UNIVERSITY OF CALIFORNIA
RIVERSIDE

Decarbonizing the Electric Grid: Computational Advances in Power System Planning and Scenario Analysis

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Osten Peter Anderson

September 2024

Dissertation Committee:

Dr. Nanpeng Yu, Chairperson
Dr. Matthew Barth
Dr. Daniel Wong

The Dissertation of Osten Peter Anderson is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to express my sincere gratitude to Dr. Nanpeng Yu for all of his guidance and support throughout graduate school. Dr. Yu has many great qualities, but the one that has always stood out to me most is the way he truly has the best interests of his students in mind. From making sure that I was researching topics that I found interesting, to making sure I was taking the right steps towards a post-graduate career that suited me, it was always clear that Dr. Yu truly cared. I cannot thank him enough for his guidance, and of course, all of the knowledge he has shared.

I would like to also thank Dr. Matthew Barth for all of his support, and for giving me my first research opportunity many years ago. I would not be here without him. I would like to thank Dr. Daniel Wong and all past committee members as well for all their time and feedback.

Finally, I would like to thank my family for all their support. My parents have supported me immensely and I owe everything to them. Drew's support and belief in me has been and will always be invaluable. Pacal's friendship over the last 20 years means everything to me and has shown me that tough times never last. Jordan has been my rock throughout the program, and is the reason why I do all that I do.

This dissertation contains materials from the following publications:

- Osten Anderson, Mikhail Bragin, and Nanpeng Yu, "Optimizing Deep Decarbonization Pathways in California with Power System Planning Using Surrogate Level-based Lagrangian Relaxation," https://arxiv.org/abs/2309.07202, to appear in Applied Energy, 2024.

- Osten Anderson, Nanpeng Yu, Konstantinos Oikonomou, Patrick Maloney, and Di Wu, ”Representative Period Selection for Robust Capacity Expansion Planning in Low-carbon Grids,” IEEE PES T&D Conference & Exposition, 2024.
- Osten Anderson, Nanpeng Yu, Konstantinos Oikonomou, and Di Wu, ”On the Selection of Intermediate Length Representative Periods for Capacity Expansion,” IEEE PES General Meeting, 2024.
- Osten Anderson, Wanshi Hong, Bin Wang and Nanpeng Yu, ”Impact of Flexible and Bidirectional Charging in Medium- and Heavy-Duty Trucks on California’s Decarbonization Pathway,” https://arxiv.org/abs/2401.10194, to appear in Applied Energy, 2024.
- Osten Anderson, Nanpeng Yu, Cameron Bracken, Casey D. Burleyson, and Alex Pusch, ”Improved Decarbonization Planning through Climate Resiliency Modeling”, in IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3451957.

To my parents for all the support.

ABSTRACT OF THE DISSERTATION

Decarbonizing the Electric Grid:
Computational Advances in Power System Planning and Scenario Analysis

by

Osten Peter Anderson

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, September 2024
Dr. Nanpeng Yu, Chairperson

The push for multi-sector decarbonization in the electric grids around the world has created the need for significant investment in the electrical grid. Over the next two decades, the bulk power system will be transformed from being historically dominated by gas-fired generation to dominated by wind, solar, and energy storage resources. Due to the large costs associated with investment and operation, and the importance of maintaining reliability in this critical infrastructure, a comprehensive investment plan is crucial.

This dissertation looks at the California power system decarbonization problem through several lenses. The first concerns a novel formulation for decarbonization planning which incorporates more detailed modeling of gas-fired generation, and an advanced solution algorithm which enables its practical utility. Then, several novel methods for temporal sampling in these models are proposed. Finally, two case studies are considered. The first examines the impact of smart charging for medium- and heavy-duty battery electric vehicles on the decarbonization plan. The second concerns scenario analysis around climate

change pathways. At the same time, a novel method for ensuring reliable power system investments, with specific regard to decarbonized portfolios, is proposed.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

As the impacts of climate change worldwide are accelerating, so too are the measures taken in response. Governments have taken collective action on mitigating anthropogenic climate change, such as the Paris Agreement [99]. One of the primary aims of these efforts is reducing the level of greenhouse gas (GHG) emissions. This considers emissions from all sources, such as electricity generation, transportation, buildings, and manufacturing.

The energy sector has been identified as one of the most critical components of decarbonization. Not only is the power system one of the primary sources of carbon emissions, it is also a promising candidate for decarbonization due to the associated technical considerations. In California specifically, electricity accounted for 16% of statewide GHG emissions in 2021 [89]. Whereas sectors like aviation and manufacturing currently have limited mature options for decarbonization, the power system has mature technologies for decarbonization

which have been increasingly deployed over the last decade. There is already considerable penetration of wind, solar, and other zero-emission generation, and rapidly increasing penetration of battery energy storage in California [91]. While not a trivial task, power system decarbonization is feasible for most locales.

In California alone, the cost of building, operating, and maintaining the bulk power system will exceed hundreds of billions of dollars over the next two decades. As a result, even a small percentage improvement in the total investment and operation cost for the decarbonization plan could yield savings of hundreds of millions of dollars in investment and operational expenses. This is critical, as California has some of the most expensive electricity in the US [101], and these potential savings would be passed on to ratepayers. In practice, most load in California is served by a number of investor-owned utility companies (IOUs), and these utilities make individual investment decisions. However, these utilities are regulated by the California Public Utilities Commission (CPUC). IOUs work closely with the CPUC to plan investments, as the proposed investments are paid through electricity rate increases which must be approved by the regulatory agency. Studies carried out by the CPUC and the California Energy Commission (CEC) thus inform how IOUs are regulated. These studies can also help shape broader policy goals, concerning adoption of new technologies, planning for extreme weather events, and so on.

California's legislature has passed several laws governing the transition of California's electrical grid to a low-carbon future. Senate Bill 100 2018 (SB100) mandates that all retail electrical sales come from non-carbon sources by 2045 [58]. Senate Bill 350 2015 (SB350) requires that the greenhouse gas emissions associated with the operation of the

electrical grid be reduced to 80% below 1990 levels by 2050 [57]. Consequently, a comprehensive investment plan is essential to identify the most cost-effective approach to achieve these goals.

1.2 Contributions

This dissertation focuses on the topic of power system decarbonization in California through multiple thrusts. Each thrust either carries the goal of improving on the status quo in power system planning in general, or examining the impact of a specific real-world scenario with respect to decarbonization planning. With each, a case study of the California electrical grid decarbonization problem is examined.

This problem belongs to the fields often referred to as capacity expansion planning, generation expansion planning, or bulk power system planning. Although these fields are not new, they have received increased attention over the recent decade. Across the world, power systems are in rapid flux, due both to economic development and decarbonization, and these fields are important aids to decision making. Simultaneously, increasing computational performance has opened new opportunities for improving the modeling of these systems. Accordingly, there has been considerable research interest on these topics over recent years. Although there is substantial variation in the approaches found in the literature, the most common formulation presents a co-optimization of power system investment and operation. Computational burden remains a significant bottleneck towards improving these models further.

A primary contribution of this dissertation is the development of a decarbonization planning model for California's electrical grid that incorporates more detailed modeling of gas-fired power plants than the status quo. Traditional capacity expansion models neglect the complex operating characteristics of such generators due to computational complexity. However, accurate modeling of these generators is crucial, especially over the next two decades. The operation of these generators is a primary source of GHG emissions associated with electricity generation, and carbon emission limits imposes a decreasing upper bound on their operation. Inaccurate modeling of these generators could lead to inaccuracies in the investment in clean technologies needed to meet climate goals. In order to handle the computational complexity associated with this more detailed modeling, an advanced solution methodology is deployed. This model, and the related solution methodology, serve as the underpinning of all studies in this dissertation.

This dissertation examines two novel methods for temporal sampling in the context of power system planning. The computational complexity associated with long-term planning models often necessitates the reduction in the temporal dimension. Instead of representing all 8760 hours in a year, models typically represent fewer than 1000 hours. As such, selecting the appropriate time has a substantial effect on the output of such a model. The first novel method proposes sampling representative days using weather-based features (load and renewable generation) and operational features. The second method concerns sampling of representative periods of intermediate length, in order to properly model inter-day dynamics of energy storage.

Two case studies considering concrete policy and environmental scenarios are presented. The first case study looks at the potential savings from a bulk grid perspective of allowing smart vehicle charging of medium and heavy-duty electric trucks in California. Both flexible charging and vehicle-to-grid charging are considered. Two electric truck operation scenarios are examined to provide a robust view of the potential avoided infrastructure investment. In the second case study, scenarios considering climate change and socioeconomic development are examined. At the same time, a method for improving the manner in which renewable and storage technologies contribute towards power system reliability is proposed. This novel planning formulation allows for direct simulation of extreme weather events within the planning model. Doing so avoids an over-reliance on gas-fired generation for satisfying reliability needs in the context of power system planning.

1.3 Organization

The remainder of the dissertation is structured as follows. Chapter 2 introduces decarbonization planning, discusses related works in optimization and power system planning, and formulates the optimization model. Chapter 3 introduces the Lagrangian relaxation-based solution methodology. Chapter 4 presents numerical testing results as well as comparisons to the existing linearized model results. Chapters 2 through 4 represent the core decarbonization model. Chapter 5 through 7 represent extensions of this model. Chapter 5 introduces two novel approaches to selecting representative periods. Chapter 6 investigates the impacts of smart charging for medium- and heavy-duty battery electric vehicles on California's decarbonization pathway. Chapter 7 examines the impacts of climate change on

California's decarbonization pathway, and presents a novel technique for ensuring resilient power system planning in the context of climate change and decarbonization. Chapter 8 concludes the dissertation.

Nomenclature

Sets

t, T Index, set of hour $w,$

W Index, set of week y, Y

Index, set of year

u, U Index, set of thermal unit

s, S Index, set of storage resource

r, R Index, set of renewable resource

h, H Index, set of large hydro resource

z, Z Index, set of balancing authority zone

l, L Index, set of line

U_z Subset of thermal resources in zone z

S_z Subset of storage resources in zone z

R_z Subset of renewable resources in zone z

H_z Subset of large hydro resources in zone z

Loads and Generation

$L_z(t)$ Load in zone z at time t (MW)

$v_u(t)$ On/off status of unit u at time t (1, 0)

$p_u(t)$ Power output of unit u at time t (MW)

$p_r(t)$ Power output of renewable resource r at time t (MW)

$p_h(t)$ Power output of large hydro resource h at time t (MW)

UT_u Minimum uptime of unit u (hours)

DT_u Minimum downtime of unit u (hours)

RU_u Ramp up rate of unit u (MW/hour)

RD_u Ramp down rate of unit u (MW/hour)

SU_u Startup power limit of unit u (MW)

SD_u Shutdown power limit of unit u (MW)

$\underline{P}_u$ Minimum output of unit u (MW)

$\overline{P}_u$ Maximum output of unit u (MW)

$\underline{P}_h$ Minimum output of hydro resource h (MW)

$\overline{P}_h$ Maximum output of hydro resource h (MW)

RL_h Ramping limit of hydro resource h (MW/hour)

B_h Weekly energy budget of hydro resource h (MWh)

$f_l(t)$ Flow on line l at time t (MW)

$\lambda_{l,z}$ Incidency of line l on zone z

$\underline{F_l}$ Minimum (negative) flow on line l (MW)

$\overline{F_l}$ Maximum flow on line l (MW)

IC_r Installed capacity of renewable resource r (MW)

$PF_r(t)$ Production factor of renewable resource r at time t

$p_r^{curt}(t)$ Curtailment of renewable resource r at time t (MW)

c_r^{curt} Cost of curtailment of resource r ($/MWh)

$SUC_u(t)$ Startup cost of unit u at time t ($)

$SDC_u(t)$ Shutdown cost of unit u at time t ($)

GCS_u Generation cost slope of unit u ($/MWh)

GCI_u Generation cost intercept of unit u ($/hour)

Storage

$v_s(t)$ Storage charge (0)/discharge (1) status at time t

$p_s^c(t)$ Storage rate of charge at time t (MW)

$p_s^d(t)$ Storage rate of discharge at time t (MW)

$\overline{p_s^c}$ Storage max rate of charge (MW)

$\overline{p_s^d}$ Storage max rate of discharge (MW)

$\overline{C_s}$ Storage max state of charge (MWh)

$\underline{C_s}$ Storage min state of charge (MWh)

$C_s(t)$ Storage state of charge at time t (MWh)

η_s^c Storage charge efficiency

η_s^d Storage discharge efficiency

δ_s Storage self discharge

$v_e(t)$ MHD BEV charge (0)/discharge (1) status at time t

$p_e^c(t)$ MHD BEV charge at time t (MW)

$p_e^d(t)$ MHD BEV discharge at time t (MW)

$\overline{P_e}$ MHD BEV charger power rating (MW)

$C_e(t)$ MHD BEV state of charge at time t (MWh)

$\overline{C_e}$ MHD BEV maximum state of charge (MWh)

$\underline{C_e}$ MHD BEV minimum state of charge (MWh)

t_e^{depot} Hour of depot arrival

t_e^{drive} Hour of drive start

C_e^{depot} State of charge at depot arrival (MWh)

C_e^{drive} State of charge at drive start (MWh)

Investment

$IU_u(y)$ Install status of unit u in year y

$IU_u^p(y)$ Planned install status of unit u in year y

$IU_u^b(y)$ Build flag for unit u in year y

$IU_u^r(y)$ Retirement flag for unit u in year y

$IC_s(y)$ Installed capacity of storage resource s in year y

s

$IC^p(y)$ Planned capacity of storage resource s in year y

$IC^b_s(y)$ Built capacity of storage resource s in year y

$ICE_s(y)$ Installed energy capacity of storage resource s in year y

$ICE^p_s(y)$ Planned energy capacity of storage resource s in year y

$ICE^b_s(y)$ Built energy capacity of storage resource s in year y

$IC_r(y)$ Installed capacity of renewable resource r in year y

$IC^p_r(y)$ Planned capacity of renewable resource r in year y

$IC^b_r(y)$ Built capacity of renewable resource r in year y

C^{gen}_y Generation costs in year y

C^m_y Maintenance costs in year y

C^{inv}_y Investment costs in year y

Chapter 2

Decarbonization Planning

2.1 Introduction

From a modeling perspective, power system planning problems are frequently modeled as a mixed-integer linear programming (MILP) problem to exploit the capability of existing optimization solvers such as Gurobi [43], Xpress [126], and CPLEX [49]. Over the 20-year decarbonization planning horizon, decisions regarding resource dispatch are co-optimized with decisions associated with the construction and the retirement of energy resources. While many renewable and storage resources can be modeled using continuous variables, thermal unit behavior (commitment, decommission, and investment) can only be accurately captured with binary variables. This requirement, combined with the extensive time horizon, leads to an issue known as combinatorial complexity - as the planning horizon increases linearly, the associated complexity increases superlinearly (e.g., exponentially). Consequently, when using off-the-shelf commercial software, large MILP planning

problems can quickly become increasingly challenging to solve, with no guarantee that even a feasible solution can be found within a reasonable CPU time.

Currently, the lack of tractability resulting from large time-scale and the inherently discrete nature of the planning problems is typically addressed by relaxing binary restrictions, thereby reducing the MILP problem into a Linear Programming (LP) problem. One example of this is E3's RESOLVE, a resource planning model used by the California Public Utilities Commission as well as by three major California utilities: Southern California Edison, Pacific Gas & Electric, and San Diego Gas & Electric. While LP-relaxed problems may still theoretically be NP-hard, the practical performance of such methods as simplex or barrier methods leads to much-reduced CPU times. However, LP-relaxed versions cannot accurately capture the behavior of thermal units and tend to overestimate their operational flexibility. Consequently, an investment plan based on these simplifications may lead to higher costs or even reliability issues when subjected to the constraints of real-world operations.

In this chapter, we address the California decarbonization planning issue by formulating it as a MILP problem. This approach provides a more accurate model of thermal plants' operations as compared to previous simplifications. Instead of oversimplifying the model, we tailor a surrogate Lagrangian relaxation technique to decompose the problem into manageable subproblems. This method significantly reduces the combinatorial complexity and uses Lagrangian multipliers for iterative coordination of the subproblems. By using the proposed method, investment plans are more consistent with real-world power system operations. The results are compared with RESOLVE, a model used by California state

agencies for decarbonization studies. We find that RESOLVE routinely underestimates the investment required to meet intermediate emissions targets. Further, our model results in lower overall costs on the order of billions through the investment horizon.

2.2 Related Work

While the decarbonization planning problem is relatively new, it is closely connected to the generation expansion problem, which has been the subject of study for decades. Further, as the focus shifts to renewable generation resources, the lines between these problems have blurred in the literature. This section will review related literature in both decarbonization and generation expansion planning (GEP), as well as works related to the proposed surrogate Lagrangian relaxation solution methodology.

2.2.1 Generation Expansion Planning

The decarbonization planning problem is essentially a modification of the GEP, with more emphasis on the construction of green technology and subject to constraints on emissions. Thus, these problems will be reviewed together. GEP approaches generally fall into two groups: reduction in model complexity and alternative optimization methods. Reduction in model complexity refers to relaxations made to the full MILP formulation used within GEP. Within this group, there are a few common streams. The first technique is the relaxation of integrality requirements of binary commitment and investment decisions [47, 118, 113]. The general drawback of LP-relaxation-based methods is an overestimation of the flexibility of thermal and pumped storage units. The second technique is the omission

of detailed technical constraints within the unit commitment formulation. Examples of simplifications include clustered rather than individual modeling of thermal units [73] and the omission of thermal unit operational constraints such as ramping constraints [77, 122, 41]. Furthermore, models may even neglect temporal dependencies opting for load duration curves or similar metrics instead [1, 85, 55]. As a result, two situations may occur: infeasible or non-cost-effective operations. In the former case, the fleet of build-out units satisfying the forecasted peak load may still not satisfy all constraints, such as emissions limits, once more granular operational constraints are imposed on thermal units. In the latter case, feasible solutions may turn out much more expensive during realistic operations due to the need to satisfy more granular operational constraints.

Several open-source generation expansion packages such as Gridpath [87], GenX [95], and ReEDS [96] have also been developed. The packages differ in handling MILP versus LP, clustered versus non-clustered scenarios, and the granularity of unit commitment. Some offer a choice in the level of detail for unit commitment. However, despite allowing detailed generator-level integer unit commitment modeling, these packages do not enhance computational tractability, limiting their application over a significant number of time periods.

This work closely follows Energy + Environmental Economics' (E3) RESOLVE model's data and general composition [34]. The RESOLVE model is used by the California Public Utilities Commission to perform integrated resource planning to meet California's long-term energy policy goals. However, the RESOLVE model presents a linearized and clustered version of unit commitment, in which units are clustered by similar technology.

This simplification massively increases computational efficiency, but has the potential to overestimate the flexibility of thermal units. In clustered unit commitment, individual units' feasible sets are not captured by design making it impossible to accurately apply constraints such as minimum up- and down-time to each unit. Using continuous variables instead of binary ones can result in physically impossible outcomes, such as committing "half of a unit," which is not feasible in reality. This overestimated operational flexibility will likely lead to sub-optimal investment solutions when the full set of constraints are applied in real-world operations. In this work, we overcome these issues by developing a more detailed operational model and adopting a decomposition and coordination methodology for efficient problem solving.

2.2.2 Optimization Methods: Heuristics and Decomposition

Alternative optimization methods include both heuristic methods and decomposition methods, like Benders' decomposition. Heuristic methods have been used to find satisfactory solutions to complex problems where traditional optimization methods may be inefficient or impractical and have been particularly useful when solving large-scale and complex problems. Unlike exact algorithms, such as those employed in software like CPLEX or Gurobi, heuristic methods seek "good enough" solutions within a reasonable timeframe. Notable examples of these methods include genetic algorithm [53], NSGA-II [46] and particle swarm optimization [70, 67]. One of the shortcomings of heuristic methods is the lack of a lower bound to provide a measure of the solution quality - how close a solution is to the global optimal. Moreover, heuristic and metaheuristics methods such as particle swarm optimization may generally suffer from getting trapped at local optima; by the very na-

ture of heuristics, there is generally no systematic or theoretically justified way to improve solutions.

Another approach to improving computational tractability is the use of Benders' decomposition. This method has been deployed for transmission expansion planning in several studies [51, 72, 40]. In [105], Benders' decomposition is applied to generation expansion planning, although it still relies on a genetic algorithm and thus inherits issues of heuristics methods. In general, Benders' decomposition has been less common for generation expansion as compared to transmission expansion. It should be noted that these alternative solution techniques are often coupled with reduction in model complexity as described above.

2.2.3 Optimization Methods: Lagrangian Relaxation

Lagrangian relaxation is an optimization technique that is related to the techniques described above. It has been applied to a range of problems, including unit commitment. This section will review a range of techniques within the family of Lagrangian relaxation to demonstrate the advantages of the selected technique.

Traditional Lagrangian relaxation-based methods faced difficulties such as the high computational effort required to obtain subgradient directions; even if obtained, subgradient directions tend to change drastically and lead to zigzagging of multipliers and slow convergence. Additionally, standard LR utilizing subgradient methods for multiplier updates requires the optimal dual value knowledge for convergence as in Polyak stepsizing [78].

Recently, surrogate Lagrangian relaxation (SLR) [16] has addressed most of these issues, enabling multiplier updates with only "good-enough" subproblem solutions obtained at a time that satisfy the "surrogate optimality condition." This procedure essentially improves the incumbent solution of a relaxed problem (rather than finding the exact optimal solution) in a computationally efficient way due to the drastic reduction of complexity, while still guaranteeing convergence, and reducing the zigzagging of multipliers. From the subproblem-coordination standpoint, the method eliminates the need for optimal dual value knowledge. Moreover, several advancements have been made to the SLR framework, including surrogate absolute-value Lagrangian relaxation (SAVLR) [14] which accelerates convergence through piece-wise linear penalties. These methods have demonstrated success in solving various complex problems in power systems such as unit commitment [108, 123] and beyond [16, 14].

Table 2.1: Comparison of characteristics of LR methods using Polyak's stepsize as well as Polyak's seminal work

Method	Require optimal dual value	Multiplier updating directions
Polyak's seminal work (1969) [78]	Yes	Subgradient
Surrogate subgradient method [133]	Yes	Surrogate subgradient
Surrogate "level-based" Lagrangian relaxation [17]	No	Surrogate subgradient

The surrogate "level-based" Lagrangian relaxation (SLBLR) technique has been recently developed [17], which uses the Polyak stepsize formula with efficient level-value adjustments while guaranteeing convergence of level values to the true optimal dual value leading to overall convergence. This user-friendly approach is robust and reduces the need

for domain knowledge. SLBLR has successfully addressed major issues of previous methods, specifically obviating the need to know the optional function value while still exploiting the geometric (fastest possible) convergence inherent in the Polyak formula. Key features of SLBLR include the decomposition (or "partial optimization" with respect to subsets of decision variables) of the problem into subproblems and the exponential reduction of complexity, enabling efficient coordination through iterative updates of multipliers. Table 2.1 compares the key characteristics of three important Lagrangian relaxation techniques. Due to these characteristics, SLBLR is identified as a technique of particular promise for solving large-scale MILP planning problems. In general, these Lagrangian relaxation-based methods can frequently achieve performance enhancements over commercially available software by a factor of 2-3 times, and as high as 4 orders of magnitude [19].

Lagrangian relaxation-based methods have demonstrated the ability to efficiently exploit the drastic reduction of complexity upon decomposition as well as to efficiently coordinate the subproblems to obtain near-optimal solutions for operation optimization problems in a computationally efficient manner. Lagrangian relaxation is well-suited and is expected to be beneficial for coordinating multiple subsystems and supporting decision-making in complex problems like California's large-scale power system investment for decarbonization as well. For these reasons, SLBLR will be used to solve the decarbonization model developed in Chapter 3.

2.3 Technical Method

In this section, we present a two-timescale formulation for the power system planning problem. At the hourly level, we formulate a unit commitment problem, responsible for committing and dispatching resources to meet load and ancillary service requirements. Meanwhile, the yearly level modeling focuses on investment decisions, governing the construction or retirement of resources. Subsection 2.3.1 presents the formulation for single-week unit commitment. Subsection 2.3.2 then integrates the single-week unit commitment formulation into the multi-week, multi-year planning model.

2.3.1 Single-Week Unit Commitment

Unit commitment (UC) will be considered over a time period T with 1-hour resolution. Every UC variable is indexed temporally by (y, w, t) – a tuple consisting of year, week, and hour. For the sake of conciseness, in this subsection, we formulate all constraints that are generic for any year and week. Thus, we omit y, w indices and let $(y, w, t) \rightarrow (t)$. In the planning model formulation, these constraints will be enforced over all $y \in Y, w \in W$. To circumvent the need to define the initial status of units in UC, time periods will be considered to be consist of a circular set of hours t, in a fashion echoing that of the RESOLVE package [93]. That is, all constraints that link hours are enforced between the end of the period back to the beginning, as shown in Fig. 2.1.

