this report, we describe new capabilities in the Resource Planning Model (RPM)¹ to more comprehensively assess the need for and the associated value of flexibility (and the resources that can provide it) in regional power system planning, especially in the context of high penetrations of variable renewable generation. We then use RPM to assess—from a system optimization perspective—deployment of two sources of system flexibility under a range of scenarios of the western United States: interruptible load and energy storage.

Flexibility assessment has become an integral part of grid integration studies for renewables (GE Energy 2010; Lew et al. 2013; EnerNex Corporation 2011; Bloom et al. forthcoming; Venkataraman et al. 2010; E3 2014; GE Energy 2014; Brinkman et al. 2015). These studies use high-resolution models of grid operations (i.e., production cost models) to assess how the introduction of greater shares of variable renewable generation might affect production costs, emissions, and the mix of generated electricity. Covering a range of geographic areas, these detailed studies have generally found that increasing system flexibility can offer significant system benefits by reducing production costs, emissions, and renewable curtailment.

The modeling methodologies of large-scale renewable integration studies have also been used to conduct more targeted system flexibility research. In these studies, the flexibility of a system is incrementally altered—for example, by changing thermal plant configurations (e.g., minimum generation points and ramp rates), or adding direct energy storage technologies, concentrating solar power with thermal energy storage, or demand response—and differences driven by these incremental changes are evaluated to discern the operational impact of flexibility (Black and Strbac 2006; Palchak and Denholm 2014; Zucker, Hinchliffe, and Spisto 2013; P. Denholm et al. 2013; Jorgenson et al. 2013; Denholm et al. 2015; Hummon et al. 2013). Furthermore, these increments are often applied to systems with different pre-defined amounts of variable generation (VG) such that the marginal value of flexibility can be analyzed as a function of VG penetration.

Other operational approaches have been used to research power system flexibility. These include “price taker” models that can be used for revenue analysis of a candidate option under historical or simulated conditions (Graves, Jenkin, and Murphy 1999; Denholm and Sioshansi 2009; Drury, Denholm, and Sioshansi 2011).² More sophisticated approaches exist and include stochastic production cost models (e.g., Hargreaves et al. 2015).

All of these methods measure the *operational* value of flexibility options. Production cost models in particular capture a great amount of hourly and sub-hourly detail by determining unit

¹ For more information about RPM, see Mai et al. (2015) and www.nrel.gov/analysis/models_rpm.html.

² As opposed to production cost simulations that take a system-wide optimization approach.

commitment and economic dispatch, often co-optimized with operating reserves, as a function of fuel costs, variable operation and maintenance (O&M) costs, startup costs, and detailed operational constraints (e.g., generator minimum generation and ramp rates, transmission constraints, and reserve requirements). However, production cost modeling has certain important limitations. First, while operational costs are used to inform commitment and dispatch decisions in production cost simulations, fixed costs (e.g., capital and fixed O&M costs)—which make up the predominant share of costs for many flexibility options—are not. No direct comparison of cost versus value is made in production cost models. Second, flexibility options can offer non-operational value to the system that is not captured in production cost models. For example, storage and demand response can have sizeable capacity value, which would need to be analyzed separately from the grid simulations.³ Third, the approach does not directly compare multiple options simultaneously and does not find the least-cost flexibility provider. Finally, production cost models rely on predetermined portfolios that do not take into account path dependencies. Systems are likely to evolve differently depending on the amount of, need for, and availability of flexibility, and the production cost modeling methods used in many valuation studies do not directly consider these interactions.

In this study, we evaluate flexible technologies with a different class of model, the capacity expansion model (CEM), which typically determines the evolution of a power system subject to initial conditions, technology availability and costs, and select operational constraints. The key outputs from CEMs are the amounts and locations of various technologies to meet future system needs for capacity, energy, and policy compliance. They can be used to study the co-deployment of technologies like variable generation and storage. The value of variable generation could depend on system flexibility (including storage availability), and conversely, the value of storage could depend on the level of variable generation. While having a broader scope than production cost models, CEMs typically have more limited operational detail (Sullivan, Eurek, and Margolis 2014). Existing CEMs often implicitly or explicitly include some representations of flexibility—both the need for, and the potential supply of, flexibility (e.g., Ma et al. 2013; Johnston et al. 2013; Short et al. 2011; EPRI 2014)—however, the computational complexity needed to simulate investment decisions necessitates simplifications that make it difficult to fully capture operational flexibility needs and contributions. Thus, we improve these prior methods and the present report describes our approach.

In addition to research on production cost and capacity expansion models, related research exists on assessing the amount of flexibility in various generation portfolios. For example, Mills and Seel (2015) use a “screening-level flexibility inventory” based on methodology developed by IEA (2011) to characterize flexibility in four distinct timescales for portfolios developed in recent integrated resource plans for utilities and regions in the western United States. In another example, Mills and Wiser (2014) present a model to evaluate the long-run economic value of VG and how this value might be impacted by the introduction of mitigation options. They also examine the economic attractiveness of the mitigation options themselves. Others have developed new metrics to evaluate the operational flexibility of different portfolios (e.g., Lannoye, Flynn, and O’Malley 2012; Ma et al. 2013).

³ On the other hand, for systems that are long on capacity, the capacity value of a candidate option might be less important.

In this report, we describe a new approach developed for and implemented in NREL’s Resource Planning Model (RPM) to better reflect (1) the impact of variable generation on system operations and resource adequacy, and (2) the potential value of additional flexibility. The methods described could be adopted in other CEMS. Using these capabilities, we analyze two distinct flexibility investment options—storage and interruptible load—in the western United States through 2030. While the assessment is not comprehensive either technologically or geographically, it highlights potential opportunities for flexibility in power systems, particularly as the share of renewables grows. It also identifies the degree of cost reductions, and the desired configurations of storage and interruptible load, that would be needed to achieve significant deployment levels at the utility scale in the western United States.

The report is organized as follows. Section 2 briefly describes RPM and summarizes the key electric sector assumptions used in the analysis. It also describes how RPM captures the need for system flexibility, particularly at higher renewable penetrations, and the potential value of increasing flexibility. Appendix A provides detail on these methods, and (Mai, Barrows, et al. 2015) includes other model details. We describe our model characterization for interruptible load and storage in Section 3; Appendix B and Appendix C provide details. The scenario framework for this initial analysis is described in Section 4, with results following in Section 5. We conclude in Section 6 with key findings and potential future research.

2 The Resource Planning Model

NREL’s Resource Planning Model (RPM) is a recently developed capacity expansion model designed to investigate the evolution of a regional power system such as a utility service territory, state, or balancing authority area (BA). RPM co-optimizes new generation, transmission expansion, and dispatch over time in five-year increments beginning in 2010 and continuing to 2030. The least-cost optimization algorithm minimizes overall system cost, including capital costs, fixed and variable operations and maintenance (O&M) costs, and fuel costs. The model has high spatial resolution, representing individual generation units for a select subset of units; solar and wind resources are represented using geographic areas, each with similar production characteristics. RPM models hourly dispatch for a representative sample of days throughout a year. Each hourly step balances generation with load, maintains the required amount of reserve capacity, and remains within operational constraints for individual generators and transmission paths. (Mai, Barrows, et al. 2015) provide a thorough description of RPM.

In this section, we summarize some essential aspects of the model, including the spatial configuration and initial conditions used for this analysis (Section 2.1), major input assumptions related to investment decisions (Section 2.2), and dispatch modeling (Section 2.3). Section 2.3 includes an overview of our basic dispatch model and recently developed methods that augment the dispatch and reserve constraints to better capture the need for and value of system flexibility.

2.1 Spatial Structure and Initial Conditions

RPM models a single interconnection at a time. This study focuses on the Western Interconnection electricity system, which includes all or parts of 13 states in the western United States, two western provinces in Canada, and a small region of northern Mexico. This geographic boundary comprises 36 model BAs, which are the primary regional units in RPM.⁴ Embedded within this zonal structure, the model has a “focus region,” within which generation units, transmission lines, and loads are represented with a high level of detail. Within this region, the optimization is carried out nodally, while the remaining regions are treated zonally to capture power transfers into and out of the focus region and between connected BAs. For this analysis, the focus region is comprised of only one BA, the Sacramento Municipal Utility District (SMUD) – we refer to this SMUD-focused version of RPM as RPM-SMUD.⁵ This choice of a SMUD-only focus region reduces computational run times while still retaining the combined zonal-nodal structure of the model. This paper is mostly focused on new methodologies, so most results in the analysis are presented at the full interconnection-wide level or for select groups of BAs. Therefore, the choice of focus region would not likely have a significant impact on these model results.⁶ Figure 1 shows the combined zonal and nodal structure of RPM with the SMUD focus region.⁷

RPM also includes additional spatial layers to represent renewable resources. RPM-SMUD includes 615 solar and 106 wind resource areas in the Western Interconnection to describe the location-specific resource potential (developable area after accounting for various land use exclusions), performance (annual and hourly capacity factors), and grid interconnection distances (to substations, transmission nodes, load nodes).⁸ In general, resource regions are defined dynamically for each version of RPM as distinguished by the focus region selected for an analysis, with more resource regions defined within the particular focus region being studied. This ensures that the region of interest is represented with the highest resolution possible.

⁴ Model BAs are closely, although not perfectly, aligned with real BAs as designated in the data used in WWSIS Phase 2 (Lew et al. 2013). Similar to the WWSIS Phase 2 modeling, three of the smaller BAs, GCPD, CHPD, and DOPD have been combined with BPA in RPM.

⁵ In actuality, this BA corresponds to the Balancing Area of Northern California (BANC). We retain the misnomer (which derives from the WECC TEPPC 2020 Data Set) for consistency with other work. The misnomer is corrected in the WECC TEPPC 2024 Data Set.

⁶ Future analysis could use the established methodology to evaluate the potential for flexibility options within a region (i.e., at the nodal level for a focus region of interest comprised of one or more BAs).

⁷ Full and abbreviated names for the 36 model BAs include: Alberta Electric System Operator (AESO), Arizona Public Service (APS), Avista (AVA), British Columbia Transmission Corporation (BCTC), Bonneville Power Administration (BPA), Comision Federal de Electricidad (CFE), El Paso Electric Company (EPE), Far East (FAR_EAST), Imperial Irrigation District (IID), Los Angeles Department of Water and Power (LDWP), Magic Valley (MAGIC_VLY), Nevada Power (NEVP), Northwestern Montana (NWMT), Pacificorp East Idaho (PACE_ID), Pacificorp East Utah (PACE_UT), Pacificorp East Wyoming (PACE_WY), Pacificorp West (PACW), Pacific Gas and Electric Bay Area (PG&E_BAY), Pacific Gas and Electric Valley Area (PG&E_VLY), Portland General Electric (PGN), Public Service Colorado (PSC), Puget Sound Energy (PSE), Southern California Edison (SCE), Seattle City Light (SCL), San Diego Gas and Electric (SDGE), Sacramento Municipal District (SMUD), Sierra Pacific Power (SPP), Salt River Project (SRP), Tucson Electric Power (TEP), Turlock Irrigation District (TIDC), Tacoma Power (TPWR), Treasure Valley (TREAS_VLY), Western Area Power Administration Colorado/Missouri (WACM), and Western Area Power Administration Upper Missouri (WAUW).

⁸ There are 22 solar regions and 1 wind resource region within the SMUD focus region.

Underlying data for the start-year (2010) power system infrastructure are from the Western Wind and Solar Integration Study (WWSIS) Phase 2 (Lew et al. 2013).⁹

Table 1 presents the start-year capacity for the technologies modeled in RPM. It also presents the amount of capacity added or under construction since 2010 (Ventyx 2010; NREL SolarPACES 2014)¹⁰ and retirements prescribed exogenously to the model (WECC 2014b; Saha 2013).¹¹ Beyond prescribed new capacity, endogenous investment decisions for the type, amount, and location of new capacity are based on the needs of the system, policies, and economic factors.

⁹ WWSIS Phase 2 relied on data from the WECC Transmission Expansion Planning Policy Committee (TEPPC) 2020 Common Case (WECC 2012) with updates from the TEPPC 2022 Common Case (WECC 2014b) along with other revisions as described in (Lew et al. 2013).

¹⁰ Under-construction capacity includes capacity installed since 2010 and capacity that is currently under construction (WECC 2014b). A majority of this new capacity is “prescribed” to occur by the 2015 solve year.

¹¹ Retired capacity represents any plants that have retired since 2010 and any announced retirements. We supplement the (WECC 2014b) and Saha 2013 retirements with those found in recent utility planning documents (SCE 2013; PNM 2013; SCPPA 2014; Randazzo 2014; NV Energy 2014; El Paso Electric Company 2012) and the Navajo Generating Station. Prescribed retirements occur over many years but are concentrated in the 2015 and 2020 solve years. We also enable endogenous “economic” retirements based on a lower bound on the capacity factor of generators. Currently only coal is allowed to retire endogenously, with a lower capacity factor bound set at 7%.



Figure 1. Combined nodal and zonal spatial structure of the Resource Planning Model with a Sacramento Municipal Utility District (SMUD) focus region

Table 1. Start Year (2010), Prescribed, and Retiring Capacity in NREL’s Resource Planning Model

Generator Type ^a	Western Interconnection			SMUD Focus Region		
	2010 Capacity (MW)	Planned Capacity (MW)	Retirements (MW)	2010 Capacity (MW)	Planned Capacity (MW)	Retirements (MW)
Coal	38,818	405	8,884	-	-	-
NGCC	48,505	1,941	269	1,399	-	-
NGCT	16,659	3,736	1,650	374	49	-
NG-Other	23,423	8	14,985	100	-	-
Nuclear	9,681	-	2,246	-	-	-
Biomass	1,559	76	40	18	-	-
Geothermal	3,054	35	-	-	-	-
Hydropower	69,989	103	-	2,497	-	-
PHES	3,787	40	-	-	-	-
Utility PV ^b	74	2,653	-	-	103	-
Rooftop PV ^{b,c}	1,096	29,370	-	32	498	-
CSP	354	1,887	-	-	-	-
Wind	10,172	6,621	-	-	-	-
Total	227,174	46,875	28,074	4,420	651	-

^a Technology category acronyms: NGCC = natural gas-fired combined cycle, NGCT = natural gas-fired combustion turbines, PHES = pumped hydropower energy storage, PV = photovoltaic, and CSP = concentrating solar power.

^b All PV capacities are listed in MW-AC. The model uses the inverter loading ration from PVWatts 5 of 1.1.

^c Rooftop PV capacity is exogenously input based on a scenario from the NREL dSolar model (Sigrin et al. 2016).

2.2 Modeling Investment Decisions

Starting from the 2010 initial conditions, RPM finds the least-cost capacity expansion solution in five-year increments through 2030. This is accomplished by matching new builds to system needs identified exogenously based on potential growth in annual electricity consumption, peak demand, planned retirements and capacity additions, and reserve requirements. In addition, policy requirements such as state renewable portfolio standards may result in demand for new capacity. Economic factors and assumptions inform the least-cost solution found by the model. We describe some of the essential assumptions related to RPM’s investment decision-making in this section but refer readers to (Mai, Barrows, et al. 2015) for details.

Load growth assumptions are based on data from WECC (2014a), where we assume constant annual demand growth rates for each BA based on the average of estimated annual load growth between 2010 and 2016. As a result, the annual load growth rate between 2010 and 2030 is 0.9% per year for the entire interconnection, but there is a good deal of diversity in individual BA growth rates. Hourly demand (and wind and solar) *profiles* are based on 2006 meteorology for

all years (Lew et al. 2013). Exogenous retirements and prescribed capacity are described in Section 2.1.

We include many of the major existing federal and state power sector policies as of January 1, 2016. These include the extensions of federal renewable tax credits that are part of the Consolidated Appropriations Act of 2016, recent changes to state renewable portfolio standards (RPS), and the U.S. Environmental Protection Agency's Clean Power Plan.¹² We notably exclude California's storage mandate and carbon cap and trade program (de Leon, Williams, and Leno 2015; California PUC 2013; California Air Resources Board 2014). Existing demand response programs or local incentives are also excluded in our analysis. These omissions could have an impact on model estimates of future capacity, generation, and emissions, but as the current analysis is focused on methodological representation of flexibility needs and *economic* investment in flexibility options, these policy omissions likely have little impact on the findings of this report. In addition, our analysis is not intended as a direct analysis of any particular energy policy, and none of the scenarios represent predictions of the future.

Table 2 shows the assumed technology costs and performance used in our analysis.¹³ Data for new natural gas-fired, wind, and solar capacity are consistent with those found in the National Renewable Energy Laboratory's 2015 Annual Technology Baseline central case (Blair et al. 2015).¹⁴ Fuel prices are from the Energy Information Administration's Annual Energy Outlook 2015 Reference case (EIA 2015). While significant uncertainties exist for future technology cost and performance, we do not conduct a sensitivity analysis as part of this study, again, due to the study's methodological nature. RPM also includes financing assumptions that differ between technologies to account for construction periods, accelerated tax depreciation rules, and tax credits.¹⁵ Transmission expansion is also allowed, however only along existing corridors.

In addition to our core set of scenarios, which use the assumptions shown in Table 2, we simulate a set of sensitivity scenarios (Section 4) with higher natural gas prices and a carbon price. In these cases, natural gas prices are derived from the EIA's Annual Energy Outlook 2014 Low Oil & Gas Resource case, which estimates delivered natural gas prices to be \$ 5.43/MMBtu and \$ 8.20/MMBtu in 2020 and 2030, respectively (EIA 2014). Carbon prices are based on a median across non-zero prices used in multiple integrated resource plans collected in Lawrence

¹² For all scenarios presented, we model Clean Power Plan compliance as a mass-based policy for all states and full credit trading between states. This simplified assumption is not meant to assume that states will choose such a policy, but merely to demonstrate a reasonable implementation of this policy. RPM has the capability to model other compliance variants, including more restrictive trading and rate-based policies, but these variants were not included in the present analysis.

¹³ RPM allows new nuclear capacity to be built starting in 2025, using cost and performance estimates from Blair et al. (2015). This is restricted to those locations in which nuclear facilities were operating in 2010. However, we find that in no scenario does the model choose to build nuclear facilities.

¹⁴ Real 2010 dollars are used throughout this report unless otherwise noted.

¹⁵ RPM uses technology-specific fixed charge rates. Fixed charge rates for NG-CC and NG-CT are 0.117 and 0.111, respectively, for all years. Assumed fixed charge rates for wind and solar PV consider the impact of changing tax credits consistent with federal tax policies as of January 1, 2016 (Mai et al. 2016) and their impact on cost of capital (Mai, Cole, et al. 2015). For wind this leads to a fixed charge rate of 0.0993 in 2015, declining to 0.0975 in 2030. Fixed charge rates for solar are 0.062 in 2015 and 2020; 0.074 in 2025; and 0.084 in 2030. These fixed charge rates are used to calculate amortized capital over 20 years using a nominal weighted average cost of capital of 8.1%.

Berkeley National Laboratory's Resource Planning Portal.¹⁶ Assumed carbon prices in these scenarios grow over time and are \$16.06/metric ton CO₂ in 2020 and \$33.02/metric ton CO₂ in 2030, and are applied uniformly across all of WECC. These scenarios and assumptions are designed to evaluate higher renewable growth as is common in utility planning. The assumptions reflect a proxy for a number of potential future energy policies or market conditions only and should not be interpreted as a policy recommendation or prediction.

¹⁶ See resourceplanningportal.lbl.gov.

Table 2. Technology Cost and Performance Assumptions for New Generation Capacity

	2015	2020	2025	2030
Overnight Capital Costs (2010\$/kW)^{a,b,c}				
Natural Gas-Combined Cycle	900	890	880	860
Natural Gas-Combustion Turbine	770	750	740	720
Wind (best resource class)	1,530	1,480	1,440	1,430
Wind (worst resource class)	1,650	1,630	1,620	1,620
PV Fixed-Tilt	1,740	1,410	1,160	910
PV Single-Axis Tracking	1,830	1,510	1,260	1,010
Fixed O&M (2010\$/kW-yr)				
Natural Gas-Combined Cycle	13	13	13	13
Natural Gas-Combustion Turbine	7	7	7	7
Wind (all classes)	47	46	45	44
PV (all)	15	8	8	8
Variable O&M (2010\$/MWh)				
Natural Gas-Combined Cycle	3	3	3	3
Natural Gas-Combustion Turbine	12	12	12	12
Wind (all classes)	0	0	0	0
PV (all)	0	0	0	0
Heat Rate (MMBtu/MWh)				
Natural Gas-Combined Cycle	6.68	6.62	6.57	6.57
Natural Gas-Combustion Turbine	10.0	9.76	9.50	9.50
Fuel Cost (2010\$/MMBtu)				
Natural Gas	4.38	5.43	6.91	8.20
Capacity Factor (%)^{a,c}				
Wind (best resource class)	52%	54%	55%	56%
Wind (worst resource class)	33%	35%	36%	37%
PV Fixed-Tilt	12–22%	12–22%	12–22%	12–22%
PV Single-Axis Tracking	14–28%	14–28%	14–28%	14–28%

This table represents those technologies installed by the model. Nuclear builds were allowed, but never realized, and new coal plants were not allowed.

^a PV capacity is represented in DC terms. PV capacity factor reflects AC output over DC capacity. AC capacity and output are used for all other technologies.

^b PV refers to utility PV only; rooftop PV performance characteristics are from the dSolar model (Sigrin et al. 2016) and the costs are not presented here nor included in RPM's system cost estimates.

^c Wind resource classes represent different turbine technologies and rotor sizes, based on assumptions from the 2015 Annual Technology Baseline.

2.3 Modeling System Operations and Flexibility

RPM's model results are primarily aimed at informing potential future infrastructure deployment; however, the model necessarily includes dispatch modeling to inform the economics of different generation options. In other words, dispatch and investment decisions are inextricably linked in the RPM optimization to capture changes in cost (and value) of individual assets, both new and existing.¹⁷ RPM uses an hourly chronological dispatch algorithm that is similar to, but reduced from, the algorithms used in production cost models. One key difference is that RPM does not model all hours of the year; this has important implications for how the need for and value of system flexibility are captured. The limited number of hours within the dispatch periods in RPM's optimization is the primary reason additional methods have been developed (Section 2.3.2) to better capture system flexibility.

More specifically, RPM models four dispatch periods within each model year optimization. The dispatch periods (Low, Mid, High, and Peak) are named according to the amount of electricity consumption in each period.¹⁸ Each dispatch period includes 24 hours,¹⁹ with each hour weighted by how many hours in the year it represents for the purpose of accurate *annual* accounting. Furthermore, wind, solar, and load profiles are also scaled so that annual quantities (e.g., annual energy production and consumption) are retained. These measures to ensure accurate annual economics are generally carried through the entire model, applying for example to variable costs. This approach balances the model so that it can capture both seasonal, diurnal, and hourly variations in electricity supply and demand (which are important for valuing operational flexibility needs) as well as longer-term annual values needed to inform investment decisions. This approach does this all while managing the computational tractability challenges of detailed operational modeling (e.g., modeling dispatch decisions over all 8,760 hours in a year).

Figure 2 shows how the reduced dispatch periods in RPM compare with the full hourly data set. It shows the 2010 aggregate Western Interconnection load duration curve (LDC) at hourly resolution compared to the down-selected and scaled representation. Although the model dispatch period structure does well in capturing the general shape of the LDC, it is clear from Figure 2 that some of the important features of the LDC's high and low regions are missing. In fact, it is often during these times that flexibility is most desired. Figure 2 also shows net load duration curves (NLDCs)—defined as the sorted hourly load minus VG production—for the same system with 40% annual generation-based VG penetration for both the raw hourly data as well as the sampled and scaled hours used in the model.²⁰ As illustrated in Figure 2, differences

¹⁷ For example, the capacity factors for thermal power plants are the product of multiple model decision variables rather than predefined quantities.

¹⁸ (Mai, Barrows, et al. 2015; Getman et al. 2015) describe the methods used to select these days within a dispatch period.

¹⁹ We use 2006 meteorological data for wind, solar, and electricity consumption profiles. For each dispatch period, the same 24 sequential hours are selected for wind, solar, and load profiles. For all but the Peak period, those 24 hours are actually the average of a full week, to avoid putting undue weight on any particular day of the week. Using 24 hours for each of four dispatch periods yield 96 total dispatch hours modeled in this configuration of RPM.

²⁰ De Sisternes and Webster (2013) introduce an approximation method where model weeks are sampled based on NLDCs. This method could potentially improve the dispatch representation within RPM; however, it would require

between the full hourly data set and the limited sampling in RPM are even greater when VG is considered, particularly on the low and high ends of the NLDCs where flexibility is likely to be most valuable. The inability of the sampled LDC to fully capture flexibility needs as a result of these differences is the motivation behind the new methods presented in Section 2.3.2, which utilize instead the full aggregated LDC.

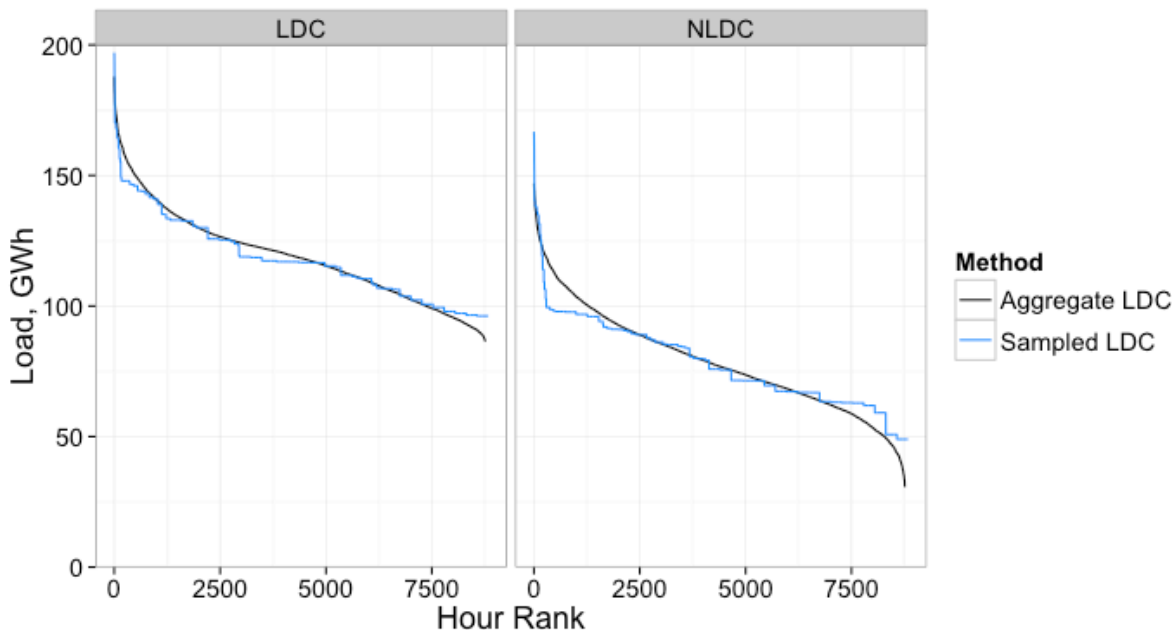


Figure 2. Comparison of 8,760 hourly and RPM 96-hour sampled load duration curves (LDCs) (left) and net load duration curves (NLDCs) (right) for the 2010 Western Interconnection for an illustrative 40% VG scenario

This load duration curve methodology is applied for each of five planning reserve regions based on the North American Electric Reliability Corporation (NERC) sub-regions (Figure 3), including California/Mexico (CAMX), Northwest Power Pool Canada (NWPP-CA), Northwest Power Pool United States (NWPP-US), Rocky Mountain Power Pool (RMPP),²¹ and Southwest Reserve Sharing Group (SRSG). These planning reserve regions have their own capacity requirements in the model. We chose to calculate the LDC and NLDC at the level of NERC sub-region due to this planning reserve requirement and in order to capture the variability in both load and renewables profiles between the regions.

dynamically changing the dispatch periods (and likely other related data) in between model years. This capability is not available in the current version of the model.

