Keep the Lights On: Ensuring Bulk-Power System Reliability in a Decarbonized Future

by

Cathy Wang

B.A., B.S. Washington and Lee University (2015)

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

Master of Science in Technology and Policy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

Author . . . .
Institute for Data, Systems, and Society
August 5, 2021

Certified by . . . .
Paul Joskow
Elizabeth and James Killian Professor of Economics, Post-Tenure
Thesis Supervisor

Accepted by . . . .
Noelle Selin
Director, Technology and Policy Program
Professor, Institute for Data, Systems, and Society and
Department of Earth, Atmospheric, and Planetary Science

Keep the Lights On: Ensuring Bulk-Power System Reliability in a Decarbonized Future

by

Cathy Wang

Submitted to the Institute for Data, Systems, and Society
on August 5, 2021, in partial fulfillment of the
requirements for the degree of
Master of Science in Technology and Policy

Abstract

Policy and market forces are ushering in a new power system, one dominated by variable renewable energy (VRE) resources (wind and solar) and energy-limited resources like energy storage. Because these resources have different characteristics than the conventional thermal generators that make up the bulk of our capacity mix currently, this capacity turnover necessitates new ways of thinking about and planning for bulk-power system reliability.

This Thesis evaluates the modeling approaches currently used in capacity planning and resource adequacy frameworks, and proposes a new iterative approach to incorporating cost-efficiency, decarbonization, and reliability goals into capacity and resource adequacy planning. As recent large-scale blackout events in California and Texas illustrate, both demand and supply can be heavily impacted by extreme weather events, contributing to more conditions of system stress in the years to come.

By carefully taking into account periods of high risks of incurring reliability shortfalls, we show that actual reliability can be greatly improved in a systems analysis, compared to separate planning and resource adequacy analyses. Going forward, we need to find better ways of capturing the variations and correlations between time-coincident VRE output, load realizations, and unplanned thermal generator outages, to appropriately characterize and communicate the risks of power supply shortfalls (i.e., duration, frequency, magnitude). This has key implications on how end-use customers think about losing power for some period of time, how much they are willing to pay for customer-side reliability, and how their preferences are reflected at the system level.

Thesis Supervisor: Paul Joskow
Title: Elizabeth and James Killian Professor of Economics, Post-Tenure

Acknowledgments

This Thesis was born out of the modeling efforts for the MIT Energy Initiative's Future of Storage (FoS) study, an inter-disciplinary undertaking across the institute. Thanks to everyone on the team – I'm much inspired by your great research and innovations in the rapidly expanding energy storage space.

My biggest thanks to the FoS Modeling team. To my advisor, Paul Joskow, whose support and willingness to push back on my sometimes pre-mature ideas allowed me to sharpen and refine the discussion points contained herein. To Dharik Mallapragada, whose mentorship and white-boarding taught me more than any class I've taken at MIT. To Cristian, Dick, Howard, Hannes, and Patrick, whose insights and discussions continuously pushed the boundaries of my thinking.

My MIT experience wouldn't have been complete without the classes and communities that challenged my preconceptions about sustainability, justice, and methods to solving "wicked" societal problems (and how they systematically exclude groups of people). To the TPP community and staff, whose relentless efforts in outreach and support helped me keep sane during the early months of quarantine (special shout-out to the stellar folks in the TPP Student Society). To Aaron, Abhishek, Drake, and Ragini, whose companionship provided sunshine and laughter in Zoom University. To my peers in the MIT Energy Club, Sloan Energy Club, Sustainability Certificate program, and MIT Policy Hackathon, whose friendship I will forever cherish.

Finally, to my family in Toronto, whose support from afar continues to fuel my personal growth. And, to Anthony, whose steadfastness and passion for self-improvement sustains me on a daily basis. Many thanks!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The power sector has a major role in the economy-wide decarbonization effort, primarily through: (1) the near-complete decarbonization of electricity production; and (2) the electrification of other sectors. These herculean efforts will necessitate overcoming integration challenges of similar magnitude, spanning areas of grid operations, infrastructure investments, and public participation in such decision-making processes. In this Thesis, we concern ourselves primarily with keeping the lights on in a highly decarbonized power system, particularly around how to improve upon the use and interpretation of Capacity-Expansion Model (CEM) results in planning and resource adequacy processes.

1.1 Background and Motivation

1.1.1 Trends in Decarbonization of the Power Sector

After decades experience, Variable Renewable Energy (VRE) resources have sufficiently gone down the learning curve to become cost-competitive against new thermal generating capacities. The ARENA 2019 Renewable Power Generating Costs report [26] showed since 2010, utility-scale solar PV power cost has declined by 82%, followed by concentrating solar power (CSP) at 47%, onshore wind at 39%, and offshore wind at 29%. More than half of the VRE capacity added in 2019 came in at lower power costs than the cheapest new coal plants [26]; and increasingly more share is coming in costs less than keeping existing coal plants online. However, entry and exit of capacities only constitute a small fraction of capacity turnover (which remains dominated by coal and gas generating capacity).

Economics alone will not be sufficient to achieve the scale and pace of the capacity turnover needed to mitigate climate change; they will need to be supported with cohesive policies and market structures. Although the U.S. (like many other large economies) does not yet have a clear climate mandate, states, utilities, corporations have taken it upon themselves to make decarbonization (often as part of sustainability) pledges. Since Iowa passed the first U.S. Renewable Portfolio Standards (RPS) goal (105 MW by 1998), 30 more states/territories now support an RPS [40]. These RPS goals differ in many aspects, eg. in terms of the timing of the emission reduction, the eligibility of what is considered "clean", and whether or not they are legally binding. Nevertheless, 13 of these are targeting an energy portfolio of more than 50% renewable electricity, including Hawaii, Puerto Rico, Virginia, and Washington DC that have increased their RPS to 100% [40]. And globally, large businesses are committing to 100% renewable electricity (e.g., through the RE100 initiative). Only time will tell as to whether or not any of these announcements carry any bite. One thing is for sure – we are on the precipice of a massive energy transition, buoyed by market and policy forces.

1.1.2 Electrification to Accelerate Economy-Wide Decarbonization

While electricity demand has leveled out (or even decreased) in the U.S. in recent years due to improvements in energy efficiency, electrification of light-duty transportation, space heating, buildings, and parts of the industry have the potential to dramatically increase the requirements for electricity in the future (Figure 1-1). In NREL's Electrification Futures Study [35], the authors assessed three plausible electrification

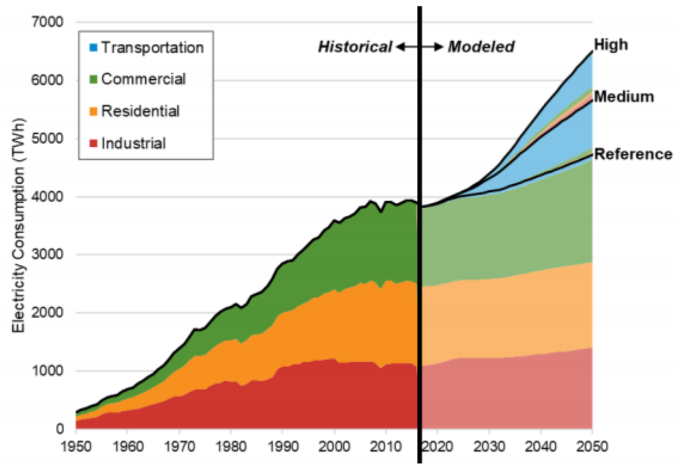

Figure 1-1: Historical and projected annual U.S. electricity demand (1950-2050)

(High, Medium, Reference) scenarios in the U.S. through 2050 using a bottom-up stock-taking tool of all infrastructures related to energy production, transportation, and consumption. The authors relied on their expert judgment to project annual sales shares of these electric technologies in various industries, which informed equipment stock turnover and consequently final energy and electricity use of: vehicle fleets, appliances, HVAC, industrial machinery, and other types of energy-consuming equipment over time [35].

Given increasing pressures to decarbonize, "high" electrification of other sectors is likely to happen, and will consequently greatly impact annual electricity consumption. The authors forecast 1,782 TWh (38%) greater total electricity demand in 2050 under their High Electrification scenario relative to the Reference (business as usual) scenario (Figure 1-1) [35]. This is driven by greater electricity demand in the transportation sector, due to greater uptake of electric plug-in vehicles. The electrification of buildings also contributes to the increase in electricity demand, though to a more moderate amount due to displacement of inefficient electric resistance heaters by high efficiency of electric heat pumps.

In addition to higher annual demand, "high" electrification of other sectors will also change the seasonal distribution of peak hours, primarily due to the increased reliance on electric heat pumps for space and water heating needs. While in 2015, most states (apart from those in the Pacific Northwest) are primarily summer peaking, growth in space heating electrification will increase the number of peak demand hours occurring in the winter times by 2050. This is a clear departure from the level and shape of electricity demand today.

1.2 Recent Experience with Large-Scale Reliability Shortages

Two recent large-scale power outages in the U.S. shed light on the growing saliency of ensuring bulk-power system reliability in a growing VRE-dominated grid: the August 2020 California outage and the February 2021 Texas outage.

In August 2020, California and its neighboring states experienced a sustained period of extremely hot weather, leading the California ISO (CAISO) to call for two successive 500 MW blocks of controlled load shedding on August 14 for an hour each, and one 500 MW of controlled load shed on August 15 for 20 minutes (implemented as rolling outages for customers) [27]. The "1-in-30" extreme weather event contributed to the outage event in three main ways: (1) higher than standard gross electricity load due to increased use of air conditioning throughout the day; (2) reduced supply of VRE, resulting in higher-than-expected net load on the system; and (3) extreme temperatures in surrounding states reduced the amount of electricity that could be imported into the states.

Actual capacity factors of VRE resources came in much below the firm capacity values determined by the California Public Utilities Commission (CPUC) for the month, making California more dependent on real-time electricity imports (which unfortunately, also declined due to increased demand elsewhere in the region). Interestingly, the shortage events occurred during the peak net load periods (Figure 1-2) [27], pointing to the inadequacies of planning based on peak gross loads.

In February 2021, Texas experienced a deep freeze event, causing 4.5 million Texans

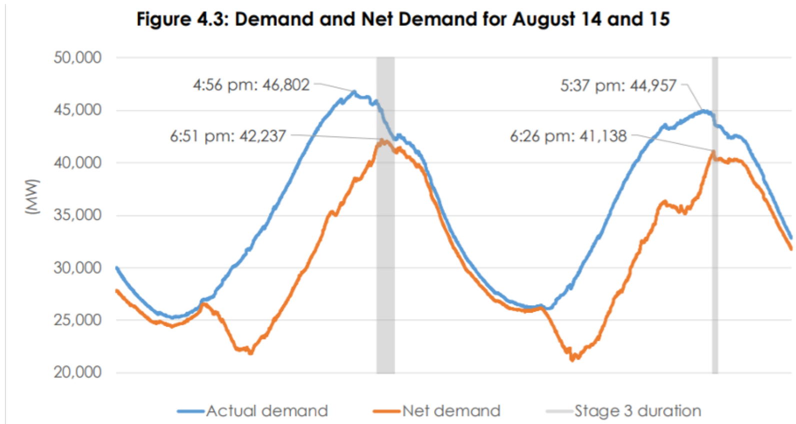

Figure 1-2: August 2020 California supply shortfall

to lose power (and sometimes water service too) for multiple days. Preliminary storm-related damages are estimated to be 151 deaths, and property damages of $130 billion [51]. Due to the electricity supply shortage, Electric Reliability Council of Texas (ERCOT) ordered rolling blackouts on February 14 through 18. At the highest point, 20 GW of load shedding was requested (one third of the system winter peak), and the whole event resulted in 70.5 hours of load shed request [43].

The ERCOT outage was different from the California one in two main ways: (1) ERCOT is not electrically interconnected to the rest of the country, meaning that it cannot depend on large amounts of electricity imports (although in this specific instance, transmission capacity wouldn't have been very helpful as the neighboring MISO and SPP also experienced shortage events); and (2) a large portion of Texas houses rely on electrical heating (61% compared to 40% nationally), making electricity demand in ERCOT particularly sensitive to cold weather events. VREs under-performed by a modest amount (relative to their firm capacities for the month in ERCOT's winter assessment provided to NERC), the thermal generators under-performed massively due to the extreme weather (e.g., well freeze-offs, processing plants curtailed during rolling blackouts). This re-emphasizes that reliability in a future power system is not simply a VRE-specific issue, but also a technology-agnostic issue.

As periods of sustained extreme weather become more frequent with climate change, their impacts on the electricity system will only grow in a high-VRE power system. Going forward, we need to rethink how we currently determine and enforce reliability standards and planning targets, so that events like those in California and ERCOT do not become the new norm.

1.3 Study Scope and Outline

The declining costs of solar and wind technologies, ambitious decarbonization targets, and the electrification of other sectors will lead to a rapid increase in VRE generating resources in the U.S. While these trends show tremendous opportunity to support the transition to a decarbonized future; they also bring about a unique set of challenges that impact on power systems' operational and investment decisions. As discussed thus far, the future of the power system will be drastically different from the one we have and understand today. It follows that legacy approaches to procuring sufficient energy supply to meet demand at the bulk power system level with a very low level of involuntary demand curtailments are no longer fit for purpose.

As such, this Thesis focuses on ensuring that reliability standards established by North American Electric Reliability Corporation (NERC) and the regional reliability councils in the U.S. are taken into account in today's planning processes. Chapter 2 describes the current capacity and resource adequacy processes, and the use of CEM and reliability assessment tools in planning decisions. Chapter 3 presents a proposed approach to preserving both a high-resolution temporal and chronological representation of the VRE output and load profiles, while maintaining computational tractability. Chapter 4 highlights the key findings from using the proposed methodology, and compares and contrasts the results observed in the U.S. Southeast and Northeast. Chapter 5 concludes, and offers suggestions for improved modeling tools, new business models to ensure customer-side reliability, and ways to center justice in the broader energy transition.

Chapter 2

Capacity and Resource Adequacy Planning Processes

Before we go further, we define three interrelated terms applied at the bulk power system level: Reliability, Resource Adequacy, and Resilience. Reliability is an umbrella term for: (1) Resource Adequacy – "the ability of the electricity system to produce and deliver energy to end-use customers at all times"; and (2) Operating Reliability – "supporting the operational needs of the grid to maintain stability and withstand sudden disturbances" [39]. In this Thesis, we don't model ancillary services or other bulk power system balancing products needed to ensure real-time frequency, voltage, and stability of the grid, so we use the terms "Reliability" and "Resource Adequacy" interchangeably. While Reliability refers to generation supply shortage events that occur under predictable system conditions on the bulk power system, Resilience refers to the performance of the grid under severe and rare stress conditions, and focuses on the grid's ability to quickly restore power in the case of unexpected disruptions of generation and high voltage transmission infrastructure.

We anticipate that going forward, with deep penetration of VRE on the bulk power system, extended periods of low solar and wind output, perhaps combined with relatively high demand realizations, are going to cause severe stress conditions, blurring the boundaries between Reliability and Resilience issues (both of which have separate handling processes). Thus, we will sometimes bring in learnings from the Resilience literature and mechanisms to inform the issue of Reliability in a highly decarbonized power system.

2.1 Reliability Standard

The power system has the particular challenge that supply must equal demand at all times for the grid to keep functioning. Because generators do not have 100% availability (e.g., scheduled maintenance, forced outages), ensuring this balance requires a safe generating capacity margin. The responsibility of ensuring this balance for all current and future demand scenarios falls on multiple stakeholders: (1) generation owners who make investment decisions on new capacities, and dispatch decisions based on existing generating capacities; (2) transmission network owners who face the same decisions for their transmission assets; (3) electricity retailers who transmit electricity to their end-use customers, subject to the amounts produced by the generation owners and transmitted by the transmission owners; (4) central planners who decide on how much capacity (and what type) to procure; and (5) reliability oversight entities.