This idea is operationalized by using the modulo operator to track hours. Generally, circular time is calculated as $\tau(t) =$ mod $(t - 1 + T, T)$, ensuring time values stay within the period's range. For constraints which link hours, such as ramping constraints,

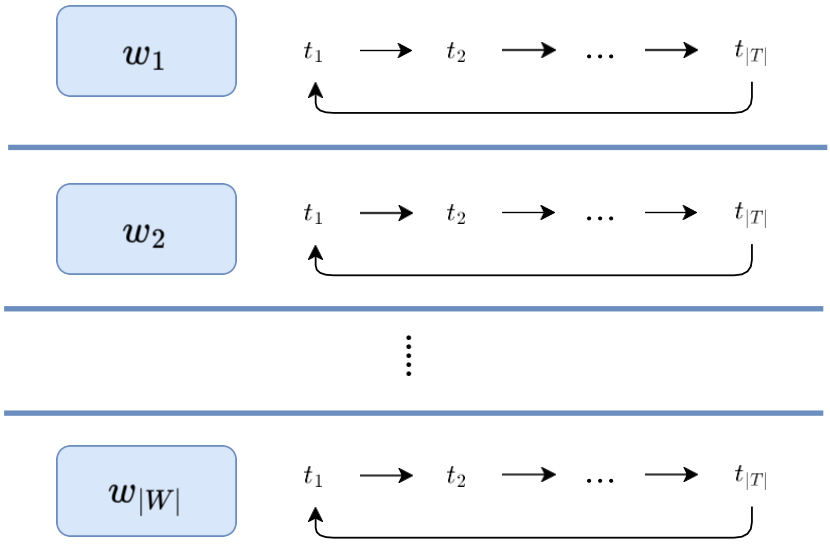

Figure 2.1: Visualization of the circular representation of time.

hours extending past the regular period loop back to the start. This circular approach is only applied to the time variable t. During normal periods (where t ranges from 1 to T), t is simply equal to itself. This time linkage is enforced within each period w, but there is no time linkage between periods and each w is considered disjoint from others.

Generation Resources

The generation fleet consists of five basic types of generation resources: thermal units, renewable resources, firm resources, storage resources, and large hydro resources. These formulations are based on the formulations presented in [93], [120], and [81]. These types of resources and their constraints will now be discussed.

Thermal Units. Thermal units include various types of gas-fired power plants such as combined-cycle gas turbines, fast-start peaking power plants ("peakers"), steam turbines,

aeroderivative combustion turbines, and coal-fired plants. The outputs of thermal units are subject to generation capacity constraints:

$$\underline{P_u} v_u(t) \leq p_u(t) \leq \overline{P_u} v_u(t), \quad \forall t \in T, u \in U, \tag{2.1}$$

minimum up- and down-time constraints:

$$\sum_{n=t+1}^{\tau(t+UT_u)} v_u(n) \geq UT_u[v_u(t) - v_u(t-1)], \ \forall t \in T, u \in U, \tag{2.2}$$

$$\sum_{n=t+1}^{\tau(t+DT_u)} [1 - v_u(n)] \geq DT_u[v_u(t-1) - v_u(t)], \ \forall t \in T, u \in U, \tag{2.3}$$

and ramp-rate constraints that include ramp up- and down-rate constraints:

$$\begin{aligned} p_u(t) \leq\ & p_u(\tau(t-1)) + RU_u v_u(\tau(t-1)) \\ & + SU_u[v_u(t) - v_u(\tau(t-1))] + \overline{P_u}(1 - v_u(t)), \ \forall t \in T, u \in U, \end{aligned} \tag{2.4}$$

shutdown ramp-rate constraints:

$$p_u(t) \leq \overline{P_u} v_u(\tau(t+1)) + SD_u[v_u(t) - v_u(\tau(t+1))], \ \forall t \in T, u \in U, \tag{2.5}$$

and startup ramp-rate constraints:

$$\begin{aligned} p_u(t) \geq\ & p_u(\tau(t-1)) - RD_u v_u(t) \\ & - SD_u[v_u(\tau(t-1)) - v_u(t)] - \overline{P_u}[1 - v_u(\tau(t-1))], \ \forall t \in T, u \in U. \end{aligned} \tag{2.6}$$

Thermal units incur fuel costs as well as startup/shutdown costs. The fuel costs are modeled as a linear function of the commitment status and generation level: $GCI_u \cdot v_u(t) + GCS_u \cdot p_u(t)$. Additionally, there are fixed costs for starting up (suc_u) and shutting

down (sdc_u). These fixed costs are applied each time the unit is turned on or off:

$$SUC_u(t) = max(0, v_u(t) - v_u(\tau(t-1))) \cdot suc_u, \ \forall t \in T, u \in U, \tag{2.7}$$

$$SDC_u(t) = max(0, v_u(\tau(t-1)) - v_u(t)) \cdot sdc_u, \ \forall t \in T, u \in U. \tag{2.8}$$

Renewable Resources. Renewable resources are utility-scale solar and wind farms, and an agglomeration of behind-the-meter customer solar. Renewable resources are described by generation shape generally represented as a certain percentage of the resource's rated capacity $PF_r(t)$ in a given hour, depending on the solar irradiance or wind speed. The power output of a renewable resource is equal to this amount minus any curtailment. Curtailment incurs a cost related to the loss of production tax credits c_r^{curt} as:

$$p_r(t) = IC_r \cdot PF_r(t) - p_r^{curt}(t), \ \forall t \in T, r \in R \tag{2.9}$$

Firm resources. Firm resources, such as nuclear, small hydro, biofuel, geothermal, and combined heat and power, are grouped together with renewable resources in our model. These firm resources generate a consistent amount of power every hour. Unlike other types of resources, they are neither schedulable nor curtailable.

Large Hydro Units. Large hydro units are dispatchable hydro resources, which are subject to weekly energy budget constraints:

$$\sum_{t \in T} p_h(t) \times 1 \, hour \leq B_h, \ \forall h \in H, \tag{2.10}$$

ramp limits:

$$p_h(t) - RL_h \leq p_h(\tau(t+1)) \leq p_h(t) + RL_h, \ \forall t \in T, h \in H, \tag{2.11}$$

and generation capacity constraints:

$$\underline{P_h} \leq p_h(t) \leq \overline{P_h}, \quad \forall t \in T, h \in H. \tag{2.12}$$

Storage Resources. Storage resources include pumped and battery storage. These resources can charge using overgeneration and discharge to serve undergeneration. Storage resources are defined and limited by their power rating (MW) and energy rating (MWh). To enforce minimum duration, particularly for pumped storage resources, storage resources have a binary discharge (1) or charge (0) status. Charge and discharge rates are modeled separately to account for efficiency losses, and are subject to minimum:

$$0 \leq \underset{s}{p^c}(t) \leq (1 - v_s)\underline{p^c_s}, \quad \forall t \in T, s \in S \tag{2.13}$$

and maximum rate constraints:

$$0 \leq \underset{s}{p^d}(t) \leq v_s\underline{p^d_s}, \quad \forall t \in T, s \in S. \tag{2.14}$$

Storage resources are subject to battery capacity limits - a minimum and maximum state of charge constraint:

$$\underline{C_s} \leq C_s(t) \leq \overline{C_s}, \quad \forall t \in T, s \in S. \tag{2.15}$$

Storage resource state of charge balance is governed by:

$$C_s(t) = (1 - \delta_s)C_s(\tau(t-1)) + \left(p^c_s(t)\eta^c_s - p^d_s(t)\frac{1}{\eta^d_s}\right) \times 1\,hour, \quad \forall t \in T, s \in S. \tag{2.16}$$

Zones and Lines

A zonal unit commitment model is employed to represent the California Independent System Operator (CAISO) and the Western Interconnection, encompassing distinct

zones: CAISO, three balancing authorities within California (LADWP, IID, BANC), and two out-of-state aggregations (NW, SW). Each zone is interconnected with at least one other zone through transmission lines, with power transfer between zones modeled as decision variables. This approach, often referred to as a 'transportation' transmission model, omits the need for detailed power flow analyses, reducing the computational complexity of the problem while facilitating a comprehensive representation of the interconnected system.

The incidence of line l on zone z is captured by $\lambda_{l,z}$ taking 0 values for non-incidence, and values of 1 and -1 to denote reference directions of line l into and out of zone z, respectively. Transmission is associated with a transmission cost c_l^{tx} which captures wheeling costs, and can be derived from Open Access Transmission Tariffs [94]. The power flows are subject to line capacity constraints:

$$\underline{F_l} \leq f_l(t) \leq \overline{F_l}, \quad \forall t \in T, l \in L. \tag{2.17}$$

Load and Reserve Requirements

The ancillary service requirements must be satisfied with resources in CAISO. Each reserve product is modeled individually. Superscripts fr, sr, $lf\uparrow$, $reg\uparrow$ denote frequency response, spinning reserve, load following up, and regulation up products, while $lf\downarrow$, $reg\downarrow$ denote the load following down and regulation down products. The exact amount of each product in this formulation is specific to CAISO, but the formulation is readily adaptable to the reserve requirements of other system operators. Products can only be supplied up to the headroom and footroom available for thermal units:

$$p_u^{fr}(t) + p_u^{sr}(t) + p_u^{lf\uparrow}(t) + p_u^{reg\uparrow}(t) \leq P_{\bar{u}} v_u(t) - p_u(t), \quad \forall t \in T, u \in U, \tag{2.18}$$

$$p_u^{lf\downarrow}(t) + p_u^{reg\downarrow}(t) \leq p_u(t) - \underline{P}_u v_u(t), \quad \forall t \in T, u \in U, \tag{2.19}$$

and hydro units:

$$p_h^{fr}(t) + p_h^{sr}(t) + p_h^{lf\uparrow}(t) + p_h^{reg\uparrow}(t) \leq P_{\overline{h}} - p_h(t), \quad \forall t \in T, h \in H, \tag{2.20}$$

$$p_h^{lf\downarrow}(t) + p_h^{reg\downarrow}(t) \leq p_h(t) - \underline{P}_h, \quad \forall t \in T, h \in H. \tag{2.21}$$

Thermal provision of frequency response is limited to 8% of the current output as:

$$p_u^{fr}(t) \leq 0.08 p_u(t), \quad \forall t \in T, u \in U. \tag{2.22}$$

For products other than frequency response, ramping limits must also be obeyed for thermal units:

$$p_u^{sr}(t) + p_u^{lf\uparrow}(t) + p_u^{reg\uparrow}(t) \leq RU_u/6, \quad \forall t \in T, u \in U, \tag{2.23}$$

$$p_u^{lf\downarrow}(t) + p_u^{reg\downarrow}(t) \leq RD_u/6, \quad \forall t \in T, u \in U, \tag{2.24}$$

and hydro units:

$$p_h^{sr}(t) + p_h^{lf\uparrow}(t) + p_h^{reg\uparrow}(t) \leq RL_h/6, \quad \forall t \in T, h \in H, \tag{2.25}$$

$$p_h^{lf\downarrow}(t) + p_h^{reg\downarrow}(t) \leq RL_h/6, \quad \forall t \in T, h \in H. \tag{2.26}$$

Storage can provide each product up to the headroom and footroom of both power capacity:

$$p_s^{fr}(t) + p_s^{sr}(t) + p_s^{lf\uparrow}(t) + p_s^{reg\uparrow}(t) \leq \underline{p}_s^d(t) - p_s^d(t) + p_s^c(t), \quad \forall t \in T, s \in S, \tag{2.27}$$

$$p_s^{lf\downarrow}(t) + p_s^{reg\downarrow}(t) \leq \underline{p}_s^c(t) - p_s^c(t) + p_s^d(t), \quad \forall t \in T, s \in S, \tag{2.28}$$

and energy capacity:

$$p_s^{fr}(t) + p_s^{sr}(t) + p_s^{lf\uparrow}(t) + p_s^{reg\uparrow}(t) \leq C_s(t) - \underline{C}_s, \quad \forall t \in T, s \in S, \tag{2.29}$$

$$p_s^{lf\downarrow}(t) + p_s^{reg\downarrow}(t) \leq \overline{C_s} - C_s(t), \quad \forall t \in T, s \in S. \tag{2.30}$$

Up to half of the load following down can be provided via curtailable renewable resources:

$$p_r^{lf\downarrow}(t) \leq 0.5LF\downarrow(t), \quad \forall t \in T, r \in R \tag{2.31}$$

up to the available footroom:

$$p_r^{lf\downarrow}(t) \leq IC_r \cdot PF_r(t) - p_r^{curt}(t) - p_r(t), \ \forall t \in T, r \in R. \tag{2.32}$$

For every hour, 770MW must be held for frequency regulation. Regulation up, regulation down, and spinning reserve each require 1% of the CAISO load. Load following up and down requirements are based upon renewable penetration scenario analysis carried out by E3:

$$\sum_{u \in U_z} p_u^{fr}(t) + \sum_{h \in H_z} p_h^{fr}(t) + \sum_{s \in S_z} p_s^{fr}(t) \geq 770MW, \quad t \in T, z = 0, \tag{2.33}$$

$$\sum_{u \in U_z} p_u^{sr}(t) + \sum_{h \in H_z} p_h^{sr}(t) + \sum_{s \in S_z} p_s^{sr}(t) \geq 0.01L_z(t), \quad t \in T, z = 0, \tag{2.34}$$

$$\sum_{u \in U_z} p_u^{reg\uparrow}(t) + \sum_{h \in H_z} p_h^{reg\uparrow}(t) + \sum_{s \in S_z} p_s^{reg\uparrow}(t) \geq 0.01L_z(t), \quad t \in T, z = 0, \tag{2.35}$$

$$\sum_{u \in U_z} p_u^{reg\downarrow}(t) + \sum_{h \in H_z} p_h^{reg\downarrow}(t) + \sum_{s \in S_z} p_s^{reg\downarrow}(t) \geq 0.01L_z(t), \quad t \in T, z = 0, \tag{2.36}$$

$$\sum_{u \in U_z} p_u^{lf\uparrow}(t) + \sum_{h \in H_z} p_h^{lf\uparrow}(t) + \sum_{s \in S_z} p_s^{lf\uparrow}(t) \geq LF\uparrow(t), \quad t \in T, z = 0, \tag{2.37}$$

$$\sum_{u \in U_z} p_u^{lf\downarrow}(t) + \sum_{h \in H_z} p_h^{lf\downarrow}(t) + \sum_{s \in S_z} p_s^{lf\downarrow}(t) + \sum_{r \in R_z} p_r^{lf\downarrow}(t) \geq LF\downarrow(t), t \in T, z = 0. \tag{2.38}$$

Each zone within the ISO area must satisfy the zonal power balance constraints (2.39) as:

$$\sum_{u \in U_z} p_u(t) + \sum_{s \in S_z} [p^d(t) - p^c(t)] + \sum_{r \in R_z} p_r(t) + \sum_{h \in H_z} p_h(t) + \sum_{l \in L} \lambda_{l,z} f_l(t) = L_z(t), \ t \in T, z \in Z. \quad (2.39)$$

Unit Commitment Objective

The objective of unit commitment is to minimize the startup and shutdown costs, fuel costs, transmission costs, and renewable curtailment costs as:

$$C^{gen} = \sum_{t \in T} \sum_{u \in U} \{ SUC_u(t) + SDC_u(t) + (GCI_u \cdot v_u(t) + GCS_u \cdot p_u(t)) \times 1\ hour + \sum_{t \in T} \sum_{l \in L} f_l(t) \cdot c_l^{tx} + \sum_{t \in T} \sum_{r \in R} c_r^{curt} \cdot p_r^{curt}(t) \times 1\ hour. \quad (2.40)$$

With this, we can write the unit commitment optimization problem as the minimization of the cost 2.40 subject to all operational constraints.

$$\min C^{gen} \quad (2.41)$$

$$s.t., (2.1) - (2.39)$$

In the next section, we take this basic formulation and transform it into a broader planning model by optimizing over a set of years and weeks, and allowing for resource investment alongside dispatch.

2.3.2 Decarbonization Planning

The objective of decarbonization planning is to minimize the total cost associated with meeting carbon emissions and renewable generation goals from power generation

through 2045. The total cost encompasses both annual energy expenses (including maintenance) and the capital costs of constructing new capacity of zero-carbon resources and lower-carbon power plants.

In the present study, it is assumed that the development of new resources will be restricted to the CAISO territory. However, the problem formulation remains broadly applicable. The portfolio of potential resources encompasses wind, solar, and energy storage at various sites, as well as geothermal, biomass, and several categories of gas-fired power plants. Decisions regarding the retirement of existing thermal units may also be considered, with certain technologies, such as coal and nuclear, already possessing predetermined decommissioning schedules. In this section, all constraints will be enforced for each year $\forall y \in Y$, and for CAISO only $z = 0$ where zonal subsets of resources are concerned.

First, let us define the build status of thermal units. Let $IU_u(y)$ represent the binary operational status of unit u in year y, where 1 indicates the unit is operational and may be turned on. $IU^p_u(y)$ denotes the planned status of unit u, where 1 signifies the unit is operational, and 0 indicates the unit is decommissioned or not yet constructed. $IU^b_u(y)$ and $IU^r_u(y)$ define whether the unit is built and retired, respectively, in year y. Consequently, the relationship between the planning layer and the unit commitment layer is expressed as:

$$IU_u(y) \geq v_u(y, w, t), \ \forall u \in U, \ w \in W, \ t \in T, \tag{2.42}$$

which constrains the unit commitment status v_u to turn on only if it is operational as:

$$IU_u(y) = IU^p_u(y) + \sum_{Y=1}^{y} (IU^b_u(Y) - IU^r_u(Y)). \tag{2.43}$$

Installation of additional capacity of renewable generation units (indexed by r) and storage units (indexed by s) is considered to be a continuous variable and the logic capturing the installation capacity follows that of (2.43) as:

$$IC_s(y) = IC_s^p(y) + \sum_{Y=1}^{\underline{y}} (IC_s^b(Y) - IC_s^r(Y)), \tag{2.44}$$

$$ICE_s(y) = ICE_s^p(y) + \sum_{Y=1}^{\underline{y}} (ICE_s^b(Y) - ICE_s^r(Y)). \tag{2.45}$$

New capacity of these types can be installed in discrete amounts on the order of tens of watts, which is effectively continuous compared to the scale at which these resources are installed. Storage capacity has two components, one each for energy capacity (MWh), denoted as ICE, and power capacity (MW), denoted as IC. The total installed capacity of each renewable resource is defined in a similar way to that of the thermal units, with the chief difference being the decision variables become continuous instead of binary:

$$IC_r(y) = IC_r^p(y) + \sum_{Y=1}^{\underline{y}} (IC_r^b(Y) - IC_r^r(Y)). \tag{2.46}$$

The installed capacities of these units impact the unit commitment formulation in different ways. The maximum rate of charge/discharge is equal to the rated capacity, represented as $p_s^{\underline{c}}(y) = p_s^{\underline{d}}(y) = IC_s(y)$. The maximum/minimum state of charge corresponds to the rated energy capacity, multiplied by a percentage factor associated with the operational range, denoted as $\overline{C}_s(y) = ICE_s(y) \cdot \epsilon_s^{max}$. For batteries, these values typically range between 0.1 and 0.9 for degradation considerations [132], while for pumped storage, they are closer to 0 and 1 [114]. Regarding renewables, $IC_r(y)$ defines IC_r for the specified year in (2.9).

Let the cost of energy generation for year y be denoted as C_y^{gen}. This cost is composed of the same components as the function being minimized in the unit commitment given by (2.40). Within the planning problem, unit commitment is performed for a sample of several weeks per year. Each sampled week is assigned a weight ω_w that conveys its representative factor to the annual load profile, with the sum of these weights amounting to 52, corresponding to the number of weeks in a year. The yearly unit commitment cost is calculated as the weighted sum of the weekly unit commitment costs. It is also weighted by the yearly weight ω_y, which encodes the number of years represented by y. Consequently, the cost of generation in year y and week w in (2.40) is expressed as $C_{y,w}^{gen}$, and the annual generation costs can be written as:

$$C_y^{gen} = \omega_y \sum_{w \in W} \omega_w \cdot C_{y,w}^{gen}. \tag{2.47}$$

Yearly maintenance costs are considered as a function of the installed capacity and the cost of maintaining a given technology. Renewables have a single cost component expressed in \$/MW. Thermal units have a single cost component in \$/unit. Storage has two maintenance cost components, for rated energy $c_{s,E}^m$ and rated power $c_{s,P}^m$, expressed in \$/MWh and \$/MW, respectively. The cost of maintenance for the year y is then:

$$C_y^m = \omega_y \cdot \sum_{u \in U} IU_{u,y} \cdot c_l^m + \sum_{s \in S} ICE_{s,y} \cdot c_s^{m,E} + \sum_{s \in S} IC_{s,y} \cdot c_s^{m,P} + \sum_{k \in K} IC_{k,y} c^m{}_k + \sum_{h \in H} IC_{h,y} \cdot c^m{}_h \ . \tag{2.48}$$

Lastly, let us consider the investment costs for constructing new resources. Annualized costs are assessed for every year after a resource is constructed. Each thermal technology is associated with an annualized capital cost per unit, denoted as $c_{y,u}^{cap}$. Similarly, storage and

renewable technologies have an annualized capital cost per megawatt ($/MW), represented as $c_{y,s}^{cap,P}$ and $c_{y,r}^{cap}$, respectively. Additionally, storage has an annualized capital cost for energy capacity, expressed as $c_{y,s}^{cap,E}$:

$$C_y^{inv} = \sum_{\gamma=y}^{|Y|} \omega_\gamma \cdot \left(\sum_{u \in U} (IU_{u,y}^{b}(y)) \cdot c_u^{cap} + \sum_{s \in S} (IC_s^{b}(y)) \cdot c_s^{cap,P} + \sum_{s \in S} (ICE_s^{b}(y)) \cdot c_s^{cap,E} + \sum_{r \in R} (IC_r^{b}(y)) \cdot c_r^{cap} \right). \quad (2.49)$$

The objective function of decarbonization planning is:

$$\min O = \min \sum_{y \in Y} C_y^{gen} + C_y^{m} + C_y^{inv} . \quad (2.50)$$

The planning process in decarbonization is subject to various constraints, which are central to formulating effective strategies. These constraints comprise emissions targets, renewable energy penetration, and system reliability. In this study, the analysis focuses solely on the constraints utilized by CAISO, excluding other balancing authorities. Thus, for notational clarity, the subscript z denoting zone will be used, with the specification that $z = 0$ for these planning constraints.

Emission Limits. Carbon emissions are generated when energy is produced by thermal plants. CAISO is subject to an emissions constraint specifying that the emissions associated with all generation within CAISO, as well as emissions associated with imports, must be less than the emissions target for year y, E_y. Given the emissions associated with unit u in tons/MW as e_u, and the emissions associated with imports e_l, we can examine the relationship between emissions and energy generation:

$$E_y \geq \sum_{w \in W} \omega_w \cdot \sum_{t \in T} \sum_{u \in U_z} e_u \cdot p_u(y, w, t) + \sum_{l \in L} e_l \cdot \max(0, \lambda_{l,z} f_l(y, w, t)) \Big) . \quad (2.51)$$

Only imports count towards the emission constraint, and exports should not count to reduce the emissions. Thus, the contributing emissions are lower-bounded by 0.

Renewable Portfolio Standards. In addition to the net-zero emissions target by 2045, renewable portfolio standards (RPS) impose interim requirements on the proportion of electricity generation from carbon-free resources, based on a percentage of CAISO's annual load. The majority of renewable energy sources qualify for RPS, with the notable exceptions of combined heat and power (CHP) and nuclear power, which are grouped with renewables due to their similar generation attributes. The following constraint ensures that the renewable portfolio standards (RPS) are met for each year $y \in Y$ as:

$$RPS_y \cdot \sum_{w \in W} \omega_w \sum_{t \in T} L_z(y, w, t) \leq \sum_{w \in W} \omega_w \cdot \sum_{r \in R} \sum_{t \in T} p_r(y, w, t) \cdot RPS_r^{eligible}, \tag{2.52}$$

where binary variable $RPS_r^{eligible}$ indicates whether a renewable source in the set R meets the RPS criteria. Accordingly, $RPS_y \cdot \sum_{w \in W} \omega_w \sum_{t \in T} L_0(y, w, t)$ represents the required amount of electricity generation from eligible renewable resources for the year y, based on a percentage (RPS_y) of the total annual load in CAISO.

Planning Reserve Margin. CAISO must also satisfy reliability requirements, particularly the planning reserve margin (PRM). These requirements ensure that the portfolio, even with high renewable penetration, can meet energy demands. The PRM guarantees that the peak load, with some additional headroom, in a given year is satisfied by the installed capacity. Each resource contributes to the PRM by a fraction of its installed capacity.

For thermal units and large hydro capacity, it is modified by the net qualifying capacity fraction (NQC). Wind and solar are modified by the effective load-carrying capacity (ELCC), a fraction that decreases as renewable penetration increases. Essentially, ELCC

encodes the complex nature of variable generation, such as the behavior that increasing installation of renewables, particularly solar, may have little effect on the peak net load, which occurs in California in the early evening when solar generation is rapidly decreasing. The ELCC is approximated by a 3-dimensional piece-wise linear surface, with axes representing the capacity of wind and solar, respectively. This ELCC surface is visualized in Fig. 2.2. Each solar or wind resource in CAISO contributes to the total axis value by its capacity and a multiplier, denoted as $mult^{axis}$.