²¹ RMPP is also referred to as the Rocky Mountain Reserve Group (RMRG), which is the official NERC sub-region name; however we use the RMPP abbreviation in RPM and throughout the remainder of the report.



Figure 3. Western U.S. NERC sub-regions used for RPM planning reserve regions (NERC 2015a)²²

2.3.1 Co-Optimization of Operations and Investments

Within each dispatch period, RPM’s energy balance equations ensure that supply and demand match in each hour for every node and zone. These equalities account for imports, exports, transmission losses, storage losses, and renewable curtailment. In addition, RPM endogenously co-optimizes energy and reserves. Operating reserves include frequency regulation, spinning contingency reserves, and renewable forecast error reserves. These reserves are referred to as regulation, spinning, and flexibility reserves, respectively, for the remainder of this report. We only model reserves in the “up” direction.²³ Reserve provision is restricted by generator-specific ramp rates and the timescale of the different reserve products. As is common in production cost simulations, we model the need to hold reserve capacity available but do not model reserve events (e.g., contingency events) or operations explicitly.

Renewable curtailment is a decision variable in the model within the optimal dispatch algorithm, but it is constrained by and interactive with the other constrained variables and parameters. In RPM and in real system operations, local curtailment can be caused by transmission limits or by generator inflexibility. We also constrain curtailment on an hourly basis to not exceed the amount of available renewable energy during that hour.²⁴ In addition to this *upper* limit on *hourly* curtailment, we apply a *lower* limit on *annual* curtailment, allowing the model to choose when this curtailment occurs during the four dispatch periods. The lower limit is based on system conditions outside of the model dispatch periods, particularly during low net load period hours where curtailment is expected to occur. Section 2.3.2 describes our new method for calculating this lower limit. Finally, we use this methodology to model curtailment of both the existing renewable capacity and, more importantly, the *marginal* curtailment of potential new renewable energy to inform investment decisions.

²² This information from the North American Electric Reliability Corporation’s website is the property of the North American Electric Reliability Corporation and is available at http://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/2015_Summer_Reliability_Assessment.pdf. This content may not be reproduced in whole or any part without the prior express written permission of the North American Electric Reliability Corporation.

²³ We only model up reserves for several reasons. First, simulations from production cost models frequently show much lower costs for down reserves; oftentimes there is zero cost for providing these products. Additionally, a least-cost model will preferentially have more units at maximum capacity, such that down reserve constraints are typically non-binding. And finally, our model does not in most cases have minimum generation levels, making it even more likely that down reserves would be non-binding.

²⁴ We allow curtailment of utility-scale PV, CSP without thermal energy storage, wind, and fixed-dispatch hydro (e.g., run of river facilities). We recognize that curtailment can be strongly influenced by market rules and other local factors that are not easily modeled. Therefore, our analysis and reporting does not focus on differential curtailment, for example between wind and solar; however our methods do differentiate between technologies.

Relatedly, the capacity value assigned to new investments when evaluating the planning reserves constraint is another aspect of the model in which operational considerations and investment decisions interact. For fully dispatchable resources (e.g., gas capacity), full nameplate capacity is assumed. However, the capacity value of variable generators depends on the time-varying and system-dependent correlation between hourly production and peak net load hours. Conceptually and based on our approximate capacity value method, VG capacity value is estimated based on the upper end of the NLDC as it evolves through time – this is explained in greater detail in Section 2.3.2.

Other model configurations used for the analysis include linear representation of generators, a pipe-flow transmission representation, and must-run coal²⁵ and nuclear plants. Details about these configurations can be found in (Mai, Barrows, et al. 2015), which finds that model capacity expansion outcomes are generally insensitive to many of these configurations.

2.3.2 Net Load Duration Curve-Based Flexibility Adjustments

RPM’s dispatch algorithm captures some of the impacts that increasing VG deployment would have on systems operations, and some of the more significant sources of operational value for flexibility options. These value opportunities include operating reserve provision, energy arbitrage, and reduction in renewable curtailment.²⁶ Operational costs of flexibility provision are also modeled, including efficiency losses and variable costs. However, optimal dispatch in RPM is restricted to a limited number of hours in the year and so does not fully capture the impact of VG integration and value of flexibility during all hours. In this section, we describe our method of estimating the capacity value of VG²⁷ and a new method that better captures VG curtailment. Appendix A provides details on these methods.

Estimating Capacity Value

To capture a more complete range of variability in system operation, we use the LDC and NLDC to capture the tail events, which are not fully represented in RPM’s dispatch of sampled hours. The LDC reflects the sorted hourly demand over all hours in the year. We define the NLDC as the sorted hourly demand net of variable generation. Additionally, for new capacity an incremental load duration curve (ILDC) is calculated independently for each type of potential incremental VG technology installed in each region in which the model could build a resource. The ILDC reflects the NLDC, but it is perturbed by an additional incremental amount of VG, storage, or interruptible load that the model might build, and sorted after the perturbation. Capacity value of renewable resources is estimated by accounting for their average ability to contribute to the capacity needs of the system during the top 100 hours of the year²⁸

²⁵ We assume all large coal-fired units over 300 MW in the focus region and over 550 MW elsewhere are operating during all modeled hours.

²⁶ To be more precise, renewable curtailment reduction is a form of energy shifting or arbitrage where an increase in low- or zero- variable cost renewable energy production in low-demand times replaces production from higher-variable cost sources at high demand times.

²⁷ This method is consistent with the method used in Mai et al. (2015) but refined for VG and extended to apply to storage and interruptible load.

²⁸ These top hours are calculated separately for each LDC, NLDC, and ILDC, such that the hours under consideration may actually shift between the various load duration curves. This allows us to consider changing net load shapes, as for example increased solar power moves the peak hours of a day from the afternoon to evening.

Capacity value is estimated for the aggregate *existing* VG fleet, as well as *marginal* values for potential new VG capacity—separately by technology and for each region—considered. This is calculated on a per-MW basis using the difference in duration curves. For calculating *existing* VG capacity value, differences between the LDC and NLDC are used, whereas for calculating marginal capacity value, differences between the NLDC and ILDC yield the appropriate capacity value estimate. For example, the capacity value of the *existing* VG resources is calculated as:

$$CV = \frac{\sum_{h=1}^N LDC_h - NLDC_h}{N \cdot C}$$

where CV is the capacity value; N is the number of peak hours (100 in this case); LDC and $NLDC$ are the load duration and net load duration curves, respectively; and C is the installed capacity of the resource. The marginal capacity value of a potential resource is calculated similarly but using $NLDC - ILDC$ in the numerator and an assumed amount of incremental capacity C in the denominator. Throughout this work, the assumed increment is 500 MW.

A graphical representation of this method is shown in Figure 4. More robust methods can be used to estimate capacity value; however, this method has been demonstrated to be an effective approximation (Madaeni, Denholm, and Sioshansi 2012).

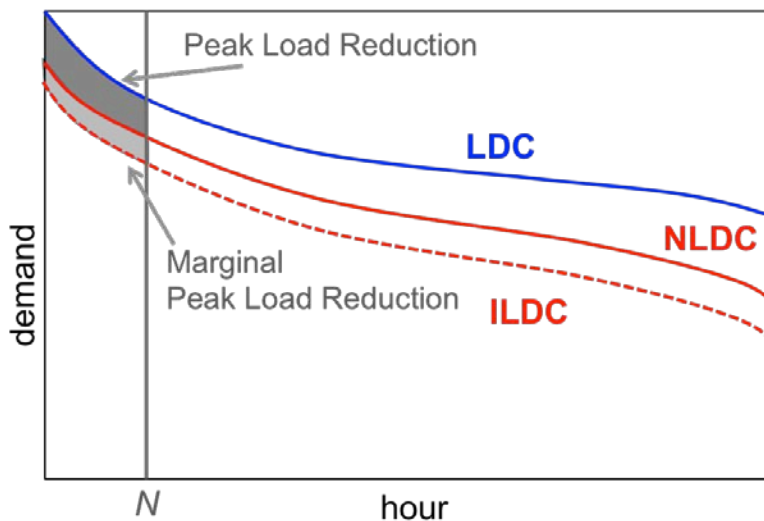


Figure 4. Conceptual depiction of the NLDC-based capacity value estimation method

This method estimates the capacity value of the existing VG fleet collectively at a given point in time. For example, during each solve period, the aggregate output profile²⁹ of all VG capacity installed to that point is used to derive the NLDC, which in turn is used to estimate the average capacity value of this existing VG fleet. This aggregate capacity value, along with the nameplate capacity of the non-VG units, peak demand, and the planning reserve margin, are used to estimate any additional firm capacity needs of the system. Marginal capacity value is similarly

²⁹ We estimate capacity value for each reserve-sharing group separately. For this purpose, we assume five reserve-sharing groups in the Western Interconnection, comprised of five NERC sub-regions. We assume contracts for firm capacity are possible between the reserve-sharing groups.

calculated for each wind and solar technology in each resource region. If the planning reserve constraint is binding, these marginal capacity values can influence the type, location, and amount of VG deployed. All else being equal, the optimization would choose the technology and region with the highest marginal capacity value. This method ensures that estimated—average and marginal—capacity values change with a changing system. Figure 5 shows estimated capacity values from an example RPM run for those technologies currently existing or installed by the model in that year. In this scenario, the estimated average VG capacity value declines over time primarily as a result of greater reliance on wind generation, which typically has lower capacity value than solar. It also shows slightly declining marginal capacity value for solar as a result of net peak hours shifting away from peak solar output hours.

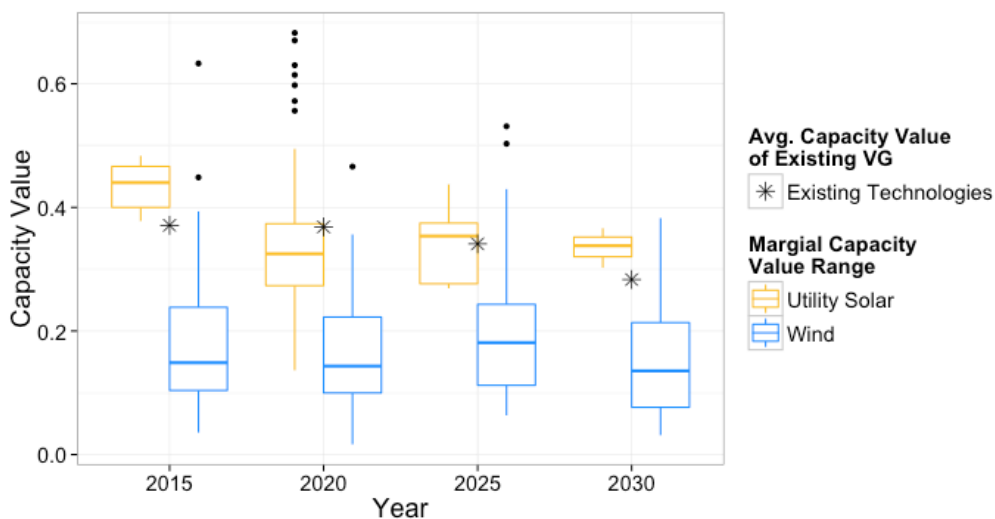


Figure 5. Marginal and existing capacity value from utility-scale solar and wind by year from an example of an RPM scenario that reaches 40% VG penetration by 2030

The box-and-whiskers plots show quartiles including medians. Outliers are shown with black dots. These ranges reflect the capacity that was installed in a given year for the example scenario. Greater ranges would be expected in plots of all possible wind and solar builds. The existing capacity value is an average of all wind, solar, and fixed-dispatch hydropower.

Estimating Curtailment

The method for estimating curtailment is similar to the capacity value method in that it utilizes load duration curves to capture the hourly variability of the system over an entire year. However, it focuses instead on the lower tail of the curve, and it is complicated by the need for an effective system minimum generation level below which the system must curtail energy.³⁰ Figure 6 depicts the methodology for calculating curtailment. The effective minimum generation line represents the limit of the system’s flexibility to de-commit generators. It must be evaluated for a particular system, and it will change as the system’s generator fleet changes and as the penetration of variable generation increases. While curtailment can occur in higher load hours, particularly due to forecast errors, this methodology is focused on capturing curtailment from system flexibility constraints and better representing the impacts of a more flexible system.

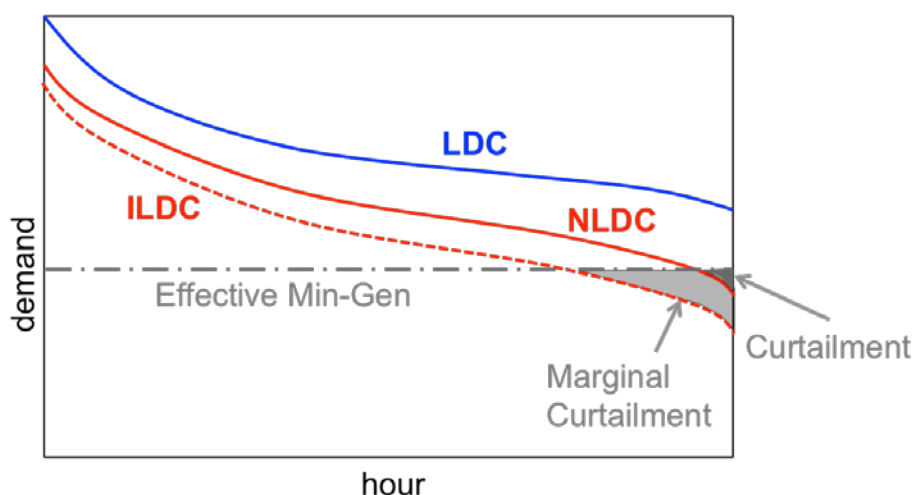


Figure 6. Conceptual depiction of the NLDC-based minimum curtailment estimation method

To implement this curtailment method, we generated a regression model for the effective minimum generation of each NERC sub-region as a function of variable generation capacities. This regression is calculated using production cost model results from two WECC-wide studies (Lew et al. 2013; Brinkman et al. 2015), which are used as training data for the regression. This method of calculating the effective min-gen does not explicitly account for transmission between regions, nor does it account for upgrades in the transmission network that the model might build. However, transmission between regions is implicitly accounted for due to the transmission flows in the training data from the production cost simulations. The regression methodology and results are described in detail in Appendix A.

Once the effective “min-gen” level is determined, the total curtailment of all existing variable generation is calculated as the area underneath the effective min-gen level and above the NLDC. The marginal curtailment is calculated using the incremental area between the effective min-gen line, the ILDC, and the NLDC. Figure 7 shows the existing and marginal curtailment fractions,

³⁰ The minimum generation capabilities of a system are influenced by the flexibility of the system, including physical minimum generation levels, commitment statuses of generators, transmission congestion, and ramping capabilities. The duration curve-based method does not contain commitment decisions, and so is unable to fully capture the actual minimum generation capability of a system.

curtailed VG energy over available VG energy, installed in the system for each model solve year for an example scenario with 40% variable generation penetration in 2030. Only utility-scale resources are allowed to be curtailed (i.e., utility-scale PV, CSP without storage, wind, and fixed-dispatch hydro). Rooftop PV is not included in the curtailment calculation, which is consistent with current practices in the western United States, however these resources are incorporated into the load duration curves so their energy provision is accounted for. In this example, we observe curtailment fractions of less than 5% through 2025, but significantly higher marginal curtailment rates in 2030 due to growing renewable penetrations. This nonlinear behavior is consistent with other studies (e.g., Palchak and Denholm 2014).

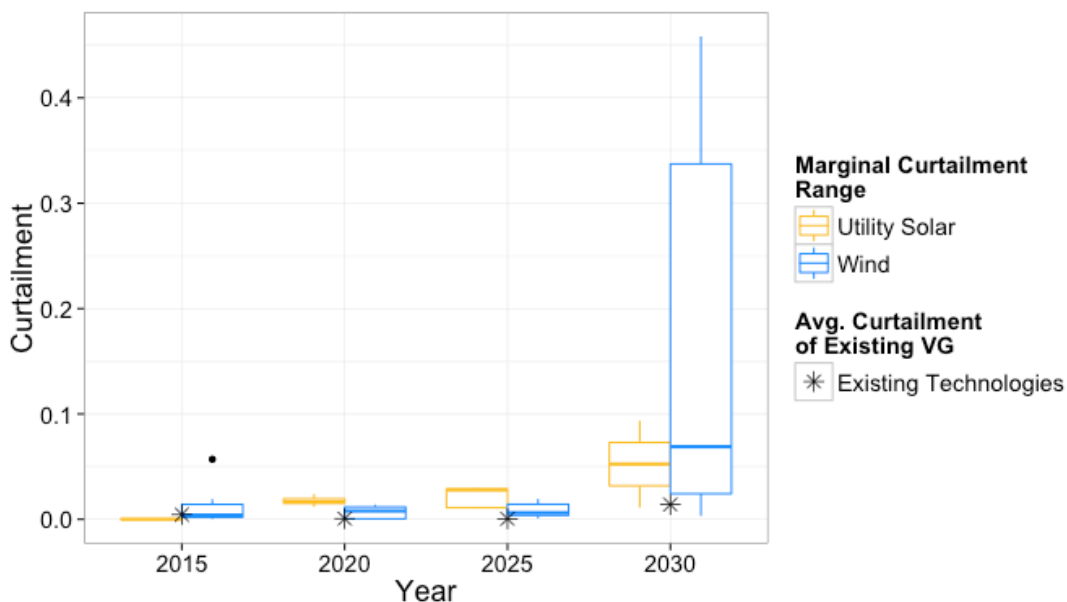


Figure 7. Existing and marginal curtailment fractions for variable resources installed in RPM for each solve year for an example of a scenario with 40% VG penetration by 2030

The box-and-whiskers plots show quartiles including medians. Outliers are shown with black dots. The marginal ranges reflect the capacity that was installed in a given year for the example scenario; greater curtailment ranges are possible across the full range of wind and solar regions and technologies, however these were not economical to the model. The existing curtailment is an average of all wind, solar, and fixed-dispatch hydropower.

RPM uses the calculated existing and marginal curtailment fractions as constraints on the amount of energy that may be provided from variable resources,³¹ enforced separately for each of five NERC sub-regions. The marginal curtailment fraction in particular contributes to the model's decision of where to build additional resources, as locations with higher marginal curtailment levels will be unable to provide as much energy, and so will offer less value, than other locations.

³¹ More specifically, RPM applies a lower limit on estimated annual curtailments based on the method presented, but RPM's dispatch constraints allow for even higher curtailments as a potential source of additional flexibility.

Appendix A provides details on the new methods designed to better reflect curtailment and capacity value. In addition, Section 3 and Appendix B describe how these methods are extended to estimate parameters for interruptible load and storage resources.

3 Modeling Flexibility Investments

The primary purpose of this work is to improve flexibility investment modeling in RPM. We focus on *technologies* whose primary purpose is flexibility, that is, utility-scale energy storage and demand response, rather than institutional or market changes that improve the flexibility of generators. Although the latter are important focuses of study in their own right, as well as likely lower cost to deploy in the near-term, they are more difficult to reflect endogenously in a least-cost optimization model. Note that some incremental improvements to generator flexibility (e.g., increased ramp rates, allowing reserves from wind and solar) can already be modeled within RPM, such that the current work positions the model more generally for future studies that allow a choice between improving existing-fleet flexibility versus investing in flexibility-focused technologies. For this initial work, we further limit scope to two particular technologies: energy-constrained interruptible load and utility-scale storage. Although other flexible technologies, including thermal storage-based demand response are potentially low-cost and thus noteworthy potential resources, their operation is more complex and their costs are highly uncertain. We leave this and modeling of demand response from other highly flexible end uses as future work. In this section, we describe our models for interruptible load and utility-scale storage.

3.1 Interruptible Load

Demand response is currently most often used as a capacity resource, that is, load is briefly reduced either automatically or manually during peak time periods, under direction of the utility (or, less commonly, in response to a price signal). Direct load control of air conditioners can be used in just this way, and many utilities also offer interruptible load utility rates to large customers. In the realm of public data, demand response is primarily reported as a capacity for reducing annual or seasonal peak load (e.g., as “annual potential peak reduction” [in MW] in FERC’s annual Assessment of Demand Response and Advanced Metering reports), and as a capacity subtracted from overall peak demand in the NERC reliability assessments (Lee et al. 2014; NERC 2014).

In RPM, we therefore model interruptible load as a highly energy-limited resource that can be dispatched as part of load balancing, contributes capacity value to the planning reserve requirements, and can be held as spinning contingency reserve when it is not being actively used.³² That is, it can provide three services to the system: capacity, energy, and contingency reserves. Consistent with the notion of interruptible load primarily being a capacity resource, we do not model make-up energy, but rather assume that dispatch results in an overall load reduction.³³ Interruptible load is also resource-constrained in a way that differs from other technologies in RPM—it is built at the BA level, and its capacity is restricted to be less than or

³² The model assumes that contingencies are rare enough that the re-dispatch of interruptible load following a contingency event would not significantly affect its capacity or energy value.

³³ Admittedly for some end-uses (including residential air conditioners), this assumption is likely too simple, however, based on the prevalence of direct load control DR, it appears that operators are able to utilize DR resources as capacity without rebound effects negating the overall effort.

equal to a fraction of the BA’s peak load. For the purpose of integrating with the rest of the optimization model, interruptible load is assumed to be able to ramp fully over an hour (the finest time resolution modeled in RPM); all costs are modeled as annual costs, similar to generator Fixed Operating & Maintenance (FOM); and the BA-level capacity is distributed to individual nodes in proportion to each node’s load participation factor.

The energy-limitedness of these resources is modeled at two timescales—each sub-class of interruptible load is characterized by a daily energy capacity and an annual energy capacity, each expressed in hours. For example, a simplified version of the annual energy capacity constraint for each resource is:

$$\sum_{d,h} P_{d,h} \cdot s_d \leq C \cdot e^{annual},$$

where $P_{d,h}$ is the power “produced” during each modeled hour (denoted by the pair d, h , where d is the dispatch period and h is the hour within that dispatch period) in MW, s_d is the number of hours in the full year represented by that modeled hour, C is the capacity of the resource in MW, and e^{annual} is the annual energy capacity of the resource (again, in hours).

Together, the daily and annual energy constraints capture the premise that most demand response resources agree to terms in which they will only be used for some small portion of any given day, and only a handful to a few dozen times throughout the year. Note that this formulation does not restrict the number of hours in which the resource is used on an absolute basis but rather insists that the sum of the dispatch is less than or equal to a particular amount of energy. It thus accurately represents aggregated demand response resources, albeit (1) only at the hourly timescale (and not at the sub-hourly timescale, which could, for instance, allow the same amount of interruptible load to provide a greater amount of spinning reserves) and (2) only to the extent that it is reasonable to assume constant (rather than time-varying) demand response capacity.

Capacity value for interruptible load is assumed in the planning reserves constraint, and it is calculated using a methodology similar to that used for variable generation. In particular, the contribution of interruptible load to meeting demand in the top 100 hours of net load is determined based on a heuristic dispatch given its daily and annual energy capacities. These energy constraints typically result in capacity values less than one, though they can be very high for products with greater annual and daily energy limits. To calculate the capacity value, we assume operation of these products at times of high demand, and we create a heuristic dispatch profile that respects the resource’s energy constraints. We therefore produce heuristic dispatch profiles that correspond to an approximate best case in terms of operating interruptible load as a capacity resource, which is consistent with the notion that a resource built for its capacity value is more likely to be operated as such. These example profiles are then used in the same way that the variable generation profiles are used to calculate the fractional capacity value of existing and incremental resources. However, unlike variable generation, the dispatch in the optimization model is not required to match these heuristic profiles in any way. The methodology is explained in detail in Appendix B.

Because we model interruptible load as a capacity resource without explicit energy shifting, it is not able to contribute to reducing curtailment, and is therefore only able to support renewables in

limited ways. For example, the use of interruptible load for contingency reserves can free up other resources to provide flexibility reserves. Interruptible load capacity can also make it possible to develop higher-value wind and solar resources in non-capacity constrained regions instead of developing poorer resources in capacity constrained regions.

3.2 Storage

Utility-scale storage is in many ways a more straightforward flexibility option than demand response in that it is built simply to provide grid services, whereas demand response by definition simultaneously provides an end-use service, provision of which must be retained, *and* grid services. However, modeling storage remains significantly more complex than modeling conventional generation, and because its costs are currently typically higher than generation or demand response, value stacking is a relatively more important factor in its deployment.

In RPM, we allow storage to provide the following services:

- **Capacity**, estimated through a dynamically calculated capacity value that depends on the energy capacity of the technology³⁴ and characteristics of the net load curve at and near the peak hours of the year.
- **Energy arbitrage** modeled as part of RPM dispatch. Each storage technology is characterized with a roundtrip efficiency, which implicitly determines the gap in hourly electricity prices needed to make energy arbitrage worthwhile.
- **Curtailement reduction**. The ability to shift variable generation enables a reduction in curtailment, measured through net-load duration curve methods. Curtailment reduction is a specific example of energy arbitrage.
- **Spinning reserves**. Just as is done with our model of interruptible load, storage is modeled with a ramp rate that is sufficiently fast to not be binding on an hourly timescale. This means that storage is able to provide spinning reserves up to its full capacity rather than a smaller, ramp-limited amount. Only resources with at least 20 minutes of storage are allowed to provide this service.
- **Flexibility reserves**. As they are meant to help balance longer-term (1–4 hour) uncertainty in variable generation, storage resources must have at least one hour of storage to provide this service.
- **Regulation reserves**. All storage resources are allowed to provide this short-term service.

Utility-scale storage is generally achieved by technologies such as pumped hydro, compressed air energy storage (CAES), batteries, flywheels, and thermal storage coupled to concentrating solar plants (CSP). Due to their entanglement with other technologies (process heat sometimes generated by a gas-fired combustion turbine and solar field plus steam turbine, respectively), we

³⁴ Equal to energy capacity in MWh divided by power capacity in MW, and thus expressed in hours (h).

do not address CAES or CSP any further and instead focus on the others, which are characterized by an energy capacity (e in MWh/MW), a roundtrip efficiency (η), and a time-varying storage level (L).

Storage level is modeled by letting the model choose the starting storage level L_d^{start} for each dispatch period d , and then calculating changes to that level based on generation P^{gen} , charging P^{chg} , and use for flexibility and regulation reserves (R^{flex} and R^{reg} , respectively):

$$\begin{aligned} L_{d,1} &= L_d^{start} - P_{d,1}^{gen} + P_{d,1}^{chg} \cdot \eta - 0.25(R_{d,1}^{flex} + R_{d,1}^{reg})(1 - \eta) \\ L_{d,h} &= L_{d,h-1} - P_{d,h}^{gen} + P_{d,h}^{chg} \cdot \eta - 0.25(R_{d,h}^{flex} + R_{d,h}^{reg})(1 - \eta), \quad \forall h > 1 \end{aligned}$$

where each operational variable is indexed by the model hour (d, h). The storage level $L_{d,h}$ must always be less than the total energy capacity, and the total dispatch ($P_{d,h}^{gen} + P_{d,h}^{chg} + R_{d,h}^{flex} + R_{d,h}^{reg}$) must always be less than the total power capacity. The initial storage level is allowed to float freely because the dispatch periods do not have an obvious sequential order, and we assume that the long time periods and variability represented by most of the dispatch periods would facilitate the gradual shifts in storage level needed to accommodate 8,760 operations. The use of storage to provide flexibility and regulation reserves results in a partial need to recharge the resource due to the assumption that these signals will be energy neutral, thereby leaving a deficit proportional to one minus the roundtrip efficiency. The proportionality constant of 0.25 was chosen based on the multiplication of two factors, with one factor of 0.5 representing energy-neutrality and another factor of 0.5 representing a uniform distribution of actual dispatch between 0 MW and the amount of capacity committed to that service. For example, on average a 100 MW storage facility providing one hour of flexibility or regulation reserves would spend half of its time charging and half discharging, and would on average be operating at 50 MW, resulting in a change in level of $-50\text{MW} \cdot 0.5\text{h} + 50\text{MW} \cdot 0.5\text{h} \cdot \eta_c = -0.25 \cdot 100\text{MWh}(1 - \eta)$.