In the U.S., the NERC establishes and enforces reliability standards [39]. Many utilities across the U.S. use a 1-in-10 reliability standard, though no broad consensus can be reached as to what a single "event" means. Common interpretations of the 1-in-10 standard includes: 0.1 hour of lost load per year, 2.4 hours of lost load per year, or one loss of load event per 10 years (independent of severity or duration) [11]. In this Thesis, we will interpret the reliability standard as limiting bulk power loss of load events to a single event of modest duration and depth to every 10 years.

Where there is a supply shortfall, demand must be curtailed involuntarily to return the system to the required balance. Planning for the appropriate amount of demand (to avoid supply shortfalls) is a bit tricky since the planning and operations of the electricity grid occurs on different timelines. In the short term, “prices” are used to establish a merit order of least-cost to highest-cost resources in order to dispatch the cheapest available resources. These prices take the form of market clearing day-ahead and real-time energy prices in restructured markets, and short-run marginal costs in vertically integrated utilities with primarily dispatchable fleets of generating plants, and help to ensure the cost-efficiency of real-time dispatch. In the long term, where investments for new generating capacities are decided upon, “prices” also exist to signal the value of providing new capacity to meet reliability criteria – Planning Reserve Margin (PRM) into the future. These prices take the form of centralized capacity market clearing prices, or proxies in the form of bilateral capacity contracts -– they convey information about how much additional generating resources are needed to meet reliability criteria in the future.

Because curtailment happens in real-time (long after investment decisions into supply have been made), and they occur somewhat at random (in that you don’t typically get to choose whether or not your electricity is shut off), there exist incentives for the under-contracting of new capacities needed to cover demand at all times into the future. Wolak (2013) coined the term “reliability externality” to explain this phenomenon [51]. Because of the reliability externality, the incentives for retailers and end-use customers are to under-purchase forward contracts to ensure reliable supply.

The current solution to the reliability externality is a capacity procurement mechanism. In California, capacity obligations are assigned to the electricity retailers (primarily distribution companies since there is limited retail supply competition), who enter in bilateral capacity contracts with generation owners to meet assigned obligations. The restructured markets in the PJM Interconnection, ISO-New England, New York ISO, and Midcontinental ISO (MISO) have a centralized capacity market for the competitive forward procurement of firm capacities to meet the forecast PRM or Installed Capacity Requirement (ICR).

In the U.S.'s only energy-only market, ERCOT, capacity procurement is incentivized through the Operating Reserve Demand Curve (ORDC) [44], which serves to provide a capacity remuneration supplement over and above pure energy market prices when operating reserves fall below specified levels needed to ensure reliability. The total price of energy in the wholesale market Locational Marginal Price (LMP) plus ORDC supplement is capped at $9,000/MWh. Prior to the events in ERCOT in February 2021, this cap was hit fewer than 10 hours per year.

It's clear that the amount of firm capacity to be procured (or the design of the ORDC curve) squarely depends on the methodology used to estimate each resource's expected firm capacity contribution for various system demand realizations. While this may be relatively straight-forward for dispatchable thermal generation (i.e., accounting for expected forced outages), it is not quite so for VRE resources whose effective generating capacity varies according to the variations in solar irradiation, wind speeds and directions, which in turn vary widely over time and location. Calculating the effective firm capacity of an energy-constrained resource like energy storage is even more complicated. While it is technically dispatchable, its availability is a direct function of its state of charge from the previous period — so actual firm capacities during stress conditions depend on the assets' mode of operation, and require highly-resolved systems analyses.

2.2 Capacity Modeling and Planning

2.2.1 Legacy Approach to Capacity Planning

Power system planning tools were developed under the Baseload Paradigm, where demand was mostly passive, and supply was dispatchable (coming mainly from coal, natural gas, and nuclear-powered generating resources). Figure 2-1(a) [12] shows the traditional merit order for economic dispatch: Nuclear and coal have high capital cost and low variable cost – thus are typically run at full capacity to fill the "baseload" component of demand. Gas-fired generation is a bit more flexible in terms of ramping

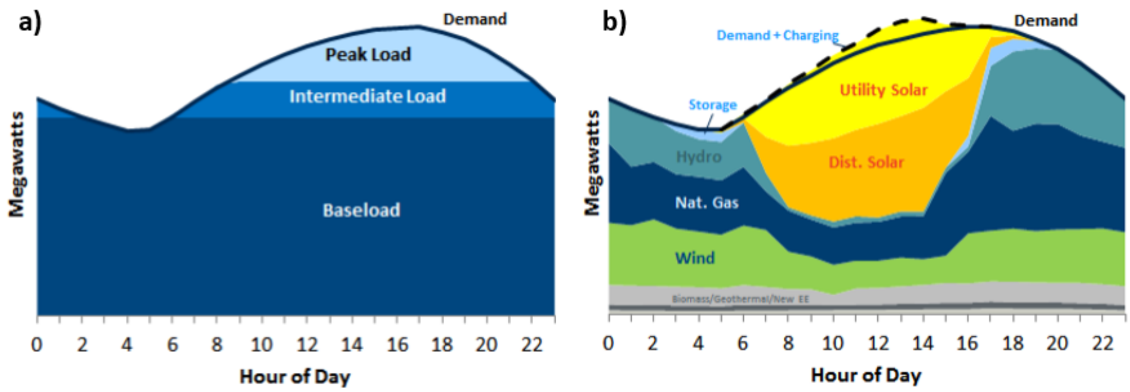

Figure 2-1: From (a) baseload to (b) flexibility

rates – thus are typically dispatched to follow load (e.g., fill in for "intermediate" load in the case of CCGTs and "peak" load in the case of OCGTs). Then, finally, VRE are dispatched when available (due to their zero-marginal cost); these are not represented in the figure due to their relatively low penetration levels.

The Baseload Paradigm lends itself well to an interpretation of the one-in-ten standard as a PRM, which is a quantity of capacity needed above the median (or 80th percentile) year peak load to meet the loss-of-load-expectation of 1-in-10 standard. The idea is that if the grid has enough dispatchable resources to meet the system peak demand (i.e.,maximum instantaneous demand plus a 15%–20% reserve margin), then it would have enough capacity to meet demand at all hours of the year except under unusual extreme conditions. Thus, if most generating capacity can be dispatched, temporal resolution and chronology would have little impact on the resulting least-cost portfolio of resources – since the expected peak demand dictates the size of the generating fleet. Considering the trade-off between more granular data and resolution and more computational and analytical capability, the legacy CEMs employed simplifications that made sense under the Baseload Paradigm.

However, such an approach is no longer sufficient as the proportion of VRE resources to total generation increases (see Figure 2-1(b) [12]). This is because solar and wind output are impacted by solar irradiation, wind speeds, and other ambient conditions that are highly variable and exogenous from the system operator's perspective. Thus,

VRE Attribute	Physical Impact	Relevance for CEMs
Variability	VRE and DF increase the variability of net load, because its availability changes through time and space based on changing weather patterns	Requires appropriate temporal and spatial resolution to capture variability and correlations with othertime-series data (eg, load)
Uncertainty	VRE and DF increase the uncertainty of net load because availability cannot be perfectly forecast at all times	Requires methods to account for adequate reserves - oftentimes, CEMs bypasses this issue by assuming perfect foresight (as we do in this Thesis)
Near-Zero Marginal Cost	VRE and DF have near-zero or zero variable production costs (relative to conventional technologies); where production-based subsidies exist, this variable cost may even be negative	Requires proper accounting of fixed costs, variable costs, and degradation costs (eg, for storage technologies)
Geographically Dispersed	VRE and DF have different capacity factors across large geographies; regional coordination may help to smooth out some of their variability and uncertainty (which are correlated with weather)	Requires appropriate representation of the transmission network, including co-optimized transmission expansion

Table 2.1: Summary of VRE and DF attributes

these resources cannot be guaranteed to generate during periods of system stress (e.g., during a hot summer day, which is typically correlated with low wind output). Thus, "Flexibility" (or the ability to quickly ramp up or down, not "Baseload") more accurately captures the type of services needed to efficiently run grid operations [38]. Likewise, on the demand side, the rapid adoption of distributed energy resources, smart appliances, and electric vehicles are making the instantaneous matching of electric supply and demand even more difficult. Table 2.1 summarizes common characteristics across these emerging technologies and their relevance for CEMs, as adapted from [13]. It's clear that next-generation CEMs need to have the appropriate temporal, spatial, and operational detail to capture the variability, uncertainty, and non-dispatchability of VRE and demand-side resources.

2.2.2 Overview of Current CEMs

British statistician George E. P. Box is often credited for stating that: "All models are wrong, but some are useful." Indeed, CEM are no exception – far from being a straightforward tool for helping decision-makers in their planning processes, they themselves embody many non-objective assumptions and worldviews that are important to parse out and recognize, to arrive at efficient and equitable outcomes.

A robust CEM should co-optimize across several interlinked power system decision layers, including: (1) investment and retirement decisions for a full range of centralized and distributed generation, storage, and demand-side resources; (2) hourly dispatch of generation, storage, and demand-side resources; (3) unit commitment decisions and operational constraints for thermal generators; (4) commitment of generation, storage, and demand-side capacity to meet system operating and planning reserves requirements; (5) transmission network power flows (including losses) and network expansion decisions; and (6) several optional policy constraints.

In practice, it may not be feasible to model all decision layers at the highest resolution, due to limitations on data availability and computational tractability. Thus, simplifications are often made depending on the nature of the problem. Figure 2-2 shows example resolutions of detail along the temporal, spatial, and network dimensions. The modeling choice to specify the level of detail/abstraction along each of these layers (or the choice to omit layers from consideration entirely) affects modeling outcomes – and it's incredibly important to understand the mechanisms and extents of such effects.

Cole et al. (2017) [13] surveyed four U.S.-based national-scale CEMs to assess inter-model differences, infer modeling best practices, and identify future research needs. The models are:

- **Integrated Planning Model (IPM)**: U.S. Environmental Protection Agency (EPA)'s IPM is often used to evaluate various emission and environmental poli-

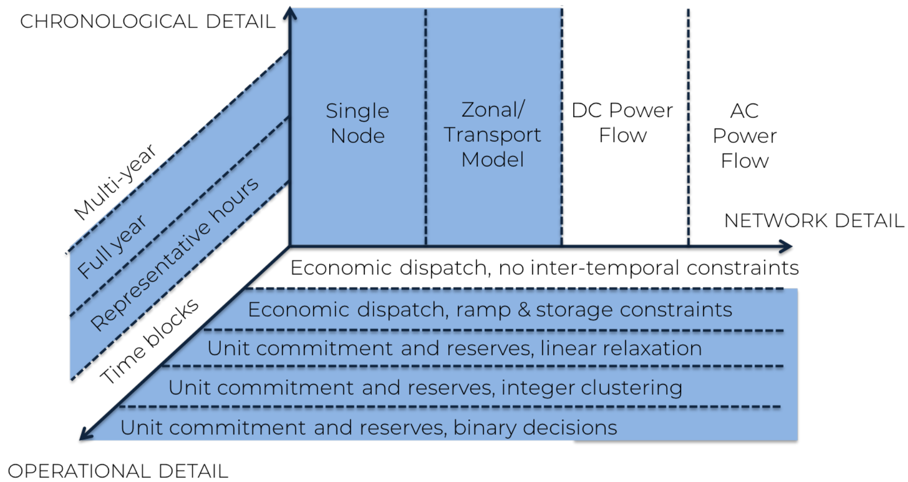

Figure 2-2: The three dimensions of capacity-expansion modeling

cies; thus it has a detailed representation of fossil-based generator technologies and associated emissions and environmental impacts.

- **National Energy Modeling System (NEMS)**: U.S. Energy Information Administration (EIA)'s NEMS is the agency's primary tool to provide projections for its Annual Energy Outlook reports (full linkage of the energy and economy sectors), which provide a baseline view of U.S. energy markets and assess impacts of future policies and market evolutions.

- **Regional Energy Deployment System (ReEDS)**: National Renewable Energy Laboratory (NREL)'s ReEDS is used to analyze scenarios of high VRE penetration levels; thus it has a highly resolved VRE representation.

- **United States Regional Economy, Greenhouse Gas, and Energy (US-REGEN)**: The Electric Power Research Institute (EPRI)'s US-REGEN is the newest of the four CEMs, and was designed to be highly customizable across temporal resolution, UC modeling, end-use demand, and linkage to rest of the economy.

Based on their survey (including modeling workshops), the authors identified three key differences in modeling spatial and temporal resolution [13]: (1) spatial resolution;

(2) temporal resolution; and (3) sampling approach.[1] I summarize these below, and expand on what why these differences matter.

2.2.3 Spatial Resolution

High spatial (along with temporal) resolution is key in capturing the variability of VRE and their correlation with load. Such modeling choices embed assumptions around resource aggregation within a region. That is, spatial resolution/aggregation affects both the detail with which transmission constraints and VRE resource quality are modeled. Lower resolution on the transmission side can lead to model results yielding lower overall system costs, because better quality (on an LCOE basis) but more remote resources can contribute to meeting intra-zonal demand with fewer transmission limitations. On the other hand, lower resolution on the VRE side can produce model results with higher overall system costs, as the best quality resources get averaged out with average- or lower-quality resources, reducing the savings associated with filling out the best sites first. These competing effects result in uncertain net impacts of spatial resolution on the build out of the system.

Selecting the appropriate spatial resolution depends on the context in which the study is used. Because ReEDS was designed to model scenarios of high VRE penetration, it has a high spatial resolution. It can represent VRE resources by 356 resource regions, while more spatially homogeneous resources are resolved into the native 134 load balancing regions. The other three models use a more generic spatial resolution – typically resolved to the modeled regions (or pre-defined smaller regional definitions).

2.2.4 Temporal Resolution

Data availability and computational limits impose a necessary trade-off between spatial and temporal resolution. Since multi-zonal models cannot typically optimize economic dispatch across a full year of operations, they instead employ time slices

[1] The authors also identify key differences in how resource adequacy and economics of production are treated in each model. I address these two topics elsewhere in the Thesis.

to account for the full year of operations. With the growing penetration of VRE resources, it's increasingly important to characterize the intra-day ramping needs associated with balancing solar generation, and inter-day ramping needs associated with balancing wind generation. To appropriate characterize these dynamics, temporal resolution has to capture both: (1) the granularity of these time slices (i.e.,the number of modeled time slices); and (2) and the coverage of these time slices (e.g., do the time slices cover a multi-hour, multi-day, or multi-week period). On the latter, the time slices need to cover a range of possible extreme events to ensure reliability over heat waves and cold snaps.

The choice of temporal resolution again depends on the objectives of the study and that data that are available. If the goal is to model a system with lots of dispatchable generation, simply representing the load changes across seasons may be sufficient. For example, IPM employs 6 stylized time blocks across the load duration curve for each of two seasons (12 total). NEMS employs 3 load-based time segments (peak, shoulder, base) per each of 3 seasons (summer, winter, fall/spring) (9 time slices). ReEDS employs 4 time slices per day per each of the 4 seasons, plus a super peak summer afternoon period (17 time slices).

If the goal were to examine the impacts of VRE on ramping and operational needs, then such an approach would not be sufficient to capture the inter-temporal dynamics of solar and wind variability. To do that, a model closer to US-REGEN, which uses a clustering method to capture both extreme events and representative hours through the year (user defined, around 100 time slices) would perform better.

2.2.5 Sampling Approach

The method for selecting the time slices can be just as, or even more important, than the selection of the temporal resolution itself. Because the time slices approach is not chronological (e.g., may fail to capture extended period of correlated low wind and high demand), there exist many challenges in interpreting model outputs. For example, a model resolved based on 17 time slices in a year may not accurately represent hourly weather and load conditions in a year (and even less across multiple years). The space in between modeled time slices thus has large implications on quantifying the role of long-duration energy storage, end-use demand response, and other technologies with strong dependencies on the state from the previous time slice.

These examples show how the temporal resolution and sampling approach can significantly impact the value of VRE to the system [13]. Insufficient temporal resolution and representation can lead to the overvaluing VRE, by underestimating their curtailment and overestimating their capacity value; and the undervaluing of longer-duration energy storage (by underestimating their flexibility and balancing services). However, because trading off temporal with spatial resolution is a necessary evil, the modeler needs to decide what dynamics are most important to capture.