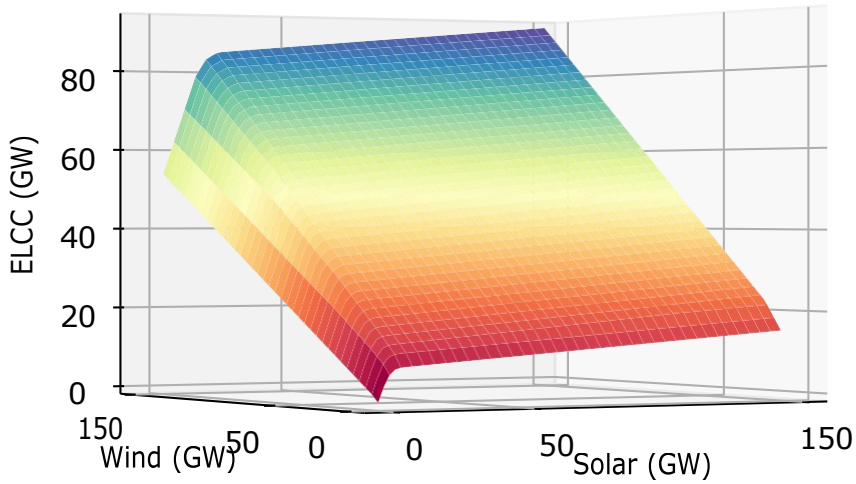

Figure 2.2: Visualization of the ELCC surface.

The resulting 3D surface is comprised of flat segments referred to as facets, which together create the overall shape of the ELCC representation. The facets simplify the complex relationships between wind, solar, and ELCC by breaking down the surface into a

series of linear segments, making it easier to analyze and understand the impact of different resource capacities on the overall system [34].

Then, the ELCC value of each facet is characterized by an intercept and slope on each axis. In the optimization, the final value for this piece-wise linear ELCC is determined by setting the ELCC as the minimum of each facet value. With R_{wind} and R_{solar} denoting the subsets of CAISO wind and solar resources, the following equation computes the ELCC for each year y based on the capacity of wind and solar resources in CAISO:

$$ELCC_y \leq \sum_{r \in R_{z,wind}} IC_{y,r} \cdot mult^{axis}_{y,r} \cdot slope_{y,wind,f} + \sum_{r \in R_{z,solar}} IC_{y,r} \cdot mult^{axis}_{y,r} \cdot slope_{y,solar,f} + intercept_{y,f}, f \in Facets. \quad (2.53)$$

Similarly, storage contributes through the 4-hour capacity. The ELCC of storage resources is characterized by a two-dimensional piecewise linear surface:

$$ELCC_{y,s} \leq intercept_{y,s,f} + \sum_{s \in S_z} \min\left(IC_{y,s}, \frac{ICE_{y,s}}{4\,hours}\right) \cdot mult_{y,s} \cdot slope_{y,s,f}, f \in Facets. \quad (2.54)$$

The following equation ensures that the planning reserve margin (PRM) for each year y is met, taking into account the contributions from different types of resources, such as thermal units, storage, wind, and solar:

$$PRM_y \leq \sum_{u \in U_z} IU_{y,u} \overline{P_u} NQC_u + ELCC_{y,s} + ELCC_y + \sum_{h \in H_z} IC_{y,h} NQC_h. \quad (2.55)$$

Chapter 3

Computational Techniques

The problem formulated in the previous chapter belongs to the class of MILP problems. MILP problems suffer from *combinatorial complexity* – because of binary decision variables, as the problem size increases, the number of possible solutions increases *super-linearly* thereby leading to a drastic increase in the computational effort. In this chapter, to efficiently solve the problem, the recently proposed decomposition and coordination approach SLBLR [17] is deployed to exploit the super-linear reduction of complexity upon the decomposition and the geometric convergence potential inherent to Polyak's step-sizing formula for the fastest coordination possible to obtain near-optimal solutions in a computationally efficient manner.

The decomposition is operationalized by relaxing coupling zonal power balance constraints (2.39). Given the additivity of the constraints and the objective function, the relaxed problem is separable into individual unit subproblems. Subproblem solutions are first coordinated through the iterative update of Lagrangian multipliers $\boldsymbol{\Lambda}$. After the mul-

tipliers have converged sufficiently, the primal problem is solved while fixing the majority of the binary variables to their *subproblem optimal* values. The process for solving subproblems is described next.

Relaxed Problem. After relaxing coupling constants, the relaxed problem is broken into subproblems decomposed by groups of thermal units. While solving each subproblem, all other thermal units' commitment status, power levels, and build status are fixed at their value in the previous iteration. Each subproblem is optimized with respect to all the variables (both discrete and continuous) associated with the group of thermal units selected, including commitment status, dispatch, reserve product supply, and investment.

After relaxing zonal power balance (2.39), which couples thermal units, the relaxed problem becomes:

$$\min \mathrm{L} = \min_{\{\mathbf{p},\mathbf{v},\mathbf{I}\}} \quad \mathrm{O} + \boldsymbol{\Lambda} \cdot \mathbf{R} + c \cdot \quad \mathbf{R} \quad _1 \quad , \tag{3.1}$$

$$s.t., (2.1) - (2.38), \ \forall y \in Y, w \in W,$$

$$(2.43) - (2.46), (2.51) - (2.55), \ \forall y \in Y,$$

where $\mathbf{R} = [r_z(y, w, t), \forall z \in Z, y \in Y, w \in W, t \in T]$ is a vector of zonal power balance constraint violations across all zones and timepoints. The violation is given by $r_z(y, w, t) = \sum_{u \in U_z} p_u(y, w, t) + \sum_{s \in S_z} [p_s^d(y, w, t) - p_s^c(y, w, t)] + \sum_{r \in R_z} p_r(y, w, t) + \sum_{h \in H_z} p_h(y, w, t) + \sum_{l \in L} \lambda_{l,z} fl(y, w, t) - L_z(y, w, t)$. $\boldsymbol{\Lambda}$ is a vector of Lagrangian multipliers, and c is a penalty coefficient acting on the absolute value of constraint violations. For notational brevity, let **p** represent all power-related variables (including line flows), **v** represent all binary commitment variables, and **I** represent all investment variables.

The solution process is presented in Algorithm 1. Subproblems are formulated by selecting Ω_k - a group of units to be optimized with respect to at iteration k and by fixing decision variables collectively denoted as $\{\mathbf{p}, \mathbf{v}, \mathbf{I}\}$ that do not belong to Ω_k are fixed at previously obtained values $\{\mathbf{p}^{k-1}, \mathbf{v}^{k-1}, \mathbf{I}^{k-1}\}$. In particular, all units are split at random into groups, and these groups are iterated through during the subproblems. In each iteration, all non-thermal-unit variables are also solved.

After one subproblem is solved, the multipliers are updated along "surrogate" subgradient directions, which are violation levels of relaxed constraints, with an appropriate stepsize as follows:

$$\mathbf{\Lambda}^k = \mathbf{\Lambda}^{k-1} + s^k \cdot \tilde{\mathbf{R}}^k. \tag{3.2}$$

Here "tilde" (˜) indicates that the constraint violations are obtained approximately without optimizing the relaxed problem with respect to all the units at a time. Penalty coefficient c acts on the absolute value of constraint violations. Care must be taken to update c since if c is initialized "too large" or grows "too quickly," it can hamper the convergence of multipliers thus severely impacting the iteration time. Further discussion of the role of c and strategies for updating it can be found in [123, 82].

The nature of the problem can contribute to slow convergence since renewable resources are generally dispatched at identical costs. While each renewable project has a unique investment cost, generation is associated with zero marginal cost. Therefore, with high penetration, a large amount of renewable resources respond identically to the multipliers. The dispatch of these resources may jump between maximum and minimum,

and multipliers may oscillate and overshoot optimal multipliers. The mechanism by which this occurs will be elucidated through an example.

Consider a single-hour simplified illustrative example where renewable generation exceeds load, and disregard the penalty coefficient c for simplicity. The optimization problem min L will minimize R by curtailing all generation, for any positive value of the multiplier Λ. Conversely, with a negative Λ, there is no curtailment, resulting in overgeneration. While this is a simplified example, similar behavior is exhibited for the dispatch and construction of resources. Essentially, Lagrangian multipliers serve as "price signals," causing renewable resources with similar costs to exhibit similar responses. This can lead to solutions jumping and, consequently, to the zigzagging of multipliers.

To alleviate the solution jumping issue, specifically, to suppress the jumping of solutions, each continuous power variable is restricted within $\Delta = 500MW$ of its value from the previous subproblem. Experimental results suggest that the value of Δ is not particularly sensitive, with values in the range of 250MW to 1000MW having similar effect in experiments. This parameter is best set experimentally, as it depends on the number of projects, size of projects, relative costs of projects, load, and so on. The suggested range for this parameter is 0.5% to 2% of the average load. If this parameter is too small, dispatch variables will lag behind updated multipliers and not react in proportion, but if it is too large, convergence of multipliers will be poor due to large dispatch oscillations. The procedure also alleviates zigzagging since with suppressed jumping of solutions, the corresponding multiplier-updating directions tend not to change drastically (i.e., become smoother) eventually leading to a smoother update of multipliers. From the multiplier

convergence perspective, to alleviate the overshooting issue, proper stepsize selection plays an important role as explained ahead.

Stepsize Update. The step size, denoted as s^k, plays a pivotal role in the proposed algorithm's convergence. Following the methodology outlined in [17], the step size is computed as follows:

$$s^k = \zeta \cdot \gamma \cdot \frac{\bar{q}^k - L^k}{\|\tilde{\mathbf{R}}\|^2}. \tag{3.3}$$

In this equation, ζ and γ are hyperparameters that are chosen to balance the trade-off between convergence speed and algorithm stability. While in the original work of Polyak [78] $\gamma < 2$, which would be appropriate for standard Lagrangian relaxation that utilizes subgradient directions for multiplier update, in further "surrogate" extensions of LR [133, 17], $\gamma < 1$. Moreover, since level value $\bar{q}^k$ represents the current overestimation of the dual value, ζ is chosen in a way to reduce stepsizes (e.g., $\zeta = 1/2$). Since the method solves one subproblem at a time, γ is chosen to be the reciprocal of the number of subproblems. Overall, as compared to standard LR, this method allows for a more frequent update of multipliers along smoother directions with smaller steps leading to the alleviation of the "overshooting" issue mentioned above.

The step size is initially set using an overestimation of the optimal dual value. The value of $\bar{q}^k$ is not static, rather, as detailed in [17], it undergoes periodic adjustments based on a level-based resetting mechanism, which detects the lack of multiplier convergence. This resetting process is designed to lower $\bar{q}^k$ to approach the actual (dual) optimal value in light of new information obtained during the iterative procedure, thereby refining the overestimation of the dual value and guiding the algorithm toward the optimal multipliers.

In essence, this approach to updating the step size is grounded in dynamism and adaptivity. By making use of gathered information and tuning the step size accordingly, we can expedite convergence and enhance the efficiency and robustness of the overall algorithm. For more information, interested readers are refered to [17, 62].

Algorithm 1: Surrogate "Level-Based" Lagrangian Relaxation

initialize $\boldsymbol{\Lambda} = \mathbf{0}$, c, $\bar{q}_0$, ζ, γ

for $k = 1 \ldots$ **do**

select subproblem units Ω_k ;

solve subproblem $min\, L^k$ with $\boldsymbol{\Lambda}^k$ (3.1) $\rightarrow \tilde{\mathbf{R}}^k$;

if $\| \tilde{\mathbf{R}}^k \| < threshold$ **then**

break ;

end

if $\boldsymbol{\Lambda}$ *not converging* **then**
reset $\bar{q}_k$

end

$s^k \leftarrow \zeta \cdot \gamma \cdot \frac{\bar{q}_k - L^k}{\| \tilde{\mathbf{R}} \|^2}$;

$\boldsymbol{\Lambda}^{k+1} \leftarrow \boldsymbol{\Lambda}^k + s^k \cdot \tilde{\mathbf{R}}^k$;

end

Solve primal problem with heuristic partially-fixed binary variables

Feasible Solution. Although the method is guaranteed to converge towards the optimal (dual) solution [62], obtaining zero constraint violations through multipliers alone is often difficult or impossible. As a result, a heuristic is necessary to find a feasible solution to the

primal problem. Once the multipliers have converged such that the constraint violations are sufficiently low, the primal problem is solved by fixing the commitment status of most units to the values obtained in the relaxed problem. The above heuristic presents a fundamental trade-off: constraining fewer variables in the primal problem requires greater computational effort but may lead to lower overall costs. Nevertheless, as empirical evidence suggests in the chapter ahead, by solving the entire primal problem but with respect to only a small number of units, the primal problem is much easier to solve and can generally be solved to near-optimality.

Chapter 4

Results

4.1 Numerical Study

4.1.1 Experimental Setup

The decarbonization model is based on the Western Interconnection. The model incorporates CAISO and 5 other zones: LADWP, BANC, IID, and aggregations of non-California balancing authorities in the Northwest and Southwest. The data used within the model is taken from the RESOLVE implementation published by CPUC [34]. To reduce computational complexity, rather than model every year, and every hour of every year, it is typical to model representative periods. We model 8 weeks per year, and biennially from 2023 to 2045. The model has 91.1 million continuous variables and 11.8 million binary variables, roughly 40x more total variables than in RESOLVE. We use Gurobi on a workstation with an AMD Ryzen Threadripper 3970X CPU to solve the subproblems and primal problem. The total solution process takes less than 48 hours. Each iteration accounts

for approximately 25 minutes, although the exact time varies considerably and tends to slow as the multipliers converge. Fig. 4.1 shows the average absolute value of the power balance constraint violation over iterations. A feasible solution is attained at 75 iterations. At this point, the constraint violations are low enough that a solution to the primal problem can be recovered. As a direct comparison, the model is also deployed on the same set of representative days as RESOLVE. The solution process in this case takes approximately 24 hours, with a feasible solution obtained at 120 iterations. Finally, the models are compared over representative days with a higher emissions scenario, representing an emissions limit of 38 MMT by 2030, instead of 30 MMT as in the other scenarios. A feasible solution is obtained at 90 iterations and 18 hours. Due to the number of modeled timepoints being substantially reduced, at 37 days per year as opposed to 56 days per year in the case of representative weeks, the time for each iteration is much shorter. Although these solution times are slower than continuous relaxation-based solution models, which can often be run in less than 12 hours, these solution times are not unreasonable in the context of long-term planning.

To reduce computational complexity, rather than modeling unit commitment for full years, representative weeks are sampled and used instead. The sampling method follows the method used in RESOLVE studies [48]. We diverge from RESOLVE by modeling representative weeks rather than representative days, and the advantages will be demonstrated in the results. Histogram bins $b \in B$ are created from features of the data, most importantly the distributions of hourly loads. Then, an optimization problem is solved to select weeks and corresponding weights which minimizes the Manhattan distance of bin frequency in the

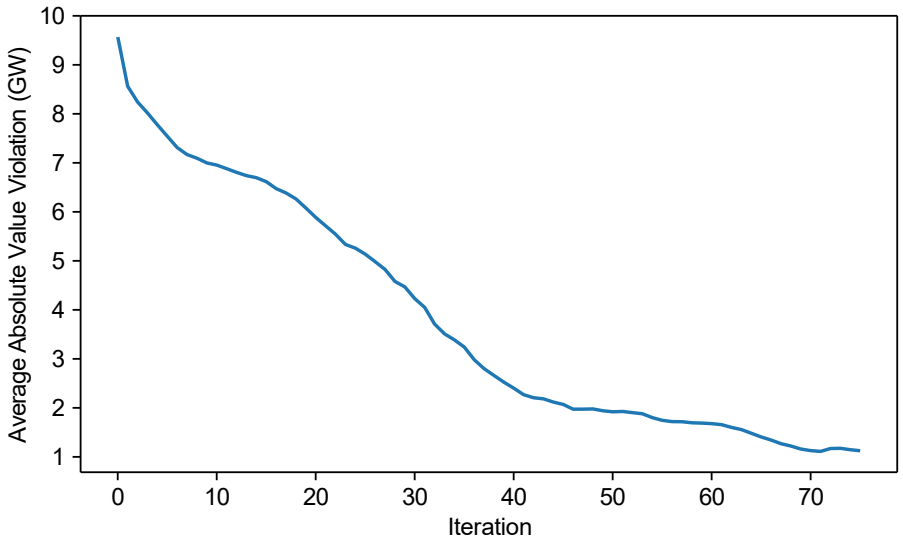

Figure 4.1: Minimization of constraint violation as Lagrangian multipliers converge.

full year to the representative weeks.

$$\min \sum_{b \in B} YearlyFreq_b - \sum_{w \in weeks} \omega_w \cdot WeeklyFreq_{w,b} \quad (4.1)$$

The optimization horizon is through 2045, and financing through 2065. Due to the sampling of weeks and years, it is necessary to weight the weekly and yearly components to ensure that costs assessed over different timescales are accurately balanced. In essence, weights ensure that the costs according to representative periods and years are scaled to approximate full timespace costs. The capital costs in (2.49) are amortized. Then, these costs are assessed for every subsequent year to the decision in the optimization horizon. Note that this also accounts for residual value of new capacity. All costs are in real dollars, but the yearly weight should capture the time value of money, with an assumed discount

rate of 5%, and the number of real years represented by the sampled year:

$$\omega_y = \left(\frac{1}{1.05}\right)^{(Y_y - Y_0)} \cdot (Y_{(y+1)} - Y_y). \tag{4.2}$$

We compare the investment decisions of the proposed model with those of RESOLVE. Maintenance and investment cost comparisons can readily be made. However, for operational cost comparisons, it is necessary to analyze RESOLVE's investment decisions under the detailed MILP unit commitment model. For both models, in each year, investment decisions are fixed and unit commitment is solved, with a modified emissions constraint as described in (4.3). This modified constraint allows for violations of the emissions limits E_y^{vio}, penalized by a large value M. The exact value is not critical, and the optimization should tend towards minimizing emissions even without this penalty, because investment decisions are not considered here and carbon-free resources generate with zero marginal cost. The same SLBLR method is used to solve these dispatch problems. In each case, the difference between the integer feasible solutions and the linear-relaxed objective value ranges from 0.5% to 1.5% of the linear-relaxed objective value, and no pattern observed towards any set of solutions achieving lower gaps. This metric can be considered similar to the MIP gap, used by commercial solvers like Gurobi to quantify the quality of a mixed-integer solution. However, it is not equivalent, because the lower bound used by the MIP gap is updated during branch-and-bound, while the lower bound is fixed in the above

metric.

$$\min_{y} C_{y}^{gen} + M \cdot E_{y}^{vio} \tag{4.3}$$

$$s.t., (2.1) - (2.39), \forall w \in W,$$

$$E_y + E_y^{vio} \geq \sum_{w \in W} \omega_w \cdot \left(\sum_{t \in T} \sum_{u \in U_z} e_u \cdot p_u(y, w, t) + \sum_{l \in L} e_l \cdot \max(0, \lambda_{l,z} f_l(y, w, t)) \right)$$

4.1.2 Results

We first compare the fleet of RESOLVE to the fleets of our model when optimized over both the same representative days as RESOLVE, and the proposed representative weeks. Cost comparisons are presented using both the proposed representative weeks and RESOLVE's representative days. The proposed method with representative weeks will be generally referred to as SLBLR, with the suffix of "day" or "week" added where the distinction is necessary. In each of these cases, operational costs are determined using (4.3) over representative weeks. Including the SLBLR Day case in these comparisons allows for a better demonstration of which component of the improvement comes from more detailed MILP modeling and which component comes from more extensive modeling of weeks.

With the goal of decarbonization by 2045 in mind, a comparison of CAISO fleets in 2045 is shown in Figure 4.2. Compared to RESOLVE, our model builds fewer solar and storage resources. However, this is not the case throughout the optimization horizon. Figure 4.3 shows the fleet composition from 2023 to 2045. Our model begins investing in

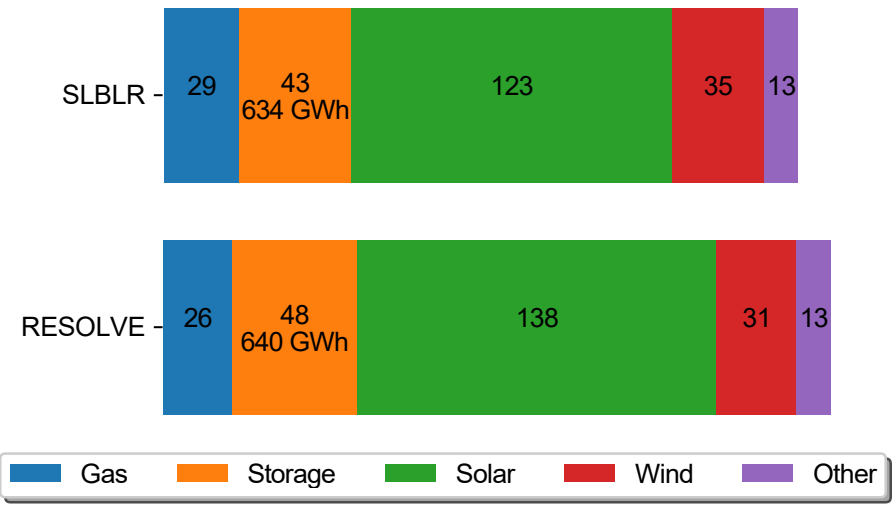

Figure 4.2: Comparison of CAISO fleet in 2045.

additional capacity, especially solar, much earlier than RESOLVE, and holds more total capacity than RESOLVE until 2045.

The likely explanation for the lower build of renewable resources in RESOLVE during this transitional years is that the overestimated flexibility of thermal resources underestimates the running costs of thermal units, as well as their emissions. In fact, RESOLVE's investment decisions, when applied to the SLBLR model, cannot satisfy emissions constraints between year 2027 and 2045, and overemits anywhere from a few thousand to several million tons of GHG per year. Due to this underestimation, RESOLVE chooses to defer investment in renewables to later years. As evidence of the cost underestimation of running gas generators, Table 4.1 shows the shutdown, startup, and fuel costs for CAISO gas generators in 2022, before substantial investment occurs. The costs shown for RESOLVE are with respect to their linearized, clustered unit commitment formulation. RESOLVE underestimates fuel costs by roughly 20% and, vastly underestimates startup and shutdown

costs. Thus, this may explain why RESOLVE's investment plan leans more heavily on gas units. Similarly, RESOLVE drastically underestimates the emissions in 2022, although it is still within emissions limits. California Air Resources Board estimates emissions of roughly 40MMT from in-state electricity production in 2020, which is much more aligned with our model than RESOLVE. The takeaway is that if RESOLVE results are used to inform policy decisions, it may be difficult to meet intermediate emissions targets due to under-investment in renewable energy and storage.

By a similar token, the differences in wind vs solar investment may be explained by the more accurate modeling of gas generators. When gas generators are modeled more faithfully, wind may have a more complementary load shape. Even today, ramping poses difficulties during early evening hours in which load is increasing and solar generation is rapidly decreasing. With solar capacity more than doubled, this effect will become even more pronounced.

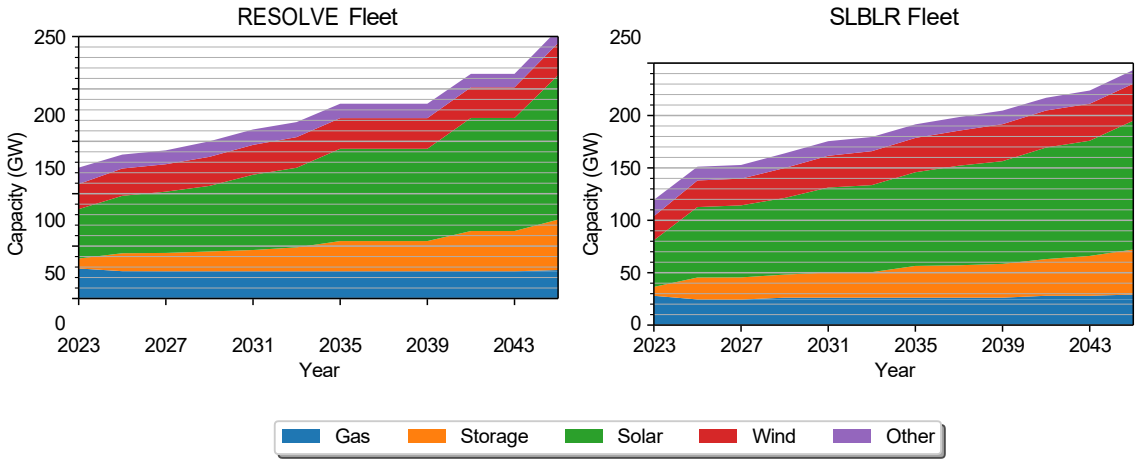

Figure 4.3: Comparison of CAISO fleet over time.