The storage level is then constrained to stay within the energy constraint both on an hourly, and a seasonal basis. The seasonal energy constraint is necessary because we allow imbalances over each of our four dispatch periods, and we only require circular balancing to take place over the year. Practically, this lets our model over-generate during the peak demand day, for instance, and make up the balance of the charging at other times, subject to the storage levels remaining within bounds. The detailed model equations can be found in Appendix C.

Capacity value is calculated similarly to interruptible load, except that now storage level and charging must also be accounted for. An example profile is created for each storage technology, again using a heuristic algorithm that in this case charges during periods of low net demand, discharges during peak net load hours, and ensures that the constraints on the energy level are respected. A cycle period of $e(1 + 1/\eta)$ is assumed for each technology, where e is the energy capacity, with a minimum cycle time of one day. The example profile is used to calculate the capacity value of each technology based on dispatch during the top 100 net-load hours of the year. Further methodological information may be found in Section B.2 (Appendix B).

In addition to providing capacity value to the system, storage technologies can also reduce curtailment of variable generation. Similar to the calculation of capacity value, the heuristic

charge and discharge profile informs the ability of storage technologies to reduce curtailment from existing variable generation sources. The methodology provides a curtailment reduction fraction from existing and a marginal curtailment reduction factor from new generators, which is included in the investment decision. Details are provided in Section B.2 (Appendix B).

4 Scenario Framework

We demonstrate these new investment choices in RPM by modeling high and low renewables assumptions, first without either technology investment, then with each new technology type enabled separately. This evaluation is conducted for the Western Interconnection, using SMUD as the focus region. As this work is primarily methodological—and the costs of flexible technologies, especially projected to the 2030 timeframe, are highly uncertain—we do not aim to determine likely deployment scenarios. Rather, we investigate at what price points these technologies would be built, when and where they might be built, and why. Further, we do not model scenarios with both interruptible load and storage (or other flexibility options) available at the same time; therefore, our scenarios may overestimate deployment of these options compared to a situation with greater competition for flexibility provision. On the other hand, the model takes a system-wide optimization approach, which might over-represent the current flexibility of the system. More-constrained operations might offer more demand for flexibility and yield greater deployment of storage and interruptible load than our estimates indicate.

Different approaches are used to determine price points for deployment of the two technologies. Both make use of technology subclasses, which allow for some technology parameters to vary while keeping the overall model structure the same across all subclasses under a given technology category. In this work, for interruptible load, subclass characteristics include the amount of resource (as a fraction of BA peak load) allocated to the subclass, annual and daily energy constraints, and cost. Storage subclass characteristics include hours of energy capacity, roundtrip efficiency, and cost.

Price points for interruptible load deployment are investigated by keeping the amount of resource available per subclass constant and very small. Then within a given scenario, the cost of each subclass is different, such that which subclasses are enabled indicates the price the model was willing to pay to obtain access to that resource. Again due to the methodological nature of this study, the subclass energy capacities and costs are chosen to bracket the problem and make it possible to determine why and to what extent interruptible load might be attractive in particular locations and at particular times; we do not model particular demand response programs or products. On the other hand, we model particular storage technologies for the storage scenarios, with data selected from the DOE/EPRI 2013 Electricity Storage Handbook, from which energy capacity (h), roundtrip efficiencies, and costs (capital, FOM, and VOM) are easily extracted (Akhil et al. 2013). Price points for deployment are in this case determined by applying different assumed cost reductions to storage capital costs for each scenario, and comparing deployment across the scenarios.

4.1 Interruptible Load Scenarios

For interruptible load, we explore the parameter space of daily and annual energy capacities, which roughly correspond to how much a given day's operations could be interrupted and how many times a year the resource could be called. We then ask at what price points these different sub-classes of interruptible load would be enabled on an annual cost (\$/MW-yr) basis.

Our scenarios are consistent with the current understanding of demand response as primarily a capacity resource, that is, load is assumed to be interrupted only occasionally, generally at peak times or during a contingency event, and is compensated primarily for its capacity. However, we are not attempting to model current demand response programs, which FERC estimates to provide capacity resource of approximately 6% of peak demand across the ISOs and RTOs in its jurisdiction (Lee et al. 2014). Instead, each scenario identifies 0.25% of non-coincident peak BA load as potential interruptible load resource. This small quantity is chosen to isolate the marginal value of interruptible load providing capacity, energy, and/or spinning reserves in particular locations and at particular times.³⁵ Further, to determine which service is most important in these different situations, the daily energy capacities are varied within individual scenarios, and the annual energy capacities are varied between scenarios. This provides differentiation in capacity value and energy value; the former both within and between scenarios, the latter mostly between scenarios.

Within each scenario, the 0.25% of peak load is split into 28 equal slices, each of which is assigned to a different technology subclass. The set of subclasses spans four daily energy capacities: 1, 2, 4, and 8 hours; and seven annual costs. The subclasses are then defined as all possible combinations of daily energy capacity and cost—that is, 28 subclasses altogether. Table 3 lists the annual costs and provides an illustration of how the subclasses are constructed. Taken as a whole, the subclasses form a supply curve such that the model is expected to build the least expensive and least constrained resource first (8h, \$22,000/MW-yr), and proceed from there. The relative value of resources with different daily energy capacities depends on which service is most coveted by the model. For example, if spinning reserves are most needed, we do not expect to see much difference in deployment across the different daily energy capacity subclasses, but capacity hungry systems are expected to have a preference for the subclasses with higher daily energy capacities as they tend to have higher capacity values.³⁶ Unlike other infrastructure investment decisions in RPM, interruptible load capacity built by the model in previous years is not retained and must be re-purchased on the same annual cost basis in subsequent model years.

³⁵ We chose this quantity to be (1) small enough to give a good measure of price points (by not shielding more expensive but still valuable-to-the-system resource with abundant less-expensive resource) and (2) large enough to be significant in the numerical decision-making process.

³⁶ For one particular scenario in 2030, the capacity value of the subclasses with 1h daily energy capacity ranged from 0.17 to 0.31, while that for the 8h subclasses was 0.82 to 1.00. We chose this quantity to be (1) small enough to give a good measure of price points (by not shielding more expensive but still valuable-to-the-system resource with abundant less-expensive resource) and (2) large enough to be significant in the numerical decision-making process.

Table 3. Full Factorial Design of Interruptible Load Technology Sub-Classes

Daily Energy Capacity (h)		Annual Costs (\$/MW-yr)		Example Sub-classes
1		22,000		1h, \$22,000/MW-yr
2	X	28,600	=	1h, \$28,600/MW-yr
4	(full factorial)	37,180		...
8		48,334		1h, \$62,834/MW-yr
		62,834		...
		81,684		8h, \$62,834/MW-yr
		106,189		

Interruptible load is modeled as 28 technology subclasses in RPM, one for each combination of daily energy capacity and annual cost shown above.

The cost levels for this analysis are based on the notions that (1) customers will need to be incentivized on an energy basis at a rate greater than retail rates and (2) capital and fixed costs are likely to dominate for this technology, but they are highly uncertain. We therefore model all costs as annual costs (for simplicity's sake), have a floor LCOE of \$0.15/kWh, and then scale up the costs by a factor of 1.3 over seven levels to bracket the highest costs at which these resources may be valuable to the system. The highest level in particular was chosen to exceed the annualized costs for combustion turbines in our model, which are about \$90/kW-yr, under the assumption that demand response more expensive than CTs is unlikely to be competitive. Indeed, the model never chooses to build these highest priced resources, giving us confidence that we have bracketed the problem.

The scenario framework ultimately consists of three different scenarios that vary only in the annual energy capacity assigned to their interruptible load resources. In this case, we do not try to bracket the range of plausible capacities but instead choose values that help discern whether BAs enable interruptible load for capacity, energy, spinning reserves, or some combination thereof. The three scenarios have annual energy capacities of 50h, 100h, and 150h, and are listed in Table 4 along with rough indications of how much value they provide for each of the three aforementioned services. For example, for energy value, the 50h scenario provides significantly less energy per unit of capacity than does the 150h scenario. Since these scenarios are modeled using the same annual costs, interruptible load energy is much more expensive in the 50h scenario than in the 150h scenario (a minimum LCOE of \$0.44/kWh in the former case, compared to \$0.15/kWh in the latter), and thus not as much of the supply curve from the 50h scenario is expected to be used if energy is the primary service that the system needs. Briefly touching on the other services, maximum capacity value also varies across the scenarios, but the availability of each resource to provide spinning contingency reserves is essentially flat (as reflected in the Min. Contingency Capacity column).

Table 4. Interruptible Load Scenario Framework

Scenario Name	Annual Energy Capacity (h)	Max. Capacity Value	Min. Contingency Capacity (h) ^a
50h	50	0.5	8,710
100h	100	1.0	8,660
150h	150	1.0	8,610

The three scenarios vary only in annual energy capacity, the values of which were chosen to discern the relative importance of particular grid services in individual deployment decisions.

^a Minimum contingency capacity refers to the minimum number of hours that can be used for spinning contingency reserves. This is calculated as the total number of hours in the year minus the annual energy capacity. If the full energy capacity of a particular resource is not used, it could instead provide additional reserves beyond what is listed in the table.

The scenario framework is duplicated for a central assumptions case and a high renewables case (implemented with high gas prices and a carbon price as described in Section 2) to explore the impact of having more renewables in the system. Also, in Section 5.1.2 we use the mathematically regular structure of our supply curves to quantitatively analyze deployment drivers, and so take a moment to highlight that structure here. By grouping the interruptible load sub-classes into those with the same daily energy capacity, each scenario contains four supply curves per BA, each with seven cost levels. If we denote the fraction of all resource with a particular daily energy capacity deployed in a given BA as d , we expect d to be 0, 1/7, 2/7, 3/7, etc., up to a maximum value of 1. These values (when they are non-zero) can further be mapped to the maximum annual cost paid by the model with the equation

$$\text{maximum annual cost paid} = \bar{p} = \$22,000/\text{MW} \cdot 1.3^{(d-1/7)/(1/7)}$$

where the factor of 1.3 is the multiplicative difference between each of the annual cost levels listed in Table 3. In Section 5.1.2 we combine this expression with capacity and energy values to derive expectations of deployment difference between and within scenarios depending on which grid service is most important in a particular BA and a particular year, and use those relationships to analyze results.

4.2 Storage Scenarios

For utility-scale storage, we build on the DOE/EPRI 2013 Electricity Storage Handbook to select a dozen technologies of varying energy capacity to represent in RPM (Akhil et al. 2013). These were chosen by sorting all the pumped hydro, battery, and flywheel technologies listed in Appendix B of that report by energy capacity, and then selecting one or two technology types in each energy capacity bin based on relative costs and efficiencies. The resulting technologies are listed in Table 5.

Table 5. Storage Technology Sub-Classes Modeled in RPM

Storage Subclass	Description	Energy Capacity (h)	Roundtrip Efficiency	Capital Cost (\$/kW)	FOM (\$/MW-yr)	VOM (\$/MWh)
c0	Li-Ion for regulation	0.25	0.9	1,038	6,403	1.04
c1	Lead-acid for reserves	0.5	0.9	1,633	5,917	0.49
c2	Li-Ion	1	0.9	1,426	6,984	3.72
c3	Li-Ion	2	0.93	4,570	26,800	2.70
c4	NaCl-Ni	2	0.85	2,040	10,137	2.62
c5	Lead-acid	4	0.9	4,261	13,320	0.97
c6	Iron-chromium	4	0.75	1,500	8,924	1.36
c7	Zinc-air	6	0.8	1,663	7,404	0.74
c8	Pumped hydro	8	0.81	2,350	6,603	0.30
c9	Lead-acid	10	0.9	5,023	9,200	0.50
c10	Iron-chromium	10	0.75	2,484	7,596	0.40
c11	Pumped hydro	16	0.8	2,200	6,130	0.30

To examine deployment potential as a function of price point, this same suite of storage technologies is retained, but the capital cost is multiplied by a fraction less than or equal to one. (Fixed and variable O&M costs are kept constant.) We do not attempt to model realistic learning curves or time-varying cost reductions for any technology. Before 2020, the values in Table 5 are used. In 2020 and beyond for each cost scenario, a fractional cost between 1.0 and 0.05 of the capital cost values in Table 5 are used to explore the effect of different cost levels on storage deployment.³⁷ However, these multipliers are not applied to the pumped hydro technologies (c8 and c11) because of the relative maturity and site-dependence of that technology, and we further limit the reduction to a minimum cost of \$100/kWh. The scenarios studied are listed in Table 6.

Table 6. Scenario Framework for Storage Technologies

Scenario	Capital Cost Multiplier Ranges for Subclasses with at Most 2h of Storage	Capital Cost Multiplier Ranges for Non-Pumped Hydro Subclasses with More than 2h of Storage	Capital Cost Multiplier for Subclasses c8 and c11 (Pumped Hydro)
1p00	1.00	1.00	1.00
0p50	0.50	0.50	1.00
0p10	0.10	0.10 – 0.40	1.00
0p05	0.05 – 0.10	0.09 – 0.40	1.00

³⁷ The latter value is loosely based on optimistic price projections for electric vehicle battery packs, for instance see Nykvist and Nilsson (2015).

4.3 Overall Framework

Taken altogether, the scenarios analyzed in this work are listed in Table 7. In particular, the labels in the main body of the table are the scenario names we use in the remainder of the report to specify which RPM run or runs we are discussing. Differences under low and high renewable penetrations are analyzed by duplicating each technology-specific set of scenarios across two sets of baseline assumptions. The Base assumptions achieve 32% renewable penetration on an energy basis by 2030, whereas the High RE assumptions achieve 40% penetration by the same metric.

Table 7. Summary of Scenario Framework

	Base Assumptions	High RE Assumptions
No New Flexible Technologies	Base	High RE
Interruptible Load Scenarios	Base-Interruptible Load	High RE-Interruptible Load
Annual energy capacity of 50h	50h	High RE-50h
Annual energy capacity of 100h	100h	High RE-100h
Annual energy capacity of 150h	150h	High RE-150h
Storage Cost Scenarios	Base-Storage Cost	High RE-Storage Cost
Capital costs at 100% of 2013	Base-1p00	High RE-1p00
Capital costs at 50% of 2013	Base-0p50	High RE-0p50
Capital costs at 10% of 2013	Base-0p10	High RE-0p10
Capital costs at 5% of 2013	Base-0p05	High RE-0p05

5 Results

The scenarios described in Section 4 are used to explore the potential value of flexibility investments to the Western Interconnect from 2015 to 2030. The Base assumptions include a central gas price trajectory and current policy as of the beginning of 2016. The High RE assumptions include a high gas price trajectory and an effective carbon price, which does not reflect any specific policy in particular but is meant to represent a generic future in which carbon emissions are penalized. Details are provided in Section 2.2.

For reference, the capacity expansion for the Base scenario is shown in Figure 8. The difference in the capacity expansion for the High RE scenario as compared to the Base scenario is in Figure 9, which clearly shows additional wind and solar deployment, especially in the years 2025–2030. Regionally, CAMX mimics the overall pattern seen across the interconnection: economic builds of wind and solar result in less need to provision CT capacity. RMPP and NWPP-U.S. build additional renewables without reducing investments elsewhere, because for those regions the planning reserve constraint is not binding for any of the model years. SRSR, on the other hand, builds additional wind and solar in 2025, but by 2030 swaps a capacity-value equivalent amount of wind for solar, netting an overall difference of +1.4 GW wind and -0.7 GW solar.³⁸ Although

³⁸ In the remainder of the report, we will continue to discuss many results in the context of NERC sub-regions, which are the geographical unit of accounting for planning reserves. These results are not meant to be interpreted as

the model allows new transmission to be built along existing corridors after 2021, this investment option is not exercised in either reference scenario. This finding is likely influenced by the primarily zonal structure of the RPM-SMUD model, along with the hurdle rates applied along many inter-BA lines; RPM configurations for other focus regions may identify more need for transmission.

The methodology described in Section 2.3.2 and Appendix A finds some curtailment for both the Base and the High RE scenarios. This is depicted, for instance, in the dispatch plots of Figure 10. By 2030, the Base scenario reaches a renewable penetration of 32% and a variable generation (wind and solar) penetration of 28%, whereas the high renewables case has penetrations of 40% and 37%, respectively, all on an annual generation basis. Average curtailment of wind and solar in 2030 is 1.0% in the Base scenario and 1.3% in the High RE scenario; however, marginal curtailments are often higher.

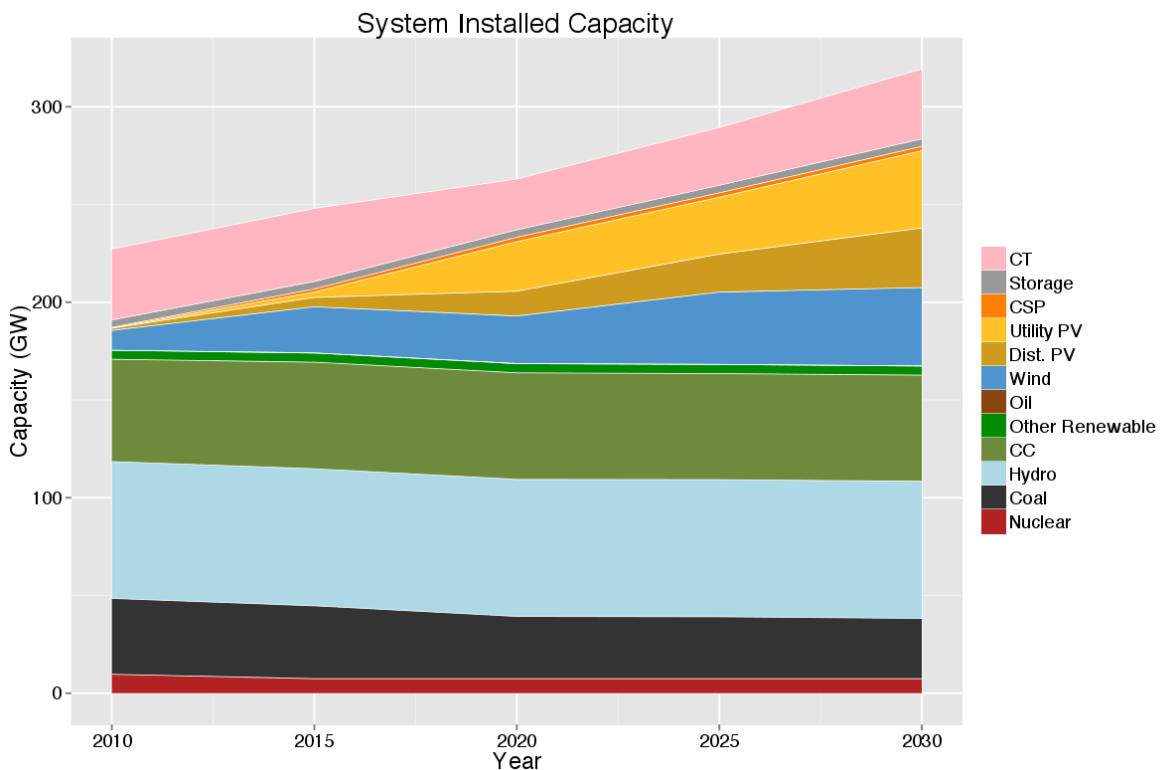


Figure 8. Base scenario system-wide capacity expansion over time
Retirements are visible as declining capacity.

precise analyses of these sub-region’s capacity needs over the study period of interest, as they are highly sensitive to model inputs, some of which are significantly uncertain from the point of view of the authors. A prime example of this uncertainty is the row labeled “External & Imports” in Table 8.

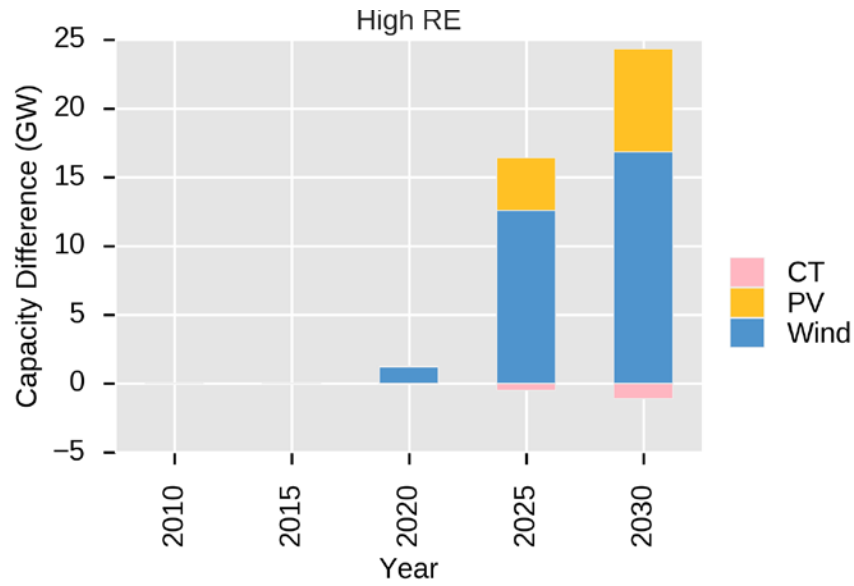


Figure 9. Difference in capacity between the High RE and Base scenarios

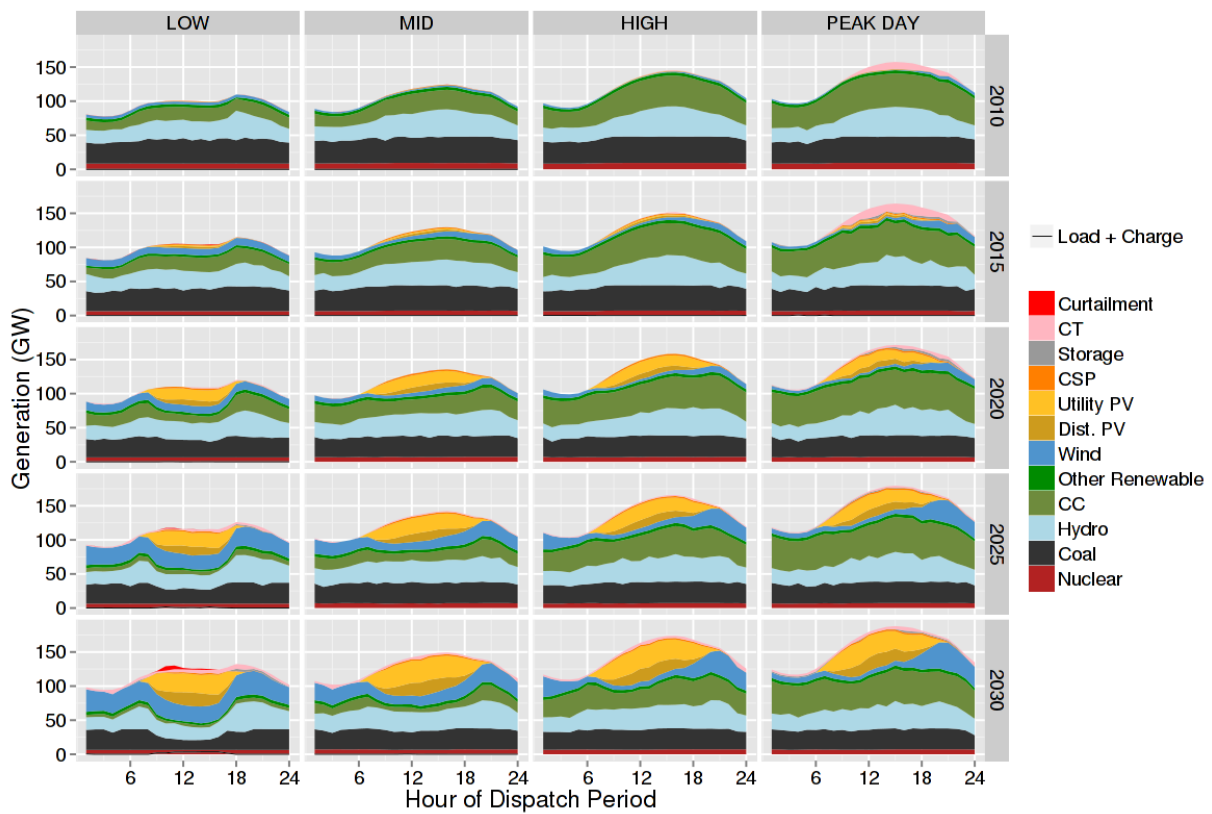


Figure 10. Hourly dispatch for the High RE scenario

The scenarios presented in Figures 7–9 do not include new interruptible load and storage capacity, which are the focus of this report, because they were specifically excluded from the investment options. The following sections describe their impacts in detail. As will be shown, the availability of interruptible load and storage amount to minor perturbations of the overall scenario results presented here, with a minor exception for the lowest cost storage scenarios. In particular, the dispatch plots here can be taken as largely indicative for all modeled scenarios. The most extreme storage cases still do not impact the results as much as the High RE assumptions do compared to the Base case, with, for instance, the former impacting renewable energy penetrations by less than 0.5% of annual generation and the latter representing an 8 to 9 percentage point swing.

5.1 Interruptible Load

Interruptible load provides three services to the system: capacity value via the planning reserves constraint, spinning reserves, and energy dispatch (with load typically interrupted at peak times in fulfillment of its capacity-providing role). The three scenarios described in Section 3.1, taken together and then duplicated for both Base and High RE conditions, show (1) at what prices these services are acquired in different parts of the Western Interconnect as it evolves through the model years 2015–2030 and (2) how different levels of energy availability, on a daily basis and an annual basis, affect the results. In addition to presenting the system-wide perspective, we present NERC sub-region specific results to highlight capacity built to satisfy planning requirements, but these results are not meant to be interpreted as precise analyses of capacity needs, which are in practice highly sensitive to uncertain model inputs (e.g., the row labeled “External & Imports” in Table 8, as well as load growth assumptions). Similarly, we also discuss BA level results, because they allow us to represent a wider diversity of power system conditions than that seen in the aggregate, even at the NERC sub-region level.

5.1.1 General Trends in the Least Constrained Scenarios

To first examine the overall usefulness of interruptible load in RPM, we start with the least-constrained 150h scenario in which all sub-classes are allowed to provide up to 150 hours of energy over the course of the year. Figure 11 shows the difference between the capacity expansion of the 150h scenario and the Base scenario (whose detailed capacity expansion is shown in Figure 8). Recall that we require the model to re-purchase its interruptible load each year, such that while the capacity differences for all other technologies shown in Figure 11 are cumulative, the interruptible load shown is the amount that the model chose to enable (by paying an annual cost) for that year. Interruptible load is also modeled as a limited resource with the maximum amount available in the model equal to 0.25% of non-coincident load by BA. In particular, Table 8 shows the total amount of resource per energy capacity class in each NERC sub-region in 2030.

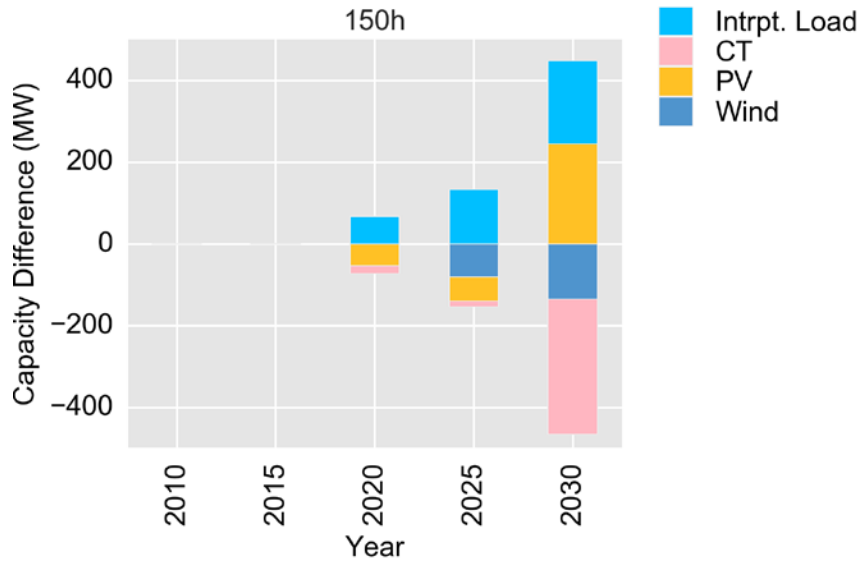


Figure 11. Capacity differences for the 150h scenario relative to the Base scenario

Interruptible load capacity is repurchased in each model year, while all other builds persist to the next model year.