2.3 Resource Adequacy Assessments

2.3.1 Effective Load Carrying Capability

So far, we have discussed the deterministic modeling for capacity expansion – this is akin to identifying the least-cost portfolio of resources needed to meet a specific demand profile (subject to various policy constraints). In reality, future demand profiles are uncertain – particularly when considering the net demand (net of VRE output, which also has inter-temporal and spatial variations). Thus, resource adequacy assessments are performed to explore a range of operating states in a predetermined system, using probabilistic methods to quantify unmet demand under a variety of scenarios involving variability in peak load, VRE production, and unplanned thermal generator outages, and correlations between these three sets of variables, and to add in enough capacity into the system to achieve the acceptable level of forced outages (frequency, depth, and duration) on average.

The determination of a resource's resource adequacy contribution involves a quantitative analysis of the effective firm capacity that can be expected to be available during times of expected peak demand. However, the analysis itself embeds non-objective assumptions that influence a priori probability distributions of load and coincident VRE and thermal generator availability. For example, basing capacity values of VREs based on historical output assumes no technological advances in efficiency. Similarly, basing capacity values of thermal generators on optimal output holds no consideration of fuel security during extreme weather events. Because of the non-transparent nature of such processes, different attitudes towards the practical, financial, and political risks of losing power on the bulk-power system level [8] could affect the development of the time series data inputs, and thereby exert over-sized influence over the model results.

Effective Load Carrying Capability (ELCC) is the industry's best-practice approach to estimating the resource adequacy contribution of a technology. First introduced as a methodology in the 1960s, ELCC is now used in CAISO, MISO, PJM, etc to calculate the firm capacity contributions of VRE and energy-limited resources [47, 50]. In simplest terms, the ELCC is a stochastic approach that considers many realizations of load, VRE output, and thermal outages [24]. Assuming an equivalent "perfect" dispatchable load that is available 24/7, the approach calculates how much of this resource is needed to restore the system to the same level of reliability after introducing a new resource [21]. For example, if 100 MW of solar can be replaced by 30 MW of perfect firm capacity, then the solar's ELCC would be 30%.

Computing the ELCC of a resource is highly data-intensive, and the value needs to be updated often depending on the state of the rest of the system. This is because resources do not operate in a vacuum, but rather have synergistic or antagonostic synergies with other resources on the grid (thereby affecting their overall value to the system) [47].

2.3.2 Metrics of System Reliability

At the system level, a common metric of reliability is the loss of load probability (LOLP). This is a measure of the probability that a shortage event (a period of

	Frequency	Duration	Magnitude
LOLH	No	Yes	No
LOLE	Yes	Yes	No
LOLP	Yes	Yes	No
EUE	Yes	Yes	Yes

Table 2.2: Summary of the described reliability risk metrics

consecutive hours of load shedding) will occur as demand exceeds firm capacity availability in the system in a given period. Related LOL metrics are the Loss of Load Expectation (LOLE), which is the number of days of outage in a period, and the Loss of Load Hours (LOLH), which is the number of hours of outage. Importantly, the LOL metrics do not inform of the frequency or magnitude of shortage events, only of a measure of their combined duration (Table 2.2). Frequency is the count of the number of loss of load events over a period of time (assuming that events don't last more than a day). Duration is the length of time of a loss of load event.

An alternative reliability metric is the expected unserved energy (EUE), which is the amount of electricity demand in MWh that is expected not to be met by generation in a given time period. This reflects the combined frequency and magnitude of the shortage events (Table 2.2). However, it does not distinguish between a large number of small events vs. few numbers of large events, which has significant impacts on the type of response needed. For example, if the magnitude were large, it is hard to implement rolling blackout that allow the scarcity to be "shared" across many customers. This was an issues in the ERCOT shortfall event in February 2021, so that many customers ended up being blacked out for 70 hours straight.

2.3.3 Economically Efficient Level of Reliability

Reliability issues can be avoided by over-building the system – such that there is always excess supply to absorb unexpected increases in demand. However, least-cost considerations have to balance out the costs of additional installed capacities against the costs of incurring bulk-power system level outages. This is where the Value of

Lost Load (VoLL) concept comes into play, where we approximate how much economic value end-use customers attach to having reliable power.

Because of our relatively limited exposure to large-scale outages caused by supply shortages in the U.S., our ability to relate to such events are informed by our experiences with distribution-level outages (more frequent, generally less severe). Understanding the various dimensions (frequency, duration, and magnitude) of shortage events is crucial in informing appropriate responses to such events. For example, there is a clear difference between losing power for a few hours in a residential home, when the occupants are out, vs. losing power for a few days at a critical services facility. Breaking down reliability metrics into their separate components has direct consequences on how customers perceive these risks, and consequently, what investments they are willing to make to avoid such outages.

Baik et al. (2020)'s study on U.S. residential customers' willingness to pay (WTP) for resilience to large and long-duration (LLD) power outages [2] highlighted three main shortcomings of previous elicitation methods used in WTP studies: (1) they do not ensure that respondents fully understand the implications of large and long-duration (LLD) outages (where few people have experienced or thought much about LLD outages); (2) they focus on brief outages that last only a few hours, and; (3) they do not consider partial back-up services even though "there exists a considerable amount of consumer surplus for small amounts of electricity". Filling this gap in the WTP (and relatedly, VoLL) literature thus requires better characterization of how truly bad the shortage events are likely to be.

2.4 Implications on Investments and Operations

Recent inter-model comparison studies have qualified the impacts of the model simplifications discussed above. Mallapragada et al. (2018) [36] compared installed capacity and operational metrics under a CEM with a time slice representation (akin to seasonal averages) to a CEM with retained chronology (using 12 representative days).

The authors found that lower temporal and operational resolution overestimated solar PV capacity (by 35% in one case), and underestimated wind and the supporting NG capacity requirements. This is partly because reducing the temporal resolution leads to an undervaluation of the positive correlation between wind and load, and an overvaluation of solar generation during system peaks. The differences in the capacity mix have reliability implications, as the CEM with retained chronology results in lower unserved energy when tested in an hourly simulation of annual grid dispatch and operations. Similarly, Pina et al. (2011) [45], Mai, Barrows, et al. (2015) [33], Poncelet et al. (2016) [46], Bistline et al. (2017) [4], and Cole, et al. (2017) [13] showed that limited temporal resolution produces inaccurate resource mixes in a high VRE system. Thus, increased temporal resolution and preservation of chronology are needed to enable an improved characterization of temporal variability of VRE generation, load, and the inter-temporal dynamics of thermal generators and energy storage technologies.

In addition to high temporal resolution, studies have shown that the inclusion of multiple weather years serves to better represent resource needs and costs. Zeyringer et al. (2018) [52] compared the performance of the CEM optimized over 2001-2010, each year at a time vs. the whole ten-year period together, and found that if the U.K. power system were planned on the basis of one weather year, it could lead to severe supply shortages and failures to meet long-term carbon reduction targets. They found that in an 80% renewable system, up to 5% of annual demand (33 GW) could be unmet if basing the investments on a single year. A recent decarbonization study from Brown and Botterud (2021) [7] showed that inter-annual variability can lead to large differences in average System Cost of Electricity (SCOE) — one-year solutions can vary up to 2x depending on the weather year – though this effect can be smoothed out be more geographic aggregation.

A common theme is that spatial and temporal resolution in modeling is highly important for high-VRE systems. If not done correctly, model simplifications will lead to non-optimal system build-outs, with an over-building of baseload capacity (e.g., dispatchable power, overbuilding of VRE), and under-building of flexible capacity (e.g., energy storage, demand response). Long-duration energy storage is particularly sensitive to the selection of temporal resolution. Because most of the value is in shifting energy from a period (e.g., on the order of days, weeks) to another, insufficient temporal resolution will under-value the benefits of such a technology. On the operations side, a fleet based on these non-optimal capacities will lead to: (1) an over-reliance on gas-powered capacity (increasing both fuel costs and emissions in actual operations) and (2) an under-reliance on clean alternatives (such as energy storage and demand response). The lack of flexible resources available to meet sharp ramps throughout the day will likely cause reliability issues on the bulk power system.

Additionally, the current processes for capacity and resource adequacy planning occur somewhat independently: (1) CEMs target an PRM informed by the reliability standard (in a non-transparent analysis); and (2) ELCC analyses take as given the system – with no way of co-optimizing for new capacities and load shedding. Moreover, ELCC for RA purposes typically focuses on a single resource, ignoring its effect within a broader system (where the penetration of similar resources will impact the resource's value). This approach precludes from the simultaneous optimization of capacity investments and resource adequacy procurements, and the time delay can lead to either over-investments in antagonistic capacities, or under-investments in synergistic resources.

This is why, we propose an integrated model that optimizes capacity deployment, while considering issues of reliability. This allows us to assess the robustness of the hourly grid operations approximated by the CEM, and provide a transparent view into: (1) the number of unserved energy events (frequency); (2) how long each of the events occurred for (duration); and (3) how much energy was shed during each event (magnitude). We describe this process in further details in Chapter 3.

Chapter 3

Experimental Approach for Optimizing Reliability within CEMs

Power system planning tools seek to minimize total system costs while adhering to an "acceptable" level of involuntary load shedding or "blackouts". One way to do this is to define a reserve margin above expected system peak load, to ensure that sufficient capacity is available to serve load at all times. However, as discussed in Chapter 2, this approach is quickly becoming irrelevant in a system dominated by intermittent VREs and energy-limited resources (e.g., energy storage, demand flexibility). We discuss below the iterative approach we have adopted for the MIT Future of Storage study. Much of the text below comes from Chapter 8 of the upcoming publication, of which I am a co-author.

3.1 An Introduction to GenX

This Thesis uses an open-source capacity expansion model (CEM), GenX that takes the perspective of a welfare-maximizing centralized planner to determine the cost-optimal generation, storage and transmission investments needed to meet a pre-defined system demand, while adhering to various grid operational constraints, resource availability limits and other imposed policy/environmental constraints. Notably, GenX incorporates a detailed temporal resolution of power sector operations, either based on modeling representative periods or one or more years at an hourly resolution, depending on the model configuration. As discussed previously, the increased spatial and temporal resolution and preservation of chronology enables an improved characterization of temporal variability of demand, VRE generation, and the inter-temporal dynamics of various generators and energy storage technologies.

The major grid operating constraints activated in GenX for this Thesis include: (1) demand and supply balance for each time step at the zonal level, considering inter-zonal imports and exports as well as the option of shedding load in each zone at a value of lost load (VoLL) equal to \$50,000/MWh. Such a high VoLL was chosen to minimize instances of load shedding and incentivize addition of more capacity to meet demand within the energy-only market framework implemented in the model. (2) linearized unit commitment (start-up/shut-down) decisions, minimum up/down times and hourly ramping limits for thermal generators, (3) transmission capacity limits and linear line losses, if applicable (4) inter-temporal constraints governing storage state-of-charge and capacity constraints on maximum hourly charge/discharge and stored energy, and (5) zonal and site-level VRE resource availability limits in each time step.

To model system evolution to meet the decarbonization targets mentioned previously, we include constraints to enforce upper limits on annual average CO_2 emissions intensity that accounts for generation and storage discharge as well as storage losses. Further details on the model formulation can be found elsewhere [30], along with prior publications using GenX [23] and the open-source model itself, available at: `https://energy.mit.edu/genx/`.

While most CEM studies tend to overlook the operational impacts of their capacity mix projections, we focus here on solving for reliability – that is, arriving at a least-cost capacity mix that ensures a minimum level of unserved energy across multiple realizations of coincident load and VRE output. Rather than developing our own projected distributions of load, VRE and thermal generator availability, we use seven years of hourly VRE capacity factors (based on historical weather conditions) to demonstrate a range of plausible grid conditions. The non-coincidence nature of our load and thermal availability data assumes that the time series are independent of one another.

3.1.1 Selection of Model Regions

Our modeling work focuses on two regions in the U.S. in 2050: the Southeast and the Northeast. Rather than seeking to provide detailed projections of the evolution of the resource mix in these regions, which is affected by turnover of the existing generation fleet, market design, state incentives, permitting rules, etc., we focus on the effects of differences in VRE resource quality and the availability of long-lived existing low-carbon hydro, nuclear, and pumped hydro storage assets, assuming cost-efficient investment and operation. The two selected regions differ across several key factors that affect the potential costs and benefits of achieving various decarbonization goals, including: (1) wind speeds and solar irradiation, land availability and resulting installed costs of wind and solar generation, (2) hydroelectric resources, and (3) market constructs and their implication for nuclear power development. We also assume that the existing stock of fossil generating capacity retires by 2050, so that our analysis basically examines a "greenfield" system developed to meet 2050 demand, with some regional differences (as detailed below).

The **Southeast** (Tennessee, Alabama, Georgia, the Carolinas, and Florida) is characterized by the presence of regulated, vertically integrated utilities, a winter-peaking system for some regions, and an extensive nuclear generation fleet, which contributed 28% of power generation in the region in 2018. Unlike in restructured electricity markets, where nuclear plant economics have recently been adversely impacted, reliance on vertically integrated utilities in the U.S. Southeast combined with greater public acceptance of nuclear energy makes it more likely that nuclear plants in this region will apply for, and be granted, a second license renewal that extends their remaining life, for a total unit life of up to 80 years. Nuclear, as a low-carbon dispatchable resource, could mitigate the need for VRE resources and storage technologies and has the potential to lower the system cost of deeply decarbonized grids [9, 48]. The Southeast region is also endowed with relatively good quality onshore wind and solar resources.

The **Northeast** (New England and New York) is characterized by strong legislative and regulatory support for renewable generation, offset by difficulty in siting that in some cases translates into increased infrastructure costs. Most of Northeastern states have pledged to reduce their economy-wide greenhouse gas (GHG) emissions by 80% by 2050, with a few states committing to more stringent targets. The region has relatively low-quality solar, but high-quality onshore and offshore wind. There are also non-trivial hydro imports from Canada, as well as domestic hydro resources that can support VRE integration. While the Northeastern electricity demand profile peaks in the summer currently, the high electrification of space heating anticipated may transform this into a winter-peaking region [35, 49]. Due to their poor economic viability and political stigma, we assume that all existing nuclear units retire by 2050.

3.1.2 Cost and Operational Data Inputs

We use the latest mid-range EIA fuel price projections for 2050, and NREL Annual Technology Baseline 2020 (ATB) [41] for characterizing capital cost of various generation technologies as well as Li-ion battery storage. For the 2050 demand projections, we use the High Technology Adoption – Moderate Technology Advancement ("High Electrification") scenario from the NREL Electricity Futures Study (EFS) study [34]; these projections assume a high degree of electrification in space heating and transportation and are based on the weather year 2012.

As noted above, we conducted (mostly) greenfield modeling, to reflect the fact that most existing generation capacities will have to be replaced for either economic or policy reasons by 2050. Thus, we restrict investment to the following technologies:

utility-scale solar, onshore wind as well as offshore wind and distributed solar in one instance (the Northeast), natural-gas fired plants (OCGT, CCGT) with and without amine-based carbon capture and storage (CCS) technology, and hydro resources where they play a major role.

For high spatial and temporal resolved representation of PV and wind resources, we followed (Brown and Botterud 2021)'s approach [7] in developing supply curves of available land area for PV and wind development, excluding water bodies, national parks, urban areas, mountain ranges, and Native American territories from development, and in quantifying the interconnection cost of spur lines to connect new VRE generation to existing transmission infrastructure. For each site, the hourly capacity factor (CF) for PV is simulated using a horizontal one-axis-tracking PV over 2007–2013 satellite data from the National Solar Radiation Database (NSRDB) [42]; and the hourly CF for wind is simulated using climate reanalysis data from the WIND Toolkit [15] and manufacturer power curve data for the Gamesa: G126/2500 turbine at 100-meter height [20]. Different quality bins of VREs (based on the levelized cost of energy, considering generation and interconnection costs) are developed by aggregating over these individual sites. Further details about the modeling data inputs and assumptions are provided in Appendix A.