Table 4.1: Comparison of CAISO 2022 Baseline Generation Costs

	Shutdown (Millions $)	Startup (Millions $)	Fuel (Millions $)	Emissions (MMT)
SLBLR	29.71	70.46	2594.53	35.7
RESOLVE	5.28	5.28	1935.45	21.4

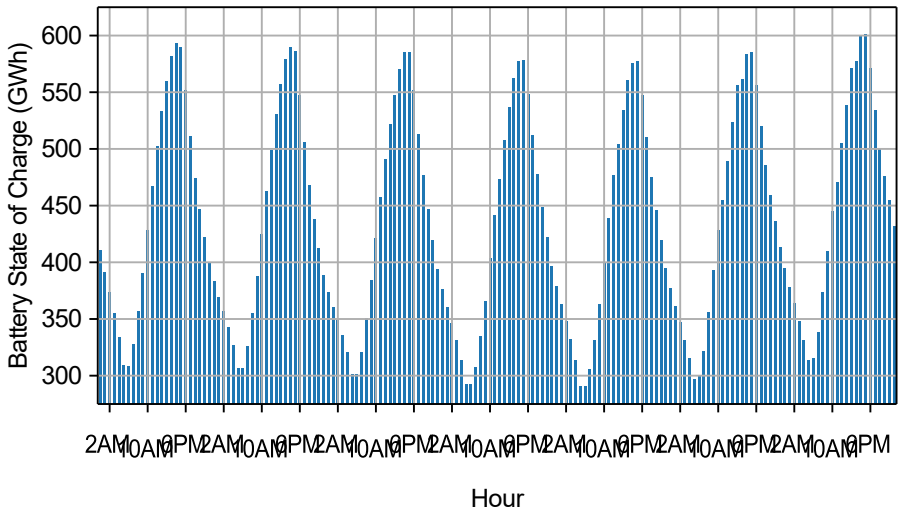

Figure 4.4: Battery state of charge for an exemplary week in 2045.

The storage state of charge for an exemplary week in August 2045 is shown in Fig 4.4. As expected, the state of charge is at its maximum in the late afternoon and its minimum in the early morning. A feature of note is that the state of charge maximum and minimum vary by roughly 30GWh. This indicates that energy is being shared between days, behavior which is enabled by modeling longer representative periods. In contrast, if energy sharing is not allowed, as in the case of modeled days in RESOLVE, the state of charge must be equal at the beginning and end of the day. Also of note is that our model builds 5 less GW of storage, but only 6GWh less energy capacity. In 2045, our model builds battery storage with approximately 8 hour duration vs 7 hour in RESOLVE, and pumped storage with 100 hour duration vs 90 hour in RESOLVE.

Yearly costs are shown in Fig 4.5. Our model maintains lower operation costs in almost every year. Investment spending is higher in early years, but increases at an overall lower rate, resulting in lower investment costs in the second half of the study period. Costs shown in Fig 4.5 are yearly, including financing of investment from earlier years, and not adjusted for discount rate. As shown in Table 4.2, with generation, investment, and maintenance costs included but neglecting emission violations, our model presents a savings of 1.2%, or 4 billion dollars through 2065. This includes all costs for CAISO as well as operating costs for the other WECC zones. As a reminder, investment in non-CAISO zones is exogenous, and CAISO investment has a limited impact on outside operating costs. When considering CAISO costs alone, our model saves 1.9%.

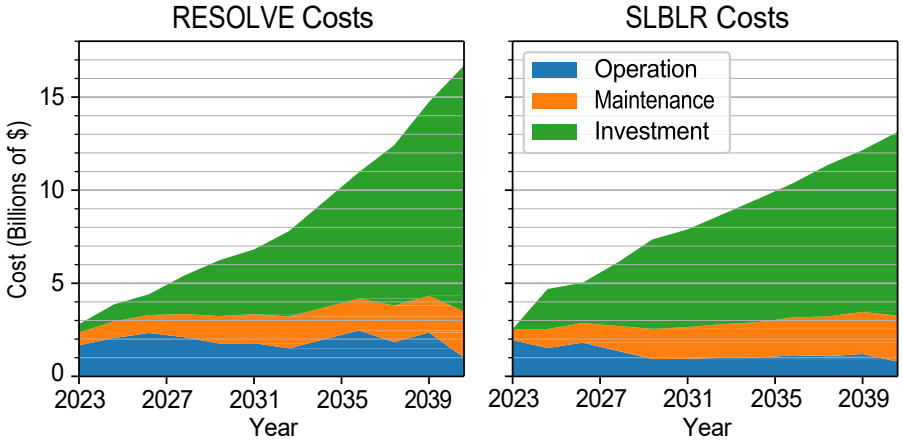

Figure 4.5: Yearly cost breakdown for CAISO.

Table 4.2: Comparison of Total System Costs. 2022 $, Billions.

	System Cost	CAISO Costs			
	Total	Total	Op.	Maint.	Invest.
RESOLVE	341.35	205.86	32.42	65.38	108.06
SLBLR, Day	336.47	202.53	26.23	52.96	123.35
SLBLR, Week	337.36	201.93	22.73	56.01	123.19

Table 4.3: Comparison of total CAISO costs under range of per ton carbon costs. 2022 $, Billions.

	Total CAISO Costs		
Tax	$0	$30	$100
RESOLVE	205.86	207.51	211.38
SLBLR, Day	202.53	203.07	204.31
SLBLR, Week	201.93	201.93	201.93

As exceeding emissions targets comes at the direct benefit of avoided investment in renewables and storage, it is critical to examine the total cost with some component for the cost of over-emissions. These costs are calculated using the optimal objective values and the emission violations E_y^{vio} with a tax c^{vio}, as:

$$\sum_{y \in Y} C_y^{gen} + C_y^{m} + C_y^{inv} + \omega_y \cdot c^{vio} \cdot E_y^{vio} \tag{4.4}$$

Table 4.3 shows the total CAISO costs under different values of a carbon tax per ton of GHG emissions over the limit. Two carbon tax rates are shown to essentially bookend the cost of exceeding emission targets. A carbon tax of $30 roughly corresponds to the 2022 California cap-and-trade clearing price [100]. A carbon tax of $100 roughly corresponds to the Department of Energy's Carbon Negative Shot goal for direct air capture cost per ton [102]. Due to sampling days vs weeks and increasing the total number of modeled days roughly 50% from 37 to 56, the average load in the representative weeks scenario is slightly higher than the representative days. Emissions constraints are binding constraints in most years in all three models. Thus, it is unsurprising that our own SLBLR Day investment decisions have emissions violations in the representative week evaluation. Crucially, the SLBLR Day scenario has less than 1/3rd of the total violation of RESOLVE.

Table 4.4: Comparison of total CAISO costs with representative days under range of per ton carbon costs. 2022 $, Billions.

	Total CAISO Costs		
Tax	$0	$30	$100
RESOLVE	204.34	205.96	209.73
SLBLR, Day	194.84	194.84	194.84

To demonstrate that these results are consistent across varied problem instances, two further comparisons will be made. In each, all input parameters are identical between RESOLVE and SLBLR. The first compares the investment plans of RESOLVE and SLBLR when operated over representative days in (4.3). The comparison of RESOLVE to SLBLR optimized over representative days provides a more isolated comparison of the value of more rigorous generator modeling. The results of this experiment considering a range of carbon taxes are presented in Table 4.4. In this scenario, again our model produces no violation of emissions, while RESOLVE investments result in several million tons of GHG overemitted each year. Even neglecting any cost of overemission, our more detailed model has a nearly 5% lower total cost, primarily due to increased investment and correspondingly lower operating costs. With a $100 carbon cost, the gap grows to over 7%. The second comparative scenario is again over representative days, but with a higher emissions limit. A similar pattern is present in this case. RESOLVE has several million tons of overemissions, and SLBLR has savings of 2.0% to 4.2% depending on carbon tax cost. As is intuitive, the costs with a less aggressive emissions limit is less expensive in both cases. Again, the message is the same: the simplifications to generator modeling underestimate both emissions and fuel costs. This both underestimates the requirement for and value of renewable and storage resources.

Finally, the performance of the proposed method is validated against the performance of other optimization techniques towards solving the full MILP model. First, a state-of-the-art package enabling Dantzig-Wolfe and Benders' decomposition of MILP problems [35] was used. These methods are the closest competitors to surrogate Lagrangian

relaxation methods, as they also allow for decomposition of difficult MILP into more manageable subproblems. Even a reduced model, with 5 instead of 12 modeled years, fails to find a feasible solution within 48 hours. Next, the model was implemented using the Opytimizer Python package [31]. This package implements many metaheuristic optimization algorithms. In a time limit of 48 hours, genetic algorithm, particle swarm optimization, and simulated annealing are unable to find a feasible solution. As a further point of comparison, the original surrogate Lagrangian relaxation [133] fails to find a feasible solution within 48 hours. As such, we conclude that the SLBLR method is particularly suited to solving this type of large planning model, over many of the other advanced optimization techniques.

4.2 Conclusions

This concludes the discussion of the core decarbonization model. In Chapter 2, a MILP decarbonization model for California is developed. To overcome the issue of combinatorial complexity with integer variables, SLBLR is implemented in Chapter 3, and allows us to optimize over nearly 100 million variables including 12 million binary variables in under 48 hours. In Chapter 4, we show that the existing, linearized model underestimates operational costs of gas generators, leading to a substantially different investment plan. By doing so, we develop an investment plan that saves California 4 billion dollars over the investment horizon. Further, our model suggests more substantial and early investment in renewable generation and storage is required to meet intermediate emissions targets. This result may inform policymakers that a more aggressive approach is needed than previous work sponsored by state commissions.

The following chapters will extend this decarbonization model to examine novel approaches to representative period selection, integration of smart charging for medium- and heavy-duty vehicles, and climate change scenarios.

Chapter 5

Novel Approaches to Representative Period Selection

A key component of any planning study is representative period selection. Planning models, including the one developed in Chapter 2, usually involve two-time scales: one for annual investment decisions and another for hourly generation dispatch simulations. The complexity of solving the optimization problem increases significantly when attempting to model all 8760 hours in a year, potentially leading to slow or intractable computations.

To mitigate this challenge, one solution is to use representative periods, which can significantly alleviate the computational load. For instance, rather than modeling all 8760 hours annually, one can solve 37 representative days, reducing the modeled hours by almost 90%. The main goal in selecting these representative days is to capture the essential aspects of the entire yearly system behavior while substantially reducing the computational intricacies involved.

This chapter will present two novel approaches to representative period selection. These approaches have similar goals but different focuses. The following section will review existing works on the topic of representative period selection, and identify two gaps in the literature which these works seek to fill.

5.1 Related Work

The problem of representative period selection in power system planning has received much attention in the literature. The majority of the work on this topic has been built around the framework of time series clustering. The authors in [110] compare a variety of clustering methods. The authors in [75] present a comparison of clustering and downsampling approaches. In [117], a clustering method is proposed which requires that each cluster consist of a contiguous set of days, while a downsampling-based approach is proposed in [64]. These clustering-based approaches all share the common drawback that they are only suitable for the selection of representative days, and not periods of multiple days in length, as will be discussed in detail later.

Some works seek to model a year continuously to better model long-term storage [76, 112]. In these works, the temporal reduction is achieved by holding an operating state for multiple hours. However, within the CEM, this approach severely affects ramping modeling. The ramping requirements, particularly those originating from the so-called 'duck curve' due to high solar penetration, are a critical aspect that needs to be modeled. Representative hour approaches have also been proposed [65, 9], but suffer the same loss of chronology and thus the ability to track energy storage and ramping.

The method used in Chapter 4 is a histogram-based approach [48]. This method was used in this study to isolate the deviations from the reference study to the selection of weeks and the improved generator modeling. This method ensure that the sampled periods accurately represent the loading levels of a full year, but neglect the importance of variability of renewables. This method also neglects chronology, and accounts for loading levels as equivalent regardless of when they occur during the day.

There are limited studies that utilize other features for clustering. In [59], clustering based on investment cost is proposed, along with a technique for extreme period selection. The authors in [79] examine the trade-off between temporal and technical modeling detail and propose period selection based on RE variability.

While a considerable amount of research effort has been devoted to the representative period selection problem, the majority of it has focused on clustering algorithm design. In contrast, there is little work dedicated to the selection of power grid features to be used within the clustering algorithm. Indeed, the majority of representative period selection methods use a greenfield approach, assuming no existing capacity, which is an impractical assumption for real power systems. Furthermore, these approaches overlook features highly relevant to capacity expansion planning, such as transmission congestion, RE curtailment, and load shedding. The first study, in Section 5.2, addresses this gap.

On the other hand, these works are all dedicated towards selection of representative days. The selection of longer representative periods poses unique challenges due to the clustering framework relied upon. With the interday variability of renewable energy and

growing importance of energy storage, the ability to model longer periods is vital, as it enables energy to be shared between days. Section 5.3 presents a study addressing this gap.

5.2 Representative Period Selection for Robust Capacity Expansion Planning in Low-carbon Grids

This study integrates additional operational features into a general time series clustering framework to select representative periods, thus improving the efficiency and accuracy of the clustering process. By enhancing the clustering framework with these tailored operational features, we strive to provide a more robust and insightful analysis of the underlying temporal patterns used in capacity expansion in the presence of high RE penetration and extreme events.

5.2.1 Technical Method

Feature Selection

In practice, CEMs do what the name suggests: determine optimal strategies for adding energy resource capacity. The current resource fleet within a system might be sub-optimal or insufficient for future years, as a result of a variety of drivers including the following examples. Environmental regulations may limit the use of gas-fired generators. Increasing loads could necessitate additional capacity. High fuel costs may also make it more economically viable to invest in additional energy storage or renewable resources. Therefore, it is crucial to consider the characteristics of the existing capacity when selecting representative days. While existing methods for representative period selection have disre-

garded this aspect, our objective is to incorporate the attributes of the existing capacity into the process of choosing representative days.

To this end, we propose including features that encode the existing capacity, but are typically not considered during clustering. This section identifies these features and justifies their inclusion. These features are referred to as operational features, as they are outputs of system operation, in contrast with demand and generation, which are inputs to operation. Namely, these features are load shedding, transmission congestion, and renewable curtailment.

Load shedding, sometimes known as rolling blackouts, refers to interrupting some portion of loads, generally as a last resort for load balancing. Load shedding events are increasingly associated with extreme weather events, like heat waves or cold snaps. Load shedding is associated with considerable economic cost, as well as potential loss of life, as evidenced by the 2021 Texas power crisis [26]. Load shedding is a key metric for resource adequacy and can be avoided with proper planning. Inclusion of this feature in representative day selection could help select periods that stress the existing resource fleet, thus leading to more robust capacity sizing.

Curtailment refers to disconnecting RE generation to prevent overgeneration. Curtailment occurs when renewable generation exceeds demand, and this excess cannot be exported or used to charge energy storage systems. Including curtailment will better account for days where renewable capacity is already sufficient, as well as select periods which demonstrate the value in expanding storage capacity.

Congestion refers to a $/MWh transmission cost. Congestion exists when there is more demand for transmission capacity than there is physical capacity. Including this feature should help select periods which would be relevant to expanding transmission capacity by identifying periods with excess RE generation which could be exported to other areas.

Each of these operational features can be readily obtained from the output of a production cost model (PCM). Because these features can be obtained by running a PCM in discontinuous days, the computational complexity associated with long timescales that necessitates the use of representative days in the CEM is irrelevant.

Dimensionality Reduction

Often, capacity expansion models consider wide geographic areas. Thus, there are load forecasts at many nodes, as well as renewable generation profiles at different locations. Each time step further inflates the dimensionality of each sample. Attempting to cluster without reducing the spatial dimensionality of these features could produce sub-optimal results. For load, renewable generation profiles, and congestion, this is done in the straightforward method of averaging the time series over the spatial dimension. Some operational features should be highly sparse. In particular, curtailment should be zero in most hours, and load shedding should be even less common. For this reason, the spatial dimension of these features is reduced by taking the maximum. Each feature is normalized to zero mean, unit variance before it is clustered. With each feature reduced to a time series of length 8760, we can construct the feature matrix $x \in R^{365 \times 24 \times 6}$ by reshaping the time series and concatenating each feature. Each element $x_n \in R^{24 \times 6}$ represents the multivariate time series for a given day.

Extreme Events

The goal of representative day selection is to find a subset of days that best capture the annual behavior of load and generation. However, this goal conflicts with the need to simulate extreme periods. Extreme weather events may only occur for a small fraction of days each year, and thus are unlikely to be selected during typical representative period selection. Within power system planning, extreme events are an extremely important consideration. If enough generation capacity is not held, the reliability of the system during extreme weather patterns could be compromised. Similarly, if there are periods of abnormally low renewable generation, the system could struggle to cope with demand. With climate change and an increasing push for decarbonization, these extreme events will become even more important. Several works have proposed methods for selecting extreme events based on peak load, peak ramping, or other features [104, 59]. Through the numerical study, we will show that inclusion of operational features in clustering implicitly selects these representative periods. In particular, the inclusion of load shedding as a feature in representative period selection effectively captures the inadequacy of existing capacity.

Clustering

To select representative days, the periods are clustered using k-means. Given N samples, the goal of k-means is to generate K clusters with centers μ_k and assign a cluster label to each multivariate time series x_n.

$$\min \sum_{k=1}^{K} \sum_{n=1}^{N} \gamma_{n,k} ||x_n - \mu_i||^2 \tag{5.1}$$

where $\gamma_{n,k}$ is a binary variable that indicates that sample x_n belongs to cluster k and must satisfy (5.2),

$$\sum_{k=1}^{K} \gamma_{n,k} = 1, \ \forall n \in N \tag{5.2}$$

Although the optimization problem in (5.1) is NP-hard, computationally efficient heuristics exist and are implemented in widely-used Python packages, such as scikit-learn [74]. Clusters will be represented by their medoid in the capacity expansion model. This is necessitated by the dimensionality reduction discussed in Section 5.2.1, precluding a backwards map from the low-spatial dimension representation used in clustering to the high-spatial dimension used in the CEM. The medoid is selected after clustering as the sample with the smallest Euclidean distance to the centroid. The weight of each representative period is chosen as the cluster cardinality.

5.2.2 Experimental Validation

Experimental Setup

The proposed clustering technique is validated using a CEM and PCM for the Western Interconnection. Only single-year planning is considered, and the year modeled will be referred to as the target year. First, the PCM is solved for the target year using the existing generation capacity. Then, the features described in Section 5.2.1 are extracted and used within the proposed clustering technique. Finally, the representative days corresponding to the cluster medoids are used to run the CEM for the target year. To evaluate the performance of the investment decisions made with the representative periods, the PCM is

then run again with updated investments. Fig. 5.1 demonstrates the flow of the numerical study.

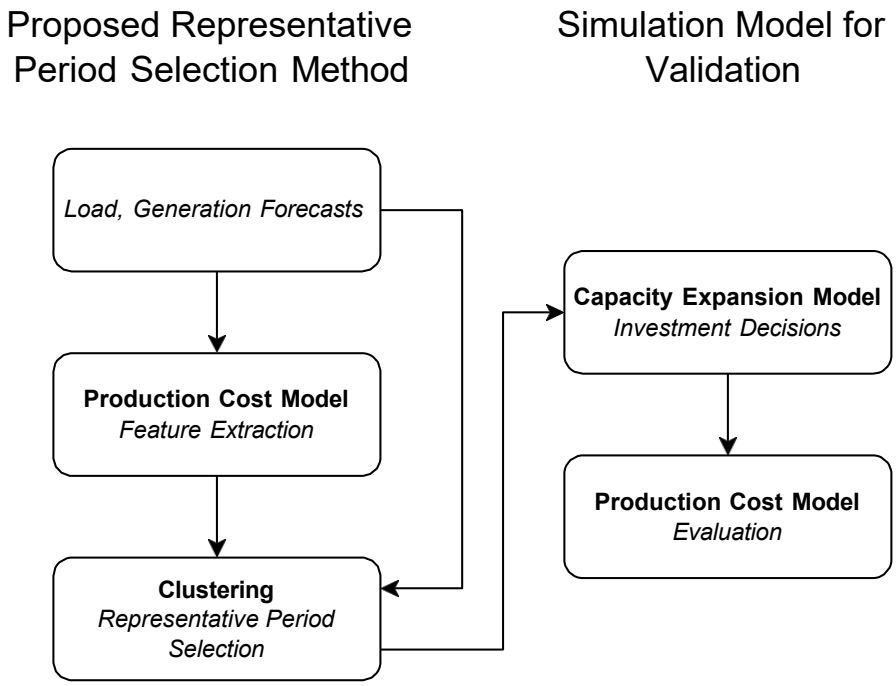

Figure 5.1: Flow of proposed method and numerical validation.

The CEM is the planning model derived in Chapter 2, and the PCM is equivalent to the single-week unit commitment model derived in 2.40.

Planning models, like the one derived in Chapter 2, typically include a planning reserve margin (PRM) constraint, requiring that the fleet be able to provide some margin greater than the maximum projected load. As a result, regardless of clustering performance, the investment decisions will produce a fleet that can most likely satisfy all load requirements. In some cases, if this constraint is particularly tight, the final resource investment plan may be predominantly determined by investment costs, thus diminishing the value of improved representative day selection. In other words, system operation may determine

the optimal fleet up to some MW of capacity, but any requirement for capacity above that may be determined only by which resource is cheapest per capacity. In the numerical study, results will be shown both with PRM omitted and with a minimal PRM, equal only to the maximum load and not a margin above, as is typical.

To demonstrate the performance of the proposed technique, we compare the investment decisions obtained using the proposed clustering method to those using only traditional features, namely the load and renewable generation profiles. The performance of investment decisions is evaluated using two metrics: cost and reliability. Reliability is evaluated using both the number of load shedding events and MWh of load shedding. Cost is the combined cost of maintenance, investment, and operation. Within the PCM, load shedding is available at $50,000 per MWh. Investment costs are amortized to allow for single-year planning.

Results

We evaluate the performance of the algorithm through two lenses: load shedding and total cost. We also look to the blend of resources to evaluate how representative day selection affects valuation of one resource group versus another. The proposed method is compared to a base method, in which the clustering step only accounts for load, solar, and wind. This method is representative of what is commonly utilized in the literature, and performs similarly to many other selection methods [110]. Fig. 5.2 shows the behavior of the clustering method with respect to the net load in 2030. The proposed method captures a more diverse set of net load conditions, whereas the base method tends to select fewer extreme points.

To better understand the impacts of the proposed method, results will be shown under several planning scenarios, with several target years. K-means heuristics are not deterministic and are highly dependent on initialization. To address this, most implementations run the algorithm multiple times and choose the result with the lowest cost function. Still, the selected days and weights can vary, so the numerical study is repeated three times for each scenario and averaged. Unless stated otherwise, 20 representative days are selected.

The two PRM scenarios will be shown with target years of 2025, 2030, and 2045. We also show the results for 2025 in which economic retirement is not allowed and for 2045 in which economic retirement is required for 50% of all in-CAISO units. No such constrained retirement will be demonstrated for 2030. The justification for including these scenarios is as follows. Without PRM, the effect of representative day selection on resource adequacy should be more obvious. With PRM, there should be a smaller effect on resource adequacy and a greater effect on system costs.

Table 5.1: Total cost comparison of base and proposed method investment decisions (Millions $)

Year	Pen	Method	PRM	No PRM	Constrained Retirement
2025	Yes	Base	13,372	13,354	12,468
		Proposed	12,412	12,407	12,442
		Improvement	7.18 %	7.09 %	0.21 %
	No	Base	12,461	12,472	12,466
		Proposed	12,412	12,407	12,442
		Improvement	0.39 %	0.52 %	0.19 %
2030	Yes	Base	14,507	14,666	-
		Proposed	14,503	14,255	-
		Improvement	0.03 %	2.80 %	
	No	Base	14,448	14,141	-
		Proposed	14,470	14,239	-
		Improvement	-0.16 %	-0.69 %	
2045	Yes	Base	23,713	23,649	23,481
		Proposed	23,414	21,323	21,351
		Improvement	1.26%	9.84%	9.07%
	No	Base	23,545	22,011	21,952
		Proposed	23,414	21,323	21,351
		Improvement	0.56%	3.13%	2.74 %

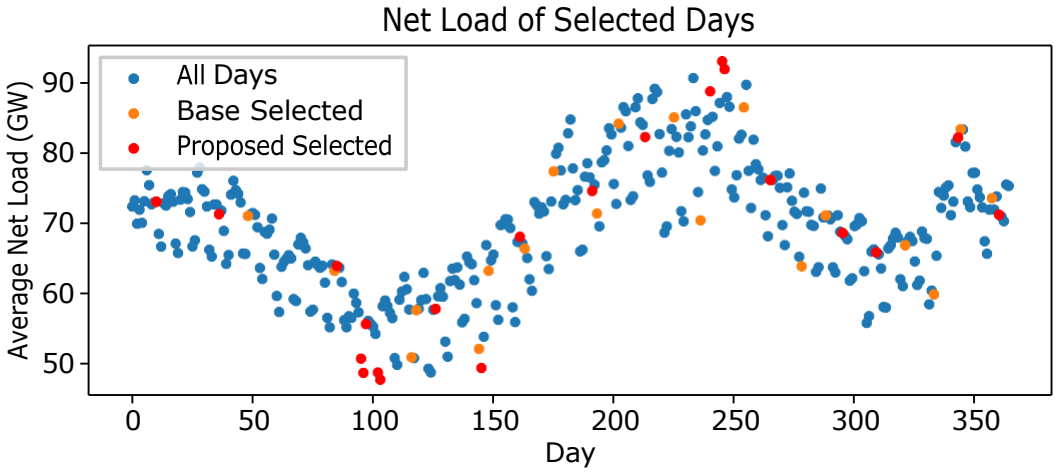

Figure 5.2: Average hourly net load of full year and selected representative days.