Unsurprisingly, interruptible load appears to become a generally more attractive resource over time. Figure 11 also shows that interruptible load mostly offsets renewables in 2020 and 2025, and new thermal peaking capacity (i.e., CTs) in 2030. The main drivers in these results are the NERC sub-regions that need additional capacity to meet their planning reserves constraint: CAMX, SRSG, and NWPP-Canada. California’s ability to meet its 50% RPS requirements via REC trading is also a major factor, as will be seen momentarily.

Table 8. Interruptible Load Resource Available in each NERC Sub-Region

NERC Sub-Region	Amount of 8h Resource Available in 2030 ^a (MW)
CAMX	45.4
NWPP_CAN	18.1
NWPP_US	34.7
RMPP	9.1
SRSG	18.8

^a These quantities are equal to 0.0625% of non-coincident peak load by BA, summed to the NERC sub-region level. The same amount of resource is available no matter the daily or annual energy capacity of the resource. However, it does vary by model year (based on load growth).

In general, the time varying deployment of interruptible load in CAMX, NWPP-Canada, and SRSG, shown in Figure 12 as the fraction of available interruptible load resource deployed in each year, is driven by the planning reserves constraint, the components of which are shown in Table 9 for the years 2020–2030 (next page). The data in that table clarify that for all cases of NERC sub-region deployment greater than 15% of available resource, interruptible load in that case is contributing positively to satisfying a binding planning reserves constraint. This need is highly (sub)-system dependent, with NWPP-Canada and CAMX needing capacity in all years, NWPP-U.S. and RMPP needing capacity in no years, and SRSG needing capacity in 2025 and 2030.

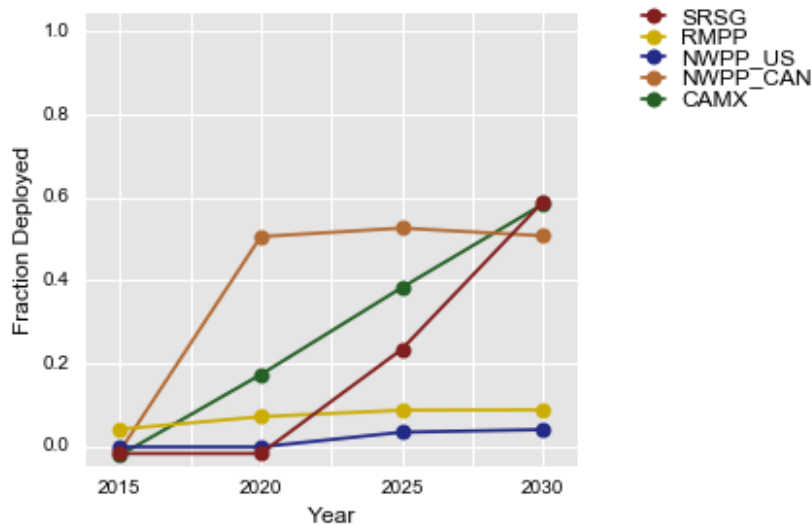


Figure 12. Deployment of interruptible load for the 150h scenario by year and NERC sub-region

Viewing capacity differences by NERC sub-region starts to explain the use of interruptible load to replace renewables in 2020 and 2025, and to enable wind and solar switching in 2030 (see Figure 13). In particular, larger sub-region deployments for those regions that build interruptible load specifically for capacity, namely CAMX, NWPP-Canada, and SRSG, are apparent. In NWPP-Canada, the predominant effect is the enablement of interruptible load instead of building CT's. However in CAMX, in addition to replacing CT's in 2030, capacity from interruptible load allows California to defer its own renewable builds in favor of procuring cheaper renewables to meet its 50% RPS, through increased REC trades with NWPP-U.S., RMPP, and SRSG in 2020, 2025, and 2030, respectively. A technology preference switch from wind to solar is also apparent between 2025, in which RMPP wind is preferred, and 2030, in which additional SRSG solar fills the gaps left by CAMX renewables and SRSG CTs. The technology switching is likely influenced by the changing relative prices of wind and solar over the modeling horizon.

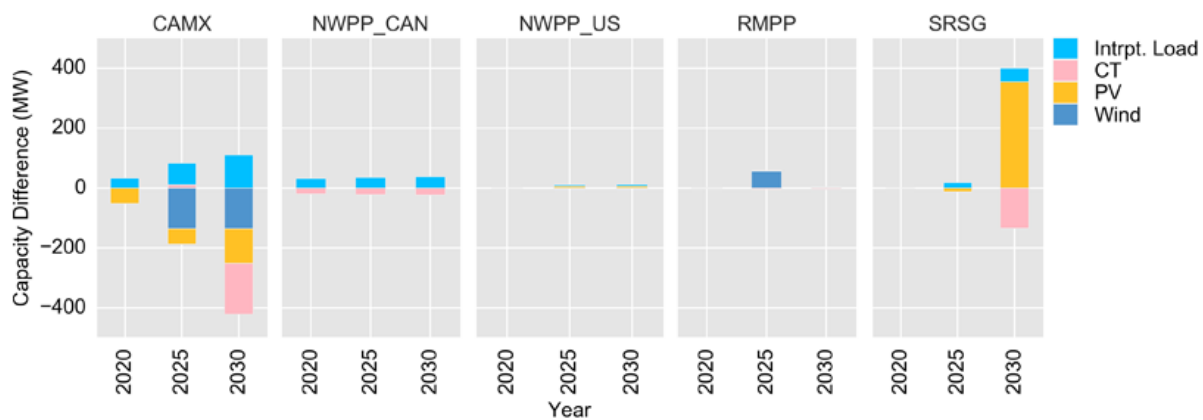


Figure 13. Capacity differences by year and NERC sub-region for the 150h scenario

As an example of other possible uses for interruptible load, we explore RMPP's nearly constant (over time) deployment of some interruptible load even though it is not needed for capacity. In that case, the average deployment level of 7% to 9% for 2020 through 2030 masks a severe difference in deployment by BA. In particular, RMPP is composed of three BAs: PSC and WACM form the Colorado operating reserves sharing group, and PACE Wyoming is the sole member of the Wyoming operating reserves sharing group. It is this latter BA in which all of the interruptible load is being deployed, at levels from 42.9% in 2015 to 85.7% in 2030. Looking at its dispatch, we see that in 2015, 2020, and 2030, all interruptible load capacity in PACE-Wyoming, when it is not being actively interrupted, is being held for spinning reserves, that is, its capacity factor including reserves is 1.0. In 2025, this picture only moderates slightly with the smallest capacity factor including reserves being 0.980. In contrast, there are other BAs that deploy interruptible load but do not use it to provide spinning reserves at all (their capacity factors with and without reserves are equal).

Table 9. Planning Reserve Constraint Summary for the 150h Scenario

Meaningful capacity contributions from interruptible load are highlighted in green.

Capacity (MW)	CAMX			NWPP-Canada			NWPP-U.S.			RMPP			SRSG		
	2020	2025	2030	2020	2025	2030	2020	2025	2030	2020	2025	2030	2020	2025	2030
Existing & Planned ^a	56,430	63,380	67,343	28,788	32,183	34,692	58,832	59,052	59,477	18,103	18,186	18,822	34,720	34,787	35,995
External & Imports ^b	16,250	16,250	16,250	(4,500)	(4,500)	(4,500)	(1,000)	(1,000)	(1,000)	(2,500)	(2,500)	(2,500)	(4,000)	(4,000)	(4,000)
Capacity Need	82,927	85,071	87,307	27,737	30,209	32,996	52,281	54,480	56,808	14,816	15,506	16,310	29,906	31,963	34,194
Capacity Shortfall	10,248	5,441	3,714	3,449	2,526	2,804	(5,551)	(3,573)	(1,670)	(787)	(181)	(12)	(814)	1,177	2,199
New Dispatchable	-	1,338	3,222	3,429	2,504	2,781	-	-	-	-	68	-	-	-	112
New VG	10,223	4,050	412	-	-	-	148	372	490	102	209	145	-	1,163	2,053
New Interrupt. Load	24	53	80	20	22	23	-	3	3	1	2	2	-	13	34
Excess Capacity	-	-	-	-	-	-	5,699	3,947	2,162	890	459	159	814	-	-

^a Includes 2010 capacity, prescribed builds through the model year, and capacity built by the model in previous years; This is capacity for planning reserves purposes, so capacity value fractions are applied to non-dispatchable and energy-constrained resources.

^b This row lists our estimates of capacity physically located within one NERC sub-region, but for contractual or other reasons counts toward another sub-region's planning reserves, combined with estimated transfers at time of peak load. These estimates were made using several resources (NERC 2015b; NERC 2015a) combined with modeling judgment, and they are only indicative and are not highly accurate, especially in out-years. Note that CAMX, RMPP, and SRSG are summer-peaking systems whereas NWPP-Canada and NWPP-U.S. are winter peaking systems, and RPM models the planning reserves constraint accordingly.

High Renewables Scenarios

Comparing Figure 14 with Figure 10 shows that interruptible load deployment is reduced and delayed under conditions that favor more rapid deployment of renewables. Interruptible load deployment steadily increases over time, with noticeably less deployment in 2025 as compared to the Base case. Again, the primary driver here is capacity value. Figure 15 shows that while NWPP-Canada's deployment is unchanged, CAMX has delayed and reduced deployment, which is directly tied to their delayed and reduced capacity needs. This effect is likely exaggerated by RPM's limited ability to endogenously retire underutilized capacity³⁹; however, it is generally true that interruptible load will be less attractive as a capacity resource the more capacity there already is on the system.

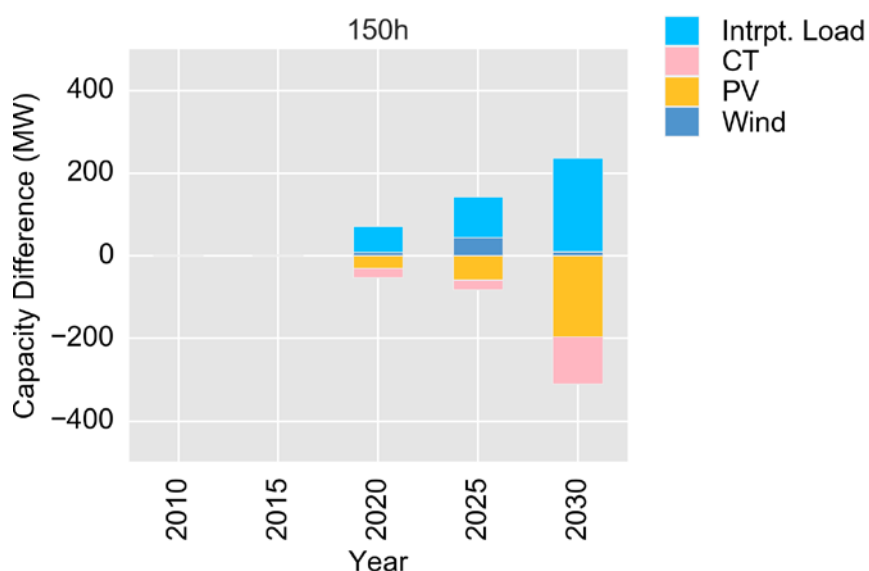


Figure 14. Capacity differences for the High RE 150h scenario relative to the High RE scenario

Interruptible load capacity is repurchased in each model year.

At a regional level, the overall patterns for CAMX, NWPP-Canada, and RMPP are similar between the High RE and Base scenarios, with the exception that RMPP wind is preferred in 2020 and CAMX wind is preferred in 2025 in the High RE case. There is a noticeable change in deployment patterns for NWPP-U.S. and SRSG, however, reflecting a difference in relative attractiveness for the next increments of renewables in the Base and High RE cases. In particular, NWPP-U.S. builds more interruptible load in High RE conditions, and also provides additional wind in 2025 and 2030; whereas SRSG goes from using interruptible load to provide more solar in 2030 under Base conditions, to using interruptible load to replace solar in 2025 and 2030 under High RE conditions. These changes are a continuation of the trends seen by directly comparing the High RE Scenario to the Base Scenario (both without interruptible load). In particular, NWPP-U.S. economically builds more wind and solar under High RE conditions, while SRSG shows a preference for more wind, first in addition to more solar in 2025, but then

³⁹ In the configuration of RPM used here, coal units are required to maintain a capacity factor of 0.07 or retire some capacity. This option is not exercised in any of the interruptible load scenarios.

at the expense of solar in 2030. NWPP-U.S.’s additional appetite for interruptible load is curious, however, as it still has excess capacity according to the model. We thus turn now to discerning the interruptible load grid services that are of most interest, when and where.

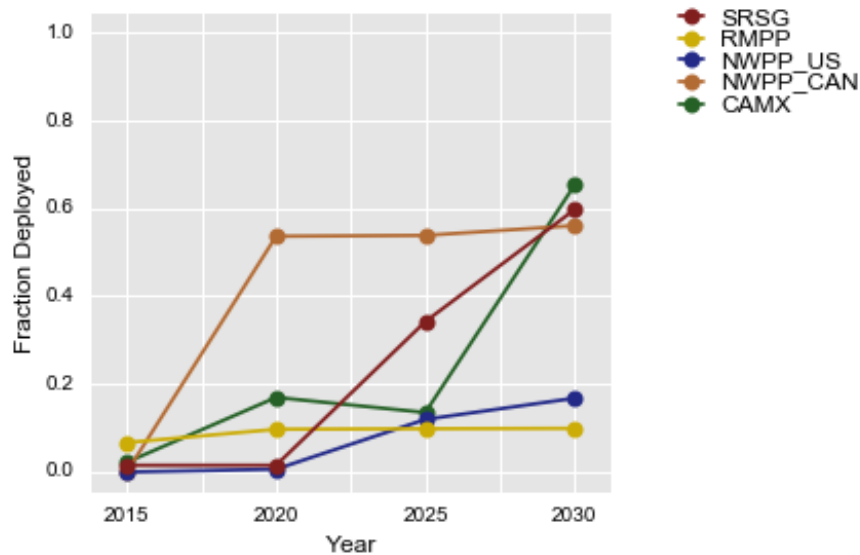


Figure 15. Deployment of interruptible load for the High RE 150h scenario by year and NERC sub-region

5.1.2 Discerning the Primary Grid Service for which Interruptible Load is Enabled

The two contrasting cases of CAMX and RMPP in the Base scenarios highlight the fact that interruptible load (and flexible technologies in general) can and do provide multiple grid services, and that which service or services are most important is highly technology- and system-dependent. The scenario framework described in Section 3.3.1 was designed to make it easier to discern exactly which service is most important in the particular cases under study.

Interpretation of the Scenario Framework

The annual energy capacities in the three interruptible load scenarios listed in Table 4 reflect the RPM capacity value calculations. In particular, because the top 100 hours of net load are used to capture resources’ ability to contribute firm planning capacity, the 50h annual energy capacity scenario yields a maximum capacity value of 0.5 for all of the interruptible load resources, but it is possible for them to capture full capacity value in the 100h and 150h scenarios. On the other hand, the three scenarios obviously differ in the amount of annual energy available from each resource, and they are largely the same with regard to how much spinning contingency service they can provide. This is illustrated somewhat more clearly by taking the annual cost tranches applied to each resource and then dividing them by the maximum (or minimum, in the case of spinning reserves) amount of each service that they can provide. The results of these calculations are shown in Table 10.

Table 10. Minimum Prices for Grid Services from Interruptible Load as a Function of Annual Cost and Annual Energy Capacity

Annual Costs (\$/MW-yr)	Min. Capacity Cost (\$/MW)			Min. Energy Cost (\$/MWh)			Max. Spinning Reserves Cost (\$/MW-h)		
	50h	100h	150h	50h	100h	150h	50h	100h	150h
22,000	44,000	22,000	22,000	440	220	147	2.5	2.5	2.6
28,600	57,200	28,600	28,600	572	286	191	3.3	3.3	3.3
37,180	74,360	37,180	37,180	744	372	248	4.3	4.3	4.3
48,334	96,668	48,334	48,334	967	483	322	5.5	5.6	5.6
62,834	125,668	62,834	62,834	1,257	628	419	7.2	7.2	7.3
81,684	163,368	81,684	81,684	1,634	817	544	9.4	9.4	9.5
106,189	212,378	106,189	106,189	2,124	1,062	708	12.2	12.3	12.3

The minimum capacity cost columns assume the resource is able to achieve the maximum capacity values shown in Table 4, which is not the case for resources with small daily energy capacities.

Notice that, for the same annual cost on a nameplate capacity basis, the effective cost for each service actually varies by scenario, such that how the depth of deployment across the different scenarios varies for a given BA provides some indication as to which service that BA values most. In particular, beyond the stylized service costs shown in Table 10, we can calculate the maximum price paid by the model for a given service in a given BA using the fact that fractional deployment d can be directly mapped to the annual costs paid (a detailed derivation is available in Section 4.1):

$$\text{maximum annual cost paid} = \bar{p} = \$22,000/\text{MW} \cdot 1.3^{(d-1/7)/(1/7)}.$$

Then, we can divide this quantity by the amount of service provided or available per unit of nameplate capacity. For providing firm capacity, this quantity is the capacity value v^{cap} (MW/MW) calculated per the methods described in Section 3.1 and Appendix B; for energy it is the annual energy capacity e^{annual} (MWh/MW) equal to 50 h, 100h, or 150h; and for spinning reserves it is $8760 - e^{annual}$ (MW-h/MW). For the moment, we represent these factors generically as v , and then we consider comparing deployment across resources with different daily or annual energy capacities, or both. Because the amount of interruptible load available in the system is small, we expect a given BA to pay a similar price for the service it most needs independent of the particular interruptible load resource, but dependent on underlying conditions (Base or High RE) and year. However, due to our coarse discretization, and also, for instance, differences in the relative ability of resources with different daily energy constraints to contribute to the highest value hours, we do expect some variation. To this end, consider a resource 1 and a resource 2 and suppose that the price paid for resource 2 is a multiplicative factor α times that paid for resource 1:

$$\bar{p}_2 = \alpha \bar{p}_1.$$

Then, for the same BA, underlying conditions, and year we expect α to be close to one and for the deployment of the two resources to be related to each other via:

$$\bar{p}_1 = \frac{1.3^{(d_1-1/7)/(1/7)}}{v_1} = \frac{\bar{p}_2}{\alpha} = \frac{1.3^{(d_2-1/7)/(1/7)}}{\alpha \cdot v_2}.$$

By rearranging we get:

$$d_2 - d_1 = \frac{1/7}{\log 1.3} \cdot \log \frac{v_2}{v_1} + \frac{1/7}{\log 1.3} \cdot \log \alpha. \quad (1)$$

Now, if we plot deployment fraction differences against, for instance, $\log v_2^{cap} / \log v_1^{cap}$, and the BA is primarily building interruptible load for capacity, we would expect all comparisons of this sort for that BA to lie along a line with slope $0.1428 / \log 1.3$ and y-intercept close to zero. For BAs enabling interruptible load for another purpose we would expect little to no correlation with $\log v_2^{cap} / \log v_1^{cap}$, that is, a slope of zero, and a y-intercept value close to that calculated by the first term of the right hand side of Equation 1 expressed for the service that is of most interest to that BA (i.e., v replaced with e^{annual} in the case of energy service, or $8760 - e^{annual}$ in the case of contingency service). Finally, note that for BAs that primarily need spinning reserve, we expect $d_2 - d_1 \approx 0$ in all cases.

Just such a plot is shown in Figure 16 for the Base-Interruptible Load scenarios in 2020, where the green-tinted BAs belong to CAMX, AESO and BCTC comprise NWPP-Canada, and PACE Wyoming is in RMPP. Note that it confirms what we found earlier—CAMX and NWPP-Canada enable interruptible load to provide firm capacity (as those comparisons generally follow the 0-intercept, $0.1428 / \log 1.3$ -slope line shown on the plot), and PACE Wyoming enables interruptible load to provide reserves. Some subtleties also emerge, as the failure of some of the CAMX 1h daily energy capacity resources to fall along the line indicate that perhaps energy and contingency service are also valuable there. This is in contrast to NWPP-Canada, which shows a more consistent preference for capacity.

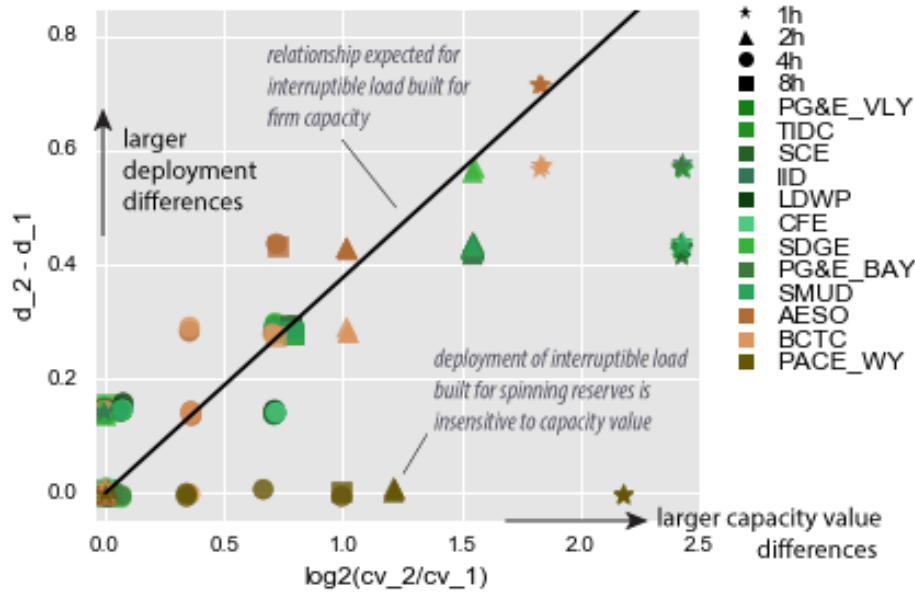


Figure 16. Deployment fraction differences versus the logarithm of capacity value ratios, Base-Interruptible Load scenarios, 2020

All possible comparisons across daily energy capacities, with resource 2 fixed to have an annual energy capacity of 150h, and resource 1 otherwise chosen to be more constrained than resource 2.

To separate resources built for energy capacity from the others, we can instead plot deployment fraction differences against $\log e_2^{annual} / \log e_1^{annual}$. Using both types of plots for the High RE scenario in 2030, shown in Figure 17, we see that CAMX, NWPP-Canada, and SRSG are using interruptible load for firm capacity in this case; the BAs in NWPP-U.S. and RMPP are more focused on its use for energy and spinning reserves. We also observe BAs taking advantage of multiple services simultaneously. For one example, NWPP-U.S. points with nonzero $d_2 - d_1$ values (more clearly visible in the capacity-value plot on the left) are for Nevada Power (NEVP), which has no within-scenario deployment differences, but does have less deployment in the 50h scenario than in the others. This is in contrast to the other NWPP-U.S. BAs that deploy interruptible load, PACE Utah and SPP, whose deployments are constant across all energy capacity variations (both between and within scenarios). This suggests that NEVP enables interruptible load for both its energy and its reserves value while PACE Utah and SPP almost exclusively focus on providing contingency reserves. As another example, the difference in deployment for the 2h daily energy resource in SRSG in 2030 for the 100h as compared to the 150h scenario appears to be driven by energy capacity concerns, as the capacity value differences are almost negligible (on the order of $1E-4$), while the deployment changes from 3/7 to 4/7 in SRP, TEP, and WALC. This hypothesis is further bolstered by a similar pattern in 2025 (for the same BAs, but deployment levels jumping from 1/7 to 2/7), when there are absolutely no capacity value differences.

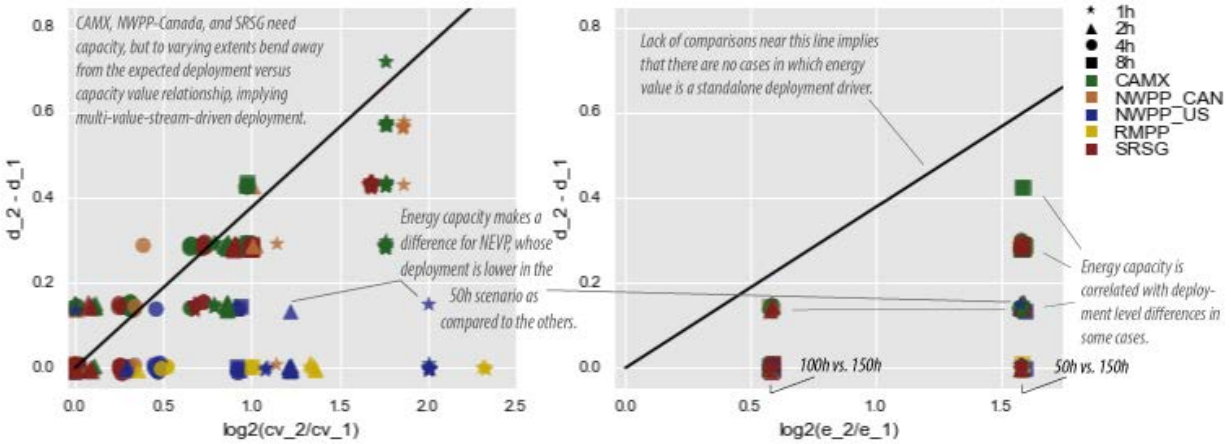


Figure 17. Deployment fraction differences versus the logarithm of capacity value ratios (left) and the logarithm of annual energy capacity ratios (right) in the High RE-Interruptible Load scenarios, 2030

All possible comparisons are plotted on the left. The right plot (energy capacity ratios) only includes comparisons between the 150h scenario and the others, with resources matched on daily energy capacity.

5.1.3 Summary of Price Point Findings

To bring the analysis back to a discussion of price points, we show in Table 11 the capacity deployed, the available resource, and the resulting fractional deployment for interruptible load as a whole for the Base 150h scenario. Note that when shown for a particular daily energy capacity in a particular BA, fractional deployments are a direct reflection of the maximum cost paid by the system for that resource (e.g., 0.1428 means that only \$22,000/MW resource was built, and 0.4286 means that the system availed itself of all of the resource through the first three price points, up to \$37,180/MW). At more aggregated levels (multiple daily energy capacities and/or multiple BAs), the fractional deployments represent a load-weighted average, and are thus indicative of the price the system was willing to pay, but they may mask large differences between technology sub-classes and/or BAs.

Table 11. Interruptible Load Deployment and Resource by Year for the 150h Base Scenario

Year	All Sub-Classes		
	Capacity (MW)	Resource (MW)	Fraction Deployed
2015	2	441	0.00
2020	68	460	0.15
2025	133	482	0.28
2030	202	505	0.40

This aggregated picture gives a rough indication that interruptible load is attractive at moderate to high prices at times. The actual price points reached by year and NERC sub-region are shown in Table 12 and Table 13 (on the following pages) by listing the number of BAs in each sub-region that deploy capacity in each price bin for the 150h scenarios, and for the 1h and 8h daily energy capacity resources in particular. The results for the Base scenarios are in Table 12; High RE results are shown in Table 13. In these tables, combinations of sub-region and year identified above as building interruptible load for capacity purposes are highlighted to help distinguish the patterns seen in those cases versus the others. In particular, the sub-regions that use interruptible load for firm capacity have greater difference in deployment between the 1h daily and 8h daily resources, and they have more uniform deployment across all of the BAs in the sub-regions. On the other hand, most of the other instances show just one or two BAs deploying interruptible load per sub-region, and less sensitivity to the daily energy constraint, which is consistent with spinning reserves being an important component of the resources' value.