3.2 Time Domain Reduction

3.2.1 Selection of Time Slices for the TS-CEM

Capacity Expansion Models (CEMs) rely on a compact temporal, spatial, and network representation of the system to maintain computational tractability. Traditional CEMs have relied on a time slices approach that is usually based on disaggregating the load duration curve based on seasonal and time-of-day blocks. The intuition of the "time-slice" approach is to represent conditions of system peak, based on the idea that if there is sufficient generating capacity to cover the peak, then reliability can be ensured at all other times too.

The drawback of employing a TS-CEM in a high VRE power system is that the low temporal resolution and operational detail fails to adequately capture fully the temporal and spatial dynamics of a highly decarbonized electricity grid. Specifically, with increasing VRE penetration, the system peak “net load”, i.e. residual load after accounting for VRE generation, is likely to more important than the system peak load may for resource planning purposes. That is why we consider different instantiations of a chronology-based CEM (C-CEM).

3.2.2 Selection of Representative Periods for the C-CEM

For the chronology-based CEM (C-CEM), we utilize an iterative hybrid approach to select the representative periods for modeling annual grid operations. But before we get into the clustering and identification of extreme periods, we need to choose: (1) the number of representative periods to cover the variability of seven weather years; and (2) the number of hours in each representative period. While the first part targets the range of variability captured (e.g., 25 periods is more than 10 periods), the second part targets the preservation of chronology (e.g., a representative week captures longer sustained correlations than a representative day). Together, these two parameters determine the temporal resolution assumed in the model. While technically exogenous to the clustering approach, the selection of these parameters have large implications for the resulting model outcomes – as we will see in the next Chapter.

3.2.3 Iterative Hybrid Approach to Time Domain Reduction

The main contribution of this Thesis is the development of an objective approach to identify representative periods for use in C-CEMs. As discussed in Chapter 2, the use of representative periods is an improvement over the time slice approach, since it is based on prevailing variability across all time series (not just load), and it also allows for preserving chronology in operations as well as inter-period energy transfer in the case of energy storage.

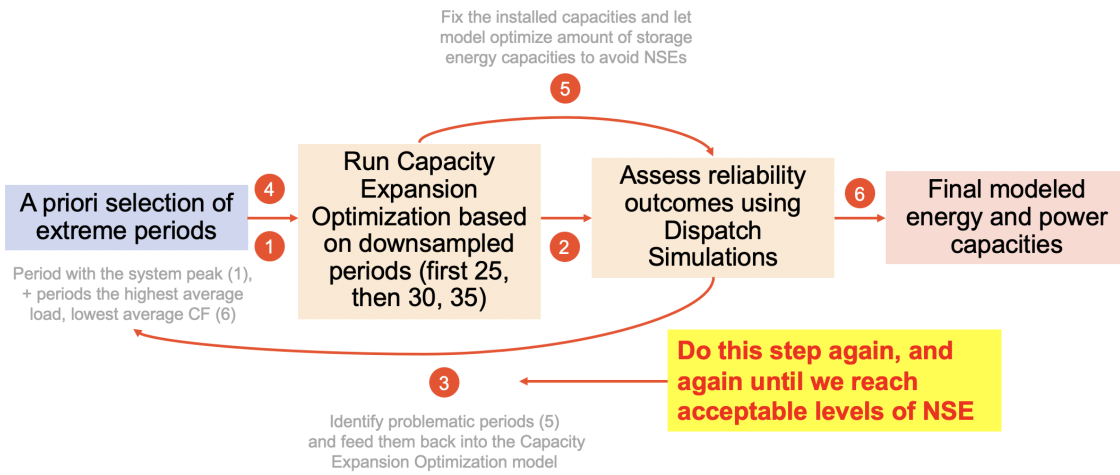

Figure 3-1: Iterative CEM-reliability simulation approach

The hybrid clustering employed here to select representative periods is adapted to provide both sufficient temporal resolution and extreme weather coverage. While the clustering procedure seeks to closely approximate the underlying temporal distributions of historical load and VRE capacity factor profiles, the extreme periods selection procedure seeks to incorporate sufficient "reliability" events corresponding to extended periods of low VRE output and high demand (e.g., heat waves, cold snaps). We outline our iterative approach to selecting the periods used in the model below (see Figure 3-2).

First, we slice the zonal load and VRE capacity factor data into n-day periods. For each period, we calculate average load, solar CF, and wind CF; and identify some "a priori extreme periods", as defined as those that have the highest system peak, highest average load, and lowest PV and onshore wind output ("a priori extreme periods") at a zonal level. We then stitch together the time series of all resources (solar, onshore wind, hydro) for each period, to create a single concatenated time series for each 10-day period; thus, each vector is of length: 24 hours/day x 10 day/period x 4 time series (load, solar, onshore wind, hydro) x number of modeled zones. Due to overlap between the extreme periods meeting the above criteria, we identify 8 extreme periods for the Southeast through this process (and 6 for the Northeast).

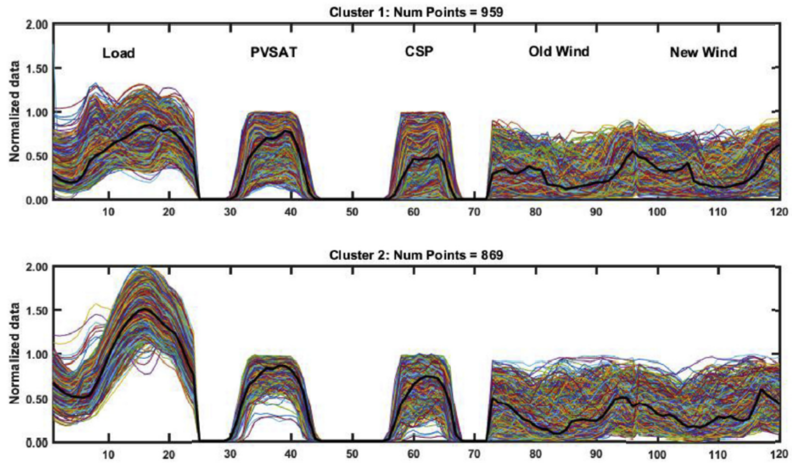

Figure 3-2: Example normalized clustered time series data

Second, we employ the k-means clustering technique to group the remaining (non-extreme periods, 255.5 - number of extreme periods) into 25 clusters. Figure 3-2 shows an example of such normalized time series data [36]. For each cluster, we select the historical period closest to the centroid of the cluster as the most representative period. This is because the centroid of the cluster may not reflect actual conditions that may exist in the system; thus, the most representative periods are indeed based on actual data. We then weigh each representative period based on cluster size, to achieve a total weight of 8,760 hours to approximate annual grid operations. To preserve the system peak, we did not scale the weighted time series to match the annual load in the original data; this results in a 2–3% increase in annual load relative to the NREL EFS data.

Third, the iterative process starts by inputting the C-CEM outputs into a simplified Production Cost Model to simulate overall reliability under the proposed portfolios. This process is more effective at identifying periods of great reliability importance, which were not already flagged to be an “a priori extreme periods” or “representative periods”. We call the periods causing the most reliability issues (i.e., frequent and long-lasting non-served energy events), “reliability periods”. We repeat this last step one or two times to ensure that we’re optimizing for system reliability at each hour (i.e., no significant load shedding due to capacity shortages).

Chapter 4

Impacts of Co-Optimizing Capacity and Reliability

In this Thesis, we seek to improve upon bulk-power system reliability of capacity mixes identified by optimal CEMs. Since we're looking at a future power system in 2050, which is drastically different from the power system we have today, it is not sufficient to apply historical intra-day variability in load, VRE, and thermal generator availability to the future system's load (as would be done in ELCC-type analyses). Instead, we focus on the broader trends in inter-annual variability, captured through our use of VRE capacity factors simulated over seven years of weather data. By testing the capacity mixes optimized for a simplified representation of the system (i.e.,through reducing the temporal resolution and chronology of the load and VRE availability data) over the whole seven years of "actual" VRE availability and load, we can see just how well a highly decarbonized power system deals with extended periods of low solar and wind output.

We consider the following formulations of the CEM's representation of the time series data (Table 4.1). The numbers in the scenario name reflect the number of periods used to characterize the time series data. TS-CEM refers to the time-slice approach, C-CEM refers to the chronology-based approach, and S-CEM is the variant on C-

Scenario	Coverage	Links to Next Period	Temporal Representation
TS-CEM	2012	No	Seasonal averages (4 seasons, 4 times of day)
S25-CEM	2012	Yes	25 representative periods, selected over 1 year
C25-CEM	2007-2013	Yes	25 representative periods, selected over 7 years
C30-CEM	2007-2013	Yes	25 representative periods + 5 reliability periods, selected over 7
C35-CEM	2007-2013	Yes	(30-n) representative periods + (5+n) reliability periods, selected

Table 4.1: CEM temporal representation assumptions

CEM considering only one single weather year. We do not explicitly evaluate the time slice approach (TS-CEM) discussed in Chapter 3, as the TS-CEM does not account for the inter-temporal dynamics of unit commitment, ramping, and energy exchange (e.g., storage state of charge).

The analysis starts with a simple chronology-based approach that selects 25 periods across one single year of weather data (2012) to resolve the optimal capacity mix (S25-CEM). Then, we consider a suite of chronological C-CEMs resolved over seven years of weather data, while incrementally increasing the number of selected periods (therefore, number of hours the model "sees" to make its investments and dispatch decisions).

4.1 Impacts on Systems without Long-Duration Energy Storage

We define our "Base Case" as one in which only today's commercially available technologies are available. As discussed above, we had already made the simplification of assuming mostly greenfield development (except for hydro resources, transmission capacity in both regions, and nuclear capacity in the U.S. Southeast). This means the model can choose to build new generation and storage capacities across: Li-ion battery storage, wind and solar generating capacity (VRE), and NG with and without carbon capture and storage (CCS). Long-duration energy storage is not considered in this set of experiments, as none of the technologies are commercially viable at scale yet. However, these emerging technologies with long-duration storage attributes are going to become increasingly important for balancing the grid on a longer time scale, and as we will see in the next section, modeling these long-duration energy storage technologies are particularly prone to the types of biases discussed in Chapter 2 (e.g., under-estimating system need for technologies with inter-temporal dynamics).

In this Thesis, we model three emission constraints, 0 gCO_2/kWh, 5 gCO_2/kWh, and "no limits" (NL). To contextualize these average emission intensity levels, 0 gCO_2/kWh serves as a strict definition of "net zero" scenarios, where we do not consider negative emission technologies. Thus, wind, solar, and storage technologies (as well as nuclear in the Southeast) have an expanded role to provide capacities around the clock, including through the over-building of VRE resources. 5 gCO_2/kWh is about 99% in the U.S. Southeast (387 gCO_2/kWh), and 98% decarbonization relative to today's generation mix in the U.S. Northeast (249 gCO_2/kWh). Finally, the NL case represents what would happen even if no emission constraints were applied (though this is highly likely). The investments in VRE under this edge case are solely driven by "least cost" economics ignoring the social cost of carbon emissions. We think that the 0 gCO_2/kWh case is appropriate for identifying problematic periods in the reliability simulations; the 5 gCO_2/kWh case works well for making planning decisions; and the NL case serves as a clear benchmark for comparing installed capacities and system costs.

To reiterate, we run GenX in two stages: (1) as a capacity expansion model with limited temporal resolution; and (2) as a production cost (i.e., dispatch) model with only operational decision variables (i.e., capacities fixed to the levels optimized for in the first stage). For the results in the first stage to make sense, we need to ensure that reliability is sufficiently covered in the second stage – hence the iterative approach we proposed in Chapter 3. This section follows the same order, showing first the reliability results, then the final installed capacities as presented in the FoS study.

4.1.1 Reliability Key Findings

As expected, a capacity-expansion model based on one year of weather data (S25-CEM) yields a lower level of reliability than one based on multiple years of weather data (C25-CEM), particularly around extended periods of low VRE output (particularly that of wind, since solar follows a fairly consistent diurnal pattern). As discussed in Chapter 2, we need to understand not only the total amount of involuntary load shedding (implemented to keep electricity demand equal to available supply, or Non-Served Energy (NSE)) across a period of time, but also how the individual events break down in terms of their duration (e.g., consecutive hours of NSE) and magnitude (e.g., as a proportion of total system load during each hour).

Figure 4-1(a) shows the largest NSE event (red) under the capacity mix resolved for one year of weather data (S25-CEM). Due to an extended period of low wind output (green), the total amount of firm power capacities (charged Li-ion and existing nuclear) are running flat out, until the energy-limited Li-ion resource gets depleted (resulting in large amounts of NSE).[1] While there is non-zero VRE contributions during the day, this is not enough to fully recharge the Li-ion resources throughout the region, leading to a single NSE event lasting for 36 consecutive hours, and shedding 1,356 GWh of load in total. At the hourly level, the NSE event is shedding 0.5% to 59% of load on a regional basis (Table 4.2). Clearly a 59% shedding of hourly load is of immense magnitude – of even greater magnitude than the February 2021 ERCOT outage (which shed up to a third of winter peak on an instantaneous basis).

By optimizing over a larger and more diverse set of weather conditions (C25-CEM), the NSE event triggered by lower VRE output is slightly improved. Figure 4-1(b) shows that with more visibility into diverse weather conditions, the model at 0 gCO_2/kWh deploys 3% more VRE, 17% more Li-ion power capacity (though 15%

[1] There is no gas with or without CCS in the 0 gCO_2/kWh case.

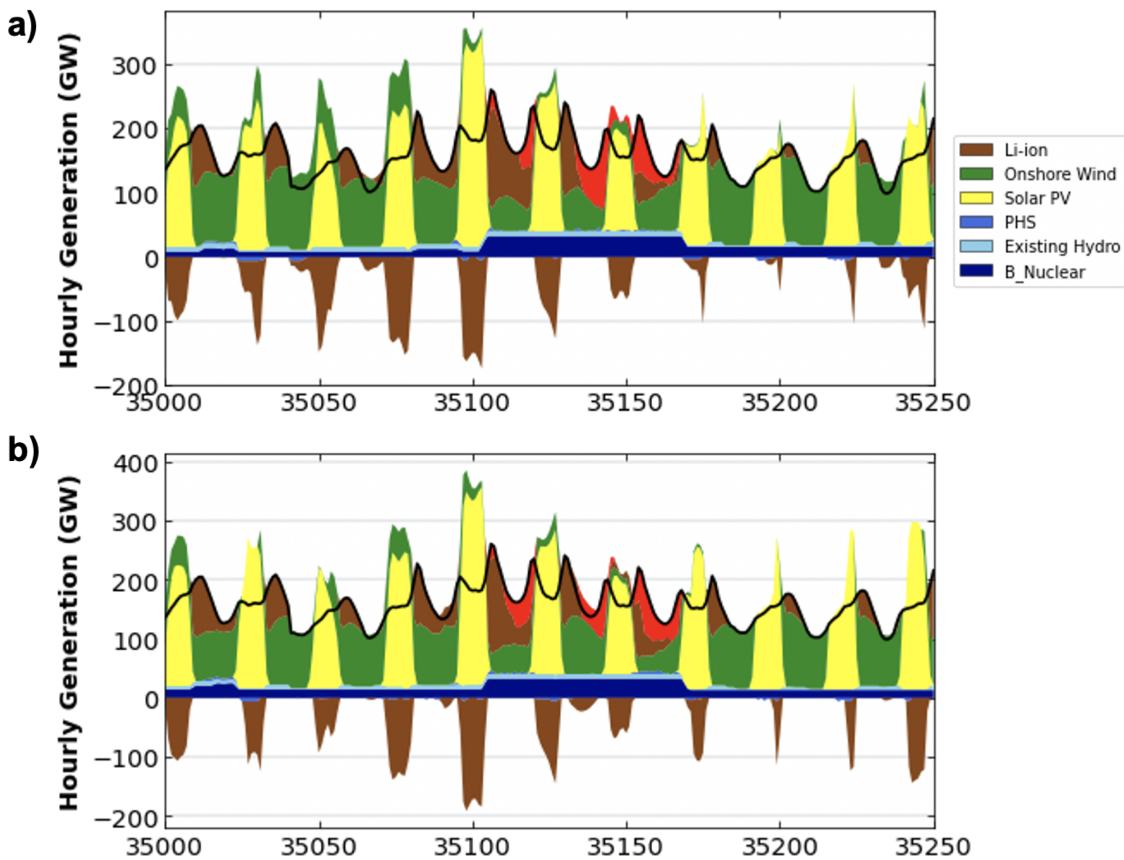

Figure 4-1: Example hourly dispatch across a selected period of system stress in the U.S. Southeast: (a) S25-CEM; (b) C25-CEM