Table 5.1 shows the total cost for each target year with the proposed method and the base method. The base method often requires considerable load shedding, which greatly increases the total cost. To give a point of comparison without this effect, costs are shown both with and without the load shedding penalty component, denoted by 'pen'. Note that for the proposed method, the penalty and no penalty values often are equal as there is no load shedding. In every case with load shedding penalty, the proposed method leads to lower or nearly identical costs. Omitting the load shedding costs, the proposed method generally has comparable costs to the base case. This indicates that the proposed method produces more realistic capacity plans, which, in turn, lead to a reduction in load shedding, with only moderately higher investment costs.

Fig. 5.3 shows the costs for 2030 without PRM, both with and without load shedding penalty. As expected, the capacity plan resulting from representative days with the base features requires more load shedding in the PCM. This load shedding is a result of lower capacity, which in turn has lower investment costs. However, the capacity plan resulting from the proposed method, has only slightly higher investment costs. With load shedding costs ignored, the total costs are only 0.69% higher for the proposed method. Investment costs are 2.6% higher, but are offset by lower operational costs and much lower load shedding costs. With load shedding costs accounted for, the total cost is drastically lower.

Fig. 5.4 shows components of capacity expansion by resource class for 2030. Intuitively, the scenarios with PRM have increased investment regardless of the representative

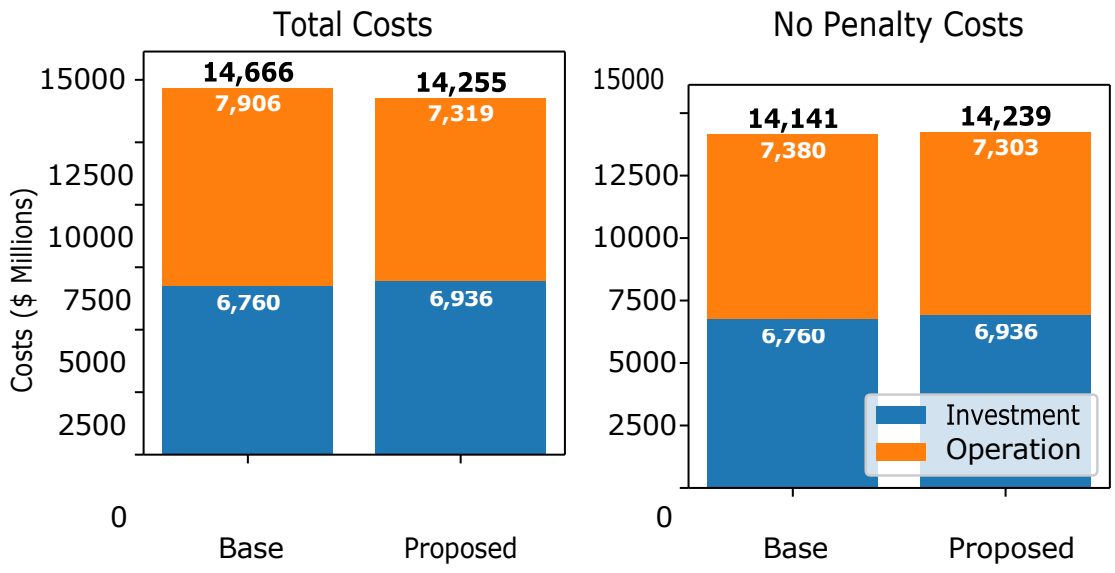

Figure 5.3: Operation and capital costs for 2030 without PRM.

day selection. The proposed method leads to greater investment in both energy storage and renewable generation. The methods have roughly equal retirement of gas-fired generators.

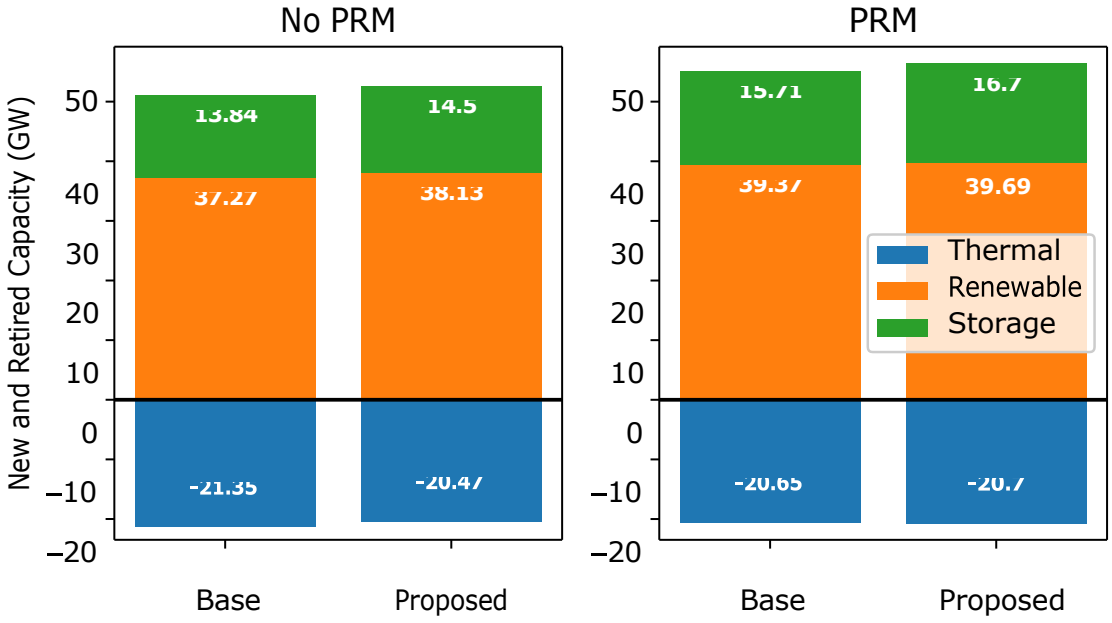

Figure 5.4: Capacity expansion decisions in 2030.

Fig. 5.5 shows the sensitivity of the number of representative days for 2030. Regardless of the number of representative days, the proposed method leads to lower load shedding, both in the number of events and the average MWh of shedding per event. As previously suggested, this indicates that the inclusion of the proposed operational features implicitly selects extreme conditions more effectively.

5.2.3 Conclusion

In this section, we proposed a novel method to select representative days that can be used in capacity expansion models. The proposed method better accounts for existing capacity by considering key novel operational features during the clustering step. By including these features, the resulting capacity expansion plan exhibits improved load-serving

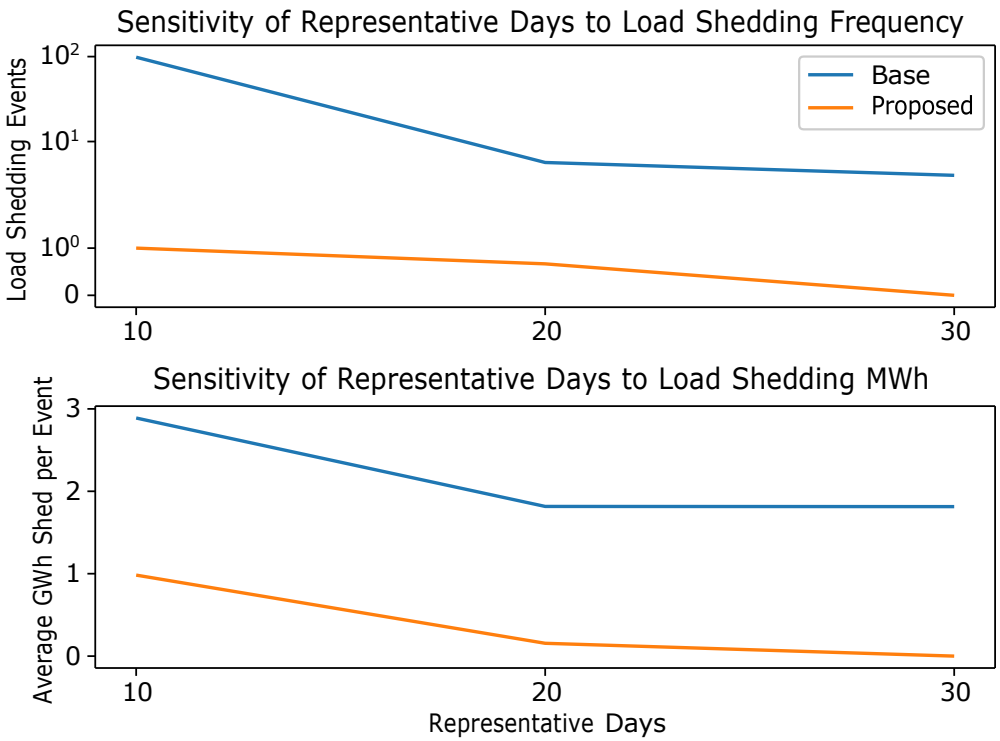

Figure 5.5: Sensitivity of number of representative days on load shedding.

capability and cost savings as compared to the base-feature case. The proposed method was validated on a capacity expansion model based on decarbonization goals in CAISO. Several planning scenarios and horizons were studied. In all scenarios, the proposed method resulted in lower load shedding in the full-space production cost model, as well as lower or comparable costs even when the cost of load shedding is neglected.

5.3 Selection of Intermediate-Length Representative Periods

The references cited in 5.1 primarily concentrate on the selection of representative days and offer general assertions regarding the algorithm's ability to choose periods of varying lengths, such as a representative day or week. However, let us consider the case where the desired period is of an intermediate length, such as 3 days. The time series clustering framework upon which the majority of representative period selection algorithms are built, requires the full time series to be divided into subsequences. Clustering these subsequences becomes highly dependent on the starting point. In particular, the load exhibits significant differences in both shape and magnitude between weekdays and weekends. For example, the Euclidean distance of a Friday-Saturday-Sunday subsequence to a Saturday-Sunday-Monday subsequence would likely be large because the loads of Friday would be compared to Saturday and the loads of Sunday would be compared to Monday. On the other hand, clustering of overlapping subsequences, obtained by sliding a window across the full time series with a stride shorter than the subsequence length, has been established to return essentially random results [54]. Even in the case of representative weeks, there are considerable drawbacks. Generally speaking, capacity expansion problems tend to reduce the

annual temporal coverage to roughly 10% or less. In the case of days, this permits 37 days, but in the case of weeks, this permits only 5 weeks. Intuitively, one would expect that representing a full year is more difficult given 5 choices than 37 choices.

Considering the limitations of existing clustering methods, along with the need to capture interday variability in CEMs subject to high levels of energy storage and renewables penetration, the selection of representative periods longer than one day becomes imperative. Indeed, enabling interday sharing of energy through storage modeling is a crucial, yet often ignored, aspect that would become particularly valuable during days of low generation from renewable resources.

While the selection of a representative period length is fundamentally an experimental design decision, there exists a noticeable gap in research when it comes to effectively choosing a period longer than a day but shorter than a week. To bridge this knowledge gap, we introduce a novel snippet algorithm specifically designed for selecting representative periods that extend beyond a single day. By comparing subsequences instead of full sequences, the proposed snippet algorithm is able to select representative periods of arbitrary length from complex datasets. The proposed algorithm draws significant inspiration from [50]; however, we have made several tailored adjustments to accommodate the unique domain to which our proposed algorithm is applied.

5.3.1 Technical Method

Overview of time series snippets

The discussion of time series snippets below provides a concise overview of the algorithm that served as the inspiration for the proposed method in Section 5.3.1. The time series snippets algorithm is built on top of the matrix profile distance (MPdist) [37] measure, which is in turn built on top of the matrix profile [131]. MPdist compares two-time series and considers them to be similar if they have similar subsequences. At its most basic level, the distance is the j-th smallest Euclidean distance between subsequences. More specifically, the goal of time series snippets is to select, from a time series T with length t, subsequences of length s that best generalize the full-time series. First, the full-time series is separated into non-overlapping subsequences S_i with $i \in [0, t/s - 1]$. Each of these subsequences then has an MPdist profile $MPdist_i$ compared to the full-time series. If each $MPdist_i$ were plotted, the goal would be to select the k profiles that minimize the area under the curve of the combined profiles, as shown in Fig. 5.6. To select these representative subsequences, a greedy algorithm is proposed, choosing the subsequence that gives the greatest reduction in the cumulative distance in each iteration.

The key contribution of these matrix profile-related methods is that they scale well to extremely long time series. The problem of representative period selection for power system planning typically considers one year of data at hourly frequency, for a time series of length 8760, which is extremely short in that context. Further, we do not need to calculate the distance measure for every subsequence, as we can exploit the known daily periodicity

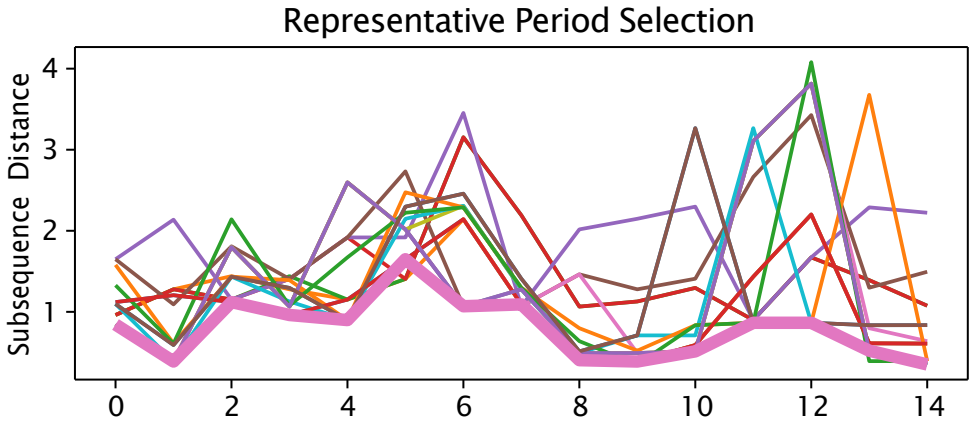

Figure 5.6: Visualization of distance profiles and minimization of area under the curve of selected profiles.

of our time series. For this reason, we can calculate a distance measure similar to the *MPdist* without relying on algorithms related to the matrix profile. This also enables us to utilize overlapping subsequences as S_i. Finally, the problem size allows us to select the representative snippets through convex optimization rather than relying on a greedy algorithm.

Proposed Method

Let $T = \{T[0], \ldots T[htbp!], \ldots, T[t-1]\}$ represent the yearly multivariate time series of length t and $T[htbp!]$ be the tuple of measurements at hour $h \in [0, t-1]$. This tuple typically encompasses load, solar generation, and wind generation information; however, the proposed method remains agnostic to the input features, providing adaptability in the analysis. Let also $S = \{S_0, \ldots S_j, \ldots S_{m-1}\}$ be the set of subsequences, and u the stride of the window that generates subsequences of length s. The subsequence S_j is then defined

as:

$$S_j = T[j \cdot u : j \cdot u + s] \tag{5.3}$$

There will be $m = \frac{t-s}{u} + 1$ total subsequences, thus $j \in [0, m-1]$. These will be the candidate subsequences used for selecting representative periods. Similarly, we can define non-overlapping subsequences of T, which we will refer to as segments. There will be $n = \frac{t}{u}$ such segments, thus $i \in [0, n-1]$. The segment T_i is then defined as:

$$T_i = T[i \cdot u : (i+1) \cdot u] \tag{5.4}$$

A visual representation of the definitions can be seen in Fig. 5.7. For clarity, i will be reserved to index the time series segments T_i and j to index the subsequences S_j.

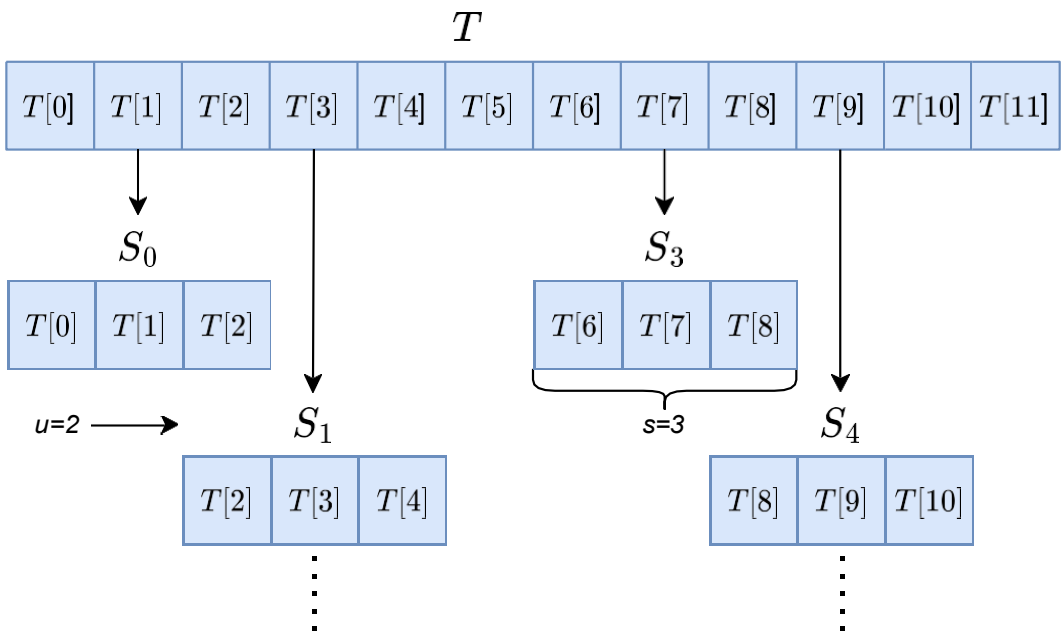

Figure 5.7: Visualization of an example time series T and subsequences S_j with $u = 2$, $s = 3$.

Inspired by the MPdist, let D be a matrix of distances with shape $n \times m$. For our case, we will assume u and s are chosen such that n, m, and s/u are integers. We also have domain knowledge of periodicity. Each of the features, especially load and solar generation, have strong 24-hour cycles. It is unlikely that an afternoon subsequence from one day would be similar to a nighttime subsequence from another day. Further, this is physically meaningless in the context of capacity expansion. For this reason, whereas the MPdist compares subsequences for every timestep, we apply stride $u = 24$ in calculating the distances. In essence, this compares each day in the multivariate subsequence to each day in the full multivariate time series, and assigns a distance correspondingly.

$$D_{i,j} = \min_{x} \left| \left| S_j[x \cdot u : (x+1) \cdot u] - T_i \right| \right|, x \in \left[0, \frac{s}{u}\right) \tag{5.5}$$

The goal is then to find a subset of those candidate days which best captures the patterns for the year as a whole. This goal is the same as the one visualized in Fig. 5.6. Time series snippets were originally proposed with a greedy algorithm that iteratively selects the subsequence which minimizes the cumulative sum of distances to the full time series, necessitated by the long time series the algorithm was designed for. Because our time series is rather short, we can instead formulate this problem as a mixed integer linear

program and find the solution using any suitable optimization solver.

$$
\begin{aligned}
&\min_{\alpha} \sum_{i=0}^{n-1} dist_i \\
&s.t. \sum_{j=0}^{n-1} \alpha_j = k \\
&\quad dist_i = \sum_{j=0}^{n-1} md_{i,j} \cdot D_{i,j}, \ \forall i \in [0, n) \\
&\quad \sum_{j=0}^{n-1} md_{i,j} = 1, \ \forall i \in [0, n) \\
&\quad md_{i,j} \leq \alpha_j, \ \forall i \in [0, n), \ \forall j \in [0, m) \\
&\quad dist \in \mathrm{R}^n, \quad md \in [0, 1]^{n \times m}, \quad \alpha \in [0, 1]^m,
\end{aligned}
\tag{5.6}
$$

where k is the desired number of representative periods; α_j is a binary indicator selecting subsequence S_j as a representative period; $dist_i$ is the minimum distance between the selected representative periods to T_i; and $md_{i,j}$ is a binary indicator signifying that subsequence S_j has the smallest distance to day T_i. Within CEM, representative periods are typically weighted by the amount of the year that they account for. The weights associated with each representative period are a function of $md_{i,j}$, and can be written then as:

$$w_j = \sum_{i=0}^{n-1} md_{i,j}/(s/u), \tag{5.7}$$

where s/u ensures the weights sum to 365.

5.3.2 Experimental Validation

Experimental Setup

To the authors' knowledge, no work has made a dedicated attempt to address the sampling of intermediate-length representative periods in capacity expansion planning.

This absence poses a challenge when comparing the proposed method with widely used state-of-the-art approaches. We will compare the performance of the proposed algorithm to a popular method for selecting representative days, and show that our algorithm is at least comparable with the state-of-the-art for this task. The proposed method will also be used to compare single-day planning to multi-day planning. The goal of this comparison is to show the value in simulating representative periods longer than one day, particularly in sizing energy storage. Our goal is not necessarily to show the optimality of a particular representative period length, but rather to demonstrate the differences between period lengths on investment plans and operational cost.

The following general experimental design will be used to validate the proposed method. First, the representative days are selected and used within the CEM. Then, the investment decisions are fixed, and the model is solved again as a production cost model. The production cost model (PCM) is ran for the full year in two-week stages, and results from this will be referred to as fullspace results. The choice of two weeks is somewhat arbitrary, with the key being that this length is considerably longer than each of the candidate representative period lengths to avoid giving bias towards any particular length. The models described in 5.2.2 will be used.

Results and Discussions

First, we compare our method to a popular approach for representative day selection: k-means clustering using load, wind, and solar profiles with medoid cluster center representation used in the CEM. We use our proposed algorithm with subsequence length $s = 24$ to select 21 representative days. Our algorithm has a total cost (investment, main-

tenance, fullspace operations) of 14.560 billion US dollars as compared to 14.611 billion US dollars. [110] compared many clustering techniques for power system planning, and established that there are not clear patterns on which technique is best, and many have comparable performance. With this in mind, we can suggest that even for representative day selection, our proposed method is at least comparable with one of the most commonly used representative day selection approaches.

With the validity of the proposed method established, we now seek to defend the motivation behind selection of longer representative periods. This will be explored via investment, fullspace operation cost, emissions, and the investment portfolio, considering representative periods of 1 to 5 days, i.e. with subsequence length $s \in [24, 48, 72, 96, 120]$. In each case, periods are adjusted to model a total of 35 or 36 days, nearly 10% of the annual days.

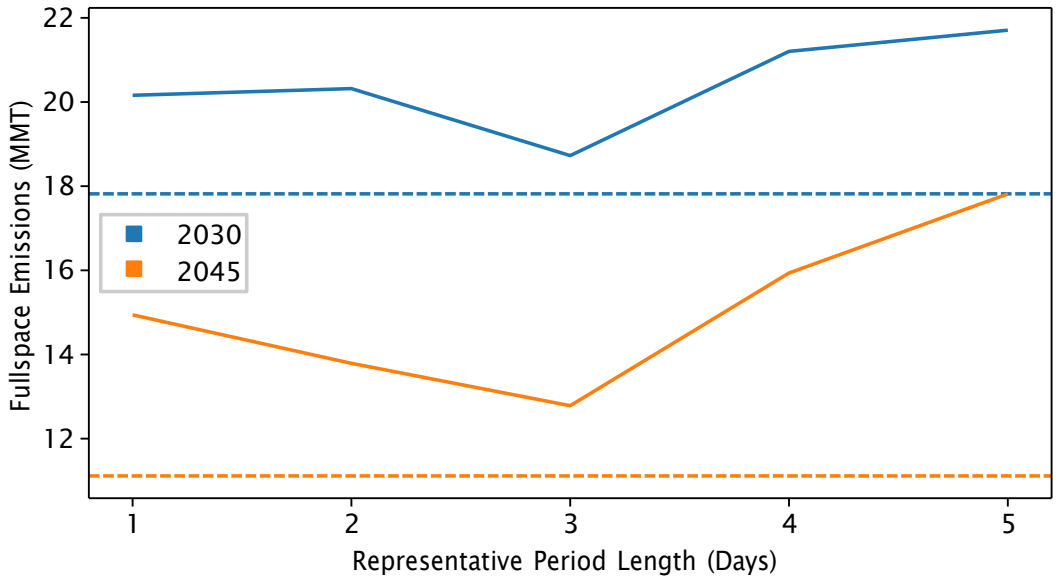

Figure 5.8: Fullspace emissions

Fullspace emissions are shown in Fig. 5.8. None of the scenarios meet fullspace emissions limits. The primary driver of emissions is investment in renewable generation and energy storage. Once these technologies are purchased, their use incurs no additional operational costs in the unit commitment model. However, the PCM is myopic in terms of emissions, and may take actions which lead to lower costs but higher emissions, such as export generation from thermal units in CAISO. As the fullspace model is run in discontinuous segments for reasons of computational tractability, it is impossible to effectively enforce emissions limits. It is difficult to say which, if any, of these fleets would be able to satisfy the emissions limits. Still, it is notable that 3-day representative periods present the lowest emissions, and longer, and thus fewer, periods have substantially higher emissions. This suggests that by modeling an intermediate-length period, interday energy storage can be

leveraged to lower emissions. However, as the length of period grows, the number of periods must shrink. Without a sufficient number of periods, it is difficult to select periods which represent the annual behavior sufficiently well. This is illustrated by the higher emissions in the 4 and 5-day cases, which sample 9 and 7 periods respectively.