In comparing the results from the two sets of scenarios, Base and High RE, we again see how the additional capacity in the system under High RE conditions generally lessens the value that interruptible load can provide. On the other hand, there are still times and locations for which \$81.68/kW-yr is a price worth paying for the capacity, energy, and/or spinning reserves interruptible load is able to provide. This finding is generally consistent with the notion that interruptible load's main competitor is CT capacity, which in our model has an annualized cost of about \$90/kW-yr. CAMX in the High RE scenarios and the years 2025 and 2030 also provides an interesting case study in that the highly various results by BA (despite the fact that the sub-region-wide planning reserve constraint is binding), point to multiple value streams being important to this region as it achieves high levels of wind and solar generation.

Table 12. Price Points Achieved in the Base 150h Scenario

Price points are depicted as number of BAs in each NERC sub-region deploying resource at the given cost. Sub-region years with binding planning reserves constraints are highlighted.

Annual Cost per MW-yr		\$22,000		\$28,600		\$37,180		\$48,334		\$62,834		\$81,684		\$106,189	
Year	Sub-Region	1h	8h	1h	8h	1h	8h	1h	8h	1h	8h	1h	8h	1h	8h
2015	CAMX (of 9)	-	1	-	-	-	-	-	-	-	-	-	-	-	-
	NWPP-Can. (of 2)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	NWPP-US (of 16)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	RMPP (of 3)	1	1	1	1	1	1	-	-	-	-	-	-	-	-
	SRSG (of 6)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2020	CAMX (of 9)	1	9	-	9	-	9	-	3	-	-	-	-	-	-
	NWPP-Can. (of 2)	2	2	-	2	-	2	-	2	-	2	-	1	-	-
	NWPP-US (of 16)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	RMPP (of 3)	1	1	1	1	1	1	1	1	1	1	-	-	-	-
	SRSG (of 6)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2025	CAMX (of 9)	9	9	1	9	1	9	1	9	-	1	-	1	-	-
	NWPP-Can. (of 2)	2	2	1	2	-	2	-	2	-	2	-	1	-	-
	NWPP-US (of 16)	2	2	1	1	1	1	-	-	-	-	-	-	-	-
	RMPP (of 3)	1	1	1	1	1	1	1	1	1	1	1	1	-	-
	SRSG (of 6)	2	6	2	6	2	6	2	2	-	2	-	-	-	-
2030	CAMX (of 9)	9	9	9	9	-	9	-	9	-	9	-	9	-	-
	NWPP-Can. (of 2)	2	2	1	2	-	2	-	2	-	2	-	-	-	-
	NWPP-US (of 16)	2	2	1	2	-	-	-	-	-	-	-	-	-	-
	RMPP (of 3)	1	1	1	1	1	1	1	1	1	1	1	1	-	-
	SRSG (of 6)	6	6	6	6	-	6	-	6	-	6	-	6	-	-

Table 13. Price Points Achieved in the High RE 150h Scenario

Price points are depicted as number of BAs in each NERC sub-region deploying resource at the given cost. Sub-region years with binding planning reserves constraints are highlighted.

Annual Cost per MW-yr		\$22,000		\$28,600		\$37,180		\$48,334		\$62,834		\$81,684		\$106,189	
Year	Sub-Region	1h	8h	1h	8h	1h	8h	1h	8h	1h	8h	1h	8h	1h	8h
2015	CAMX (of 9)	-	1	-	-	-	-	-	-	-	-	-	-	-	-
	NWPP-Can. (of 2)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	NWPP-US (of 16)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	RMPP (of 3)	1	1	1	1	1	1	1	1	-	-	-	-	-	-
	SRSG (of 6)	2	2	-	-	-	-	-	-	-	-	-	-	-	-
2020	CAMX (of 9)	1	9	1	9	-	6	-	1	-	-	-	-	-	-
	NWPP-Can. (of 2)	1	2	1	2	-	2	-	2	-	2	-	1	-	-
	NWPP-US (of 16)	1	1	-	-	-	-	-	-	-	-	-	-	-	-
	RMPP (of 3)	1	1	1	1	1	1	1	1	1	1	1	1	-	-
	SRSG (of 6)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2025	CAMX (of 9)	2	8	2	4	1	2	1	1	1	1	-	1	-	-
	NWPP-Can. (of 2)	2	2	1	2	-	2	-	2	-	2	-	1	-	-
	NWPP-US (of 16)	3	3	3	3	2	2	1	1	-	-	-	-	-	-
	RMPP (of 3)	1	1	1	1	1	1	1	1	1	1	1	1	-	-
	SRSG (of 6)	6	6	2	6	2	6	2	2	-	2	-	-	-	-
2030	CAMX (of 9)	8	9	6	9	4	9	4	9	-	8	-	4	-	-
	NWPP-Can. (of 2)	2	2	1	2	-	2	-	2	-	2	-	1	-	-
	NWPP-US (of 16)	3	3	3	3	3	3	2	2	2	2	-	-	-	-
	RMPP (of 3)	1	1	1	1	1	1	1	1	1	1	1	1	-	-
	SRSG (of 6)	6	6	6	6	2	6	-	6	-	6	-	2	-	-

5.2 Utility-Scale Storage

Utility-scale storage technologies provide a wide range of services to the system, including capacity, energy shifting, reserves provision, and curtailment reduction. Different types of storage technologies are able to meet these needs in different ways, and in particular, different technologies become more valuable at varying price points for different regions. Note however, that the high cost of storage leads to no storage being installed at full price in either scenario and only small amounts installed when the costs drop to 50% of the current costs. However, as costs drop further, an increasing amount of storage is installed across the Western Interconnection. The following sections describe in detail what the price points and motivations are for installing storage in various regions under Base and High RE assumptions.

5.2.1 General Trends in Storage Deployment

To examine general trends in how the model deploys storage, we examine the end conditions in 2030 for the Base scenario and identify the price points at which various technologies are cost-effective. We then compare these results to what is seen in the High RE case. By examining the year 2030, we get an overall picture of cost effectiveness as a function of renewables deployment and we can observe changes in the deployment of other technologies enabled by the presence of these flexible resources. We then examine one price point in detail to identify the model years in which storage technologies are deployed.

In the Base scenarios, no storage is deployed at current costs and only a very small amount of storage is deployed at 50% of current (capital) costs.⁴⁰ However, at 10% of current costs (or \$100/kWh, whichever is greater, see Table 6), storage begins to be installed in large quantities. Figure 18 shows the capacity differences in 2030 for each of the price scenarios examined relative to the Base scenario in which no storage builds are allowed. We can see that in general, once it is deployed, storage replaces gas CT and solar generators, and it enables increased deployments of wind. Note that 10% of current costs is the point at which the firm capacity cost of some storage technologies, on a \$/kW-yr basis and accounting for the capacity value calculated for planning purposes, falls below that of CTs, which is \$90/kW-yr in 2020 and \$87/kW-yr in 2030. Additionally we found that in many cases wind is more constrained by lack of reserves than solar, such that increased storage, which can provide low-cost reserves, enables increased wind deployment.⁴¹

⁴⁰ Per Section 4.2, only capital costs are reduced; fixed and variable O&M costs are kept constant. In the remainder of this section, we will simply refer to “cost reductions”, but these should be taken to only apply to capital.

⁴¹ Traditional generators are required to be generating when providing reserves. Wind more than solar has high generation patterns when most generators are offline, i.e. at night. Therefore the relative cost of reserves is higher for wind than solar.

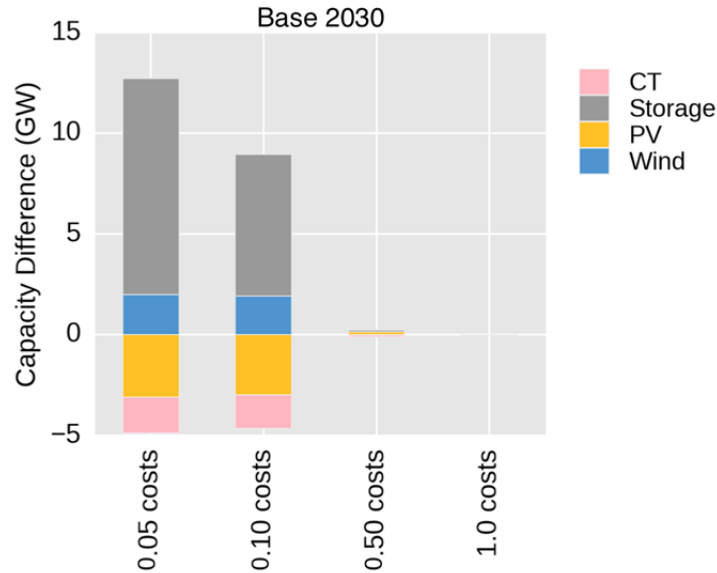


Figure 18. Differences in installed capacity in 2030 for the four Base-Storage Cost scenarios relative to the Base scenario

This analysis indicates that for most applications the price point for economic deployment of storage lies between a 50% and 90% cost reduction from current costs. Further, cost reductions of up to 95% enable even more deployment. While for many of the technologies we examined such a cost reduction would be unrealistic, this may be feasible in the future for some, particularly lithium ion batteries.

Further examination of the results at a finer resolution show additional trends based on the needs of individual NERC sub-regions. We focus in particular on four of the five sub-regions: CAMX, NWPP-US, RMPP, and SRSG. Figure 19 shows the capacity differences in 2030 for each of these four NERC sub-regions. NWPP-Canada was not allowed to build storage in these scenarios,⁴² so results for that region do not vary between scenarios and are not shown.

⁴² This is due to anomalies in results when NWPP-Canada is allowed to build storage, primarily based on its general need for capacity, the fact that our scenario framework lets storage costs get below those of gas CTs on a capacity value basis, and the further factor of reduced resolution wind and solar modeling in Canada. In fact, AESO and BCTC are modeled as having no solar resource.

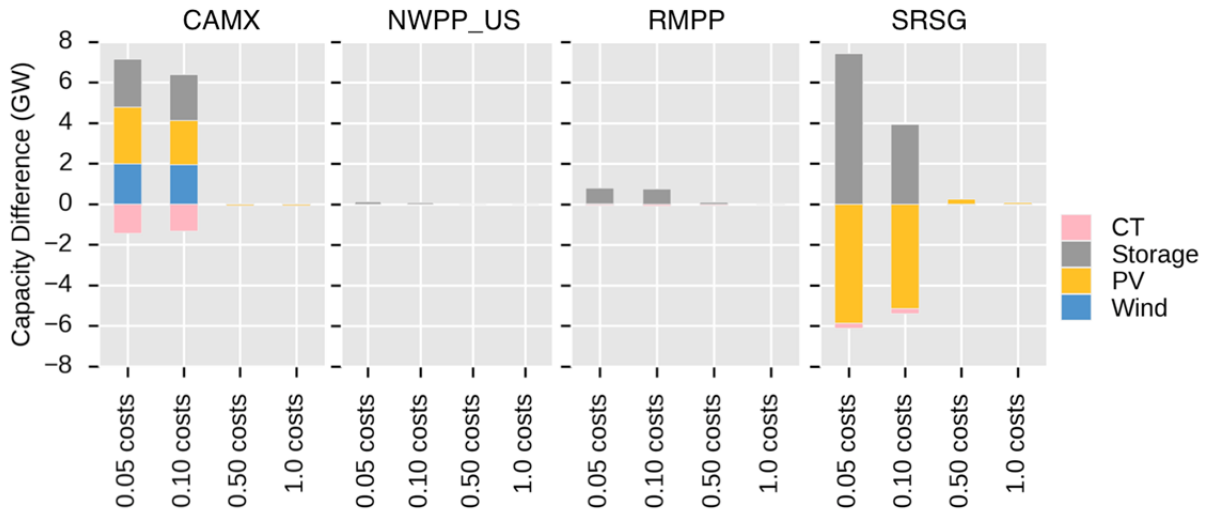


Figure 19. Difference in capacity installed for each of the Base-Storage Cost scenarios compared to the Base scenario by NERC sub-region in 2030

The different NERC sub-regions as modeled⁴³ demonstrate a variety of binding constraints, which lead to different patterns of installed storage technologies. In particular, the various motivations to install storage technologies range from a need for reserves (as seen in RMPP), a need for capacity (in SRSR), and a need for enabling increased renewables (as seen in CAMX).

Regions that show a need for reserves (e.g., RMPP) may be able to tolerate higher storage costs. As noted in Section 5.1.1, RMPP, and in particular PACE Wyoming, which comprises the Wyoming operating reserves sharing group, have a strong need for reserves in the low and mid dispatch periods in our model under the Base assumptions due to a lack of flexible generation. With regard to storage, this BA installs storage at 50% of capital cost and is the only BA to do so under Base assumptions. Additionally, this region only installs short-term storage to provide these reserves, as can be seen in Figure 20.

The need for capacity can also lead to storage installations. In particular, SRSR shows a need for capacity in 2025 and 2030 in our model. In the Base scenario, a combination of PV and CT's were installed to provide significant firm capacity for the sub-region's planning requirements, however as the cost of storage drops, its cost of peak-capacity becomes lower than either PV or gas CT's. The Base-0p10 and Base-0p05 scenarios both install storage technologies with equivalent capacity-value as the displaced technologies. The 2h storage technology built in the Base-0p10 has a similar capacity value to PV, and in the Base-0p05 scenario the cost of Li-Ion batteries have dropped enough such that it can provide capacity at a lower cost. These storage technologies can therefore effectively compete to provide the required capacity when technology costs are sufficiently reduced.

⁴³ The caveats expressed at the beginning of Section 5.1 apply here as well. In particular, we present NERC sub-region and a limited number of BA-level results because they are interesting and demonstrate a diversity of deployment drivers. They are not necessarily accurate representations of actual capacity (or reserves) needs or excesses.

Finally, storage may be installed to increase variable generation penetration. An increase in VG penetration requires both increased reserves and energy shifting to match generation with load. In particular, in the Base-0p10 scenario, CAMX typifies this use pattern and installs storage with one and two hours of energy capacity, which is able to provide reserves for the increase in installed wind and also helps shift this new energy to times of higher demand.

Figure 20 shows the breakdown by region of which types of storage are installed. As noted above, the NERC regions have different needs from storage and therefore install storage types with different energy capacities. It is of interest that only three of the twelve storage technologies available in the model were installed, and all are largely differentiated from the technologies not installed by their low cost. Additionally, due to the \$100/MWh lower cost bound, the Base-0p05 case only has a significant cost reduction for technologies with 1 hour or less of storage. The much cheaper cost of these technologies overshadowed the benefits of longer energy capacities, reiterating that cost is a very important driver in storage installations.

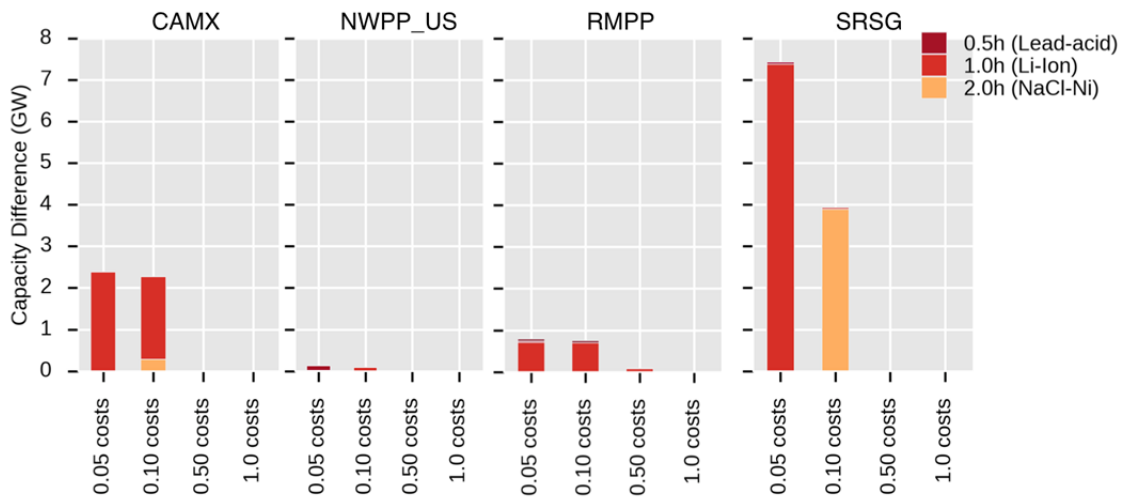


Figure 20. Storage capacity in 2030 by technology sub-class and NERC sub-region for the Base-Storage Cost scenarios

High Renewables Scenarios

The scenarios with more renewables (High RE) deploy and use storage in similar ways as the Base scenarios, however more storage is installed, and also is installed at higher price points in some locations. Figure 21 shows the changes in capacity installed by 2030 for the Western Interconnection in the High RE-Storage Cost scenarios as compared to the High RE scenario with no new storage allowed. Again storage generally replaces gas CT units and solar PV while slightly increasing wind; and the impacts on renewables vary by region. As will be discussed below in detail, the initial installations of storage are short-term storage used to provide reserves, predominantly flexibility reserves for the higher renewables present in these scenarios compared to the Base assumptions. Later these storage products are used for energy shifting.

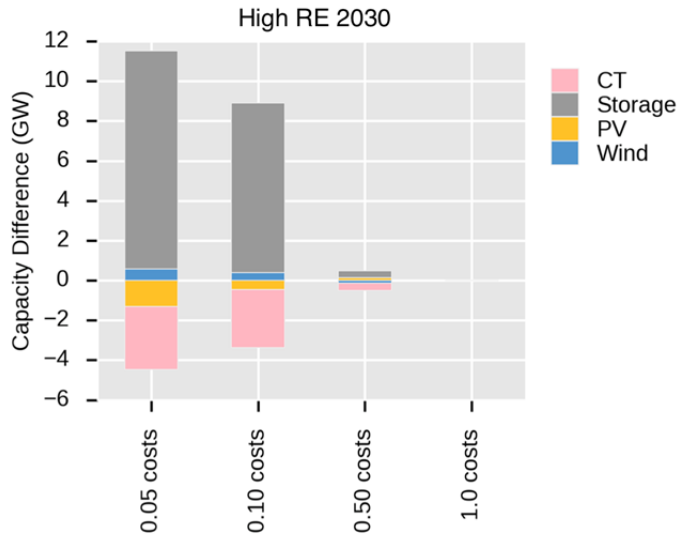


Figure 21. Differences in installed capacity in 2030 for the four High RE-Storage Cost scenarios relative to the High RE scenario

Again we examine the NERC sub-regions in detail to provide a description of the diverse motivations for model deployment of storage. Figure 22 shows the changes in installed capacity for each of four NERC sub-regions.

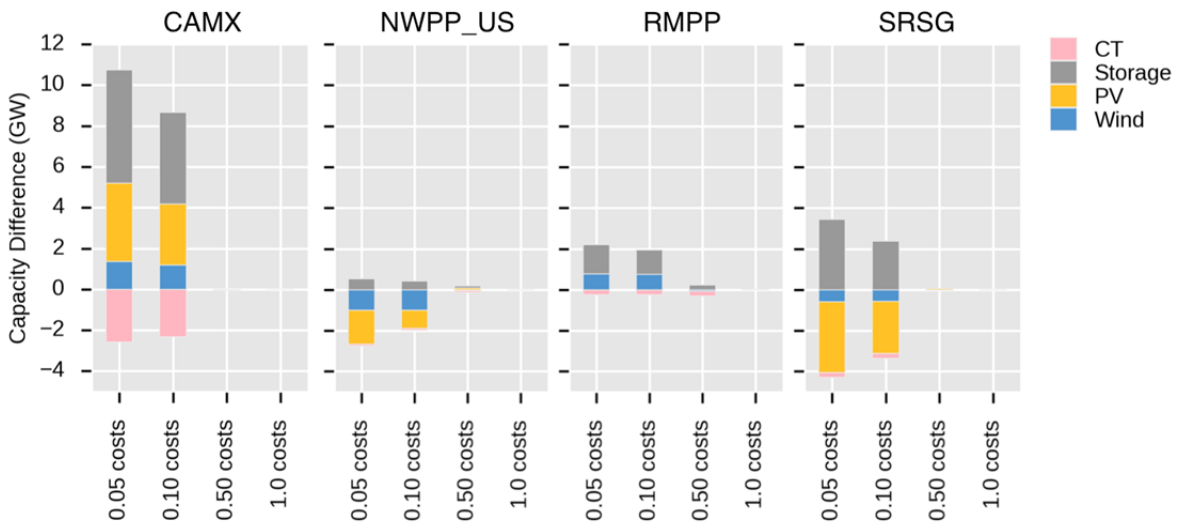


Figure 22. Difference in capacity installed compared to the high renewables scenario for each of the storage scenarios and NERC sub-region in 2030

Again in the High RE case RMPP installs storage technologies initially for reserves. In particular as storage costs drop below 10% of current costs, additional storage installations enable increased wind installations due to the capability of storage to provide more flexibility reserves. RMPP installs a significant amount of wind energy in the high renewables case, reaching 52% annual penetration of wind on a generation basis. Longer-term storage could enable additional

energy shifting of these wind resources to peak time periods, however these storage technologies remain too expensive for the model to build.

In the High RE case, NWPP-US installs storage to replace gas CTs when costs are 50% of current costs. These installations are largely used to provide additional flexibility reserves that are required with increased renewable installations. In this 0.50 cost scenario there is an increase in solar. This further indicates that storage used for reserves can often justify a higher price point, which makes sense with regard to value stacking of storage. In the Base scenarios, NWPP-US did not install appreciable amounts of storage; however, at higher renewable energy penetrations and low storage costs, this becomes economical.

As costs drop further, storage and renewables deployments both increase, predominantly in CAMX, which has a high renewables penetration due to the 50% RPS requirement in California. In this region, as in the base scenario, storage is used to further enable renewable growth.

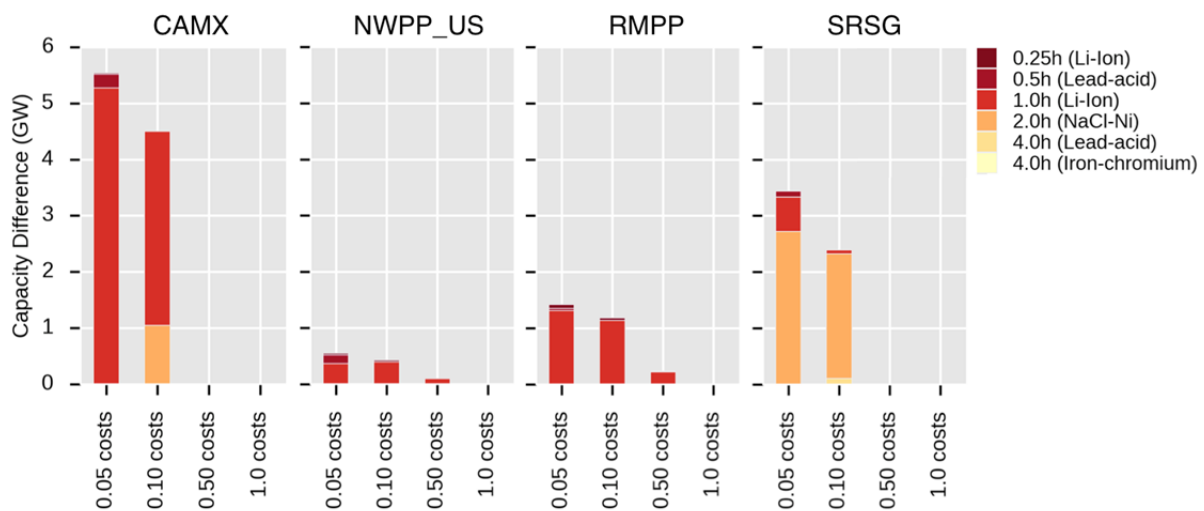


Figure 23. Storage capacity in 2030 by technology sub-class and NERC sub-region for the High RE-Storage Cost scenarios

Cost by Year

In addition to trends in installation by price point, the timing of deployments varies based on the underlying system conditions (Base versus High RE). To see these trends, we focus on the 0p10 cost scenarios, as 10% of current costs are the highest storage costs at which significant model deployment is observed.

Figure 24 depicts deployment through time for the Base-0p10 scenario and the High RE-0p10 scenario. The High RE case installs more storage than the Base case, but has less of an impact on renewables than the Base case. A large part of this is likely the already high renewable penetrations in the High RE scenario. Both scenarios see a large increase in deployments in 2030, again when the costs of variable generation have dropped further, and there is more need for capacity due to load growth. In all years it is installed, storage replaces gas CT units.

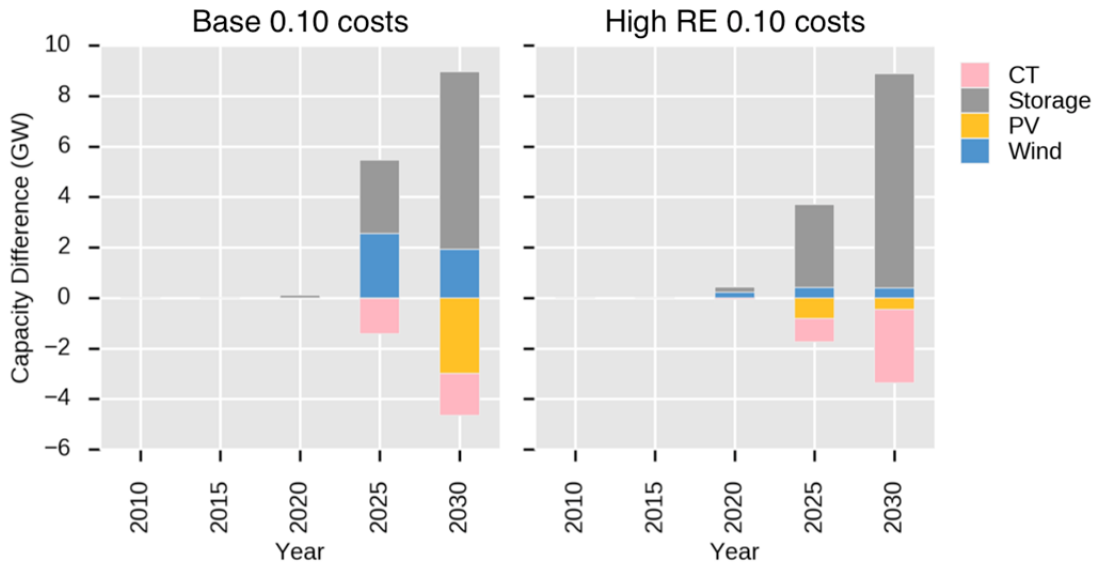


Figure 24. Yearly capacity differences at a system level for the 0p10 scenario for both the Base and High RE assumptions

Figure 25 shows the storage classes installed by year for both sets of underlying assumptions. Interestingly, 2hr storage is installed to a higher degree in the Base case, and 1hr storage is more important under the High RE conditions. The High RE scenarios however also install a small amount of 4hr storage.



Figure 25. Cumulative system-level storage capacity by technology sub-class for the Base-0p10 and High RE-0p10 scenarios

5.2.2 Description of Storage Operations

The value of storage to a system is in large part based on its ability to provide reserves, to shift energy to times of higher (net) demand, or reduce curtailment. In this section, we analyze how

the storage resources deployed by the model are used in the model dispatch, and how this corresponds to the value of storage to the system.

Figure 26 shows the changes in dispatch for the entire interconnection on an annual basis, both in the year 2030 for all of the Base-Storage Cost scenarios, and for all years for the Base-0p10 cost scenario, with all relative to the Base scenario dispatch. These results show that the use of storage for energy arbitrage is similar across scenarios with significant storage deployment and is largest in 2030, with accompanying significant curtailment reductions.⁴⁴ In general, storage enables the increased use of low-variable cost generation, including renewable energy and coal, at the expense of higher-variable cost technologies such as gas combined cycle (CC) and gas steam units, both through actual energy arbitrage and as a consequence of the capacity changes analyzed above. Energy arbitrage does appear to be more important in 2030, when more renewables are in the system than in previous years. Significant curtailment is also mitigated in this year. One factor may be storage units that are built in 2025 to provide reserves are in 2030 already in the system and available to respond to the increased need for energy arbitrage and curtailment reduction implied by VG deployment.



Figure 26. Differences in annual generation for the Base scenarios.