SE_S25

ID	Starting Hour	Total GWh	Duration (Hours)	Min Hourly %	Max Hourly %
1	t35133	1,356	36	0%	59%
2	t52620	929	21	1%	45%
3	t17608	834	18	1%	55%
4	t17584	827	20	2%	51%
5	t58048	677	15	12%	35%
6	t27305	553	16	3%	37%
7	t13721	522	14	6%	41%
8	t52673	447	17	2%	36%
9	t25409	425	14	6%	33%
10	t17200	318	16	2%	27%
Total		**6,887**	**187**	**0%**	**59%**

SE_C25

ID	Starting Hour	Total GWh	Duration (Hours)	Min Hourly %	Max Hourly %
1	t17585	904	18	3%	39%
2	t27293	900	28	1%	38%
3	t52623	812	18	4%	40%
4	t35152	738	17	1%	43%
5	t17608	625	17	1%	44%
6	t25409	545	14	11%	36%
7	t35110	453	11	6%	35%
8	t35135	422	12	3%	40%
9	t58050	419	13	2%	31%
10	t1006	409	11	1%	37%
Total		**6,226**	**159**	**1%**	**44%**

SE_C30

ID	Starting Hour	Total GWh	Duration (Hours)	Min Hourly %	Max Hourly %
1	t44762	142	9	0%	21%
2	t44776	122	6	7%	18%
3	t25181	24	2	4%	14%
4	t13675	5	1	2%	2%
5	t58578	3	1	1%	1%
6	t44754	3	2	0%	1%
7	t40507	1	1	0%	0%
8	t52629	0	1	0%	0%
9	-	-	-	-	-
10	-	-	-	-	-
Total		**301**	**23**	**0%**	**21%**

SE_C35

ID	Starting Hour	Total GWh	Duration (Hours)	Min Hourly %	Max Hourly %
1	t13732	20	2	6%	8%
2	t40507	1	1	0%	0%
3	t17611	0	1	0%	0%
4	t52630	0	1	0%	0%
5	t44782	0	1	0%	0%
6	-	-	-	-	-
7	-	-	-	-	-
8	-	-	-	-	-
9	-	-	-	-	-
10	-	-	-	-	-
Total		**20**	**6**	**0%**	**8%**

Table 4.2: Top 10 hours of system stress under each CEM configuration

less energy capacity), and importantly 11% more transmission capacity to allow for imports during periods of system stress. The largest single NSE event now lasts for 17 hours, shedding a total of 738 GWh, which is an improvement over the NSE event of 1,356 GWh lasting 36 hours – but we can do better.

Repeating this analysis for C25-CEM yields a list of periods with low reliability, as can be seen in the top 10 events of system stress in Table 4.2. C30-CEM (and by similarly C35-CEM), iteratively adds these problematic periods, so that the capacity-expansion model can make investment decisions after taking into consideration these exogenous extreme periods. Thus, C30-CEM shows no more reliability issues for the selected time period.

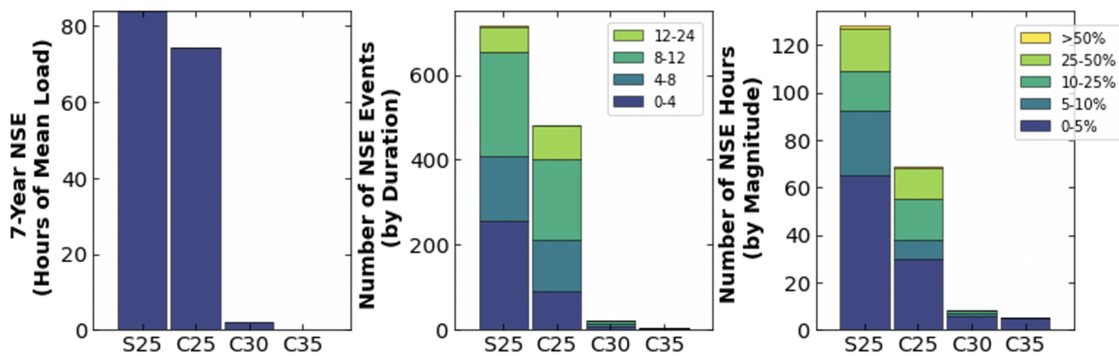

Figure 4-2: Reliability results for the U.S. Southeast: Without LDES

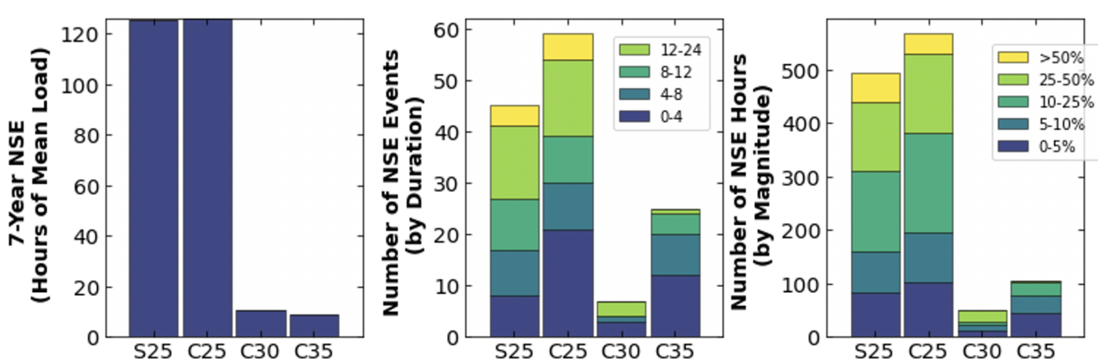

Figure 4-3: Reliability results for the U.S. Northeast: Without LDES

Running the reliability simulations allow us to quantify the benefits of increasing the temporal resolution of the CEM, as well as the incremental impact of adding in key extreme periods. Table 4.2 shows the top 10 events of system stress, for the U.S. Southeast, under the 0 gCO_2/kWh emissions case, across the four selected model specifications. Rows highlighted in yellow indicate the example described above. The inclusion of chronology, weather years, and particularly that of extreme reliability periods serve to incrementally reduce the total amount of load shed during the reliability simulations, as well as their duration and magnitude.

Similarly, Figures 4-2 and 4-3 show the breakdown of all NSE events. We see that for the Southeast (Figure 4-2), moving from the S25-CEM to the C25-CEM configuration decreases total NSE by 12% (or 1,535 GWh) across the seven weather years considered. Moving from S25-CEM to C25-CEM also decreases the total duration of such events by 33% (or 234 hours). And of note, we point to the decrease in frequency of NSE events that last for more than 4 hours (shown in different shades of green), as well as the frequency of NSE hours with more than 10% of total system load.

The reliability outcomes for the Northeast (Figure 4-3) are a bit different in that the the frequency of outage events actually seems to increase between S25-CEM and C25-CEM, and C30-CEM and C35-CEM. However, upon closer inspection, we can see that the number of severe events (i.e., with duration longer than 12 hours, or shedding more than 25% of hourly load) actually decreases, which is what we would've expected. In general, because the U.S. Northeast does not have other forms of firm capacity (e.g., existing nuclear) and has poorer VRE resource quality, we'd expect to see more instances of NSE events, relative to the Southeast.

In both regions, most striking is the large improvement in reliability that comes with the exogenous addition of extreme periods into the initial capacity expansion optimization problem, which allows the model to "see" these extended periods of low VRE output that are particularly prone to reliability issues. Moving from S25-CEM to C35-CEM decreases total NSE by almost 100% in the Southeast and by

93% in the Northeast. It also decreases the total duration of such events by 99% in the Southeast and by 78% in the Northeast. Thus, we have shown that using the proposed methodology, we are able to dramatically reduce the frequency, duration, and magnitude of resulting NSE events.

4.1.2 Planning Implications

Unlike in the reliability simulations where we want to stress test the system at 0 gCO_2/kWh (since this means there is no thermal generating capacity), we emphasize findings of the 5 gCO_2/kWh case for planning purposes, as they better reflect likely scenarios, where gas will still play a small but an important role in balancing supply and demand as we transition to a low-carbon electricity system.

Figures 4-4 and 4-5 summarize the key modeled system outcomes for scenarios with tightening CO_2 emissions limits across U.S. Southeast and Northeast. From left to right, we show the modeled optimal portfolio mix: (1) installed power capacities (relative to the region's 2050 peak electricity demand); (2) deliverable storage energy capacity to the grid (i.e., product of energy capacity and discharge efficiency, relative to the region's annual electricity demand); and (3) System Cost of Electricity (SCOE). SCOE per unit of delivered energy include: total annualized investment, fixed O&M, operational costs of generation, storage, and transmission, and any non-served energy penalty. System impacts can be observed in the trade-off between technology-level installed capacities and system costs, and between storage capacities and VRE curtailment.

We see that across the different model configurations, there are small changes to the optimal amount of installed gas generators, VRE resources, storage, and transmission capacity (transmission not shown in the summary figures) – all of which depend on the selected region's resource availability and load variations. These changes are, moreover, associated with small overall effect on the resulting system's SCOE (even leading to a lower SCOE in the Northeast). This suggests that we can seek to improve on reliability outcomes without incurring massive "over-building" investment costs.

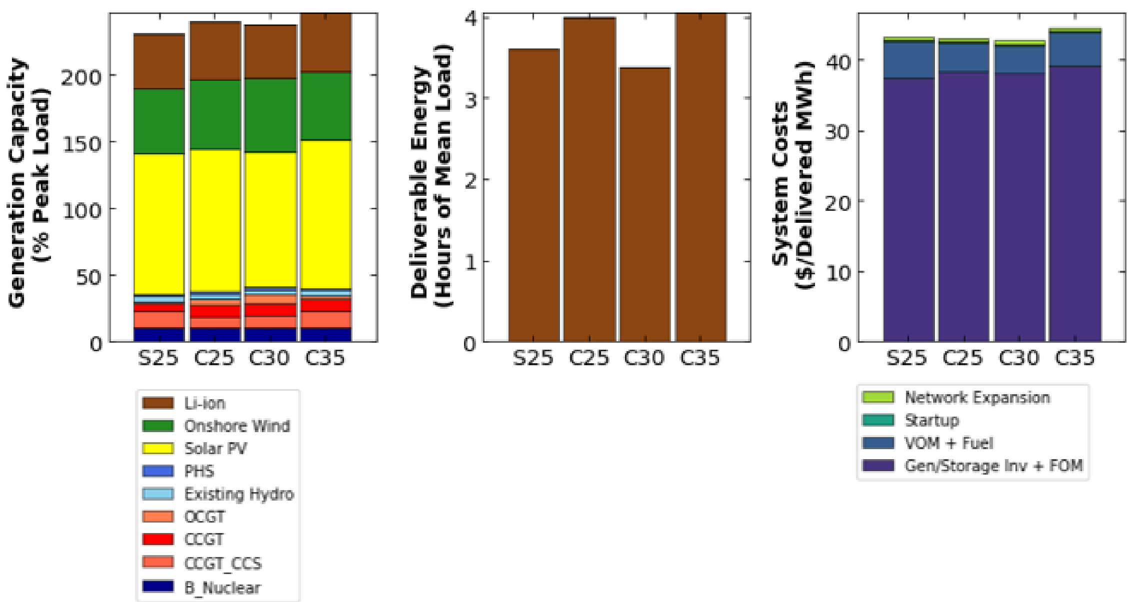

Figure 4-4: Capacity-expansion results for the U.S. Southeast: Without LDES

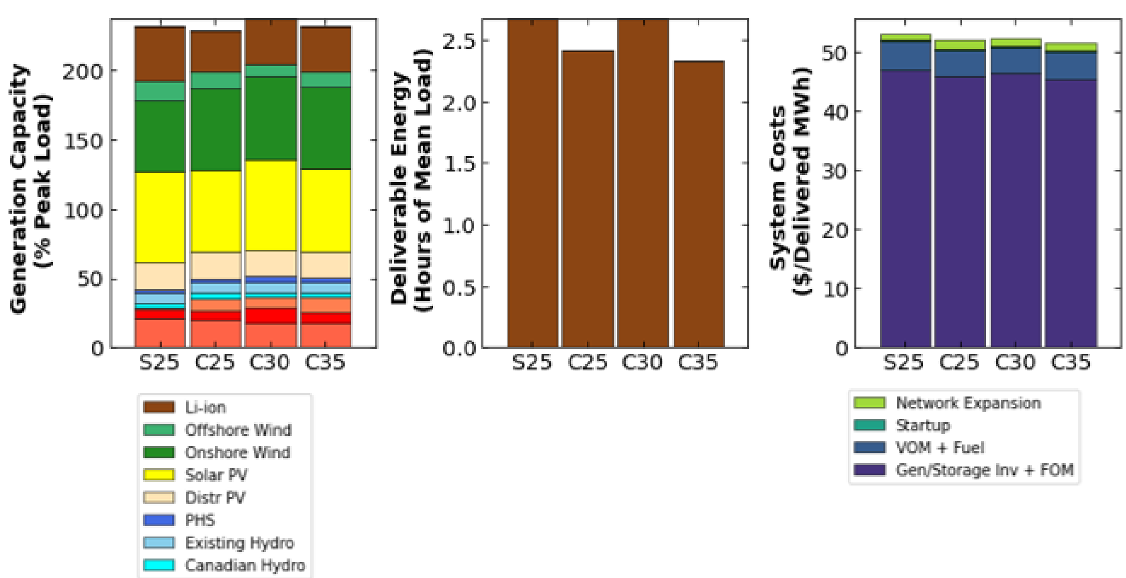

Figure 4-5: Capacity-expansion results for the U.S. Northeast: Without LDES

This is because the capacity deployment determined in the planning process has large implications on how flexibility the system can be operated in real-time, where the available solutions consist of dispatching small amounts of gas, deploying energy-limited energy storage (or demand flexibility) resources, curtailing VRE output, and utilizing imports/exports. While the solution under C35-CEM is expected to be 3% more expensive than in the S25-CEM configuration in the Southeast, results under the reliability simulations show that it will help to avoid 1,888 GWh of involuntary load shedding for the modeled year. At the assumed $50,000/MWh of Value of Lost Load, this alone is equivalent to avoiding $94 billion of non-served energy "penalties".

4.2 Impacts on Systems with Long-Duration Energy Storage

As the penetration of VRE resources increases, there will be growing needs for grid balancing on a longer time scale (i.e., days, weeks). Appropriately assessing the value of long-duration energy storage (LDES) technologies require a detailed enough modeling framework – one which appropriately captures the opportunity costs and benefits of multi-day storage.

At a high level, LDES technologies are characterized primarily by lower energy capital cost, higher power cost and lower overall round-trip efficiency (RTE) compared to available range of cost projections for Li-ion storage. The inter-temporal operation of storage technologies is parameterized using several parameters, highlighted in Table 4.3, including the hourly self-discharge rate and the variable O&M cost for charging and discharging. The flexibility of design (e.g., ability to decouple discharging capacity from storage capacity and charging capacity) is an important advantage of these emerging LDES technologies. Previous studies have also shown that in addition to energy capacity cost, discharge efficiency is another important technology design attribute impacting the value (i.e., cost reduction potential) of LDES in zero-carbon power systems [49].