The duration of installed storage, and installed power and energy capacity of storage as a function of representative period length for 2030 and 2045 are shown in Fig. 5.9. With regards to storage duration, the key takeaway is that increasing the length of representative period allows for utilization of storage for interday energy sharing, and the duration increases for lengths between 1 and 3 days. However, the tradeoff between number and length of representative seems to impact the ability of the surrogate days to effectively represent the full year, leading to less predictable effects with lengths over 3 days. With regards to the power and energy capacity of storage, a similar pattern is evident. Between lengths of 1 and 3 days, the installed capacities generally increase, and then begin to decrease again. This result is in line with the emissions result.

Fig. 5.10 shows the cost by year for each scenario. As one would expect from the emissions violations visualization, $d = 4$ and $d = 5$ have the lowest overall cost due to less build of renewable technologies. Most notable is that the $d = 3$ result is very close to the $d = 1$ result despite larger investment. This suggests that by representing longer periods, it is better able to capture the fullspace value of interday energy sharing. Thus, the cost of additional investment is offset by lower operating costs. Specifically, the total costs for 2030 are 1.1% higher but have 7.1% lower emissions.

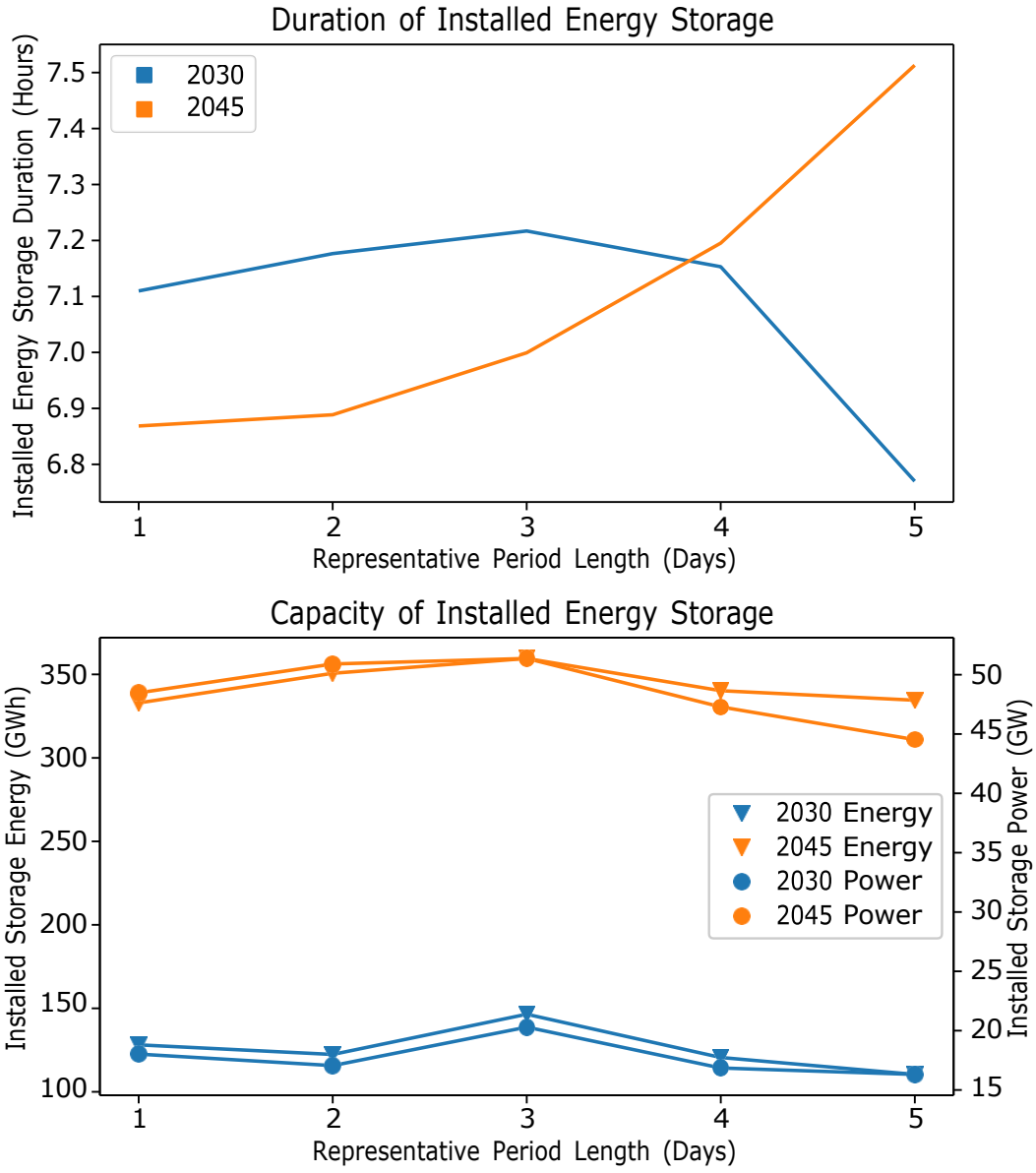

Figure 5.9: Impact of representative period length on duration and capacity of installed energy storage.

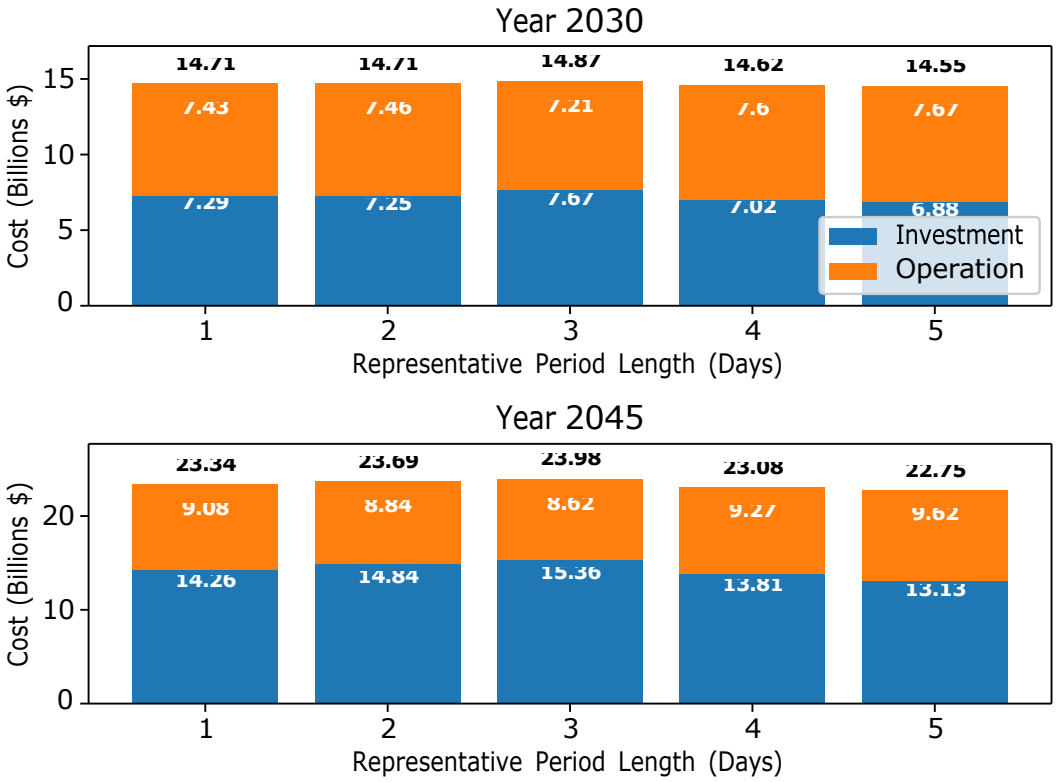

Figure 5.10: Investment and operation costs of differing representative periods

Fig. 5.11 shows an elbow plot of the objective function of (5.6). Intuitively, for a given number of total modeled days $k \times s/u$, the objective is best for more, shorter representative periods. The gap between the lines is larger at the lower total modeled days and begins to converge at higher. This characteristic explains why, for a fixed total modeled days, the representation degrades with higher period length.

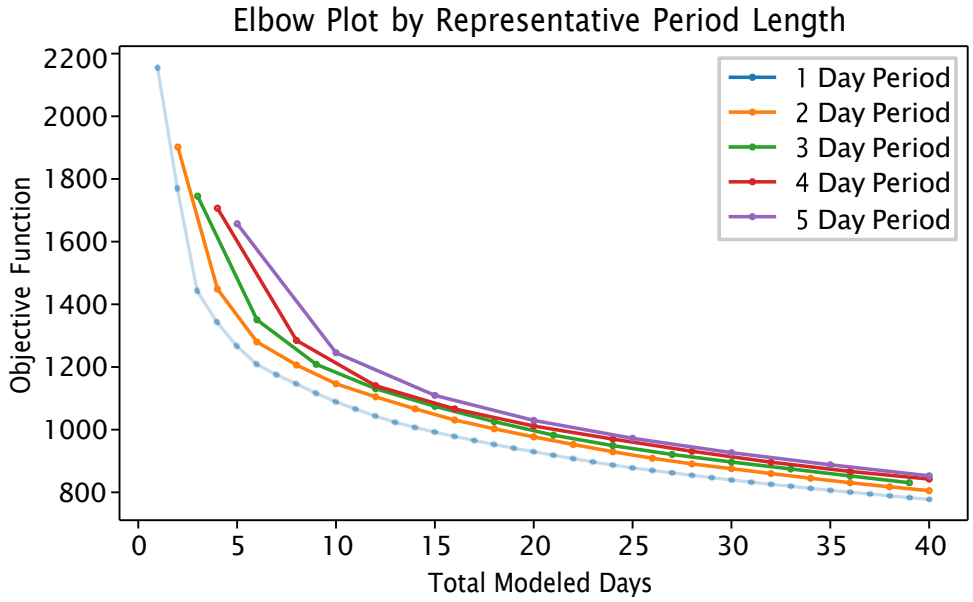

Figure 5.11: Elbow plot of objective function (5.6) at different representative period lengths.

5.3.3 Conclusion

In this section, we proposed a novel algorithm for selecting representative periods. The algorithm is particularly directed towards selecting periods longer than a single day, and is well suited for planning in systems with high penetration of variable renewable energy and reliance on energy storage. The method chooses representative days which minimize a distance measure to the timeseries of the full year. The proposed method was validated on

a CEM based on California's decarbonization targets. The proposed method is competitive with the state-of-the-art for representative day selection, and we demonstrate the impact of representative period length on investment strategy.

5.4 Conclusion

This chapter proposed two novel methods for representative period selection. The first method focuses on unique operational features within the framework of time series clustering to support more robust capacity expansion planning. The second method concerns the selection of representative periods of intermediate length, with the goal of better modeling of interday energy sharing. The next chapter will utilize the second method, and present a case study considering the adoption of smart charging of medium- and heavy-duty battery electric vehicles.

Chapter 6

Impact of Smart Charging on Decarbonization Pathways

6.1 Introduction

California has set ambitious decarbonization goals across multiple sectors, including transportation with the California Air Resources Board's Advanced Clean Fleet regulation, and electric power generation with Senate Bill 100 and Senate Bill 350. As a result, these sectors are expected to change rapidly over the next two decades. It is crucial that these transitions be planned in tandem to ensure cost-effectiveness and reliable power system operation.

It is well established that transportation and energy generation are becoming increasingly linked fields as part of the response to climate change. Transportation electrification is a key component of the energy transition, and vehicle charging load is expected

to become a large share of the energy demand as penetration of electric vehicles increases. This transition is occurring together with the shift from carbon-based to renewables-based power generation.

Adoption of battery electric vehicles (BEV) is expected to both increase electricity demand as well as impact the load patterns. A topic of recent interest has been leveraging the charging flexibility of BEVs to reduce the impacts on power grid operation. A major component of this is flexible charging, or V1G, which is the ability to control vehicle charging, typically to shift charging from a peak time to an off-peak time to lower stress on the grid or to adjust the charging power with respect to pricing and demand response signals from the electric utilities. Even further is V2G or bi-directional charging. In this case, vehicles can discharge to the grid, to provide energy shifting or ancillary services.

From the perspective of large-scale implementation of flexible and bidirectional charging, there are inherent advantages to medium- and heavy-duty (MHD) BEV over light duty (LD) BEVs. The number of MHD BEVs is projected to be much fewer; in 2035 the projected LD BEV stock in California is over 15 million, whereas the MHD BEV is approximately 400,000. The smaller number of MHD BEVs and chargers makes it inherently easier to control and coordinate. Simultaneously, MHD BEVs are associated with larger battery capacities than LD BEVs. MHD BEV are also likely to be operated with more sophisticated planning in fleets, and may be less likely to be affected by the randomness of the driving behaviors. MHD BEV adoption may also be shifted towards larger logistics companies with the capital to purchase these vehicles, and operating a large number of vehicles may influence the incentives of enrolling in flexible charging or V2G operations.

These unique characteristics make MHD BEVs a more enticing candidate than LD BEVs for pursuing flexible charging and V2G operations.

In this study, the adoption of MHD BEVs is considered exogenous to the decarbonization planning problem. It is assumed that the MHD BEV stock over years aligns with the existing California policy requirements, such as CARB's Advanced Clean Fleet Regulation [88]. As such, enabling V1G or V2G services could help avoid installing additional renewable generation or storage capacity with relatively little added cost and difficulty.

A great deal of literature has focused on the economic benefit of V1G and V2G. However, the majority of these works have focused on short-term costs and the economic benefit to the BEV owner. In [121], the potential revenue for BEV owners in California is examined while paying attention to future grid behavior, including wide adoption of BEVs and future grid changes. The value of BEVs has been examined for both managing load, including V2G [8] and peak shaving [61] and for providing ancillary services, like frequency regulation [115]. The authors in [116] look at a range of potential value streams for V1G and V2G services.

Somewhat less work has been done to quantify the economic benefit of enabling V1G and V2G services from the perspective of power system planning. These works generally optimize investment planning alongside dispatch and BEV charging scheduling to provide lower infrastructure costs and avoid buildout of generation and energy storage capacity. Ramirez et al [84] present a co-optimization of power system planning with dispatch of flexible charging with LD BEVs with a UK-based test system. Yao et al [128], Suski et al [109], Hajebrahimi et al [44], and Gunkel et al [42] present similar co-optimizations

with case studies in China, the Maldives, Canada, and the EU, respectively. In [32], an analysis of the potential savings of V1G and V2G, including ancillary services, is analyzed for LD BEVs in California. Xu et al [127] look at the potential emissions reductions of these services, including life-cycle analysis of BEVs.

Similarly, most works have focused on LD BEVs, rather than MHD BEV. As discussed above, these groups have rather distinct behavior, which affects both their theoretical value and practical implementability. In this work, we focus on this gap in the literature, and examine the value of V1G and V2G in California's decarbonization pathway, specifically with respect to electrification of medium- and heavy-duty vehicles.

In this chapter, we examine the potential savings and implicit costs of V1G and V2G services through the lens of California's energy transition investment planning. We start with the mixed-integer linear program (MILP) decarbonization planning model developed in Chapter 2 and incorporate a clustered representation of MHD BEV based on the timing of charging and driving. The same surrogate Lagrangian relaxation-based technique is implemented to provide computational tractability of the large MILP model. We analyze the results of the three charging regimes under two MHD BEV driving scenarios, and show a range of potential savings as high as 16 billion dollars. We also examine some of the costs related to charging services to show that the cost savings these services provide are robust. The remainder of the chapter will be organized as follows. Section 6.2 will integrate of MHD BEV into the planning formulation. Section 6.3 will discuss results and policy implications. Section 6.4 will present the conclusions.

6.2 Technical Method

The planning formulation is broadly unchanged from the one developed in Chapter 2. This section will describe the integration of MHD BEV in the unit commitment.

6.2.1 Flexible MHD BEV Charging

MHD BEV with flexible or bidirectional charging capability, is modeled similarly to storage resources, with the major exception that a large amount of discharge happens exogenously through driving, during which these resources are not connected to the power grid. To integrate MHD BEV into the planning framework, these resources are modeled as dispatchable by a central system operator, rather than a virtual power plant controlled by price signals. Each vehicle is associated with a charge start time, charge end time, starting state-of-charge, and ending state-of-charge. It is assumed that the vehicle is plugged in and available for charging for the entire duration that it is at the depot. These values essentially determine the vehicles charging needs, as well as potential V2G provisions. Modeling vehicles individually would make the problem computationally intractable; thus, vehicles are grouped by their start and end hour to form virtual power plants. The power and energy capacity parameters of the clusters are obtained as the summation of the individual parameters of the MHD BEVs in the cluster. MHD BEVs are modeled as a demand-side resource.

The control of MHD BEV clusters within optimization is operationalized by three variables: state of charge $C_e(t)$, charge power $p_e^c(t)$, and discharge power $p_e^d(t)$. These three variables are subject to limits based on the capacity of the cluster, as well as the timing at

which the cluster is connected to the grid at the depot for charging vs disconnected from the grid for driving. Discharge through driving is exogenous, and $p_e(t) = 0$ when the vehicle is not at the depot. If V2G is not considered, discharge is not allowed and $p_e^d(t) = 0, \ \forall t \in T$.

The definition of a period T allows for multiple days to be modeled consecutively, and the same charge events occur each day. To account for this, we define the set of days in the period D, where $|D| = |T|/24$ denotes the number of days in the period. We also define a time wrap t_e^Δ, to account for charging which occurs overnight. For each day, the variable state of charge at the time of depot arrival and departure is set equal to the input state of charge at the start (6.1) and end of charging (6.2).

$$C_e(t_e^{depot} + d \cdot 24) = C_e^{depot}, \ \forall d \in D, e \in E \tag{6.1}$$

$$C_e(t_e^{drive} + d \cdot 24) = C_e^{drive}, \ \forall d \in D, e \in E \tag{6.2}$$

$$t_e^\Delta = 24 \text{ if } t_e^{depot} > t_e^{drive} \text{ else } 0 \tag{6.3}$$

While the vehicle is at the depot, bounds of charge (6.4) and discharge rate (6.5), and bounds on state of charge are enforced (6.6). State of charge is also tracked with provisions for charger efficiency (6.7).

$$0 <= p_e^c(\tau(t + d \cdot 24)) <= (1 - v_e(t))P_{\bar{e}}, \ \forall e \in E, d \in D, t \in [t_e^{depot}, t_e^{drive} + t_e^\Delta] \tag{6.4}$$

$$0 <= p_e^d(\tau(t + d \cdot 24)) <= v_e(t)P_{\bar{e}}, \ \forall e \in E, d \in D, t \in [t_e^{depot}, t_e^{drive} + t_e^\Delta] \tag{6.5}$$

$$\underline{C_e} <= C_e(\tau(t)) <= C_{\bar{e}}, \ \forall e \in E, d \in D, t \in [t_e^{depot}, t_e^{drive} + t_e^\Delta] \tag{6.6}$$

$$C_e(\tau(t + 1) = C_e(\tau(t)) + p_e^c(\tau(t)\eta_e^c - p_e^d(\tau(t))\eta_e^d, \quad \forall e \in E, d \in D, t \in [t_e^{depot}, t_e^{drive} + t_e^\Delta - 1] \tag{6.7}$$

These BEV clusters can then be integrated into the broader unit commitment, and thus planning, through the power balance constraints. Zonal power balance constraints ensure that the generation and net line flows meet the load. Each zone must satisfy these constraints as:

$$\sum_{u \in U_z} p_i(t) + \sum_{s \in S_z} [p^d_s(t) - p^c_s(t)] + \sum_{r \in R_z} p_r(t) + \sum_{h \in H_z} p_h(t) + \sum_{l \in L} \lambda_{l,z} f_l(t) = L_z(t) + \sum_{e \in E_z} [p^c_e(t) - p^d_e(t)], \quad (6.8)$$

This formulation assumes that BEV can be perfectly controlled by a central operator. There are no provisions for virtual power plant control through price signals or user decisions to opt out of certain charging actions. In line with the focus on California, we assume that only CAISO has BEV flexible and bi-directional charging.

6.3 Numerical Study

This section will quantify the impact of V1G and V2G on Decarbonization planning. Subsection 6.3.1 will introduce two MHD BEV driving and charging datasets and processing them into planning model inputs. Then, we will present the results of the study, both in terms of cost savings and the overall impact on power system investment. Finally, we will examine some of the relevant costs, namely battery degradation and charging infrastructure, associated with V1G and V2G to draw conclusions about the value of adopting these services.

The decarbonization model is a zonal representation of the Western Interconnection. The model focuses on CAISO, but also represents 3 small balancing authorities in

California (LADWP, BANC, IID) and 2 aggregations of balancing authorities outside California in the Northwest and Southwest. Data is primarily taken from the RESOLVE implementation published by the California Public Utilities Commission [93]. Representative periods are selected using the sampling method in [5]. We use 10 representative periods of 3-day length. Investment is modeled in 5-year frequency from 2025 through 2045. Financing is considered through 2065.

6.3.1 Specifications for MHD BEV

Accurate modeling of V1G and V2G services requires projections of both the number of MHD BEVs and the operating characteristics of each vehicle, such as drive duration and miles traveled. In general, there is a great deal of uncertainty associated with long term planning models, due to the reliance on projections of future load, technology costs, and so on. This is compounded by the fact that this planning model is reliant both on the adoption of MHD BEV as well as the usage characteristics. While datasets exist on the driving and parking characteristics of gas and diesel trucks, it is not known if the use cases of MHD BEV will be the same.

To address this, we examine the impact of V1G and V2G MHD BEVs utilizing the simulated trip patterns in the HEVI-LOAD tool and we build an additional scenario informed by the temporal patterns extracted from a historical truck driving dataset, FleetDNA [98].

The two scenarios share the same technical underpinnings, such as MHD BEV population, charger size, and kWh/mile driving efficiency. The principle difference between the two scenarios is the temporal distribution of charging availability, as demonstrated by

the comparison of drive start times in Fig. 6.1. By presenting both scenarios, it is possible to get a look at a larger picture of the range in potential cost savings of V1G and V2G and investigate the sensitivity with respect to the trip temporal patterns. These scenarios also raise additional questions regarding the total cost and savings associated with enabling these services.

The HEVI-LOAD scenario (Scenario HL) relies on the results of the HEVI-LOAD tool, which [119] is an agent-based driving and charging simulation tool for MHD zero-emission vehicles (ZEVs) developed by the Lawrence Berkeley National Laboratory in collaboration with the California Energy Commission (CEC). HEVI-LOAD takes multiple data sources as input and resolves the integrated driving, parking, and charging/refueling behaviors of the future MHD ZEVs. Individual trucks are referred to as agents whose behaviors are constructed and calibrated utilizing multiple data sources, such as adoption projection, travel demand, telematics data, power-train specifications, etc. Trip origin and destinations are provided at the traffic analysis zones (TAZ) level for better geospatial granularities. The overall trip statistics in terms of vehicle miles traveled (VMT), energy consumption rate (kWh/mile), and vehicle stock by segment have been validated with existing state policies. HEVI-LOAD creates a virtual environment that replicates real-world transportation scenarios with fine-grained representation of electrification scenarios. However, the high geospatial resolution that HEVI-LOAD charging profiles provide are obfuscated in this study to match the load zones as we consider only CAISO-level load.

The additional scenario with varied temporal patterns (Scenario FD) is informed by the Fleet DNA dataset. This dataset is composed of thousands of historical drives

across a variety of vehicle classes, vocations, and days. Each entry has several hundred associated fields, but for our purposes, the key information extracted is drive start time, drive end time, and VMT. Then, for each drive, the efficiency mapping in Table 6.1 is used to convert VMT to kWh consumption. We assume that each vehicle charges to 100% before departing. The SoC at depot arrival can be calculated as the difference between the capacity and consumption. This dataset is combined with the California Energy Commission's 2023 AATE3 truck adoption projections [90]. Similarly to the approach in [10], we bootstrap from the Fleet DNA dataset according to the population projections by class and vocation. There are several key assumptions. Of course, bootstrapping assumes that the distribution of drive timing and distance present in Fleet DNA is the same as future MHD BEV drives in California. We assume that every vehicle drives and charges every day. It is also assumed that all charging occurs at the depot and there is no en-route charging.