All Base-Storage Cost scenarios versus the Base scenario in 2030 (left) and differences in Base-0p10 and Base annual generation by year (right)

It is of interest to note that the use of storage does not result in greater greenhouse gas (GHG) emissions in these scenarios, despite increased coal generation, due to the significantly greater corresponding reduction in natural gas generation. The Base-0p10 and Base-0p05 cost scenarios show a very slight reduction in GHG emissions and the other scenarios are essentially equivalent to the base scenario.

⁴⁴ This may be partially due to a modeling artifact, in that new storage is unable (in our model) to reduce curtailments from new renewable builds, as that would introduce a nonlinearity (and RPM is a mixed-integer program, that is, a MIP).

The use of storage to provide reserves is an important aspect of the value of storage. Figure 27 shows the changing makeup of the generator types providing all reserve products for the entire interconnection for all Base-Storage Cost scenarios in the year 2030 (left), and for the Base-0p10 cost scenario for all years (right). Beginning in 2025, storage provides a noticeable amount of the reserves for the system, reaching 8% of reserves provision by 2030 in the Base-0p05 cost scenario and 6% in the Base-0p10 cost scenario. Interestingly, the fraction of reserves provided by storage drops in 2030 for the Base-0p10 cost scenario; this is due to a greater provision of energy in that year compared to previous years.

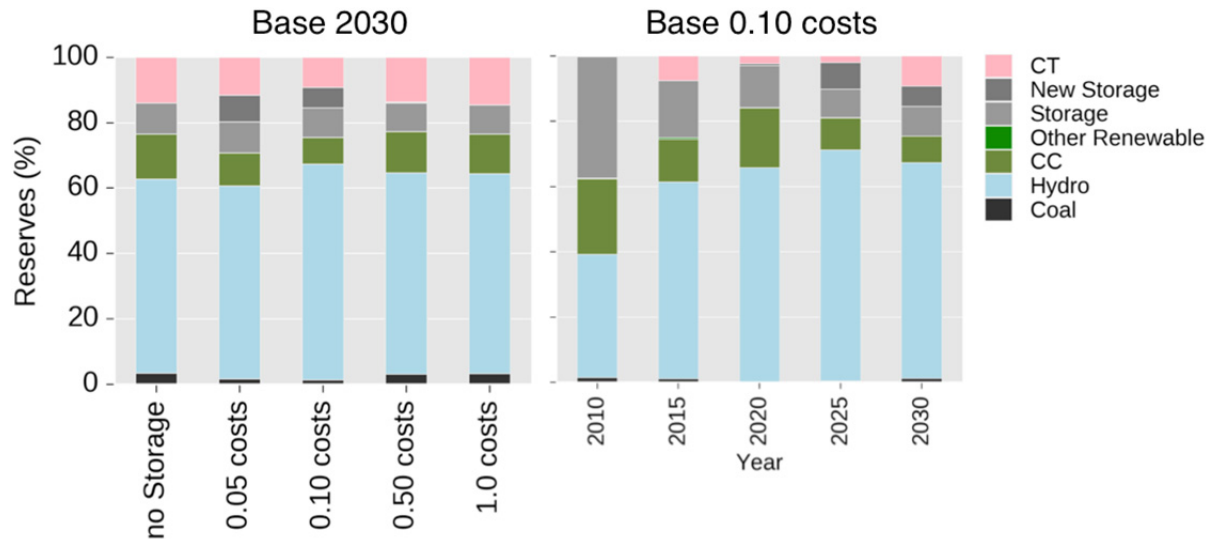


Figure 27. Flexibility reserves provision by generator type, as a fraction of all flexibility reserves provided, for the entire Western Interconnection

Note – Here “New Storage” refers to storage built according to the scenarios described above; “Storage” refers to pumped hydro (PHES) that existed as of 2010 or is added to the model as a prescribed build.

Figure 28 shows the changes in reserves provision in 2030 for each of the Base-Storage Cost scenarios and specifically for the reserve-sharing group of Wyoming, the most reserves-constrained region under Base assumptions due to its high percentage of coal generation. In this region low-cost storage displaces gas CT builds and enables coal to provide energy instead of reserves. Additionally, storage provided reserves enable an increase in renewable generation. Flexibility reserves in particular, on the left in Figure 28, are required to be held based on a percentage of the generation from variable generation sources, and this provision of reserves is a binding restraint in the Base scenario. In particular, in the Base-0p50 scenario the increase in net reserves provided exactly equals the increase from storage. At lower storage costs, the net amount of reserves required to be held increases over that needed for the Base scenario, indicating that higher renewables penetrations are being supported by storage.

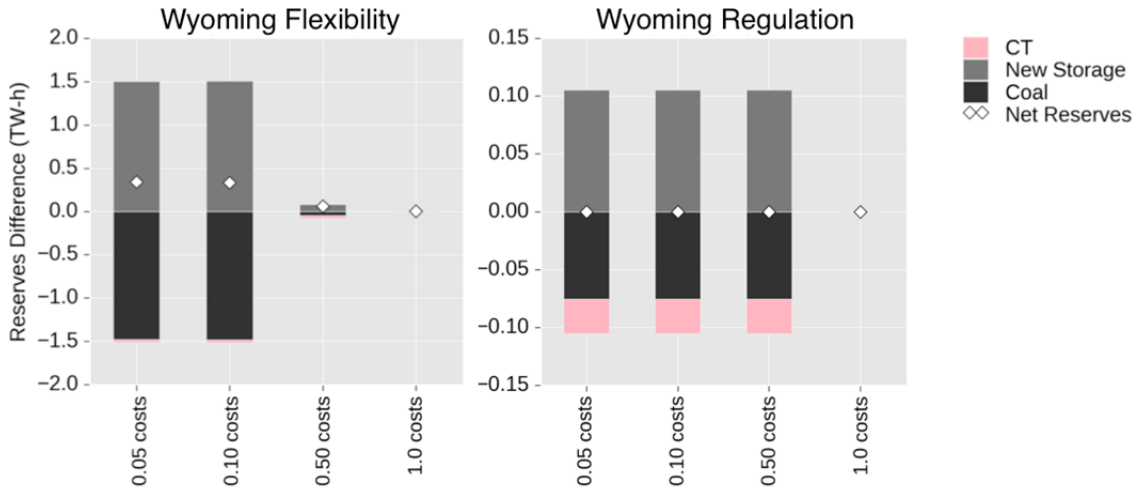


Figure 28. Differences in Flexibility and Regulation reserves for the Wyoming reserve sharing group in 2030

High Renewables Scenarios

The High RE scenarios show similar patterns as the Base scenarios, with storage enabling increased use of low variable cost generation, including wind and coal. Additionally, in 2020 there is an increase in wind generation without a large increase in storage-based energy arbitrage. This supports the idea that in the High RE scenarios, storage is initially built to provide additional capacity and reserves and only later is used to provide energy shifting, which it does in 2030.

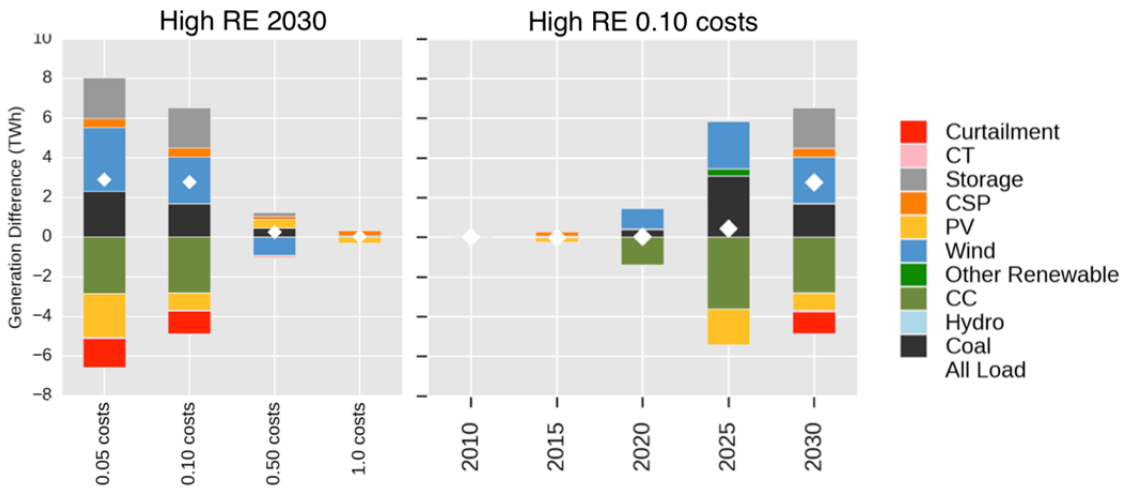


Figure 29. Differences in annual generation for the High RE scenarios, showing all High RE-Storage Cost scenarios versus the High RE scenario in 2030 (left) and differences in High RE 0.10 costs and High RE annual generation by year (right)

In the high renewables scenarios, the use of storage does lead to an increase in GHG emissions in the 0p05 and 0p10 cost scenarios. This is on the order of 1 million metric tons increase out of a total 244 million metric tons. The High RE-0p50 scenario shows a slight decrease in GHG emissions.

Fractional reserves provision for all reserves under High RE assumptions is shown in Figure 30. Due to low installed capacity, storage provides only 1% of all reserves in the High RE-0p50 cost scenario. However, in the lower cost scenarios storage provides a significant amount of reserves, reaching over 20% by 2030 in both the High RE-0p10 and High RE-0p05 cost scenarios. This provides two services, predominantly a lower need for gas CT units, which are built in several regions solely for reserves, and also reduces the need for gas CC generation.⁴⁵

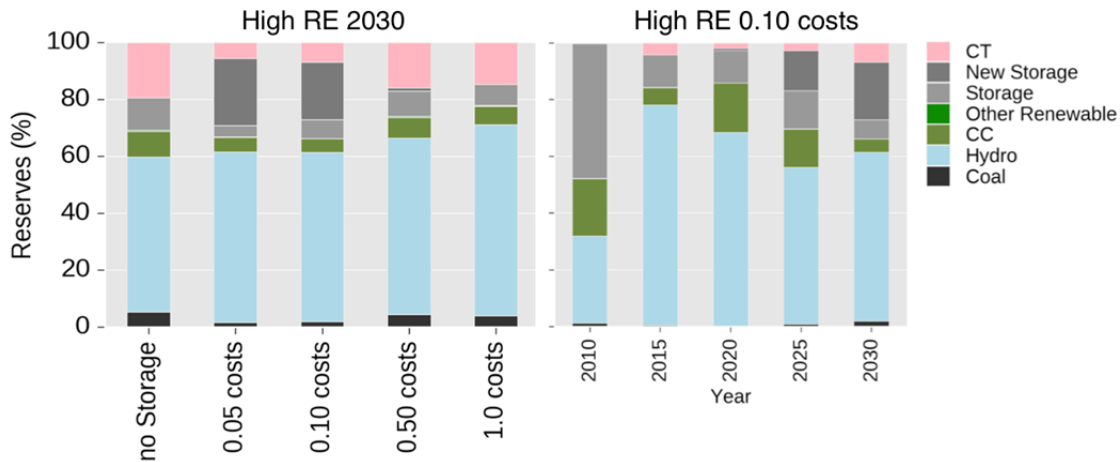


Figure 30. Reserves provision by generator type, as a fraction of all reserves provided, for the entire Western Interconnection, showing all High RE-Storage Cost scenarios in 2030 (left) and the High RE-0p10 scenario annual provision by year (right)

Note – Here “New Storage” refers to the storage built by the model according to the scenarios listed above; “Storage” refers to existing storage, consisting of PHES.

In the High RE scenarios, additional reserve sharing groups show a need for flexibility reserves with the higher penetrations seen in this scenario. In particular, the reserve sharing groups in southern California and Colorado show an increase in flexibility reserves provision, shown in Figure 31 for 2030. These groups in particular use storage to displace other generators providing reserves and show a slight increase in the total reserves provided. These slight increases enable increased renewables in both these regions, in particular wind in Colorado and solar in southern California.

⁴⁵ RPM requires generators to be operating while providing reserves, so in some cases generators will operate largely in order to provide reserves.

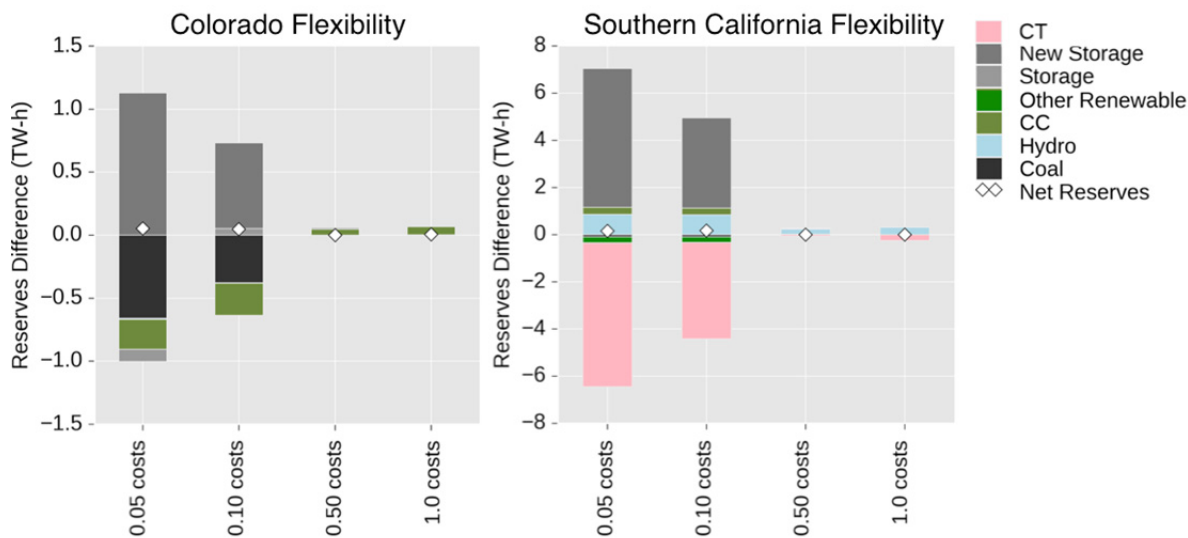


Figure 31. Differences in Flexibility reserves for the Colorado and southern California reserve sharing groups in 2030, under High RE assumptions

Storage also provides value to the system by reducing the curtailment of variable generators, including wind and solar power. The use of storage reduced or left unchanged the total amount of curtailment in the system in the year 2030 (see Table 14). While it does not appear that curtailment reduction is a dominant reason for storage deployment in RPM, this use-case certainly increases the value of storage, adding to the firm capacity and reserve provision values, and enhancing basic energy arbitrage.

Table 14. Variable Generation Curtailment in 2030

Scenario	Annual Curtailment, TWh	
	Base	High RE
-	3.0	4.9
1p00	3.1	4.9
0p50	3.0	4.0
0p10	1.1	3.8
0p05	0.9	3.5

Storage Dispatch

The dispatch of storage technologies provides additional insight into how these technologies are used in RPM. This section presents dispatch results from two BAs to illustrate how storage dispatch can support renewables: SCE in the Base-0p10 scenario and NEVP from the High RE-0p10 scenario.

Figure 32 shows the dispatch of 2h storage in SCE, which is installed in the Base-0p10 scenario starting in 2025. (SCE also has existing pumped hydro capacity, whose dispatch is not shown.) In the first year of service, the 2h storage is only used for reserves to aid further wind

installations; however, by 2030 it is being used to shift energy from periods of low net-load, mostly in the middle of the day during the low dispatch period to peak hours and non-daylight hours, particularly as wind is ramping down. Per the full dispatch picture for this region (Figure 33), this storage dispatch pattern adds flexibility to the system while wind and solar output are changing rapidly. The system chooses to use this stored energy during the non-daylight hours of the low dispatch period, and it keeps enough energy in storage to be available on the peak day. Note that the storage in this example is difficult to see, as it is only approximately 1% of the generation in this region.

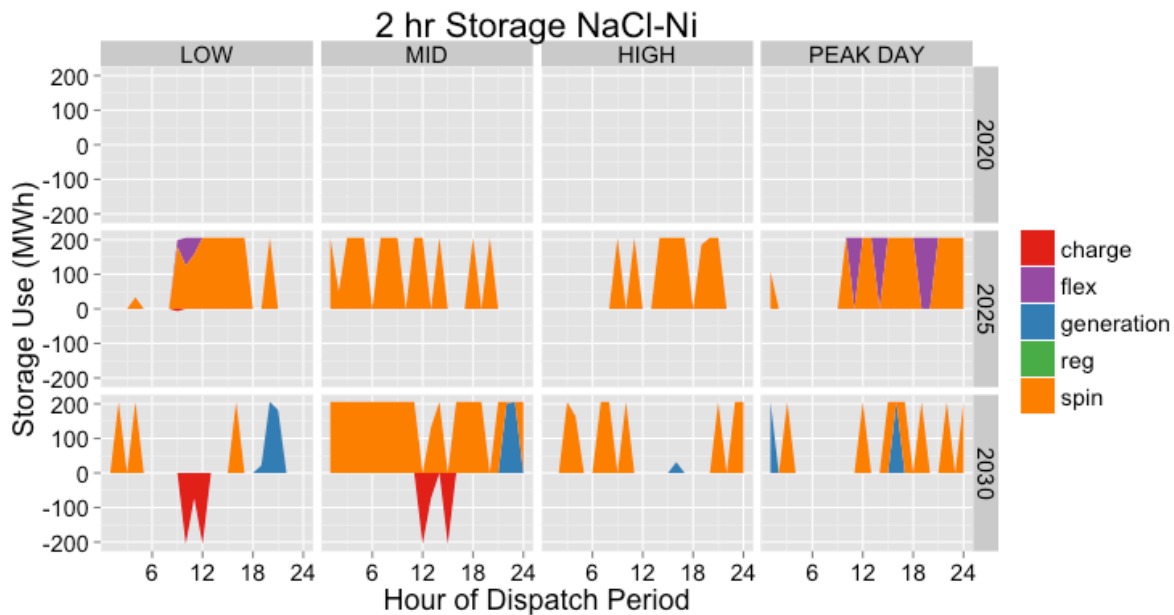


Figure 32. Dispatch of the 2h NaCl-Ni energy storage technology in SCE in the Base-0p10 scenario

Due to the discontinuous nature of the dispatch periods, storage capacities are required to balance annually, but not necessarily within a specific dispatch period. Therefore, storage may be charged during the low dispatch period, and discharged during the high dispatch period as is seen here. For more information, see Section 3.2.

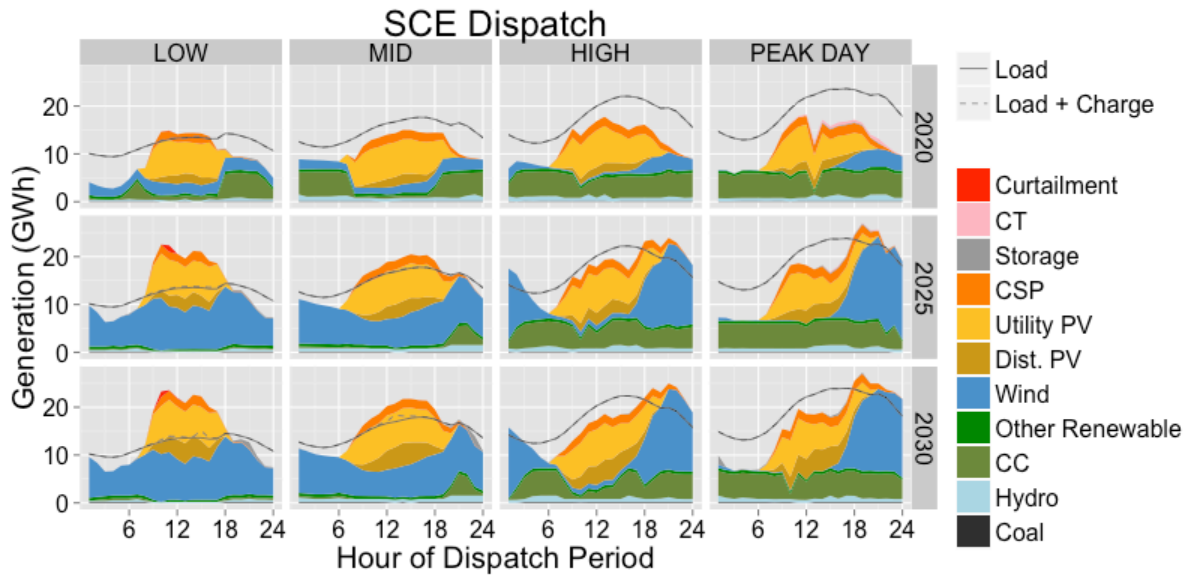


Figure 33. System dispatch for SCE in the Base-0p10 scenario

Note: The generation and load in a particular BA are not required to exactly balance, since BA's can import and export from adjacent regions if needed. In this example, SCE is a net importer in 2020 and a net exporter in 2030.

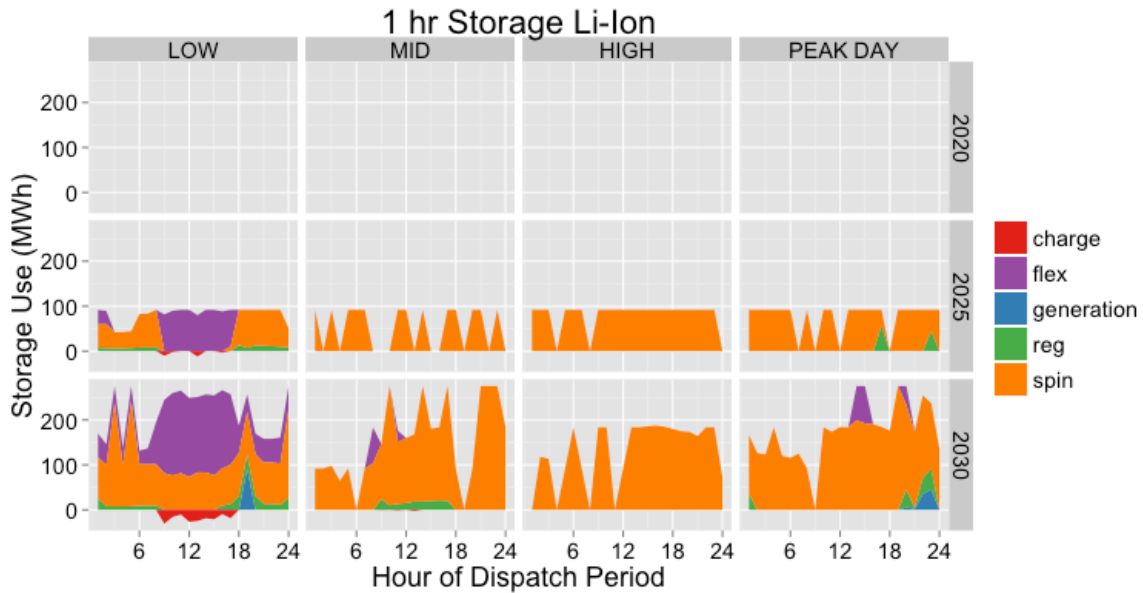


Figure 34. Dispatch of the 1h Li-ion energy storage technology in NEVP in the High RE-0p10 scenario

As noted above, the High RE scenarios have a greater need for reserves, particularly flexibility reserves in support of the increased renewable penetration. Figure 34 shows an example of the resulting reserves-heavy storage dispatch in NEVP, for its 1h energy storage resource in the High RE-0p10 scenario. This storage technology is not installed until 2025, at which time it is used exclusively for reserves, particularly flexibility reserves during high solar hours. Coincident with this increase in storage is an increase in solar deployment (Figure 35). Despite the large increase

in solar in 2025, the storage is not used for energy shifting until 2030⁴⁶ when the storage facility shifts energy in the low dispatch period from the mid-day hours to the evening solar ramp-down time, and it also holds some storage in reserve for the peak day.

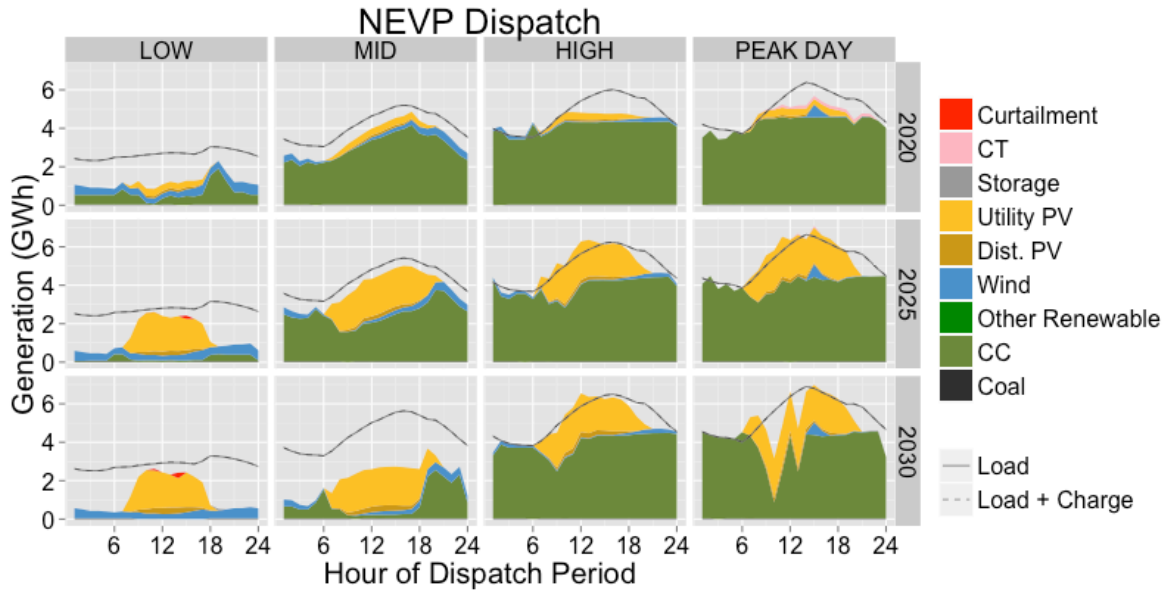


Figure 35. System dispatch for NEVP in the High RE-0p10 scenario

⁴⁶ This lack of energy shifting in the first year may in part be a consequence of energy arbitrage not being lucrative enough to justify a capacity purchase. However, once that capacity has been installed it is operationally advantageous to use the storage for shifting energy.

6 Discussion

Our analysis demonstrated methods for extending capacity expansion models to (1) better represent power system needs for flexibility and (2) ensure that the value provided by technologies designed to provide flexibility—namely storage and interruptible load—is factored into investment decisions. This was done in the context of NREL’s Resource Planning Model (RPM), which from its conception has been designed to accurately model investment decisions for variable wind and solar generation.

Aspects added to the model for this work include an extension of the dynamic capacity value methodology described in Mai et al. (2015) to energy-constrained technologies, in particular interruptible load and utility-scale storage. Methods were also developed to better estimate curtailment from variable generation, both with and without storage. The development of the curtailment estimation methods, which are based on net load duration curves that are similar to our capacity value methodology, also relies on an estimate of the system-wide minimum generation level using regression models of production cost simulation results. The capacity value and curtailment methods were used to calculate tail effects for both the existing fleet and the next (marginal) increments of new variable generation, interruptible load, and storage capacity. Used in this way, they infuse the investment decision with additional flexibility information, including both the need for it and the ability of flexible technologies to provide it; these are important aspects to model when studying the possible evolution of power systems in high renewable energy futures.

The interruptible load and utility-scale storage models both required that additional constraints be added to the RPM dispatch model, particularly because both types of options are energy-constrained. Interruptible load is constrained in how much energy it can provide daily and annually, which conceptually represents capacity-focused demand response programs. Our storage model enforces a technology subclass-specific energy capacity, expressed in MWh/MW, on both the hourly and the seasonal timescales. Overall, charging and discharging is required to balance on an annual basis subject to subclass-specific roundtrip efficiencies.

These models, in concert with the capacity value and minimum curtailment estimation methods also added for this work, were exercised in two sets of scenarios. One set of scenarios allowed new builds of interruptible load but not of storage; the other was formulated oppositely. Because this work is primarily methodological, and the cost trajectories (and in the case of interruptible load, the resource availability) of these technologies are highly uncertain, our scenarios were designed not to assess probable deployments but rather to explore possible deployment patterns in the Western Interconnection as a function of technology cost. That is, at what cost and for what reasons might flexible technologies be deployed in the western United States in the 2020-to-2030 timeframe?