	Tech		Discharging Capital Cost ($/kW)	Charging Capital Cost ($/kW)	Storage Capital Cost ($/kWh)	FOM ($/kW-year)	FOM ($/kWh-year)	VOM ($/kWh)	Efficiency Up (%)	Efficiency Down (%)	RTE (%)	Self-Discharge Rate (%-hr)
[1]	PHS	Mid	1,966	-	0.0	41.0	0.0	0.0	89%	89%	80%	-
[2]	Li-ion	Low	32	-	70.9	0.3	1.4	0.0	92%	92%	85%	0.2%
[3]	Li-ion	Mid	110	-	125.8	0.8	2.2	0.0	92%	92%	85%	0.2%
[4]	Li-ion	High	154	-	177.0	1.4	3.2	0.0	92%	92%	85%	0.2%
[5]	RFB	Low	297	-	15.5	4.1	0.0	0.0	92%	88%	80%	0.0%
[6]	RFB	Mid	396	-	48.0	4.1	0.0	0.0	92%	88%	80%	0.0%
[7]	RFB	High	530	-	102.2	4.1	0.0	0.0	92%	88%	80%	0.0%
[8]	Metal-air	Low	595	-	0.1	14.9	0.0	0.0	70%	59%	41%	0.2%
[9]	Metal-air	Mid	643	-	2.4	16.1	0.1	0.0	73%	63%	46%	0.2%
[10]	Metal-air	High	950	-	3.6	23.7	0.1	0.0	72%	60%	43%	0.2%
[11]	Hydrogen	Ultra-Low	1,190	479.3	1.1	11.0	0.0	0.0	77%	65%	50%	-
[12]	Hydrogen	Low	1,150	356.1	6.0	11.0	0.1	0.0	80%	70%	56%	-
[13]	Hydrogen	Mid	1,190	479.3	7.0	11.0	0.1	0.0	77%	65%	50%	-
[14]	Hydrogen	High	1,230	602.4	8.0	11.0	0.1	0.0	60%	60%	36%	-
[15]	Thermal	Low	494	3.3	2.9	3.9	0.0	0.0	100%	55%	55%	0.0%
[16]	Thermal	Mid	736	3.3	5.4	3.9	0.0	0.0	100%	50%	50%	0.0%
[17]	Thermal	High	1,226	3.3	9.0	3.9	0.1	0.0	100%	46%	46%	0.0%

Table 4.3: LDES cost and operational assumptions

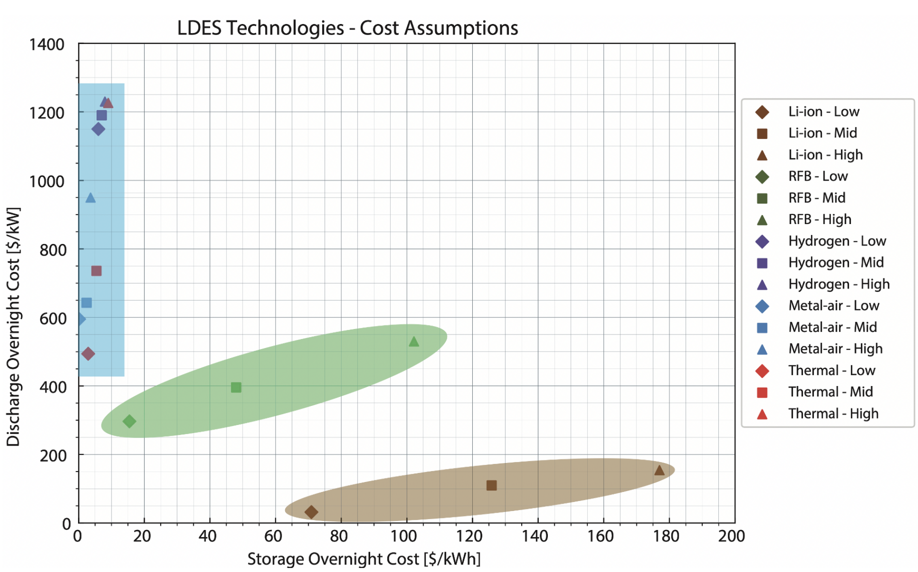

Figure 4-6: LDES power discharging and energy storage costs for 2050

For easier visualization, Figure 4-6 groups available energy storage technologies based on two specific design attributes: (1) technologies with the lowest power cost, relatively high energy capacity cost, high RTE (e.g. Li-ion); (2) technologies with mid-range power and energy capacity costs and RTE (e.g. current and future RFBs); and (3) technologies with high power costs, low energy capacity costs, and low RTE (e.g. emerging LDES options). There are other features that distinguish these technologies from Li-ion (e.g., charging costs, fixed operating & maintenance costs, variable operating & maintenance costs, self-discharge, etc.).

In this Thesis, we have selected one representative storage technology at mid-costs per group (Figure 4-6) to illustrate the impact of modeling with higher temporal resolution and inclusion of chronology. We hereby show results for three electrochemical technologies: Li-ion batteries, redox flow batteries (RFB), and metal-air systems. See the FoS study for the CEM results for the other two LDES options (i.e.,hydrogen, and thermal systems) as well as results from cost sensitivity analyses.

4.2.1 Reliability Key Findings

Figures 4-7 and 4-8 show the breakdown of all NSE events for the Southeast and Northeast, when longer-duration storage technologies are available. Similar to the case with Li-ion only, we see that for the U.S. Southeast (Figure 4-7), under the 0 gCO_2/kWh emissions case, there is a significant decrease in total NSE between the S25-CEM configuration and the C35-CEM one. Moving from S25-CEM to C25-CEM decreases total NSE by 43% (or 4,528 GWh) across the seven weather years considered, and it decreases the total duration of such events by 44% (or 229 hours). While there is a spike in NSE events between the C25-CEM and C30-CEM configurations, it is partly due to the decrease in VRE and storage deployments in the C30-CEM case — resulting in a single large-scale NSE event (starting in Hour 52623) that accounts for 1,089 GWh of load shedding. The iterative approach of selecting problematic periods helps to smooth out these outliers. We see this in the subsequent decrease of NSE events (in magnitude, frequency, and duration) moving from C30-CEM to

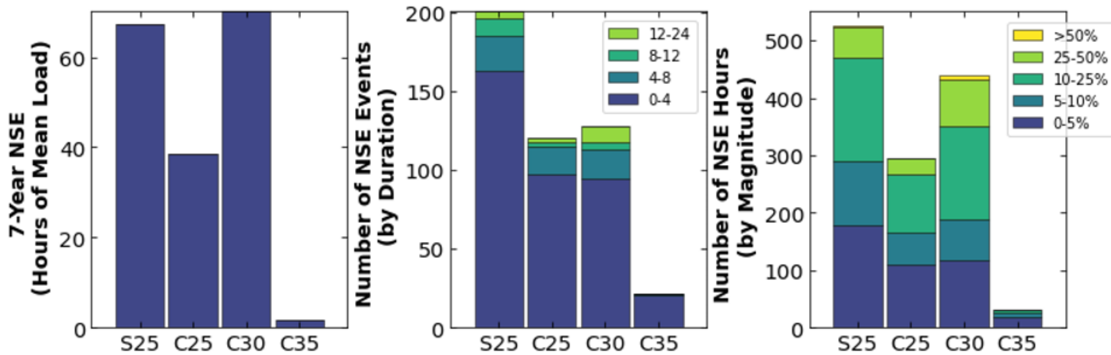

Figure 4-7: Reliability results for the U.S. Southeast: With LDES

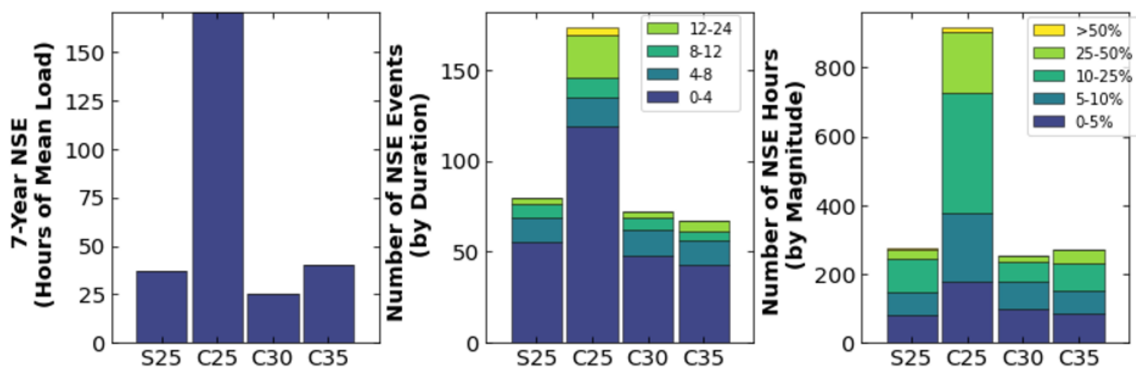

Figure 4-8: Reliability results for the U.S. Northeast: With LDES

C35-CEM. Overall, relative to S25-CEM, moving to C35-CEM decreases total NSE by 98% (Figure 4-7), and it decreases the total duration of such events by 94%.

In contrast, the reliability results for the Northeast (Figure 4-8) shows little improvement moving from S25-CEM to C35-CEM, though there is significant improving moving from C25-CEM to C35-CEM. This can partly be explained by the fact that the low reliability periods were identified as part of a Li-ion only capacity mix. Thus, the reliability results when assessing LDES are not monotonic. That is, in such a systems analysis, there are many moving parts (e.g., relative cost-effectiveness between LDES, thermal, and VRE technologies) that contribute to different optimal levels of VRE and storage deployed under the Base Case relative to the LDES case. That reliability outcomes are not improved by moving from S25-CEM to C35-CEM is a combination of all the above-mentioned interactions.

In this specific case for the Northeast, basing investment decisions on the weather year 2012 (S25-CEM) appears to produce better reliability results than looking at all weather years from 2007 to 2012 (C25-CEM) – but in practice, it's difficult to know ahead of time which historical weather years would be most similar to the future year. Thus, it is still encouraging to see that the iterative approach of adding in "reliability" periods (C30-CEM and C35-CEM) yields lower non-served energy events (in magnitude, duration, and frequency) than the traditional clustering approach (C25-CEM).

4.2.2 Planning Implications

Compared to when Li-ion batteries were the only available energy storage option (with higher relatively higher energy storage costs), we find that the availability of LDES (e.g., RFB, metal-air systems) substitutes for NG capacity and VRE capacity, and leads to reduced curtailment of wind and solar generation, along with a modest reduction in SCOE compared to the scenarios without LDES.

Figure 4-9 shows a summary of the key modeled system outcomes for different LDES availabilities for the U.S. Southeast at 5 gCO_2/kWh. Optimal deployment of LDES resources reduces the need for thermal capacity (gas with and without CCS, nuclear) by 24%-28%, relative to the case with only Li-ion. This thermal capacity is replaced by VRE and Storage capacities, as evidenced by the increase in VRE capacity by 21% in the S25-CEM configuration and 17% in the C35-CEM configuration. This is because the availability of LDES makes VRE capacity more dispatchable and thus increases their value to the power system.

Our analysis also reveals that there is a clear trade-off between installed storage capacity and VRE curtailment across the three modeled regions. When it is optimal to employ LDES in the 5 gCO_2/kWh case, it is generally optimal to have LDES durations much greater than those of Li-ion or RFB (Figure 4-9). Across the configurations analyzed, the duration of the LDES resources ranges between 41-52 hours, as compared to Li-ion duration of 1-2 hours, and RFB duration of 6-7 hours; these trans-

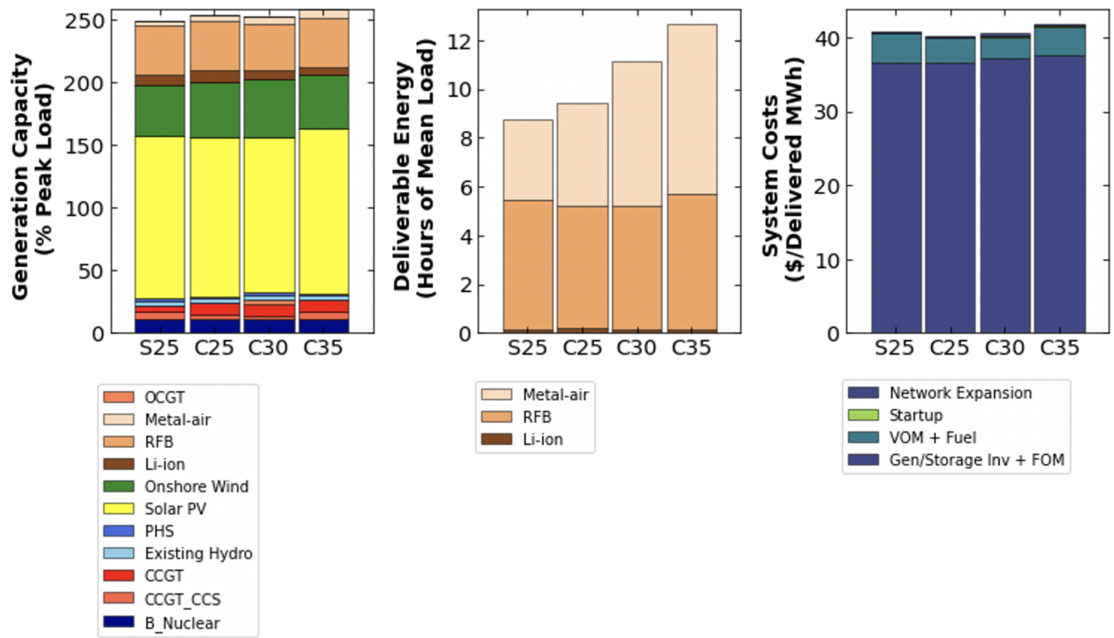

Figure 4-9: Planning results for the U.S. Southeast: With LDES

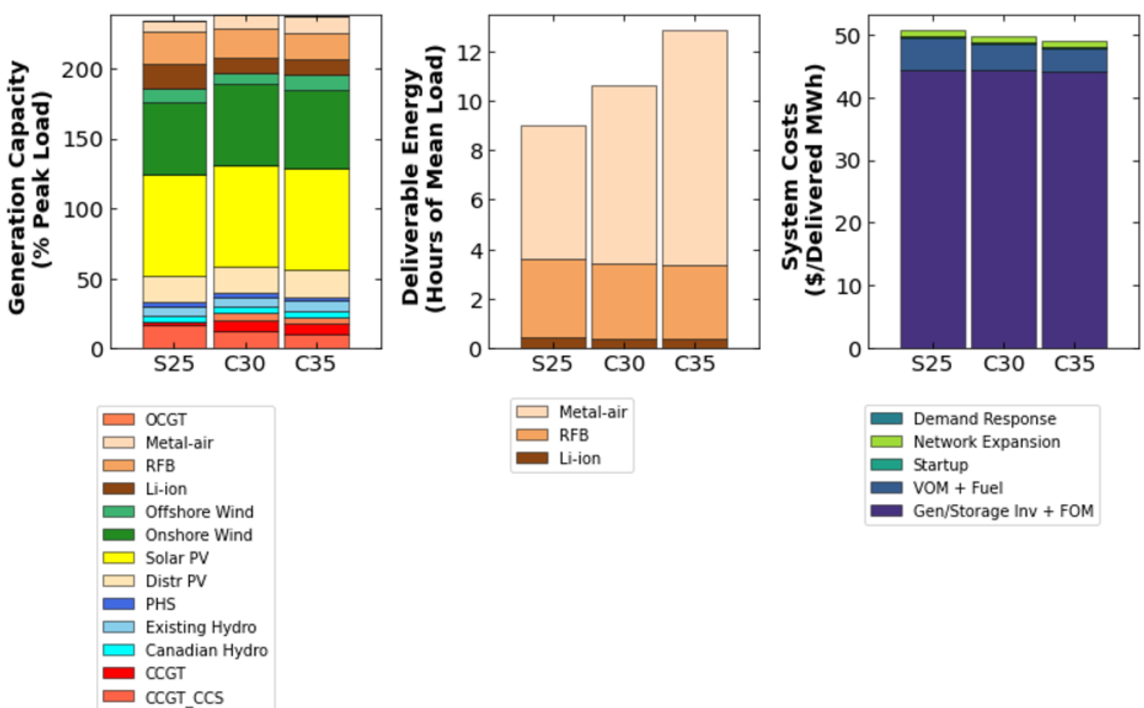

Figure 4-10: Planning results for the U.S. Northeast: With LDES

late to a total deliverable energy capacity of 8-12 hours of mean system load (across the LDES options and regions). Optimal VRE curtailment in the U.S. Southeast is reduced from 5%-7% without LDES deployment to 4%-5% with LDES deployment.