Table 6.1: MHD BEV Technical Assumptions

	Charger Size (kW)	Capacity (kWh)	Efficiency (kWh/mile)
Class 2-3	150	100	0.6
Class 4-6	150	300	1.05
Class 7	150	400	1.1
Class 8	150	600	1.8

As previously mentioned, modeling each vehicle individually would make computations intractable. For both scenarios, it is necessary to cluster the individual vehicles, and the same approach is used. We assume that if the vehicle is not driving, it is plugged in at the depot, and vice-versa. As dispatch is modeled hourly, vehicle charge start times are rounded to the next hour and vehicle charge end times are rounded to the previous

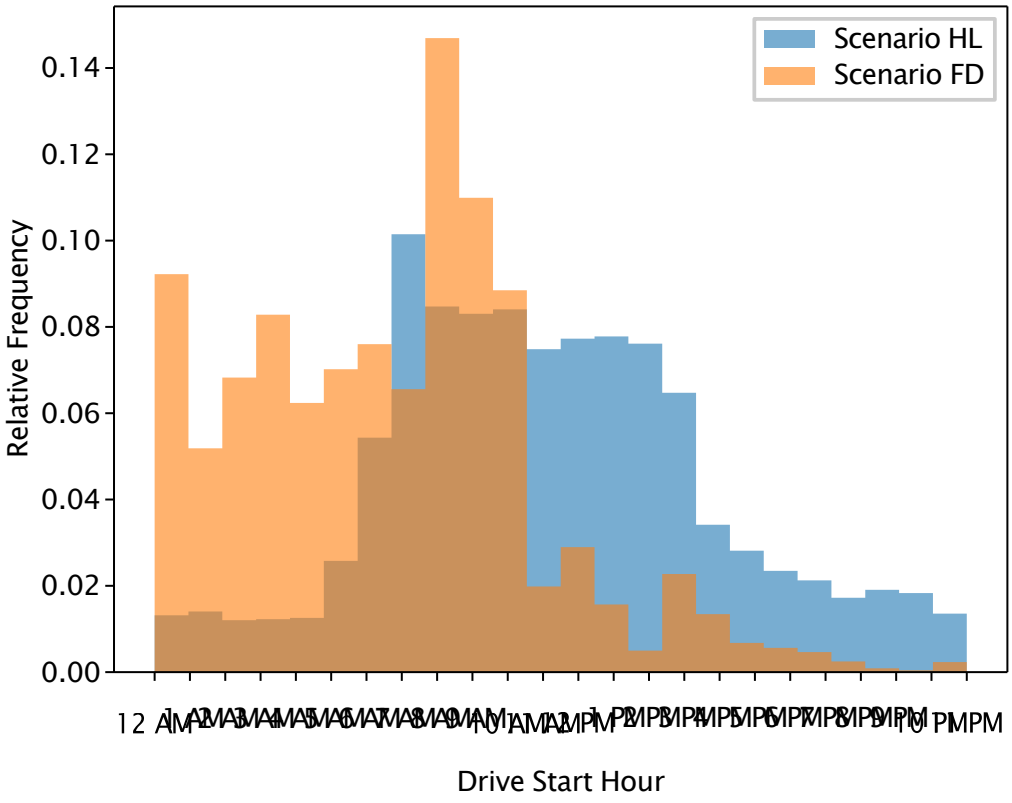

Figure 6.1: Comparison of drive start times between two scenarios.

hour. This rounding is to prevent an overestimation in the time flexibility of vehicles. First, clusters are generated by enumerating all possible combinations of start and end hour. Each vehicle is assigned to a cluster. If the cluster size accounts for less than 0.1% of all vehicles, this cluster is not modeled with V1G or V2G and left with a fixed charging profile, as this cluster would increase the associated complexity of the problem while only mildly impacting the solution due to the small number of associated controlled vehicles. This results in 87 clusters for Scenario HL, comprising in total 92% of all vehicles and 168 clusters for Scenario FD, comprising in total 94% of all vehicles.

As a result of the assumptions made in Scenario FD, and the methodology of Scenario HL, the two scenarios have some key differences in addition to the trips' temporal patterns. In Scenario HL, approximately 1 out of 3 vehicles charge each day, as many vehicles make short trips and do not need to charge. Scenario FD does not account for this, and charges each vehicle daily. However, because the underlying assumptions on VMT per day and truck efficiency are similar, the total daily MHD BEV load is extremely similar, within 1%. This means in Scenario FD, the vehicles have considerably higher starting SoC, as well as a much larger number of vehicles connected resulting in considerably higher total power and energy capacity. The results will reflect this, and the ensuing discussion will consider both the pros and cons of this detail in terms of cost.

We consider 3 charging regimes for both scenarios: a baseline case in which all charging is fixed, V1G, and V2G. For Scenario HL, fixed charging profiles are provided by HEVI-LOAD. For Scenario FD, the fixed charging profile is generated using the assumption that 50% of vehicles charge immediately at full power and 50% charge with the lowest power to fully charge by departure. For simplicity, all chargers are assumed to have 150kW rating.

6.3.2 Results

The key consideration related to V1G and V2G with respect to decarbonization planning is quantifying how enabling these services lower the cost of power system de-carbonization through lower investment, and potentially lower operation costs. Figure 6.2 shows the cumulative added capacity in year 2045. In general, V1G and V2G are associated with lower build of renewable and storage resources.

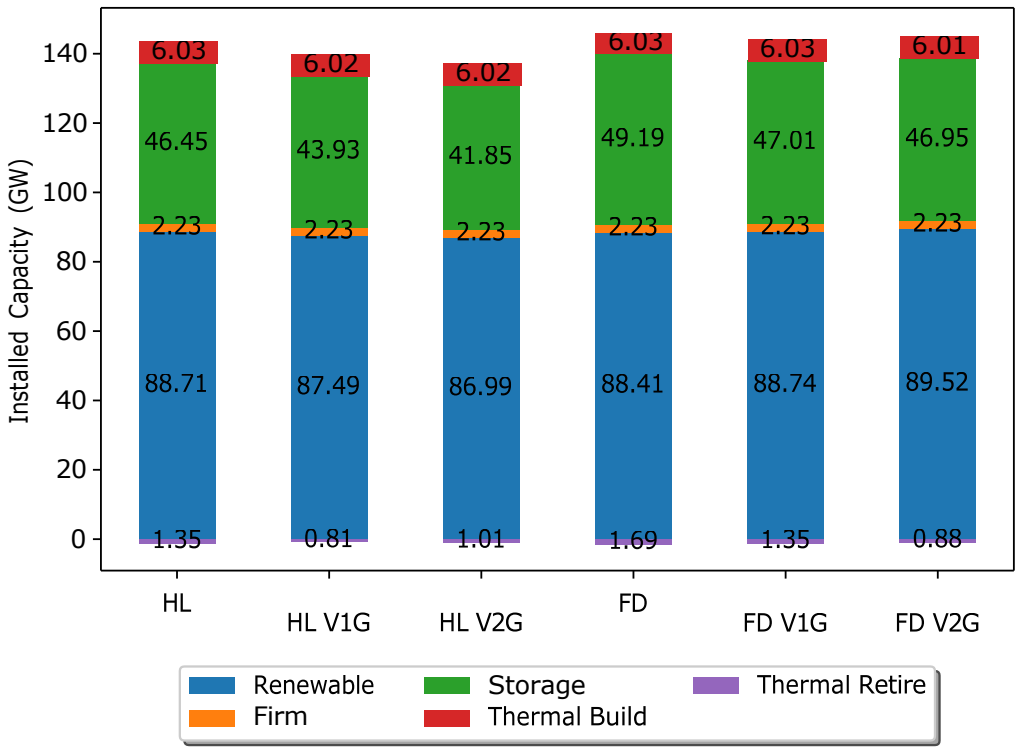

Figure 6.2: Comparison of installed resources in 2045.

By enabling V1G and V2G services, it is possible to avoid some of the installation of renewable and storage capacity that are needed in the base scenario to meet emissions targets. Accordingly, there are slightly less retirements of thermal units, which are kept online to meet the planning reserve margin.

The mechanism by which these services lower investment costs is straightforward. Figure 6.3 shows the gross load for an exemplary day in 2035 under fixed charging, V1G, and V2G. Load is shifted from hours with lower renewable generation to hours with higher renewable generation. In the case of V2G, MHD BEV are able to provide power injections at critical hours to further reduce the need for energy storage. Most MHD BEV spend the

bulk of the day driving, and thus are unable to charge when there would be most excess generation. As such, charging is mostly correlated to periods with lower variable renewable generation, and the cost savings comes mostly as avoided storage investment. This behavior is demonstrated by the visualization of net load for each regime in Fig. 6.4. V2G flattens the net load peak in the early morning, and recovers energy through the afternoon by charging batteries of stationary vehicles when renewable generation is plentiful.

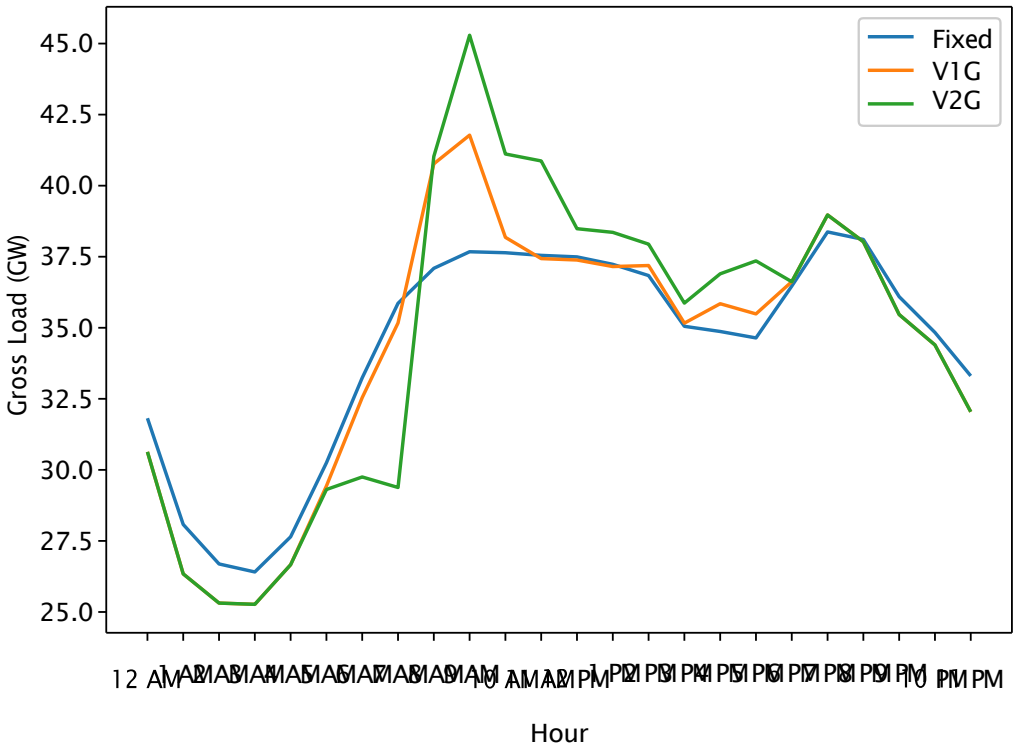

Figure 6.3: Scenario FD gross load considering fixed charging, V1G, and V2G.

Figure 6.5 shows the MHD BEV load for each hour, averaged over the year 2035. The shape of V1G and V2G load is broadly similar, with the key difference that V2G is providing power to the grid for early morning hours, between 4am and 8am, then charging

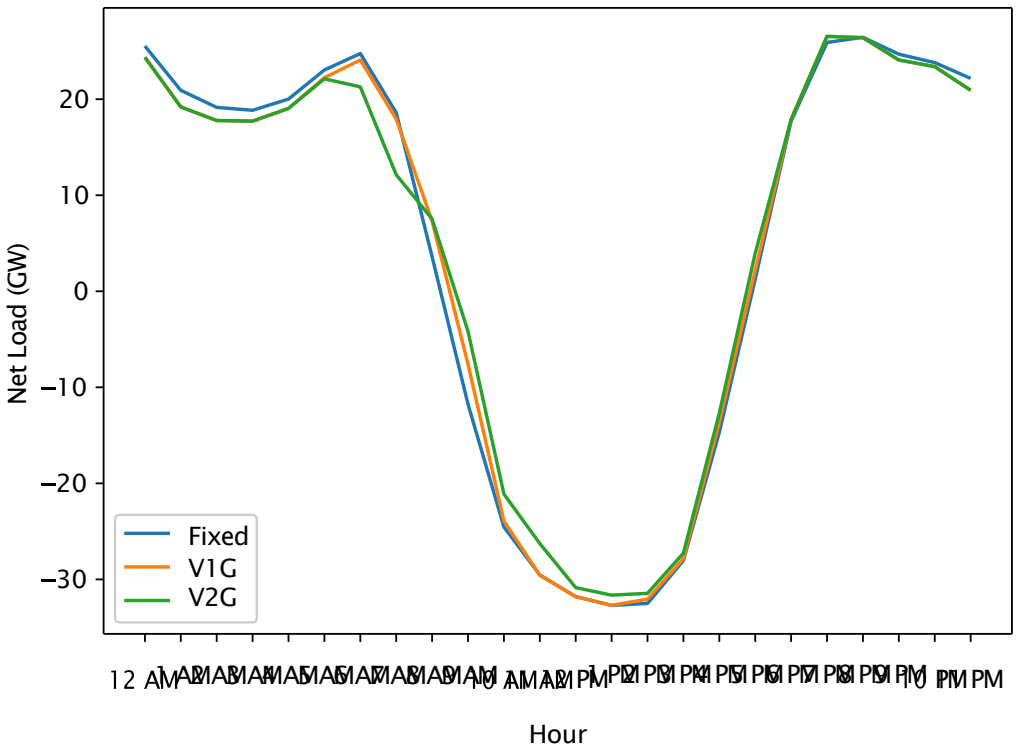

Figure 6.4: Scenario FD net load for an exemplary day in 2035.

quickly between 8am and 10am, when the bulk of vehicles are leaving. There is a large spike in charging load in the morning, as other system loads are generally lower and solar generation ramps up. This spike is even larger for V2G, as the vehicles provide power in the very early morning.

The total costs as well as costs broken down by component are shown in Table 6.2. As a resulted of the avoided investment in storage, there are substantially lower investment and maintenance costs. These services also help lower operational costs by lowering the use of thermal units. Scenario HL V1G and V2G present 3.5% and 4.6% savings over baseline

Figure 6.5: Scenario FD MHD BEV hourly load averaged over year 2035.

in California costs, respectively. Scenario FD V1G and V2G present 1.8% and 3.0% savings.

The cost savings of V1G and V2G over fixed charging are shown per vehicle, per year in Table 6.3. These costs are not discounted for the time value of money. In the worst case, V1G saves a few hundred dollars per vehicle per year. In the best case, V2G saves several thousand dollars for each vehicle each year.

Table 6.2: Costs, billions 2025$

	Scenario HL			Scenario FD		
	Fixed	V1G	V2G	Fixed	V1G	V2G
Total Cost	397.5	384.4	381.0	401.8	397.1	391.8
CA Cost	247.8	239.2	236.3	251.7	247.1	244.2
Maint. Cost	67.0	65.8	65.9	67.2	67.0	67.137
Inv. Cost	152.3	145.5	142.5	156.0	151.9	149.1
CA Op. Cost	28.5	27.9	27.9	28.4	28.2	28.0

Table 6.3: Levelized Cost Savings over Fixed Charging ($ per vehicle-year, Non-discounted)

	2025	2030	2035	2040	2045
HL V1G	2765	1337	1243	1378	997
HL V2G	4317	1840	1510	1822	1350
FD V1G	871	592	431	457	724
FD V2G	1277	1204	933	1005	999

6.3.3 Policy Implications

The decision to enable V1G and V2G services does not exist in a vacuum, and it is crucial to quantify potential costs related to these services. The main two considerations are the cost of battery degradation and the cost of charging infrastructure.

Degradation

Battery degradation is quantified using the BLAST model [97]. This model takes an input SoC time series and returns a total degradation %. We run this model for each 5 year investment interval. The goal is to understand how V1G and V2G services impact battery degradation over default operation. Understanding how the batteries degrade over this interval helps evaluate the overall cost and value of these services.

Each cluster of MHD BEV is evaluated for degradation independently. The SoC time series is created by stacking the MHD BEV SoC time series of each representative

period by their respective weights to make a yearly time series, then stacking that time series to obtain a 5-year-long time series corresponding to the investment frequency. We calculate the degradation given 3 battery chemistries (lithium-iron-phosphate, nickel-cobalt-aluminium, and nickel-manganese-cobalt) and take the average as the final degradation %. This percentage can then be converted to a total degraded kWh given the kWh capacity of each cluster.

Degradation cost is estimated by making the assumption that, at the end of the interval, batteries can be refurbished by replacement of cells to restore battery health. Thus, cost is linear with degradation. Of course, the cost of battery degradation is more complex than this, but this method provides an effective way of comparing the relative degradation between BEV charging regimes and scenarios.

We assume a battery degration cost of $100/kWh. In 2022, the cost of battery packs reached $150/kWh [92]. The cost of battery packs are expected to drop further, with projections covering a significant range. [29] predicts grid stationary battery costs will see a reduction of 16% to 47% by 2030. [39] estimates a lithium-ion battery pack cost of 72$/kWh (in 2022 $) by 2030. Thus, $100/kWh should be fairly conservative.

We examine the degradation for the Scenario FD. We present degradation for a base scenario, V1G, and V2G. Scenario FD is a good candidate for quantifying degradation because each vehicle is controlled. There is not a rigorous way of measuring degradation in Scenario HL, because during the optimization, roughly 1 in 3 vehicles charge each night. From the vehicle perspective, some MHD BEV are charging every night and some are charging less frequently. From the perspective of the grid, it does not matter which vehicles

are plugging in. As a consequence, this does not permit rigorous tracking of each vehicle's SoC.

The cost of degradation as well as the average relative battery capacity at the end of each 5-year interval is shown in Table 6.4. The impact of degradation is relatively mild. The vast majority of the degradation seems to be due to aging. Batteries experience on average an extra 0.2% of degradation for V1G vs the baseline case, and an additional 0.3% again for V2G. The critical consideration is the increase in degradation costs over baseline. Operating vehicles will necessarily incur degradation, but it is critical to understand what costs are incurred by V1G and V2G services. The cost associated with degradation is increased by 0.1 billion USD for V1G and 0.2 billion USD for V2G, as compared to the baseline. Although these costs are considerable, they are an order of magnitude less than the potential savings. As such, increased degradation is a relevant consideration, but it is not a critical risk to the business case for V1G and V2G services.

Table 6.4: Battery degradation

	Baseline	V1G	V2G
Degradation Cost (Billions)	7.5	7.6	7.7
Residual Discharge Capacity %	81.9	81.7	81.4

Cost of Chargers

In terms of BEV supply equipment costs, the most relevant factors are the cost of ensuring vehicles have sufficient access to chargers, and the cost of enabling bidirectional charging over unidirectional charging.

At time of writing, there are very limited number of V2G ready chargers on the market. Bidirectional chargers are substantially more expensive that unidirectional chargers, but it is difficult to estimate how much of that cost difference is driven by the lack of commercialization. While numerous studies examine the cost of BEV supply equipment, there are no concrete comparisons of the cost of bidirectional and unidirectional MHD BEV supply equipment. To estimate the potential cost of bidirectional chargers vs unidirectional chargers, we consider two elements which are necessary for enabling bidirectional charging. The first is an islanding switch, which can be opened to prevent energy flowing into lines, for example, when lines must be serviced. The cost of this switch is likely negligible if it is installed at the time that the charging depot is constructed. The other cost is an inverter required to convert the DC current of the MHD BEV battery to AC used by the grid. We estimate this cost using the cost of solar inverters, approximately $50 per kW [83]. The total cost of this equipment adds $1.1B to the V2G cost of Scenario HL in Table 6.2. These costs reduce substantially the potential savings of V2G. We should emphasize that the upcharge associated with V2G is purely speculative. Depending on the cost of bidirectional equipment, V2G could pose a better or worse business case.

The two scenarios are generated under different basic charging behavior assumptions, and these assumptions impact the cost related to charging in a major way. Scenario HL is an agent-based approach, in which vehicles only charge when necessary. As such, approximately 1 in 3 vehicles charge on a given day, and the number of chargers can be provided accordingly. A key assumption of Scenario FD is that each vehicle charges each day. We consider two cases which bookend the spectrum on which this could be enabled.

The first is providing every vehicle in Scenario FD with a dedicated charger. The cost of a 150kW DC fast charger is estimated at $142,200 for hardware and installation [30]. For each investment interval, we calculate the cost of installing a dedicated charger for each vehicle in Scenario FD and installing only the necessary chargers in Scenario HL. In Scenario HL, we assume that a dedicated charger is installed for each vehicle charging in a given day. In total, the cost of chargers in Scenario HL would be $20.6B and $61.6B for Scenario FD. The second is providing only the necessary number of chargers. An emerging concept is to connect multiple vehicles to a single charger. If a charger is rated at 150kW, it may be able to connect to multiple vehicles simultaneously and provide either lower power to all, or full power to individual vehicles at different times. This service could be enabled without performing substantial hardware upgrades, only by providing some additional switchgear and plugs. If we take inspiration from this, we can suggest that in Scenario FD, the number of chargers needed is proportional to the peak hourly charging demand. This brings the number of necessary chargers down substantially, to approximately 1 charger per 4 vehicles in most years. Accordingly, the cost of installing chargers drops to $14.5B. Although installing chargers is essential with or without V1G and V2G, the range in potential charger costs is extremely large, and is bigger than the potential savings associated with these services. Because of this, minimizing the number of necessary chargers is a very relevant consideration alongside lowering chargers costs with V1G and V2G.

6.4 Conclusion

In this chapter, we examined the potential costs and savings of enabling V1G and V2G services for MHD BEVs in California. Using a large scale MILP model, we calculate the savings of these services from the perspective of a central power system planner. Two scenarios are used to understand the driving and charging behavior of vehicles. We also estimate costs linked to these services. We show that battery degradation is not insignificant, but is associated with costs an order of magnitude lower than potential savings. We estimate that the cost of enabling bidirectional charging could be a very relevant element, and could weaken the business case of V2G over V1G. Carefully identifying the number of necessary chargers is of utmost importance, as costs associated with chargers could be very large.

Chapter 7

Impact of Climate Change on Decarbonization Pathways

7.1 Introduction

California has set aggressive targets for power system decarbonization through Senate Bill 100 and Senate Bill 350, establishing limits on the minimum generation from renewable resources and maximum emissions from power generation. These targets are meant to address California's contributions towards climate change mitigation. However, climate change in turn also poses adaptation challenges for these decarbonization goals.

Climate change, and the societal response to it, have a range of potential impacts on the electrical grid. The level of electrification that occurs in response to multi-sector decarbonization, such as transportation and buildings, will massively impact both the shape and magnitude of load patterns [33]. Simultaneously, climate change itself may impact

the level of heating ventilation and air conditioning (HVAC) demand and the efficiency of electric assets as temperatures rise. Weather, and thus both load and renewable generation, will likely become even more volatile towards the middle of the century. This study seeks to address some of the challenges that climate change poses to the decarbonization pathways of California's electrical grid.

We examine the impacts of climate change and associated increase in uncertainty through two lenses. The first is climate change pathways. We look at a set of scenarios that span a broad but plausible range of climate scenario uncertainty, climate model uncertainty, and socioeconomic and policy scenario uncertainty. These scenarios are derived from models in the Coupled Model Intercomparison Project Phase 6 (CMIP6) climate model archive, also utilized by the Intergovernmental Panel on Climate Change (IPCC). The second is a new approach towards ensuring resilient capacity planning, which leverages joint load-renewable generation forecasts in the face of climate change.

Large investments in electrical resources are expected over the coming decades to meet decarbonization targets. As such, capacity expansion modeling or generation expansion planning have been increasingly important for planning these investments, with models such as RESOLVE [34], Gridpath [87], and REEDS [96] used by various state agencies and load-serving entities. Due to computational limitations in capacity expansion modeling, it is typical to reduce the temporal dimension by modeling representative periods instead of all 8760 hours per year [5]. The goal of selecting representative periods is to choose a set of periods which, in tandem, best represent the year as a whole. However, some of the most stressful periods on a power grid account for only a few days per year, or may not even occur

each year. Because these fringe cases account for so little of the yearly behavior, they will not naturally be selected as representative periods. However, they still must be planned for to ensure enough generation capacity is held to reliably operate the grid. Thus, it is necessary to enforce some constraint on reliability. Typically, this is done via a planning reserve margin (PRM) constraint [86]. However, this constraint is overly rigid, as the parameters of this constraint are determined exogenously to the generation portfolio optimization. Thus, this constraint could lead to generation portfolios that are either overly cautious, or fail to respond to periods of low renewable generation.

Ensuring power system planning produces generation portfolios that are reliable has been a task of increasing importance. In the past, the generation mix was dominated by thermal units, which can generate at full-capacity except for outages or derates. This simplified reliability planning because it ensured dispatchability in resources and allowed grid operators to adjust output on demand. As the penetration of variable renewable energy increases, reliability planning is becoming increasingly difficult due to the unpredictability of meteorological patterns and the timing of generator availability (lack of dispatchable resources). A large amount of work in recent years has been devoted towards this task as well as studying the effects of rare weather events on reliability.