Overall, we found capacity and reserves shortfalls to be major potential drivers of deployment for both flexible technologies. Exactly where these shortfalls occur is highly sensitive to uncertain model inputs, so our results should not be interpreted as predictions of the potentially most valuable geographic areas for flexible technologies. Rather, they should be interpreted as descriptions of a diversity of possible conditions under which any system might start to find flexible technologies more attractive. To this end, in regions in need of capacity, interruptible

load at an annual cost approaching the annualized capital costs of natural gas combustion turbines (approximately \$83/kW-yr) was often deployed. This cost level was also supportable in select locations in need of reserves.

We also found firm capacity needs to be a key driver of new storage deployment. As such, significant cost reductions (90% below estimated 2013 costs) were found to be required for storage capital costs to dip below those of natural gas combustion turbines and result in significant storage deployment (about 8 GW). However, since storage technologies can provide multiple services—including energy arbitrage and avoided curtailments—to the system and incurs lower variable costs compared with combustion turbines (and other conventional generators), some new storage capacity was found to be competitive in scenarios with higher costs (i.e., 50% of 2013 estimates).

In general, this study found that it is possible to construct computationally tractable capacity expansion models that capture many system flexibility needs and the potential for storage and interruptible load to fulfill those needs. The analysis does not, however, consider the need for sub-hourly services or the associated value opportunities for flexibility options that could be used during these shorter timescales. In addition, while the methods use 8,760 hourly data, they rely on data from a single year (2006). A more robust assessment could include multiple years of data. Moreover, new energy policies that might be enacted in the future would change scenario results and alter the quantitative findings reported above. Finally, like many other similar models, RPM has certain limitations, including its system-wide approach that may not accurately reflect current market or institutional flexibility (or inflexibility) or the costs and value of flexibility from the perspective of developers, utilities, system operators, regulators, or others power sector participants. Nonetheless, our research demonstrates a new capability to begin to incorporate considerations of flexibility in large-scale capacity expansion models and provides initial estimates on the deployment potential for interruptible load and storage in the western United States.

References

Black, Mary, and Goran Strbac. 2006. "Value of Storage in Providing Balancing Services for Electricity Generation Systems with High Wind Penetration." *Journal of Power Sources* 162 (2): 949–53.

Blair, Nate, Karlynn Cory, Maureen Hand, Linda Parkhill, Bethany Speer, Tyler Stehly, David Feldman, et al. 2015. *Annual Technology Baseline (Including Supporting Data)*. NREL (National Renewable Energy Laboratory).

Bloom, Aaron, Aaron Townsend, David Palchak, Josh Novacheck, Jack King, Matthew O’Connell, Eduardo Ibanez, et al. Forthcoming. "Eastern Renewable Generation Integration Study." Golden, Colorado: National Renewable Energy Laboratory.

Brinkman, G., J. Jorgenson, J. Caldwell, and A. Ehlen. 2015. "Low Carbon Grid Study: Analysis of a 50% Emission Reduction in California." Golden, Colorado: National Renewable Energy Laboratory. NREL/TP-6A20-64884.

California Air Resources Board. 2014. *ARTICLE 5: CALIFORNIA CAP ON GREENHOUSE GAS EMISSIONS AND MARKET-BASED COMPLIANCE MECHANISMS TO ALLOW FOR THE USE OF COMPLIANCE INSTRUMENTS ISSUES BY LINKED JURISDICTIONS*. Title 17, *California Code of Regulations*. Vol. Subchapter 10 Climate Change, Article 5. <http://www.arb.ca.gov/cc/capandtrade/ctlinkqc.pdf>.

California PUC. 2013. "Order Instituting Rulemaking Pursuant to Assembly Bill 2514 to Consider the Adoption of Procurement Targets for Viable and Cost-Effective Energy Storage Systems." Regulatory Decision 13-10-040. Public Utilities Commission of the State of California. <http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M079/K533/79533378.PDF>.

de Leon, Kevin, Das Williams, and Mark Leno. 2015. *SB-350 Clean Energy and Pollution Reduction Act of 2015*. Vol. Chapter 547, Statutes of 2015.

Denholm, Paul, Jennie Jorgenson, Mackay Miller, Ella Zhou, and Caixia Wang. 2015. "Methods for Analyzing the Economic Value of Concentrating Solar Power with Thermal Energy Storage." Golden, Colorado: National Renewable Energy Laboratory. NREL/TP-6A20-64256. <http://www.nrel.gov/docs/fy15osti/64256.pdf>.

Denholm, Paul, Matthew O’Connell, Gregory Brinkman, and Jennie Jorgenson. 2015. "Overgeneration from Solar Energy in California: A Field Guide to the Duck Chart." NREL Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-65023. <http://www.nrel.gov/docs/fy16osti/65023.pdf>.

Denholm, Paul, and Ramteen Sioshansi. 2009. "The Value of Compressed Air Energy Storage with Wind in Transmission-Constrained Electric Power Systems." *Energy Policy* 37 (8): 3149–58.

Denholm, P., J. Jorgenson, M. Hummon, T. Jenkin, D. Palchak, B. Kirby, O. Ma, and M. O’Malley. 2013. "Value of Energy Storage for Grid Applications." Golden, Colorado: National Renewable Energy Laboratory. NREL/TP-6A20-58465. <http://www.nrel.gov/docs/fy13osti/58465.pdf>.

De Sisternes, Fernando J., and D. Webster Mort. 2013. "Optimal Selection of Sample Weeks for Approximating the Net Load in Generation Planning Problems." ESD Working Paper Series ESD-WP-2013-03.

DOE/EPRI. 2013. "Electricity Storage Handbook in Collaboration with NRECA." Albuquerque, New Mexico: Sandia National Laboratories. SAND2013-5131. <http://www.sandia.gov/ess/publications/SAND2013-5131.pdf>.

Drury, Easan, Paul Denholm, and Ramteen Sioshansi. 2011. "The Value of Compressed Air Energy Storage in Energy and Reserve Markets." *Energy* 36 (8): 4959–73.

E3. 2014. "Investigating a Higher Renewables Portfolio Standard in California." Energy and Environmental Economics, Inc. (E3).

EIA (Energy Information Administration). 2014. "Annual Energy Outlook 2014 with Projections to 2040." DOE/EIA-0383(2014). Washington, D.C.: U.S. Energy Information Association.

———. 2015. "Annual Energy Outlook 2015 with Projections to 2040." DOE/EIA-0383(2015). Washington, D.C.: U.S. Energy Information Association.

El Paso Electric Company. 2012. "Integrated Resource Plan of El Paso Electric Company for the Period 2012-2031." El Paso Electric Company.

EnerNex Corporation. 2011. "Eastern Wind Integration and Transmission Study (EWITS) (Revised)." Subcontract Report NREL/SR-5500-47078. Golden, Colorado: National Renewable Energy Laboratory. <http://www.nrel.gov/docs/fy11osti/47078.pdf>.

EPRI (Electric Power Research Institute, Inc.). 2014. "Program on Technology Innovation: US-REGEN Model Documentation 2014." Technical Results 3002004693. Palo Alto, California: Electric Power Research Institute (EPRI). <http://www.epri.com/abstracts/Pages/ProductAbstract.aspx?ProductId=000000003002004693>.

GE Energy. 2010. "Western Wind and Solar Integration Study." Golden, Colorado: National Renewable Energy Laboratory. NREL/SR-550-47434. <http://www.nrel.gov/docs/fy10osti/47434.pdf>.

———. 2014. "Minnesota Renewable Energy Integration and Transmission Study: Final Report." <http://www.minnelectrans.com/documents/MRITS-report.pdf>.

Getman, Dan, Anthony Lopez, Trieu Mai, and Mark Dyson. 2015. "Methodology for Clustering High-Resolution Spatiotemporal Solar Resource Data." NREL/TP-6A20-63148. Golden, CO: National Renewable Energy Laboratory.

Graves, Frank, Thomas Jenkin, and Dean Murphy. 1999. "Opportunities for Electricity Storage in Deregulating Markets." *The Electricity Journal* 12 (8): 46–56.

Hargreaves, J., E.K. Hart, R. Jones, and A. Olson. 2015. "REFLEX: An Adapted Production Simulation Methodology for Flexible Capacity Planning." *IEEE Transactions on Power Systems* 30 (3): 1306–15. doi:10.1109/TPWRS.2014.2351235.

Hummon, Marissa, David Palchak, Paul Denholm, Jennie Jorgenson, Daniel J. Olsen, Sila Kiliccote, Nance Matson, et al. 2013. “Grid Integration of Aggregated Demand Response, Part 2: Modeling Demand Response in a Production Cost Model.” Golden, Colorado: National Renewable Energy Laboratory. NREL/TP - 6A20 - 58492. <http://www.nrel.gov/docs/fy14osti/58492.pdf>.

IEA (International Energy Agency). 2011. “Harnessing Variable Renewables: A Guide to the Balancing Challenge.” Technical Report. Paris, France: International Energy Agency. https://www.iea.org/publications/freepublications/publication/Harnessing_Variable_Renewables_2011.pdf.

Johnston, Josiah, Ana Mileva, James H. Nelson, and Daniel M. Kammen. 2013. “SWITCH-WECC Data, Assumptions, and Model Formulation.” University of California, Berkeley. http://rael.berkeley.edu/old_drupal/sites/default/files/SWITCH-WECC_Documentation_October_2013.pdf.

Jorgenson, J., P. Denholm, M. Mehos, and C. Turchi. 2013. “Estimating the Performance and Economic Value of Multiple Concentrating Solar Power Technologies in a Production Cost Model.” Golden, Colorado: National Renewable Energy Laboratory. NREL/TP-6A20-58645. <http://www.nrel.gov/docs/fy14osti/58645.pdf>.

Lannoye, E., D. Flynn, and M. O’Malley. 2012. “Evaluation of Power System Flexibility.” *IEEE Transactions on Power Systems* 27 (2): 922–31. doi:10.1109/TPWRS.2011.2177280.

Lee, Michael P., Omar Aslam, Ben Foster, David Kathan, and Christopher Young. 2014. “Assessment of Demand Response & Advanced Metering.” Staff Report. Washington, D.C.: Federal Energy and Regulatory Commission (FERC). <http://www.ferc.gov/legal/staff-reports/2014/demand-response.pdf>.

Lew, D., G. Brinkman, E. Ibanez, A. Florita, M. Heaney, B. M. Hodge, M. Hummon, et al. 2013. “Western Wind and Solar Integration Study Phase 2.”

Madaeni, Seyed Hossein, Paul Denholm, and Ramteen Sioshansi. 2012. “Comparison of Capacity Value Methods for Photovoltaics in the Western United States.” Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-54704. <http://www.nrel.gov/docs/fy12osti/54704.pdf>.

Mai, T., C. Barrows, A. Lopez, E. Hale, M. Dyson, and K. Eureka. 2015. “Implications of Model Structure and Detail for Utility Planning: Scenario Case Studies Using the Resource Planning Model.” Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-63972. <http://www.nrel.gov/docs/fy15osti/63972.pdf>.

Mai, T., W. Cole, V. Krishnan, and M. Bolinger. 2015. “Impact of Federal Tax Policy on Utility-Scale Solar Deployment Given Financing Interactions.” Golden, CO: National Renewable Energy Laboratory. NREL/PR-6A20-65014, NREL Presentation. <http://www.nrel.gov/docs/fy16osti/65014.pdf>.

- Mai, T., W. Cole, E. Lantz, C. Marcy, and B. Sigrin. 2016. “Impacts of Federal Tax Credit Extensions on Renewable Deployment and Power Sector Emissions.” Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-65571. <http://www.nrel.gov/docs/fy16osti/65571.pdf>.
- Ma, J., V. Silva, R. Belhomme, D.S. Kirschen, and L.F. Ochoa. 2013. “Evaluating and Planning Flexibility in Sustainable Power Systems.” In *2013 IEEE Power and Energy Society General Meeting (PES)*, 1–11. doi:10.1109/PESMG.2013.6672221.
- Mills, Andrew, and Joachim Seel. 2015. “Flexibility Inventory for Western Resource Planners.” Berkeley, California: Lawrence Berkeley National Laboratory. LBNL-1003750. https://emp.lbl.gov/sites/all/files/lbnl-1003750_0.pdf.
- Mills, Andrew, and Ryan Wiser. 2014. “Strategies for Mitigating the Reduction in Economic Value of Variable Generation with Increasing Penetration Levels.” Berkeley, California: Lawrence Berkeley National Laboratory. LBNL-6590E. <https://emp.lbl.gov/sites/all/files/lbnl-6590e.pdf>.
- NERC (North American Electric Reliability Corporation). 2014. “2014 Long-Term Reliability Assessment.” Atlanta, GA: North American Electric Reliability Corporation. http://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/2014LTRA_ERATTA.pdf.
- . 2015a. “2015 Summer Reliability Assessment.” North American Electric Reliability Corporation. http://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/2015_Summer_Reliability_Assessment.pdf.
- . 2015b. “2015-16 Winter Reliability Assessment.” North American Electric Reliability Corporation. http://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/2015-16%20WRA_Report_Final.pdf.
- NREL SolarPACES. 2014. “Concentrating Solar Power Projects.” <http://www.nrel.gov/csp/solarpaces>.
- NV Energy. 2014. “Southern Service Territory IRP.” https://www.nvenergy.com/company/rates/filings/IRP/NPC_IRP/.
- Palchak, D., and P. Denholm. 2014. “Impact of Generator Flexibility on Electric System Costs and Integration of Renewable Energy.” Golden, Colorado: National Renewable Energy Laboratory. NREL/TP-6A20-62275. <http://www.nrel.gov/docs/fy14osti/62275.pdf>.
- PNM. 2013. “PNM, State Reach Agreement with EPA for Plan to Address Regional Haze at San Juan Generating Station.” *PNM Press Release*, February 15. <https://www.pnm.com/0215-san-juan>.
- Randazzo, R. 2014. “APS Plans to Close One of Four Generators at Cholla Power Plant.” *AZ Central*, September 11. <http://azc.cc/1rYiRwb>.
- Saha, A. 2013. “Review of Coal Retirements.” Concord, MA: M.J. Bradley & Associates, LLC.

- SCPPA (Southern California Public Power Authority). 2014. “Request for Information: Generation Replacement and Future Resources.” Southern California Public Power Authority. http://www.scppa.org/Downloads/RFP/GenReplacement_FutureResources_RFI.pdf.
- SCE (Southern California Edison). 2013. “Southern California Edison Announces Plans to Retire San Onofre Nuclear Generating Station.” *SCE Press Release*, June 7. <http://www.songscommunity.com/new2013/news060713.asp>.
- Short, W., P. Sullivan, T. Mai, M. Mowers, C. Uriarte, N. Blair, D. Heimiller, and A. Martinez. 2011. “Regional Energy Deployment System (ReEDS).” Golden, Colorado: National Renewable Energy Laboratory. NREL/TP-6A20-46534. <http://www.nrel.gov/docs/fy12osti/46534.pdf>.
- Sigrin, Ben, Michael Gleason, Robert Preus, Ian Baring-Gould, and Robert Margolis. 2016. “The Distributed Generation Market Demand Model (dGen): Documentation.”
- Sullivan, P., K. Eurek, and R. Margolis. 2014. “Advanced Methods for Incorporating Solar Energy Technologies into Electric Sector Capacity-Expansion Models: Literature Review and Analysis.” Golden, Colorado: National Renewable Energy Laboratory. NREL/TP-6A20-61185. <http://www.nrel.gov/docs/fy14osti/61185.pdf>.
- Venkataraman, Sundar, Gary Jordan, Richard Piwko, Lavelle Freeman, Udi Helman, Clyde Loutan, Grand Rosenblum, et al. 2010. “Integration of Renewable Resources - Operational Requirements and Generation Fleet Capability at 20% RPS.” California ISO. <https://www.caiso.com/Documents/Integration-RenewableResources-OperationalRequirementsandGenerationFleetCapabilityAt20PercRPS.pdf>.
- Ventyx. 2010. “Energy Market Data.” <http://www.ventyx.com/velocity/energy-market-data.asp>.
- WECC (Western Electricity Coordinating Council). 2012. “Assumptions Matrix for the 2020 Transmission Expansion Planning Policy Committee (TEPCC) Dataset.” Western Electricity Coordinating Council. <http://www.wecc.biz/library/StudyReport/Documents/Assumptions%20Matrix%20for%20the%202020%20TEPPC%20Dataset.pdf>.
- . 2014a. “Load and Resources Subcommittee (LRS).” http://www.wecc.biz/committees/StandingCommittees/PCC/LRS/Shared%20Documents/Public_Data/.
- . 2014b. “Release Notes for WECC 2024 Common Case, Version 1.1.” Transmission Expansion Planning Department.
- Zucker, A., T. Hinchliffe, and A. Spisto. 2013. “Assessing Storage Value in Electricity Markets.” EUR 26056 EN. European Commission Joint Research Centre Institute for Energy and Transport.

Appendix A: Net Load Duration Curve Modeling for Variable Generation

The use of just 96 hours of chronological dispatch in RPM means that the tails of the demand distribution are not adequately represented. We here present our methodology for representing the ability of a system to meet demand in these periods and more adequately capture system flexibility. The methodology uses load duration curves, or sorted system load values, to analyze the highest and lowest system dispatch periods and better represent system needs and constraints in those periods.

Three types of load duration curves, shown in Figure A1, are used. Each of these load duration curves is re-calculated every model solve year to reflect changes from the previous solve year. The load duration curve (LDC) is the curve created by sorting the total system load. The net load duration curve (NLDC) subtracts the profiles of all installed variable resources, such as solar, wind, and fixed-dispatch hydro, from the load, and then orders the subsequent net load curve. In RPM, one LDC and NLDC are created for each NERC sub-region. The incremental load duration curve (ILDC) represents the potential new net load curve if an incremental amount of variable capacity is installed in a particular renewable resource region. In particular, we iteratively assume that 500 MW of capacity is installed in each resource region, and subtract the resulting output profile for that incremental capacity addition from the NLDC of the corresponding NERC region. There is then one ILDC for each technology type⁴⁷ and renewable resource region represented in RPM. We use this ILDC to calculate incremental effects on a per-MW basis of new installed capacity. Each of these load duration curves is re-calculated each model solve year to update the impact of newly installed technologies.

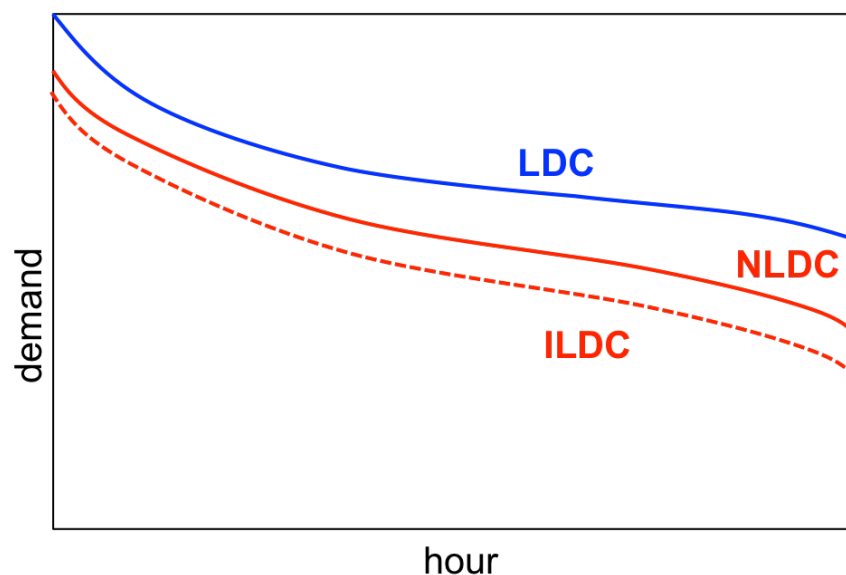


Figure A1. Sample load duration curve (LDC), net load duration curve (NLDC), and incremental load duration curve (ILDC)

⁴⁷ Separate ILDCs are calculated for each wind class and separately for the two utility-PV technologies (fixed tilt and single-axis tracking).

The choice of 500 MW for the increment of potential new capacity is somewhat arbitrary but is loosely based on large utility-scale renewable energy projects considered today. The accuracy of this assumption would depend on scenario outcomes; for example, if the model decides to deploy much greater than 500 MW of a technology in a given region, the assumption might underestimate the impact to curtailment or capacity value. However, this is partially mitigated as this installed capacity gets rolled into the “existing” VG fleet in the next solve year. In addition, we tested the impact of using incremental capacities ranging from 1 MW to 1,000 MW and found little sensitivity in the resulting capacity value and curtailment calculations until 250 MW, and then an increasing sensitivity in the results for incremental capacities up to 1,000 MW.

A.1 Capacity Value

Variable generation sources do not contribute full capacity to system resource adequacy or planning reserve needs, as they are not able to control their output or increase output during times of high load. However, these resources can reliably contribute some fraction of their capacity during these hours. In order to determine this fraction, we examine the impact of renewable resources on the capacity needs of peak hours by calculating the reduction in net demand during N peak hours. Figure A2 shows the areas used for the capacity value calculation for both existing resources and incremental resources. This method has been shown to be a good approximation technique for capacity value (Madaeni, Denholm, and Sioshansi 2012). This methodology requires N top hours of load to be analyzed. We used 100 hours for N , which is within the range examined by (Madaeni, Denholm, and Sioshansi 2012). We examined the effect of using between 1 and 500 hours for this calculation and found that between 50 and 250 hours yielded stable results. Smaller numbers of hours would lead to noisy results because the analysis is performed with only one year of data.

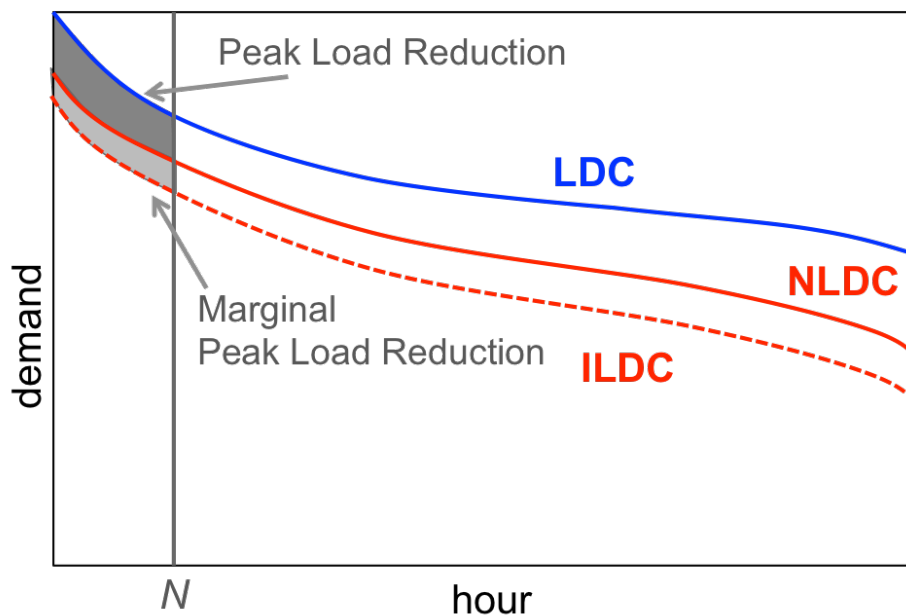


Figure A2. Example of capacity value calculation using the load duration curve methodology

The average capacity value for existing resources is calculated for each NERC region as a fraction of the installed capacity based on the area between the LDC and NLDC for the top N hours analyzed:

$$CV = \frac{\sum_{n=1}^N LDC - NLDC}{N * C} \quad (A1)$$

where CV is the capacity value, N is the number of peak hours, LDC and $NLDC$ are the load duration and net load duration curves, and C is the installed capacity of the resource. The marginal capacity value (MCV), or the capacity value on a per-MW basis of an incremental amount of potential new capacity is calculated similarly as the area between the ILDC and NLDC for the same top number of hours:

$$MCV = \frac{\sum_{n=1}^N NDC - ILDC}{N * C} \quad (A2)$$

where in this case C is the incremental capacity used to calculate the ILDC, 500 MW. The RPM model uses the marginal capacity value to better inform investment decisions. Each region and technology type will have a different marginal capacity value for the particular resource built; this capacity value is then used in the planning reserve constraint for each NERC sub-region and allows new variable generation to contribute capacity to the system. Regions that are capacity limited have an incentive to build new variable generation in regions with higher marginal capacity values. Figure A3 shows the range of capacity values in regions in which the model built capacity and in those for which it did not. Many variables contribute to the model's decision to build in different regions, however it can be seen that higher capacity values are typically preferred by the model.

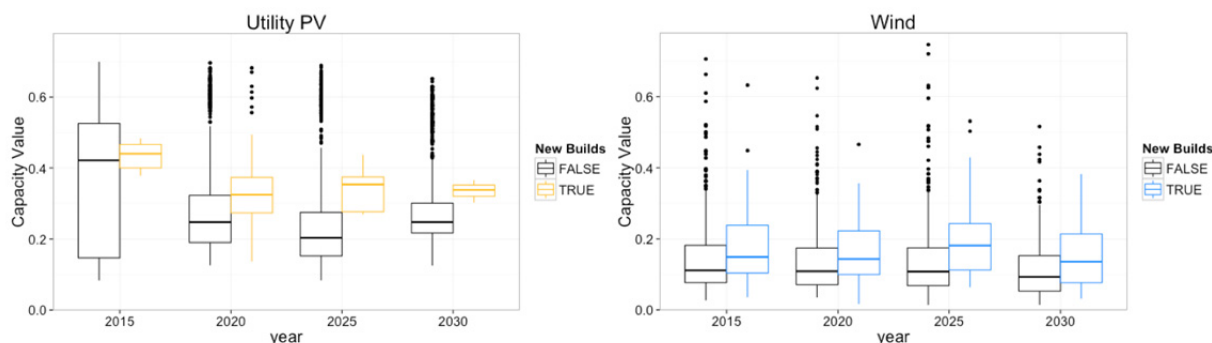


Figure A3. Marginal capacity values of utility-scale solar and wind for regions in which the model built new capacity and did not build new capacity for each of the model solve years

The existing and marginal capacity values typically decrease as additional capacity is installed, as is seen in Figure 5 of the main paper. The exact changes in capacity values are influenced by both the amount of installed capacity and also the types and locations of installed capacity. The interplay between different technology types, particularly wind and solar, and their impact on each other's capacity value is not straightforward and can in some cases lead to an increasing marginal capacity value over time for technology options in a given region. The use of load duration curves allows the RPM model to account for these complicated relationships and how they change from year to year in the model, and informs investment decisions for the specific capacity mix of the system.

The planning reserves constraint is implemented in RPM as

$$\sum_q(\text{capacity}O * cvO(q) + \text{capacity}N * cvN(q, n)) > P(1 + RM) \quad (A3)$$

where *capacityO* is the current installed capacity in the system, *capacityN* is a decision variable for the model to build new capacity, *cvO* is the existing capacity value for each technology type (for dispatchable resources this value is 1), *cvN* is the marginal capacity value by technology and region, *P* is the peak demand of the system, and *RM* is the reserve margin.

A.2 Renewable Curtailment

Curtailment of renewable generation necessitates the use of higher-variable-cost generators to replace the curtailed energy. Correctly representing curtailment in RPM leads to more accurate representation of system operations and costs. One reason for renewable curtailment includes the inability of thermal generators to reduce output due to generator inflexibility or operating reserve constraints. In other cases, transmission constraints can prevent low variable cost energy to be exported and used elsewhere. Typically, curtailment occurs during the lowest net load hours of the year, represented by the lower tail of the NLDC. We have used a similar method to the capacity value method described above to determine the amount of curtailment that would be expected throughout the year from the RPM model, and created a constraint to adequately represent this curtailment in the model.