Similar results can be observed for the Northeast (Figure 4-10)[2] – where by modeling more representative and "reliability" periods, we can see more optimal deployment of LDES technologies (particularly on the energy storage capacity front). This reinforces the claim that Baseload-oriented modeling can lead to an over-building of traditional baseload capacity, and an under-building of flexible capacity (e.g., energy storage, demand response), particularly LDES technologies.

4.3 Model Limitations

In interpreting these results, we caveat that our findings are based on a stylized model of the power system with perfect foresight over VRE availability and load variability. Perfect foresight allowed for our iterative selection of extreme periods to improve overall reliability outcomes, which served to illustrate the large differences in system outcomes based on the periods input "seen" by the model. Though in practice, we would not know the exact coincident conditions that would lead to severe system stress. More research needs to be done to construct insightful time series of VRE availability and coincident load for extreme periods.

Relatedly, our use of historical weather to simulate multi-year VRE capacity factors provides range and variation for VRE availability; however, it does not capture correlations between the effects of extreme weather events on electricity demand or VRE output, which would only partially capture events like the August 2020 California or February 2021 ERCOT outages. In practice, it's very hard to predict the co-variability of load and VRE output as a function of weather. Current ELCC best practices incorporate some linkages between load, weather, and VRE generation conditions; however, they are typically based on historical conditions – which we have

[2]C25-CEM results not shown, as they did not complete prior to the publication of this Thesis.

shown to be drastically different from a future power system (with different technology penetrations, efficiencies, and weather patterns). The forecasting of correlated load and VRE output as a function of weather is another ripe area for further research. In the next Chapter, we discuss below some key implications of this Thesis' findings.

Chapter 5

Conclusion and Discussions

The findings from Chapter 4 showed a dramatic improvement in reliability outcomes by carefully selecting representative periods for use in a CEM that considers hourly temporal resolution and preserves some type of chronology (in this case, 24 hours/day x 10 days/period). It follows that compared to the current approach of using time slices to determine optimal capacity mixes and conducting an ELCC study to evaluate resulting reliability, that there is room for improvement to combine the two modeling processes to establish fit-for-purpose planning targets for ensuring resource adequacy. In addition to recommendations on how to better incorporate extreme periods of system stress in CEMs in planning and resource adequacy processes presented above, I share below some thoughts on how to explicitly value resiliency in new business models, and how to better communicate the risks and impacts of extreme weather events on power system reliability in a just and democratic manner.

5.1 Modeling Recommendations

A main challenge in determining the cost-optimal portfolio of resources to meet future demand is data, since new builds of VRE generators and thermal plants have different operating characteristics to those in operations today. Furthermore, because weather patterns are expected to change from year to year (and more drastically in the future due to effects of climate change), demand and VRE output subject to weather are going to look drastically different in the future (and from day to day). So far, we've only looked at deterministic CEMs with perfect foresight over these time-varying variables – going forward, we need to look at probabilistic approaches that consider many realizations of these time series variables, and co-optimize for both capacity and resource adequacy. It is not sufficient to evaluate the capacity value (or effective firm capacity) of individual resources – a systems approach is needed to evaluate how these resources operate and contribute energy and capacity to relieving system stress, when embedded into a portfolio of grid-connected resources.

Some modeling-specific recommendations that follow directly from these observations are to: (1) Capture inter-temporal dynamics by using hourly (or sub-hourly) time series of VRE output and load data (especially when modeling a high-VRE power system with energy storage and demand flexibility resources). (2) Where possible, include variability of load, VRE output, and unplanned thermal generator outage across multiple weather years. (3) Conduct further research into the correlations of VRE output and load, and their relationship to weather changes (including extreme weather events) – i.e., better and more granular data on each variable. (4) Include periods of system stress (based on historical data or artificially constructed) to stress test the capacity mix resolved by capacity-expansion modeling. (5) Quantify the reliability outcomes using easy to understand benchmarks (for duration and magnitude).

5.2 New Business Models for Valuing Customer-Side Reliability

For events that cannot be adequately forecast, adequate precautions need to be taken to avoid mass power outages. Today, weather-related power outages cause \$25 to \$70 billion of economic losses in the U.S. every year (mainly on the distribution-side) [10]. Looking ahead, we expect that large power outages created by generation deficiencies on the bulk power system will become more frequent and intense, amplified by climate change-induced extreme weather events, and increasing penetration of VRE exerting pressure on antiquated power grids.

Customer-side reliability, or the ability to provide power during outage events, can mitigate the worst consequences by sustaining critical electricity applications during outages (e.g., hospitals, essential services, data centers, large commercial and industrial applications). While traditional back-up generation has been provided by diesel generators, we see an opportunity for large customers (i.e.,data centers, large commercial customers) to install a distributed Solar and Storage system to weather these increasingly more frequent outage events. During an outage, the system can be operated in an islanded mode, with both the solar and battery providing electricity to the facility.

In a class paper ("Assessing the Value of a Solar+Storage System When Considering Resilience Value to the Customer" dated May 2020), Lars Habostad and I developed a framework to assess the value of a fixed-size Solar + Storage system, by explicitly specifying a Value of Loss Load to the customer. Using an innovative hybrid approach, we varied the proportion of the battery allocated to energy arbitrage vs. resilience operations, and determined the net revenues expected under each capacity split. Figure 5-1 assumes that half of the capacity is allocated to resilience (standby to use during outage hours), averaged across ten artificially-constructed outage years. It shows that, when combined with revenues from selling electricity from solar generation and providing energy arbitrage services, the value of resilience can sometimes make unprofitable projects profitable (green). This is a strong argument for trying to reach customers particularly vulnerable to power outages, and introduce a business model to explicitly assign value to resilience.

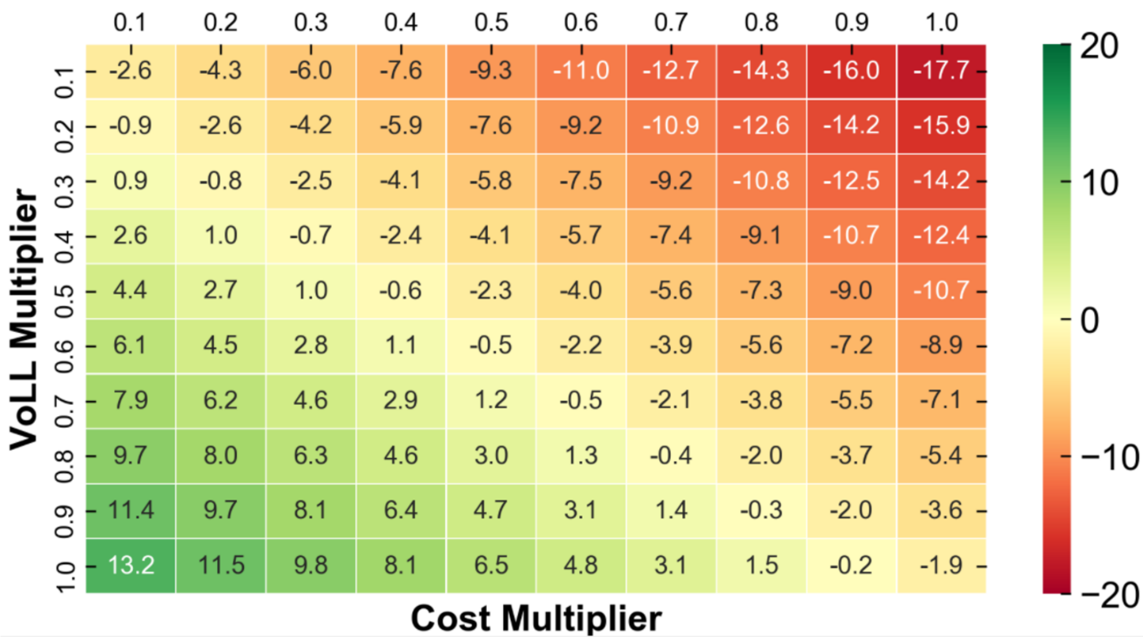

Figure 5-1: Net revenues from a solar plus storage system

5.3 Centering Justice in Planning

A common theme throughout this Thesis has been that the energy transition in the coming years will be massive in scale. Because the power system of 2050 will be vastly different from the one we know today, we need to rethink our current planning and public engagement processes to mitigate the burdens the energy system has on its surrounding communities. So far, we've focused on the technological and financial feasibility of bringing more renewable and flexible capacity onto the grid, but such infrastructure investments also have out-sized impacts on the social, political, and physical.

Evidence shows that communities of color often bear the brunt of such problems, such as shouldering higher utility bills (as a proportion of income), experiencing higher exposure to air and water pollution, and being more vulnerable to natural disasters [3]. Unfortunately, the ones who are most impacted by the siting of power plants (as part of system planning) and determination of how much reserves to contract (as part of resource adequacy) are often missing in such discussions, either through lack of compensation/time, or limited knowledge of the technical language/modeling frameworks used in such processes. For example, model behavior is often guided by expert opinion (in a non-transparent manner), and these linkages are often not up for public discussion [13, 8].

Centering justice in the planning process thus necessitates extending participation to those are most affected by the decisions. In a paper written for Professor Susskind's "Environmental Policy" class in Fall 2020, I argued that instead of the conventional public engagement model (where the "public" is only involved as a cursory box to check), the minority stakeholder groups should be involved at all stages of the process. Rather than passively vote on pre-selected issues, stakeholders should take part in setting the agenda and frame the problem (thereby be the first ones to denounce any lack of representation on the committee). This can be done through a Consensus-Building Approach (CBA); and more specifically, through a neutral mediator or facilitator, who could either: (1) add back traditionally excluded voices; and/or (2) find a proxy to represent their interests.

In power system planning and resource adequacy analyses, there are many assumptions (e.g., VoLL, future technology and fuel costs, modeled emission constraints, correlations between VRE output, demand, and weather) and model simplifications (e.g., time domain reduction) that have significant impacts on the modeling outcomes. Consensus-building offers an avenue for various parties to come together, and have repeated interactions to learn together, and build trust and collaboration. Parties with polarized views (e.g., coal plant owners, community residents, clean energy activists) may not agree on everything, but will certainly agree on some things – finding that common ground is important to benchmark the starting position. Here, an unbiased facilitator or negotiator can pre-identify potential sources of conflicts, and prepare materials to bring the polarized groups together to discuss whether their values diverge and how to still arrive at an outcome deemed reasonable by all. Finally, consensus-building can explicitly allocate a time for each party to defend and add more context to their positions (i.e.,siting challenges in specific communities, process for instituting rolling outages). When issues are assessed jointly, parties are more likely to reach a better outcome (since the non-equivalence of the issues enable trade-offs). Most importantly, consensus-building reframes how opposing parties interact with each other in a way that ensures trust-building and fosters future collaboration. This in turn enables value creation through the brainstorming of new ideas and co-development of solutions that may have not previously been considered. Because energy planning is so complex and subject to ever-evolving objectives (e.g., least-cost, green, reliable, resilient), an open and transparent process like consensus-building is much needed to align the perceived and actual well-being of the stakeholders.

5.4 Best Practices for CEMs in Planning Processes

The model survey and literature review in Chapter 2 and modeling experiment results in Chapter 4 inform a set of CEM design principles for use in planning processes. Given the pace and scale of system transformations needed to achieve carbon reduction targets, it is critically important for utilities, regulators, and public stakeholders to consider and adopt these guiding principles to ensure a smooth transition to a high VRE power system.

First, let the problem inform the complexity of the CEM analysis. In selecting the appropriate spatial, temporal, and operational resolution, the model user and consumer need to keep in mind the ultimate research questions at hand [14]. If the problem involves investment and operational decisions under a high VRE system, then high temporal resolution and operational detail would be needed. If the problem involves a system with high levels of dispatchable capacities, then simplifications along these dimensions may be acceptable. If the problem is focused on a specific zone vs. a wider geographic area, then that would have implications on how detailed the transmission network needs to be represented. When in doubt, try to make the model as simple as possible, and abstract away the pieces that are not central to the problem. Conduct sensitivity analysis to identify key levers that drive model results.

Second, if modeling a high VRE system, preserve temporal chronology as

much as possible. Given the inter-temporal of both VRE and load, chronology is important to characterize sustained periods of high load, or low VRE output, and thus the requirements for both firm capacity and flexibility. Sufficient temporal and operational details are also needed to model thermal generators' unit commitment decisions, ramping constraints, spinning reserves, quick-start reserves, and start-up costs. These all have implications on the system need for flexibility services (e.g., short-duration and long-duration storage and demand flexibility).

Third, consider both endogenous and exogenous uncertainties and how to address them. CEMs are typically deterministic models that assume perfect foresight over system conditions (e.g., load and VRE output). In reality, short-term forecasting is a ripe area for research, and variability on that time scale can introduce additional operational challenges. The consideration of extreme weather events can help to address some of this uncertainty, while a more robust approach may be to adopt a stochastic approach to modeling. Trade-offs between model selection and their impact on introducing/resolving uncertainties should be clearly communicated to the appropriate stakeholders, so that insights can be drawn given the modeling caveats.

Finally, engage with stakeholders from the beginning and throughout the study. Investments into different technology mixes will produce winners and losers. And where these investments end up (company, location) will have impacts on the local communities – raising issues of equity, in addition to the issues of decarbonization, efficiency, and reliability discussed elsewhere in the Thesis. Therefore, model users need to actively engage (including training and education) with model consumers early and continuously re-evaluate the modeling approach and objectives [14], to keep the study and results relevant.

Appendix A

Cost and Operational Data Inputs

A.1 Transmission

Existing transmission inter-zonal transfer capacity is approximated from the EPA Integrated Planning Model (IPM) model [1]. The IPM model uses a selection of 64 regions based on NERC regions, which represents fractions of states. We follow these zonal definitions to get the aggregate transfer capacities between zones within each region. Existing transmission capacity is assumed to be available at no additional cost (assuming that the network has paid itself off by 2050). When transmission expansion is enabled, new AC capacities can be added along the existing network paths. Transmission upgrades on 345 kV lines in the Northeast are assumed to be \$1,670/MW-km; upgrades on 500 kV in the Southeast are assumed to be \$960/MW-km [6].

A.2 Brownfield Capacity

We assume a mainly greenfield modeling, apart from existing hydropower in all regions and nuclear capacity in the Southeast. For existing hydropower generators, we classify individual plants into run-of-river (RoR) or reservoir generators using the ORNL HydroSource database [31]. RoR generators are modeled as must-run or non-dispatchable resources that do not have the ability to spill water (i.e., they do not respond to economic dispatch). And, hydro reservoir generators are modeled as storage devices that receive exogenous inflows to their storage reservoirs, but cannot charge from the grid. We get hydropower generators' installed capacities from Form EIA-860 [17], and historical monthly generation from Form EIA-923 [18], then aggregate them up to the zonal level.

In the Southeast, we assume that a portion of the existing fleet of nuclear generators will remain operational in "2050". That is, based on each generator's start date, we assume that plants that could run to 2055 or beyond with a Second Life Extension License (80 years from start date). These include Vogtle 3 and 4 (with a combined 2,500 MW), which are still under construction in Georgia (see Table A.1).

In the Northeast, we enforce a minimum build for distributed solar PV in the Northeast due to the strong existing policy push for it distributed PV. The minimum build is based on a projection of new capacities installed through 2050 (i.e., not including existing capacities, see Table A.2). For states in ISO-NE, we assume the 2030 installed capacities from the ISO's projections [5], then extrapolate them to 2050 using the implied EIA AEO 2030-2050 growth rate of 110%. For zones in New York ISO, the projections are directly taken and aggregated from the 2020 Gold Book [28].

A.3 Electricity Load

Electricity demand is from the NREL Electrification Future Study [34]. This demand data includes assumptions around electrification and its impacts on the load profile, and is available on an hourly basis (8,760-hour year), and on a state-by-state basis. The "2050 High_Moderate" profiles are used for the bulk of the study, except for the Reference Electrification scenario, which are based on the "2050 Reference_Moderate" profiles. These profiles reflect different levels of electrification ("High" vs. "Reference" and a "Moderate" pace of energy-efficiency improvements). To align these state-based

EIA Plant-Generator	EIA Nameplate Capacity (MW)	EIA Start Date	Date Entering Extended Operations
Browns Ferry_3	1,190	3/1/1977	7/2/2016
Brunswick Nuclear_1	1,002	3/1/1977	9/8/2016
Catawba_1	1,205	6/1/1985	12/5/2023
Catawba_2	1,205	8/1/1986	12/5/2023
Edwin I Hatch_2	865	9/1/1979	6/13/2018
Grand Gulf_1	1,440	7/1/1985	11/2/2024
Harris_1	951	5/1/1987	10/24/2026
Joseph M Farley_1	888	12/1/1977	6/25/2017
Joseph M Farley_2	888	7/1/1981	3/31/2021
McGuire_1	1,220	9/1/1981	6/12/2021
McGuire_2	1,220	3/1/1984	3/3/2023
Sequoyah_1	1,221	7/1/1981	9/17/2020
Sequoyah_2	1,221	6/1/1982	9/15/2021
St Lucie_1	1,080	5/1/1976	3/1/2016
St Lucie_2	1,080	6/1/1983	4/6/2023
V C Summer_1	1,030	1/1/1984	8/6/2022
Vogtle_1	1,160	5/1/1987	1/16/2027
Vogtle_2	1,160	5/1/1989	2/9/2029
Vogtle_3	1,250	1/1/2021	-
Vogtle_4	1,250	1/1/2022	-
Watts Bar Nuclear Plant_1	1,270	5/1/1996	-
Watts Bar Nuclear Plant_2	1,270	6/1/2016	-

Table A.1: Brownfield nuclear capacity in the U.S. Southeast

		Existing Capacity (MWdc)	Cumulative Installations through 2050 (MWdc)	Existing Capacity (MWac)	Cumulative Installations through 2030 (MWac)	Cumulative Installations through 2050 (MWac)	New Installations through 2050 (MWac)
ISO-NE Projections							
[1]	CT	-	-	682.3	1,242.8	2,607.1	1,924.8
[2]	MA	-	-	2,502.3	2,738.2	5,744.1	3,241.8
[3]	ME	-	-	68.8	320.6	672.5	603.7
[4]	NH	-	-	125.3	259.5	544.4	419.1
[5]	RI	-	-	223.8	196.7	412.6	188.8
[6]	VT	-	-	393.5	607.2	1,273.8	880.3
NYISO Projections							
[7]	A	125.0	1,276.0	108.7	-	1,109.6	1,000.9
[8]	B	63.0	371.0	54.8	-	322.6	267.8
[9]	C	169.0	977.0	147.0	-	849.6	702.6
[10]	D	5.0	101.0	4.3	-	87.8	83.5
[11]	E	123.0	974.0	107.0	-	847.0	740.0
[12]	F	299.0	1,168.0	260.0	-	1,015.7	755.7
[13]	G	251.0	705.0	218.3	-	613.0	394.8
[14]	H	34.0	70.0	29.6	-	60.9	31.3
[15]	I	46.0	109.0	40.0	-	94.8	54.8
[16]	J	210.0	791.0	182.6	-	687.8	505.2
[17]	K	537.0	880.0	467.0	-	765.2	298.3

Sources and Notes: Used a DC-AC ratio of 1.15.
Extrapolated to 2050 forecasts using 2030-2050 growth rate from EIA/AEO: 110%
[1]-[6]: ISO-NE's Final 2021 PV Forecast, pp. 23-28.
[7]-[17]: NYISO 2020 Gold Book, Table I-9a.

Table A.2: Minimum capacities of distributed PV in the U.S. Northeast by 2050 demand data to our IPM-based zonal definitions, we use the 2018 utility state-level sales data from Form EIA-861 [16] to allocate fractions of state demand to our defined zones.

A.4 VRE Supply Curves

We developed zonal VRE supply curves based on the methodology described in Brown and Botterud (2021) [7]. Hourly PV capacity factors are simulated using 2007-2013 weather data from the NREL National Solar Radiation Database (NSRDB) [42] through the PVLIB model framework [25], at a 4km x 4km spatial resolution. Hourly wind capacity factors are simulated using the same temporal and spatial resolution using the NREL Wind Integration National Dataset Toolkit [15] and power curve data from commercial wind turbines assuming a 100-m hub height [20]. To reduce the spatial resolution of the VRE capacity factor data, we aggregate sites within a zone on the basis of average LCOE (including the cost of interconnecting to the nearest substation). Thus, for each resource and zone, we get a supply curve, with each bin representing increasing resource quality with an associated maximum availability (based on land area), interconnection cost and hourly capacity factor profile (similar to the approach used in the Renewable Energy Potential (reV) Model [32]).

A.5 Generator and Storage Costs

Fossil-powered generation and VRE capital and operational costs are shown in Table 5. The gas, nuclear, VRE, and Li-ion costs are taken from the 2020 NREL Annual Technology Baseline [41] 2045 "Mid" cost projections. "Low" VRE and Li-ion costs are also taken from the NREL ATB for the sensitivity analysis. Additionally, we apply a small, non-zero VOM for wind, hydropower, and storage to distinguish their dispatch as part of the economic dispatch modeled within GenX — they do not meaningfully affect resulting system costs.

Tech	Capacity Size (MW)	Start Cost ($)	Start Cost ($/MW/start)	Start Fuel (MMBTU/ start)	Start Fuel (MMBTU/ MW/start)	Heat Rate (MMBTU/ MWh)
[1] OCGT	237	33,147	140	45	0.19	9.51
[2] CCGT	573	34,982	61	115	0.20	6.40
[3] CCGT + CCS	377	36,419	97	75	0.20	7.12
[4] Existing Nuclear	1,000	1,000,000	1,000	0	0.00	10.46
[5] New Nuclear	1,000	1,000,000	1,000	0	0.00	10.46

Tech	Min Stable Output (%)	Ramp Up (%)	Ramp Down (%)	Up Time (Hours)	Down Time (Hours)
[1] OCGT	25	100	100	0	0
[2] CCGT	30	100	100	4	4
[3] CCGT + CCS	50	100	100	4	4
[4] Existing Nuclear	50	25	25	36	36
[5] New Nuclear	20	100	100	36	36

Table A.3: Operating assumptions for existing and new thermal plants

Fuel	$/MMBtu
Uranium	0.75
Coal	2.00
Natural Gas	4.16
Coal CCS	2.00
NG + CCS	4.16

Table A.4: Assumed fuel prices for 2050

A.6 Operations and Fuel Assumptions

Table A.3 shows the assumed operational characteristics of the gas- and nuclear-powered technologies. Operational assumptions for gas plants are from best-in-class technology in the industry and academic research [22, 41, 37]. Nuclear assumptions are from the previous MIT Future of Nuclear study and related works [?, 29, 48]. Fuel price assumptions are taken from the EIA AEO 2020 Reference 2050 case [19] (see Table A.4).

Bibliography

[1] U.S. Environmental Protection Agency. Documentation for EPA's power sector modeling platform v6 using the integrated planning model.

[2] Sunhee Baik, Alexander L. Davis, Jun Woo Park, Selin Sirinterlikci, and M. Granger Morgan. Estimating what US residential customers are willing to pay for resilience to large electricity outages of long duration. 5(3):250–258.

[3] Shalanda H. Baker. *Revolutionary power: an activist's guide to the energy transition.* Island Press, Washington, 2021.

[4] John Bistline, Daniel Shawhan, and Geoffrey Blanford. Systems analysis in electric power sector modeling: Evaluating model complexity for long-range planning.

[5] Jon Black. Final 2021 PV forecast.

[6] Maxwell Brown, Wesley Cole, Kelly Eurek, Jon Becker, David Bielen, Ilya Chernyakhovskiy, Stuart Cohen, Will Frazier, Pieter Gagnon, Nathaniel Gates, Daniel Greer, Sai Sameera Gudladona, Jonathan Ho, Paige Jadun, Katherine Lamb, Trieu Mai, Matthew Mowers, Caitlin Murphy, Amy Rose, Anna Schleifer, Daniel Steinberg, Yinong Sun, Nina Vincent, Ella Zhou, and Matthew Zwerling. Regional energy deployment system (ReEDS) model documentation: Version 2019. page 140.

[7] Patrick R. Brown and Audun Botterud. The value of inter-regional coordination and transmission in decarbonizing the US electricity system. 5(1):115–134.

[8] Toby Brown, Neil Lessem, Roger Lueken, Kathleen Spees, and Cathy Wang. High-impact, low-probability events and the framework for reliability in the national electricity market.

[9] Jacopo Buongiorno, Michael Corradini, John Parsons, and David Petti. The Future of Nuclear Energy in a Carbon-Constrained World. Technical report, MIT Energy Initiative, 2018.

[10] Richard J Campbell. Weather-Related Power Outages and Electric System Resiliency. page 18, August 2012.

[11] Juan Pablo Carvallo, Nan Zhang, Benjamin D. Leibowicz, Thomas Carr, Maury Galbraith, and Peter H. Larsen. Implications of a regional resource adequacy program for utility integrated resource planning. 34(5):106960.

[12] Judy Chang, Mariko Geronimo Aydin, Johannes Pfeifenberger, Kathleen Spees, and John Imon Pedtke. Advancing past "baseload" to a flexible grid.

[13] Wesley Cole, Bethany Frew, Trieu Mai, Yinong Sun, John Bistline, Geoffrey Blanford, David Young, Cara Marcy, Chris Namovicz, Risa Edelman, Bill Meroney, Ryan Sims, Jeb Stenhouse, and Paul Donohoo-Vallett. Variable renewable energy in long-term planning models: A multi-model perspective.

[14] Joseph DeCarolis, Hannah Daly, Paul Dodds, Ilkka Keppo, Francis Li, Will McDowall, Steve Pye, Neil Strachan, Evelina Trutnevyte, Will Usher, Matthew Winning, Sonia Yeh, and Marianne Zeyringer. Formalizing best practice for energy system optimization modelling. 194:184–198.

[15] Caroline Draxl, Andrew Clifton, Bri-Mathias Hodge, and Jim McCaa. The wind integration national dataset (WIND) toolkit. 151:355–366.

[16] U.S. Energy Information Administration (EIA). Annual electric power industry report, form EIA-861 detailed data files.

[17] U.S. Energy Information Administration (EIA). Form EIA-860 detailed data with previous form data (EIA-860a/860b).

[18] U.S. Energy Information Administration (EIA). Form EIA-923 detailed data with previous form data (EIA-906/920).

[19] U.S. Energy Information Administration (EIA). Independent statistics and analysis.

[20] Gamesa. G126-2.5 MW.

[21] L. L. Garver. Effective load carrying capability of generating units. PAS-85(8):910–919.

[22] General Electric (GE). 7ha power plants.

[23] GenXProject. GenX, August 2021. original-date: 2021-05-19T17:57:57Z.

[24] G.E. Haringa, G.A. Jordan, and L.L. Garver. Application of monte carlo simulation to multi-area reliability evaluations. 4(1):21–25.

[25] William F. Holmgren, Clifford W. Hansen, and Mark A. Mikofski. pvlib python: a python package for modeling solar energy systems. 3(29):884.

[26] IRENA. Key findings: Renewable power generation costs in 2019. page 2.

[27] California ISO. Final root cause analysis: Mid-august 2020 extreme heat wave.

[28] New York ISO. 2020 load & capacity data.

[29] J.D. Jenkins, R. Ponciroli, R. Vilim, F. Ganda, F. Sisternes, and A. Botterud. The benefits of nuclear flexibility in power system operations with renewable energy. 222:872–884.

[30] Jesse D Jenkins and Nestor A Sepulveda. Enhanced Decision Support for a Changing Electricity Landscape: The GenX Configurable Electricity Resource Capacity Expansion Model. Technical report, MIT Energy Initiative, November 2017.

[31] Megan Johnson, Shih-Chieh Kao, Nicole Samu, and Rocio Uria-Martinez. Existing hydropower assets plant dataset FY20.

[32] Galen Maclaurin, Nicholas Grue, Anthony Lopez, Donna Heimiller, Michael Rossol, Grant Buster, and Travis Williams. The renewable energy potential (reV) model: A geospatial platform for technical potential and supply curve modeling.

[33] Trieu Mai, Clayton Barrows, Anthony Lopez, Elaine Hale, Mark Dyson, and Kelly Eurek. Implications of model structure and detail for utility planning: Scenario case studies using the resource planning model. page 63.

[34] Trieu Mai, Paige Jadun, Jeffrey Logan, Colin McMillan, Matteo Muratori, Daniel Steinberg, Laura Vimmerstedt, Ryan Jones, Ben Haley, Brent Nelson, Caitlin Murphy, and Yinong Sun. Electrification futures study load profiles.

[35] Trieu T. Mai, Paige Jadun, Jeffrey S. Logan, Colin A. McMillan, Matteo Muratori, Daniel C. Steinberg, Laura J. Vimmerstedt, Benjamin Haley, Ryan Jones, and Brent Nelson. Electrification futures study: Scenarios of electric technology adoption and power consumption for the United States.

[36] Dharik S. Mallapragada, Dimitri J. Papageorgiou, Aranya Venkatesh, Cristiana L. Lara, and Ignacio E. Grossmann. Impact of model resolution on scenario outcomes for electricity sector system expansion. 163:1231–1244.

[37] Dharik S. Mallapragada, Nestor A. Sepulveda, and Jesse D. Jenkins. Long-run system value of battery energy storage in future grids with increasing wind and solar generation. 275:115390.

[38] Catherine Mitchell. Momentum is increasing towards a flexible electricity system based on renewables. 1(2):1–6.

[39] North American Electric Reliability Corporation (NERC). 2020 long-term reliability assessment.

[40] Natural Resources Defense Council (NRDC). Race to 100% clean.

[41] National Renewable Energy Laboratory (NREL). Electricity annual technology baseline (ATB) data download.

[42] National Renewable Energy Laboratory (NREL). NSRDB data viewer.

[43] Electric Reliability Council of Texas (ERCOT). Preliminary report on causes of generator outages and derates for operating days February 14 – 19, 2021 extreme cold weather event.

[44] Electric Reliability Council of Texas (ERCOT). Seasonal assessment of resource adequacy for the ERCOT region (SARA) summer 2021.

[45] André Pina, Carlos Silva, and Paulo Ferrão. Modeling hourly electricity dynamics for policy making in long-term scenarios. 39(9):4692–4702.

[46] Kris Poncelet, Erik Delarue, Daan Six, Jan Duerinck, and William D'haeseleer. Impact of the level of temporal and operational detail in energy-system planning models. 162:631–643.

[47] Nick Schlag, Zach Ming, Arne Olson, Lakshmi Alagappan, Ben Carron, Kevin Steinberger, Huai Jiang, N Schlag, Z Ming, A Olson, L Alagappan, B Carron, K Steinberger, and H Jiang. Capacity and reliability planning in the era of decarbonization.

[48] Nestor A. Sepulveda, Jesse D. Jenkins, Fernando J. de Sisternes, and Richard K. Lester. The Role of Firm Low-Carbon Electricity Resources in Deep Decarbonization of Power Generation. *Joule*, 2(11):2403–2420, November 2018.

[49] Nestor A. Sepulveda, Jesse D. Jenkins, Aurora Edington, Dharik S. Mallapragada, and Richard K. Lester. The design space for long-duration energy storage in decarbonized power systems. 6(5):506–516.

[50] Mark Specht. ELCC explained: the critical renewable energy concept you've never heard of.

[51] Frank A. Wolak. Long-term resource adequacy in wholesale electricity markets with significant intermittent renewables.

[52] Marianne Zeyringer, James Price, Birgit Fais, Pei-Hao Li, and Ed Sharp. Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather. 3(5):395–403.