This methodology uses an effective minimum generation point of the system, the “Effective Min-Gen”. This min-gen level represents the lower bound on the system flexibility and is assumed to be constant throughout the year. This level depends on the penetration of the system and unit commitment decisions of the various generator types. The ability to turn down or even turn off generators will reduce the effective min-gen of a system. If the NLDC falls below the effective min-gen level, the system would not be able to reduce dispatchable generation and still meet system reserve and stability requirements, such that any renewable energy below this point would need to be curtailed. Additionally, the area between the NLDC, ILDC, and effective min-gen level represents the marginal curtailment of new variable generators. Figure A5 shows an example of the effective min-gen level and the areas representing curtailed energy.

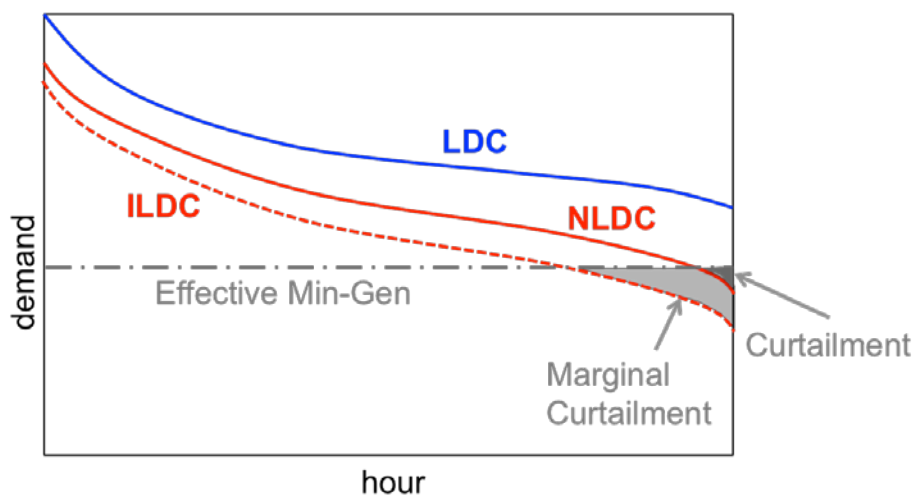


Figure A5. Example of load duration curve, net load duration curve, and incremental load duration curve and the effective minimum generation level used to calculate curtailment

The curtailment parameter for the existing fleet is represented as a fraction of the available energy from renewable energy that is below this system min-gen level:

$$Curt = \frac{\Sigma(M-NLDC)}{E} \quad s. t. \quad NLDC < M \quad (A4)$$

where $Curt$ is the total curtailment of existing resources per MWh, M is the effective minimum generation line, and E is the total available annual energy from existing variable sources. The summation is over all hours of the year. Similarly, the marginal curtailment ($MCurt$) of new resources is calculated per MWh of incremental energy from variable generation as

$$MCurt = \frac{\Sigma(M-NLDC-ILDC)}{E} \quad s. t. \quad ILDC < M \quad (A5)$$

The existing curtailment is calculated for each NERC sub-region based on all variable generation technologies installed at the beginning of each solve year, except for distributed generation sources (e.g., Rooftop PV). It is assumed that these are unable to be controlled by the dispatch process, and therefore are unable to be curtailed. Marginal curtailment is calculated for each individual variable generation technology type and region, as in the capacity value calculation. The marginal curtailment is then multiplied by the supplied energy from new generators, such that new generators are curtailed at a specific rate for each type and region, regardless of the capacity installed.

Importantly for this method, the effective min-gen line M must be calculated based on the system capabilities and will change as the system changes. In order to calculate the effective min-gen line M , we created a regression based on data from PLEXOS production cost modeling results for the Western Interconnection. Data is utilized from the Western Wind and Solar Integration Study Phase 2 (Lew et al. 2013) and the Low Carbon Grid Study (Brinkman et al. 2015). Each study has multiple scenarios with different solar and wind penetrations. For each scenario, we modified the solar and wind profiles in the RPM model to match the solar penetration, wind penetration, and fixed-dispatch hydro penetration of these scenarios. With these profiles, we calculated the min-gen level that produced the same amount of curtailment as is seen in WWSIS or LCGS. From these results we created a regression for the effective min-gen lines that calculated the correct curtailment according to the PLEXOS studies. The regression took the form of

$$M_i = a_i * (S + W + H) + b_i \quad (A6)$$

where M_i is the effective min-gen fraction, S is the solar penetration, W is the wind penetration, H is the fixed-dispatch hydro penetration, and a_i and b_i are coefficients of the regression for each NERC region i . Figure A6 shows the regression results for each of the four NERC sub-regions in the Western Interconnection. The regression model is computed on a normalized basis, with M_i calculated as the effective minimum generation level divided by the dispatchable capacity of the system.

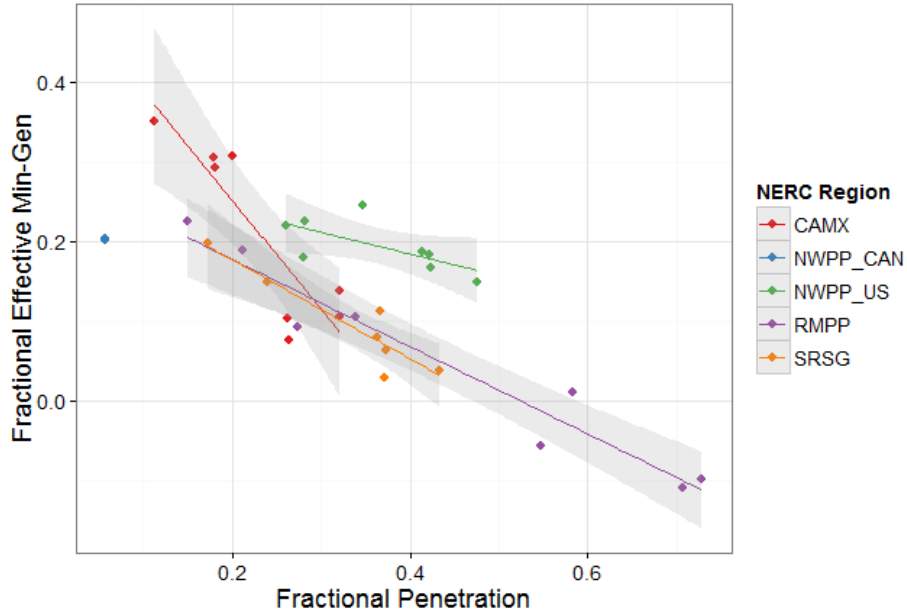


Figure A6. Effective minimum generation regression for each NERC region based on the penetration of variable generation sources

This regression is then tested against data from flexibility studies in California (Paul Denholm et al. 2015) and the Colorado Test System (Palchak and Denholm 2014) in order to ensure the regressions yield realistic results for other systems. The validated effective min-gen regression is then used with data for each model year to predict the effective minimum generation level for that particular system. This min-gen level will change as RPM installs higher penetrations of renewable sources, and is used to calculate the existing and marginal curtailment from renewables on the system.

These curtailment values are then incorporated into the RPM model as constraints on the optimization problem. RPM requires that the total curtailment from all existing renewable sources be greater than the calculated existing curtailment, and also requires that the curtailment from new renewable sources be greater than the calculated marginal value of curtailment. These constraints are represented in RPM as:

$$\sum_{h,n,q} \text{Curtailment}(h,n,q) \geq E_O * \text{Curt} + \sum_{q,n} \left(\sum_h E_N(q,n,h) \right) * \text{MCurt}(q,n)$$

$$\sum_h \text{Curtailment}(h,n,q) \geq \left(\sum_h E_N(q,n,h) \right) * \text{MCurt}(q,n)$$

$$\text{Curtailment}(h,n,q) \leq E_N(q,n,h)$$

where *Curtailment* is the total curtailed energy in the model, E_O is the total available energy from existing variable generation sources, E_N is the available energy from potential new variable generation sources, *Curt* is the existing curtailment fraction, *MCurt* is the marginal curtailment

fraction, h is a particular hour of dispatch, n is a particular resource region, and q is a particular technology type. These equations require the model to satisfy both the overall curtailment constraint throughout the model and the increased curtailment from building new generation sources in a particular area.

The curtailment constraints do not specify any time constraints on the curtailment; they only specify the yearly total curtailment. As such, the model is allowed to optimize when curtailment occurs throughout the year. However, the curtailment for any hour, region, and technology must be less than the available energy for that particular hour, region and technology. Figure A8 shows how the model distributed the curtailment in one case with 30% penetration.

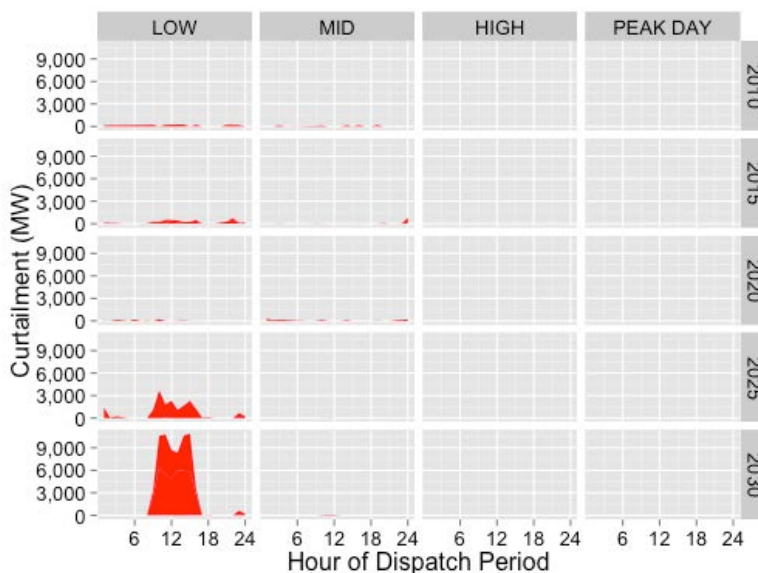


Figure A8. Hourly curtailment for WECC

The curtailment constraints influence model decision making by reducing the amount of energy that a particular resource may provide. This will influence the effective cost of these resources, potentially making one resource more expensive to the model. Figure A9 shows the marginal curtailment values for regions in which the model chose and did not choose to build new capacity. RPM does not include a cost to curtailment, so higher marginal curtailment values only impact the optimization decisions in terms of the total energy that a particular resource would be able to provide.

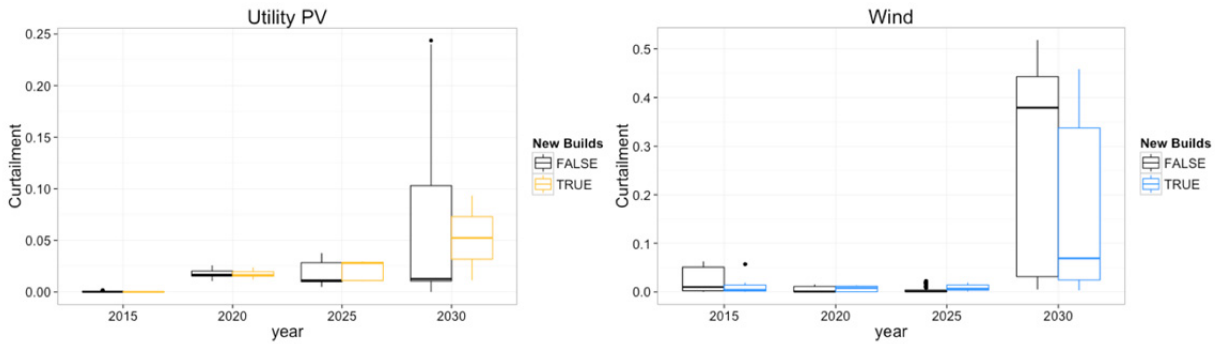


Figure A9. Marginal curtailment fractions of utility-scale solar and wind for regions in which the model built new capacity and did not build new capacity for each of the model solve years

This approach to modeling curtailment has several drawbacks, including an indirect modeling of transmission and a static regression. Typically the ability to export excess energy during periods of high renewable generation allows for reduction in curtailment from regions with high wind or solar penetration. While the aggregate imports and exports in WECC are captured within the regression model, this static method only accounts for imports and exports as modeled in these studies. Additionally, the impact of additional transmission capacity or installation of variable generation in regions not modeled in WWSIS-2 or LCGS would not be captured in this regression.

A regression model is required because RPM does not explicitly model unit commitment, so is unable to determine its own minimum generation levels based on integer commitment constraints of the generators. Additionally, the regression is based on static capacity. The retirement of base-load generators or the installation of new flexible technologies would impact the regression model; however, the regression model does not take these impacts into account, nor does the regression model change with time. However, the studies used to create the regression model represent a variety of regions and renewable penetrations such that we believe the regression adequately models potential future outcomes. This is, however, an area for future research.

Appendix B: Net Load Duration Curve Modeling for Flexibility Investments

B.1 Interruptible Load Methodology

Interruptible load technologies affect the NLDC and ILDC predominantly at the peak due to their operational pattern of reducing load during periods of high demand. As such, the only flexibility method related to these technologies is the calculation of capacity value, no curtailment impacts are modeled.

The capacity value of interruptible load is determined by a similar method to variable generation. Interruptible load resources have constraints that limit their ability to be used on demand, which reduces the resources' abilities to contribute their full capacity to the system. As such, these technologies are subject to capacity reductions similarly to variable generation sources. We assume annual and daily energy budgets for these resources, thereby constraining their capacity values. The top 100 hours of the year are used to determine the degree to which these resources are able to reduce capacity needs during peak hours. However, interruptible load resources do not have weather-determined output profiles as do variable generation resources. Before implementing the capacity value method, we create an example profile. When creating this profile, we assume these resources will be used within the full extent of their availability and that grid operators will call these resources during the times of highest demand. The created profiles do not inform model decisions; they are only example profiles for how the resources could be used by system operators.

Once these profiles are created, the same methodology is applied to the profiles as is done with variable resources to calculate a capacity value. A key difference however is that the assumed profile is subtracted from the NLDC rather than the LDC to account for the ability of these resources to mitigate variability from non-dispatchable sources. The capacity values found for interruptible resources varied greatly based on their annual and daily energy budgets and the fraction of the top hours in which they are able to participate, with the maximum capacity value of interruptible load resources with 50 hours of annual storage being only 0.5, and reduced capacity values for resources with lower daily energy limits. Figure A4 shows the marginal capacity value of the technologies that were and were not built by the model.

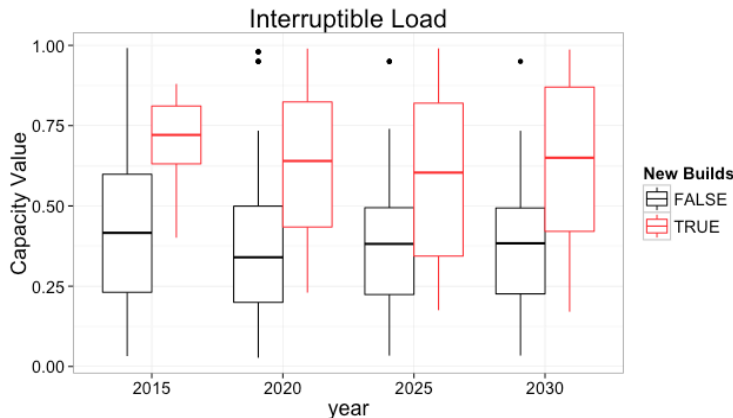


Figure B1. Marginal capacity values of interruptible load resources for resources deployed and not deployed for each of the model solve years

B.2 Storage Methodology

Storage technologies allow for additional system flexibility by shifting energy from periods of low demand to high demand. This flexibility can be captured by calculating the capacity value and the reduction in curtailment a storage technology can provide. We use similar methods to those described in sections A.1 and A.2 to determine the additional flexibility provided by storage facilities. Similar to interruptible loads, storage technologies do not have a fixed profile of available energy, but do have constraints on their ability to provide energy at peak times. We use a similar methodology to create an example profile for each storage technology, taking into account the constraints on energy storage capacity for each technology, including total storage capacity and upper and lower boundaries on storage levels. We assume that these systems would be operated to maximize their use, storing energy during periods of low demand and dispatching this energy during periods of high demand while cycling with a period of $e(1 + 1/\eta)$. While this profile does not indicate how the storage facility would be used within the RPM model, it represents an example of how the system might be operated in a manner to provide both high capacity value and high curtailment reduction. We ensure that the sample profile obeys all temporal constraints of the storage facilities.

This sample storage profile is then subtracted from the NLDC to create a new storage load duration curve (SLDC), which typically impacts both the highest and lowest hours of the NLDC. This is done for each storage technology and each NERC region. A sample SLDC can be seen in Figure A10. The SLDC is used to compute the average capacity value of all storage technologies installed in a solve year. When storage technologies of various energy capacities are installed, the SLDC is created by ordering the storage technologies from lowest energy capacity to highest energy capacity, such that fast cycling storage technologies are first used to smooth variability, followed by longer cycling storage technologies.

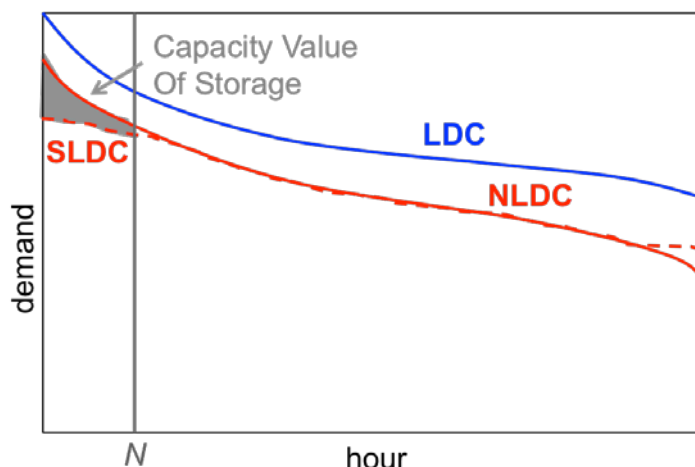


Figure B2. Example of load duration curve impacts from storage technologies on the capacity value and curtailment

New storage capacities are also investigated with the creation of an incremental storage load duration curve (iSLDC). This curve uses an incremental capacity of 500 MW and is created similarly to the SLDC based on constraints on the storage facilities. The storage capacity is defined for each class of storage technology and an iSLDC is calculated for each potential new storage technology and NERC region. We assume that the value of storage will be more closely related to the load duration curves of a particular NERC region than node, and assign the same iSLDC to each node within a NERC region. The capacity value calculated for each iSLDC informs the ability of new storage facilities to contribute to the capacity requirements of each region.

The curtailment reduction from adding storage to a system is computed in a similar manner to that described in section A.2, using the SLDC rather than the NLDC. The effective minimum generation line is assumed to be fixed from the analysis of existing and marginal curtailment, and that the storage facility would only increase the values relative to the NLDC such that curtailment would be reduced. Figure A11 shows an example SLDC and its impact on curtailment.

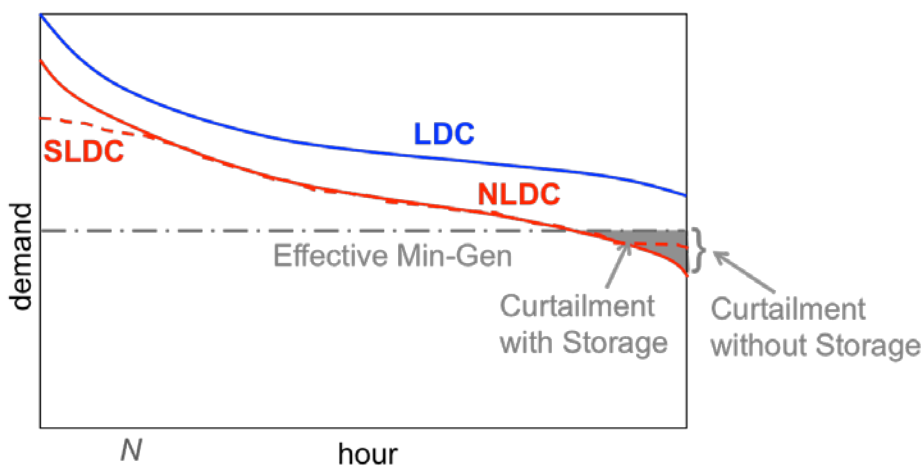


Figure B3. Example of storage impacts on the net load duration curve, and the curtailment fallout

Marginal curtailment from an incremental amount of storage is calculated as a potential reduction in curtailment when storage is added, based on the difference in the SLDC and iSLDC. Figure A12 shows the incremental reduction from curtailment.

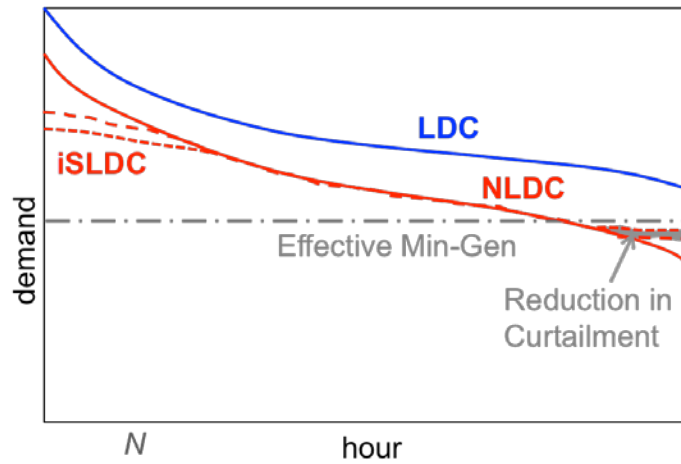


Figure B4. Example of load duration curve from an incremental amount of storage and the impact new storage technologies can have on curtailment

When storage is included in the model, Equation A7 is modified as:

$$\sum_{h,n,q} Curtailment(h, n, q) \geq E_o * (SCurt - MSCurt * S_N) + \sum_{q,n} \left(\sum_h E_N(q, n, h) \right) * MCurt(q, n)$$

$$\sum_h Curtailment(h, n, q) \geq \left(\sum_h E_N(q, n, h) \right) * MCurt(q, n)$$

$$Curtailment(h, n, q) \leq E_N(q, n, h) \quad (A8)$$

where $SCurt$ is the curtailment calculated with the SLDC instead of the NLDC, $MSCurt$ is the marginal curtailment reduction from an incremental amount of storage capacity, S_N is the amount of new storage capacity installed, and all other terms are as Equation. A7. This equation incentivizes the model to install additional storage technologies based on their ability to reduce curtailment from renewable sources.

We do not calculate an interaction term between incremental storage capacity and incremental renewable capacity, which theoretically could be used to determine the degree to which additional storage facilities could reduce curtailment of new renewable facilities. However, this term would lead to non-linearities in the model so is not included. The effect such an interaction would have is assumed to be small however, and once new technology is installed, its impacts are fully incorporated in the next model solve year.

Appendix C: Detailed Equation Listings

C.1 Interruptible Load Model

The investment decisions and dispatch of interruptible load are subject to the following constraints in the optimization portion of the model. First, nodal capacity of technology sub-class c , $C_{n,c}$, is determined by the BA capacity, $C_{b,c}$, multiplied by the node's load participation factor, p_n ; and generation P , which is a function of the model hour (d, h) defined by dispatch period d and hour h , is similarly distributed:

$$C_{n,c} = C_{b,c} \cdot p_n, \quad \forall b, c, \forall n \in N_b$$

$$P_{n,c,d,h} = p_n \cdot \sum_{n' \in N_b} P_{n',c,d,h}, \quad \forall b, c, d, h \quad (\text{C.2})$$

$$P_{n,c,u,d,h} = p_n \cdot \sum_{n' \in N_b} P_{n',c,u,d,h}, \quad \forall b, c, u, d, h \quad (\text{C.3})$$

Note that Eqn. C.2 specifies generation for new sources of interruptible load, while Eqn. C.3 specifies generation for existing sources of interruptible load, which have been assigned a unit number, u , for tracking. The sum of new and old capacity ($c_{b,c,u}$, a parameter) is restricted to be less than a fraction r_c of peak load, by BA

$$C_{b,c} + \sum_u c_{b,c,u} \leq r_c \max_t l_{b,t}, \quad \forall b, c$$

where $l_{b,t}$ is the 8,760 load time series for BA b . And then we model energy limitations at two timescales—one constraint on the amount of energy available per day a found in our 96 hour model of dispatch:

$$\sum_{(d,h) \in D_a} P_{n,c,d,h} \leq C_{n,c} \cdot e^{\text{daily}}, \quad \forall a, n, c$$

$$\sum_{(d,h) \in D_a} P_{n,c,u,d,h} \leq c_{n,c,u} \cdot e^{\text{daily}}, \quad \forall a, n, c, u$$

and another for the entire year, where s_d is the number of hours in the year represented by model hour (d, h) :

$$\sum_{d,h} P_{n,c,d,h} \cdot s_d \leq C_{n,c} \cdot e^{\text{annual}}, \quad \forall n, c$$

$$\sum_{d,h} P_{n,c,u,d,h} \cdot s_d \leq c_{n,c,u} \cdot e^{\text{annual}}. \quad \forall n, c, u$$

C.2 Storage Model

The dispatch portion of the optimization model imposes the following constraints on storage operation. Storage level is modeled by letting the model choose the starting storage level for each dispatch period, L_d^{start} (further indexed below by node n and storage sub-class c), and then calculating changes to that level based on generation $P_{d,h}^{gen}$, charging $P_{d,h}^{chg}$, and use for flexibility and regulation reserves ($R_{d,h}^{flex}$ and $R_{d,h}^{reg}$, respectively). In particular, for new resources, we have

$$L_{n,c,d,1} = L_{n,c,d}^{start} - P_{n,c,d,1}^{gen} + P_{n,c,d,1}^{chg} \cdot \eta_c - 0.25(R_{n,c,d,1}^{flex} + R_{n,c,d,1}^{reg})(1 - \eta_c) \quad \forall n, c, d$$

$$L_{n,c,d,h} = L_{n,c,d,h-1} - P_{n,c,d,h}^{gen} + P_{n,c,d,h}^{chg} \cdot \eta_c - 0.25(R_{n,c,d,h}^{flex} + R_{n,c,d,h}^{reg})(1 - \eta_c) \quad \forall n, c, d, \forall h > 1$$

and the constraints for existing resources are identical except that they are further indexed by a unit u . The roundtrip efficiency for sub-class c is η_c . Explanatory detail is given in Section 3.2.

The storage level is constrained to stay within its sub-class specific energy constraint e_c both on an hourly, and a seasonal basis. The hourly constraints for new resources are:

$$L_{n,c,d}^{start} \leq C_{n,c} \cdot e_c, \quad \forall n, c, d$$

$$L_{n,c,d,h} \leq C_{n,c} \cdot e_c, \quad \forall n, c, d, h,$$

and the ones for existing resources are nearly identical:

$$L_{n,c,u,d}^{start} \leq c_{n,c,u} \cdot e_c, \quad \forall n, c, u, d$$

$$L_{n,c,u,d,h} \leq c_{n,c,u} \cdot e_c, \quad \forall n, c, u, d, h,$$

except that $c_{n,c,u}$ is not a variable (as is $C_{n,c}$), but rather a parameter that represents the total amount of existing capacity of sub-class c at node n and assigned to unit u . The seasonal energy constraint is necessary because we allow imbalances over individual dispatch periods, and only require full balancing on an annual basis. Mathematically, these two constraints are expressed as

$$(L_{n,c,d,24} - L_{n,c,d}^{start}) \cdot s_d \leq C_{n,c} \cdot e_c, \quad \forall n, c, d$$

$$\sum_d (L_{n,c,d,24} - L_{n,c,d}^{start}) \cdot s_d = 0, \quad \forall n, c$$

for new capacity and as

$$(L_{n,c,u,d,24} - L_{n,c,u,d}^{start}) \cdot s_d \leq c_{n,c,u} \cdot e_c, \quad \forall n, c, u, d$$

$$\sum_d (L_{n,c,u,d,24} - L_{n,c,u,d}^{start}) \cdot s_d = 0, \quad \forall n, c, u$$

for existing capacity, where s_d is the number of hours in the year represented by each hour in d .