PRM requirements ensure a specified amount of generation capacity is held, and are sized according to the projected peak demand. Each resource in a system contributes towards the requirement as a fraction of its nameplate capacity. In the typical formulation of this approach, this fraction conveys how much of the resource's capacity is equivalent to firm capacity, such as in [36] and California Public Utilities Commission (CPUC) resource

adequacy requirements before 2018 [25]. However, for highly variable resources like wind, solar, and energy storage, distilling this complex variability into a single fraction is difficult. The behavior of wind, solar, and load are highly coupled due to the underlying weather dependency. The combined contributions of wind, solar, and energy storage are non-linear, as the contribution of storage is limited by both power and energy. To address this non-linearity, the effective load-carrying capacity (ELCC) has been proposed, and adopted by regulators including CPUC [20]. The ELCC allows for more accurate quantification of the load-carrying contributions of variable energy by accounting for the impact of renewable energy on net load. While the development of ELCC has improved PRM constraints, the parameters of this formulation are still determined exogenously, leading to inherent loss of accuracy. An overview of modern reliability studies focusing on the ELCC of renewables is given in [103]. In [86], the authors discuss the recent trend in exceeding PRM requirements, primarily due to techno-economic factors, as well as the impacts on planning studies accounting for the required and actual implemented margins. Ssengonzi et al. [107] analyze the ELCC of renewables across the United States, but neglect storage, which has a synergistic effect when coupled with variable renewables. Cole et al. [28] study resource adequacy contributions under a range of variable resource penetrations. Bera et al. [7] present a study of resource adequacy focusing on the sizing of energy storage in systems with high renewable penetration.

It is common to apply a Monte Carlo simulation approach to evaluating resource adequacy, as in [45] and the current approach used by the California Public Utilities Commission [23]. While these approaches are effective at evaluating resource adequacy of a fleet

after the system planning step, it is impossible to perform system planning over such a large temporal domain without significantly sacrificing the level of modeling detail used to represent dispatch, as is done in [69].

There is also a developing body of work around the selection of extreme events within representative period selection. For example, Scott et al. [104] select extreme weather periods as initial cluster centers in representative period selection. A range of approaches to representative period selection with extreme periods is examined in [56]. In [60], extreme days are added as representative periods in a second optimization step based on the costs associated with dispatch of the portfolio in the first optimization step. The authors in [129] select extreme periods as those periods with peak load. An iterative approach to ensuring reliability is proposed in [111], with each iteration adding the day with maximum lost load as a representative period, until the portfolio satisfies reliability metrics. These works, however, do not study the inclusion of extreme periods as a direct replacement to industry-standard PRM, making the comparison of these methods difficult.

In the study, we demonstrate that PRM-based reliability constraints are inflexible, and thus may lead to suboptimal generation fleets from planning solutions. In the case study of the California electric grid, PRM results in substantial overemphasis of thermal capacity for reliability needs. Many of the references above acknowledge the complex interactions between various classes of resources. We address these nuances through direct simulation of challenging dispatch conditions which we call resiliency periods. The proposed method is shown to meet the level of required reliability threshold at cost savings as high as 14 billion dollars through 2045. The remainder of the chapter will be organized as follows. Section

7.2 will present the formulation of the planning model, including the proposed resilient planning method. Section 7.3 discusses the numerical study and experimental results.

7.2 Technical Method

The decarbonization model used in this chapter is broadly the same as the one developed in Chapter 2. We will begin by repeating the formulation of the PRM constraint for context, and then describe the proposed method for enabling climate resilient planning.

7.2.1 Policy Constraints - Reliability

Due to the computational complexity associated with modeling 8760 hours per year, temporal downsampling is ubiquitous, and is often achieved by modeling representative periods. The goal of selecting representative periods is to choose a set of periods which, in tandem, best represent the year as a whole. However, some of the most stressful periods, and thus most important for reliability, on a power grid may be less frequent than a few days per year. Given that representative period selection is often limited to fewer than 37 days (i.e. 10% of the year), there is very low likelihood that these low-frequency, high-importance events will be selected as representative periods. However, accounting for these low-frequency periods is critical to ensure enough generation capacity is held to reliably operate the grid. Typically, this is integrated through a PRM constraint.

PRM constraints (7.1) ensure that the generation fleet for a given year can satisfy some factor above the forecasted peak load. Resources typically count towards the PRM requirement through a net qualifying capacity (NQC) or an ELCC. Thermal units, and firm

generation, generally have an NQC close to 1, as they can typically generate at full capacity except for rare occasions when unavailable due to maintenance or other circumstances. The contributions of solar, wind, and storage, are either modeled through NQC in the simplistic case or ELCC. These resources are generally associated with rather low NQC due to the high variability of generation. In more detailed representations, contribution of variable resources is modeled as a function of decreasing value with increasing penetration. For example, California's gross load typically experiences daily summer peak around 5pm local time. With the proliferation of both behind-the-meter and utility-scale solar PV, the net load peak has shifted closer to 7pm, at which time solar generation is rapidly decreasing. Essentially, this resource is saturated at the peak load time, and installing more solar will have little to no effect on the peak net load.

$$PRM_y \leq \sum_{u \in U_z} IU_{y,u} \overline{P_u} NQC_u + ELCC_{y,s} + ELCC_y + \sum_{h \in H_z} IC_{y,h} NQC_h, z = 0. \quad (7.1)$$

As a reminder, the optimization formulation is written as the minimization of investment, maintenance, and operation costs, subject to all operational and investment constraints:

$$\begin{aligned} &\min O \\ &s.t., (2.1) - (2.39) \ \forall y \in Y, w \in W, \\ &(2.43) - (2.46), (2.51) - (2.55) \ \forall y \in Y. \end{aligned} \quad (7.2)$$

7.2.2 Proposed Method: Resiliency Days

As discussed in Section 7.2.1, reliability requirements are often modeled by a yearly constraint on the fleet makeup. The goal of this constraint is to serve as a surrogate for

modeling periods with extremely severe load conditions. There is an inherent loss of fidelity associated with distilling a dispatch problem into a constraint weighting capacities of resources by predetermined factors, especially considering the high variability of renewable resources. This modeling approach is very rigid and it fails to rigorously account for the correlation between load and renewable generation. As demand is expected to be served dominantly by a mix of variable renewables and storage, this is an extremely important element. This approach also fails to account for complex dispatch behaviors, such as ramping of thermal units.

Instead of constraining the fleet through predetermined factors, we propose the direct simulation of extreme load serving conditions. We adopt the name "resiliency periods" to refer to these extreme periods, as a complementary to representative periods. While representative periods seek to embody the most typical behaviors of the power system, resiliency periods seek to embody the most extreme periods in order to directly ensure that enough capacity is held to meet these demands. Similar to PRM, the rarity of resiliency period events means these do not need to be considered for operating costs or emissions constraints.

To do so, we create new sets corresponding to the resiliency period W_r and hour within each resiliency period T_r. Resiliency periods do not necessarily need to have the same length as representative periods. For example, resiliency periods could model days ($\#T_r\# = 24$), while representative periods could model weeks ($\#T\# = 168$). As with representative periods, resiliency periods link time within, but not across periods. The index of year is the same, as we wish to enforce the resiliency requirements for each investment interval. The

optimization (7.2) can be rewritten to incorporate these periods and omit PRM as:

$$\begin{aligned} &\min O \\ &s.t., (2.1) - (2.39)\ \forall t \in T, w \in W, y \in Y, \\ &(2.1) - (2.39)\ \forall t \in T_r, w_r \in W_r, y \in Y \\ &(2.43) - (2.46), (2.51) - (2.52)\ \forall y \in Y. \end{aligned} \tag{7.3}$$

With this as the basic formulation behind using resiliency periods to enforce resilient planning, we can move on to discuss the selection of resiliency periods.

In general, the method described above can be used agnostic to the manner in which the resiliency periods are selected. However, we suggest that they should reflect the accepted standards which inform PRM calculations. For example, California uses the 1-in-10 standard, stating that the expectation of loss of load should not exceed 1 event in 10 years.

Due to the increasing importance of renewable resources, we propose the use of net load as the metric by which periods are selected. Taking inspiration from the recently proposed concept of compound energy droughts [12], we propose the use of net load over various timesteps to properly capture the effects of variability in renewable energy.

Net load is a combination of data (load, renewable generation factors) and model outputs (total installed capacity of resources), so an iterative approach is required. First,

the base model is solved while omitting reliability constraints (7.1):

$$\begin{aligned} &\min O \\ &s.t., (2.1) - (2.39)\ \forall t \in T, w \in W, y \in Y, \\ &(2.43) - (2.46), (2.51) - (2.52)\ \forall y \in Y. \end{aligned} \tag{7.4}$$

This provides baseline capacities of the various resource, which is necessary as it establishes a proper relationship between load in units of MW and unitless renewable generation factors. Net load is then defined as:

$$NL_z(y, d, t) = L_z(t) - \sum_{r \in R_z} IC_r(y) \cdot PF_r(y, d, t)(t) \tag{7.5}$$

For each year $y \in Y$, hourly net load metrics should be calculated. For this discussion, we will assume resiliency periods are selected as 24-hour days. The discussion here is immediately adaptable to periods of arbitrary length. We rely on a set of days $d \in D$ which allow for a probabilistic interpretation of load and generation shapes. The size, and members, of D essentially bound the loss of load expectation. The approach developed in the following section ensures that load can be served for all days present in D. For instance, to meet or exceed the 1-in-10 standard, D should include 3650 days. This ensures that no load shedding occurs in the represented 10 years of input data. This method is readily tailored to other reliability margins.

We propose calculating the net load for each year and selecting resiliency days based on the maximum net load when averaged over a specified duration. This is operationalized by, for each day, rolling a window of length n through the day, calculating the

average net load for each, and selecting the maximum (7.6).

$$NL^n(y, d) = \max\left(\frac{1}{n}\sum_{t}^{t+n} NL(y, d, t), \forall t \in [1, 24 - n]\right) \tag{7.6}$$

Then, we select the day with the maximum n-hour net load as a resiliency period. We refer to this highest net-load day as an n-hour net load peak.

$$W_r \leftarrow \arg\max_{d} NL^n(y, d), \forall n \in N \tag{7.7}$$

By selecting droughts of various lengths, a range of behaviors in the correlation between renewable generation and load can be represented. For example, the 1-hour drought corresponds to the day with highest hourly net load and the 24-hour drought corresponds to the day with highest average net load. The set of all n-hour durations used to form the set of resiliency periods W_r is given as N.

The proposed planning method is summarized in Algorithm 2 and displayed as a flowchart in Fig. 7.1. First, the model is solved without PRM or other reliability constraints to establish baseline resource capacity, enabling the calculation of net load for future years. Resiliency days are then selected based on the net load peaks (7.7). Finally, the model is solved once again with these resiliency days integrated, given as (7.3).

Algorithm 2: Decarbonization with Resiliency

Solve model without PRM (7.4);

Calculate net loads $NL^n(y, d)$ (7.5);

Select resiliency days (7.7);

Solve model with resiliency days (7.3);

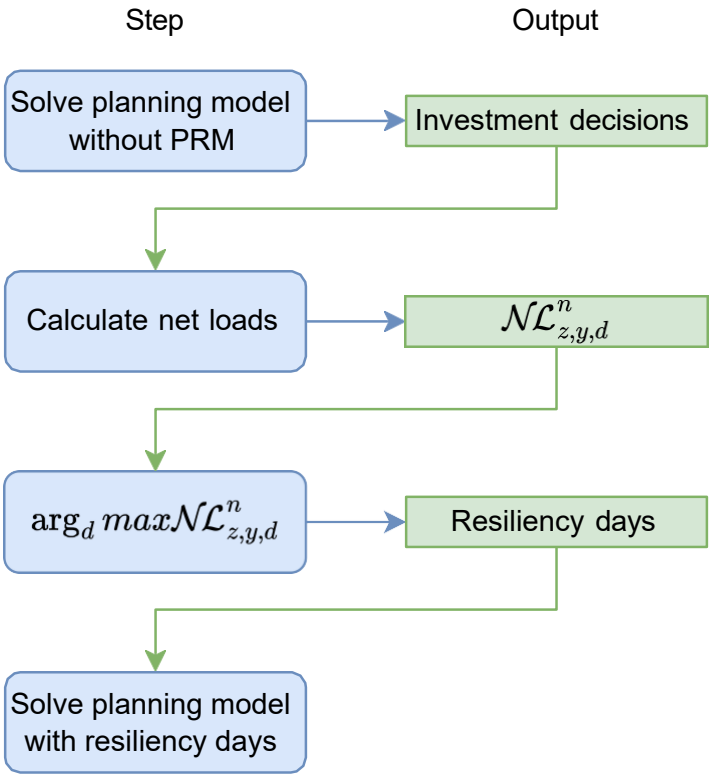

Figure 7.1: Flowchart showing the proposed method for selecting and planning with resiliency periods.

7.3 Numerical Study

In this section, we present a numerical study of the impacts of climate, socioeconomic pathways, and the proposed resilient planning method. First, we discuss the climate-generation-load dataset which enables this study. Next, we examine the planning results under a variety of scenarios. We then present a comparison of planning results under the status quo PRM reliability formulation and the proposed resiliency periods formulation. Finally, we present some of the policy implications of this study.

Outside of the climate dataset described next, the data used in this study is taken from the RESOLVE implementation published by the California Public Utilities Commissions [34]. Investment is modeled from 2025 through 2045 in 5-year intervals, with financing through 2065. We model representative periods of 3-day length and resiliency periods of 1-day length. Representative periods are selected using the approach in [5]. Resiliency periods are selected for net load peak durations of 1, 4, 12, and 24 hours. For the resiliency periods in (7.3), we also account for a 5% derate for generators in the resiliency period, based on the NQC for these units defined in [34]. The same surrogate Lagrangian relaxation technique developed in Chapter 3 is used to solve each of the models.

7.3.1 Climate, Load, and Generation Datasets

This study is enabled by recent publicly-available load and renewable generation projection datasets developed by multiple projects at the Pacific Northwest National Laboratory [11]. Load and renewable generation projections are considered in tandem, and because they are both based on the same underlying climate projections, the time series

will be properly correlated. The correlation between load and renewable generation is important, so we specifically avoid Monte Carlo approaches which generally sever this correlation. As noted in [12], there are correlations between periods of high load and low renewable generation which may have extremely large impact for future grids relying heavily on variable renewables. The renewable generation projections also account for climate impacts such as solar panel efficiency loss as temperatures increase and the suppression of wind generation under high pressure conditions.

The climate, load, and generation projections are based on 40-years (1980-2019) of historical meteorology. The use of historical meteorology results in the dataset containing actual historical extreme events. The climate methodology then takes the 40-year sequence of historical weather and repeats it twice into the future (from 2020-2059 and again from 2060-2099) for eight unique scenarios that reflect a wide but plausible range of future climate and socioeconomic conditions. The details of this approach are provided in [52]. The eight future scenarios are defined jointly by a combination of Representative Concentration Pathway (RCPs 4.5 and 8.5) and Shared Socioeconomic Pathway (SSPs 3 and 5). They also reflect a range of climate model uncertainties by using warming levels from climate models that are hotter and cooler than the multimodel mean.

The RCP scenario impacts load through both climate-related impacts, such as higher temperatures increasing the load from air conditioning, as well as the level of electrification needed to meet the emissions target. The SSPs impact load through the level of consumption, economic expansion and population growth. The RCP and the warming levels both affect the production factors of renewable generation though both technological (e.g.

solar panel derating) and climate (e.g. heat dome wind suppression), while maintaining physical consistency.

The RCP 4.5 and 8.5 pathways are denoted by R4 and R8, respectively. The SSP 3 and 5 pathways are denoted by S3 and S5, respectively. The cooler and hotter climate model outcomes are denoted by the suffixes "C" and "H", respectively (i.e. R4S3C, R4S3H). Details of the various scenarios are provided in [21].

To produce the data used in this study, the 40 years of historical and 80 years of future meteorology across the eight scenarios are then run through a series of load and renewable generation models to produce hourly time series of historical and projected load and renewable generation. The load projection model accounts for the hour-to-hour variations in demand due to weather (including extreme events like heat waves and cold snaps) and grows loads over time to reflect longer-term changes in population, economics, and energy policy (for example, electrification needs to stay on an RCP 4.5 pathway). For each 1/8th degree grid cell, a hypothetical wind and solar plant was modeled with generic assumptions about solar panel and wind turbine configurations. The details of the models used in this process are provided in [12] and [21].

7.3.2 Results

Climate Pathways

First, we examine the impacts of each of the scenarios on the 2045 fleets. The newly installed capacity in CAISO in each scenario is shown in Fig. 7.2. It is not surprising that different pathways force different levels of investment based on the amount of associated

load growth. These results do, however, emphasize the need to account for the wide range of planning outcomes of different pathways. The buildout of renewables is as low as 50GW and as high as 87.5GW, and the buildout of energy storage is as low as 32.6GW and as high as 47GW. It is vital that this range is understood, even if these extremes represent relatively less-likely scenarios. The investment over time is shown in Fig. 7.3 for both warming pathways of two of the climate scenarios. Not only do warming pathways affect the total amount of clean capacity needed to achieve decarbonization by 2045, but also the optimal intervals in which to make these investments, with both hotter scenarios favoring later investment as compared to their cooler counterparts.

Besides climate pathways, it is also critical to plan for a range of warming scenarios. Although the RCP4.5 scenarios have nearly identical fleets between the hotter and cooler warming scenarios, the RCP8.5 scenarios have notably different fleets. The hotter and cooler warming scenarios have nearly the same annual load, but the hourly loads in the hotter warming scenarios are concentrated slightly more towards the tails of the distribution. In the RCP8.5 pathway, the hotter scenarios have 16% and 21% higher renewable installations than the respective cooler scenarios. This is an early affect of climate change on load and renewable generation, and these trends will become even more severe after 2045 if the higher global emissions pathway is followed. These results also raise the issue of adaptability and mitigation. Towards the mid-century and onwards, effects of climate change are more uncertain, and so therefore are its impacts on load and generation. In turn, this can make planning more difficult, and thus require more potential investments to meet emissions targets.

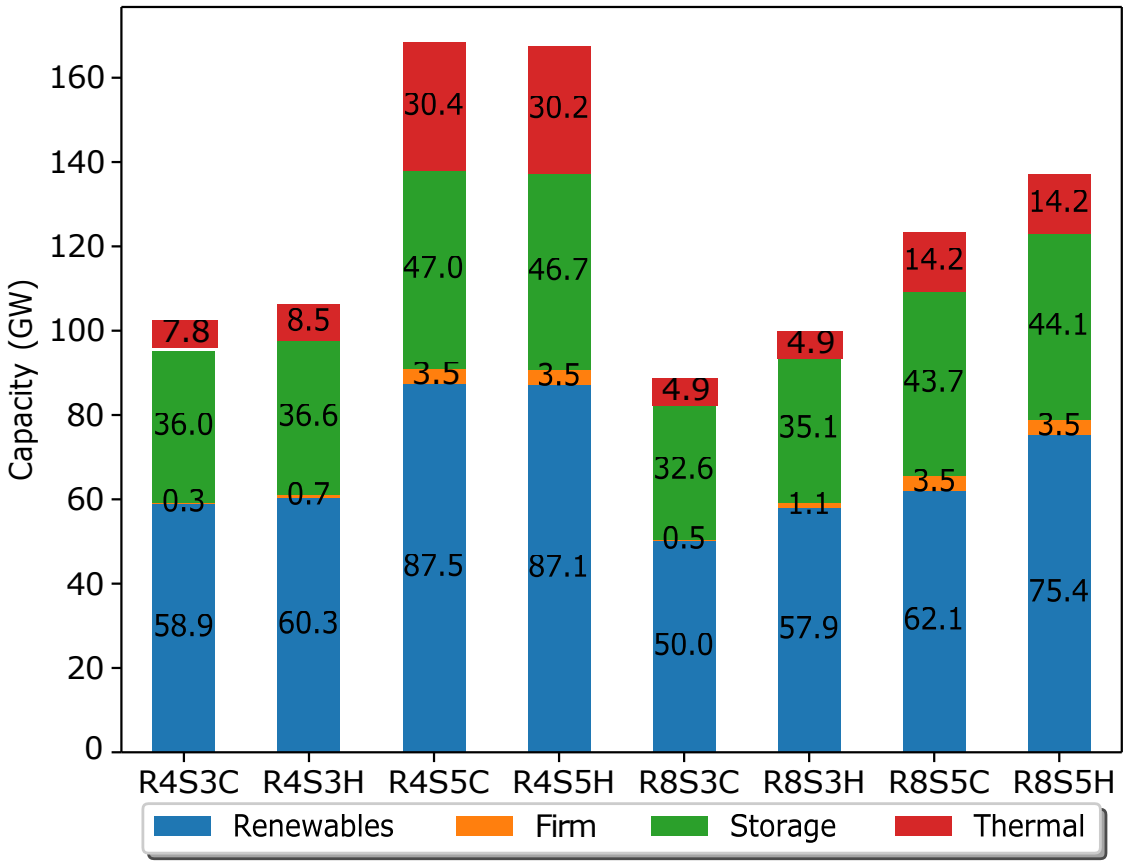

Figure 7.2: Comparison of capacity added to CAISO fleet under different climate pathways, demonstrating the extreme variation in required capacity.

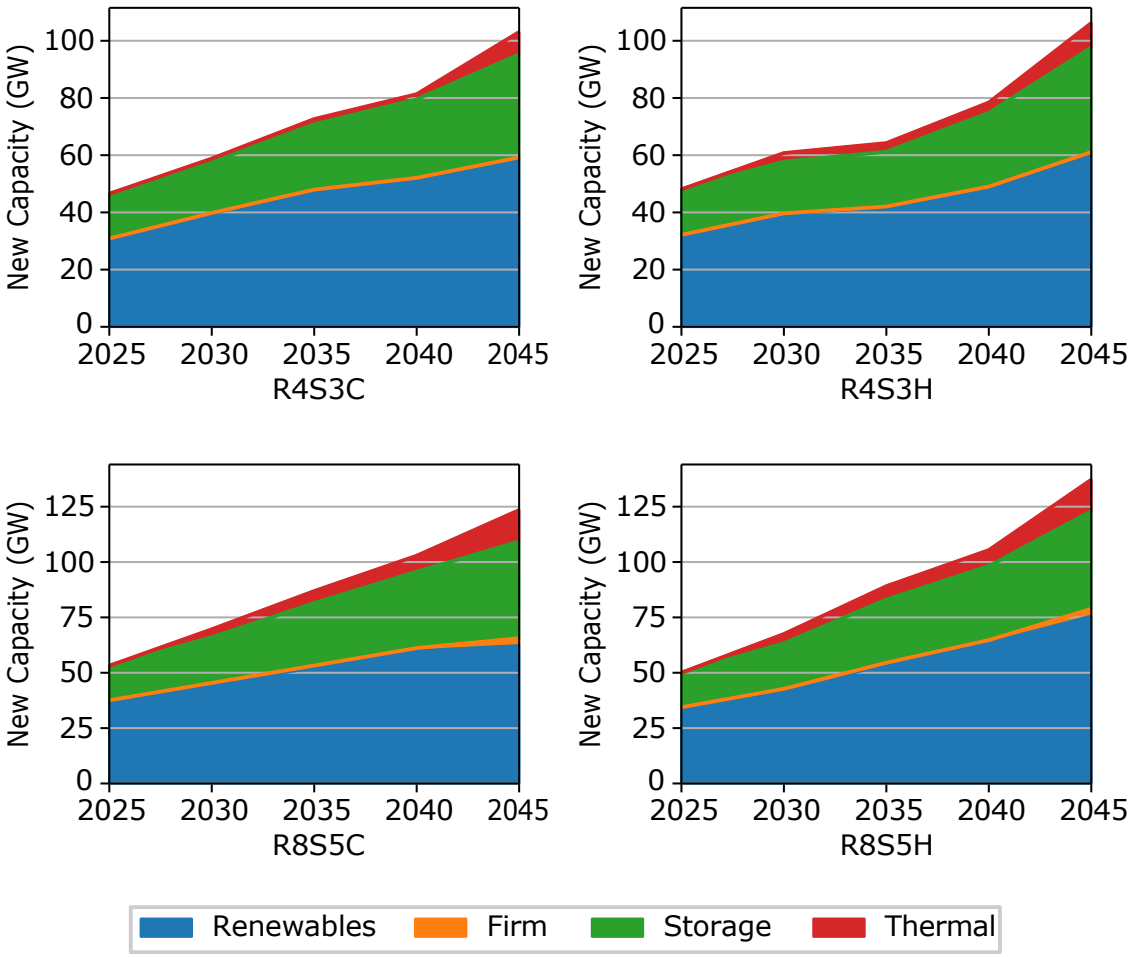

Figure 7.3: Comparison of investment over time for the R4S3 and R8S5 scenarios with both cooler and hotter warming pathways.

The evolution of the fleet can also be evaluated through its spatial distribution. Fig. 7.4 shows maps of the investment in storage and renewable projects for two years and two scenarios. These two scenarios are the most moderate of the four pathways examined. In 2030, there are relatively few differences between the two scenarios. In 2045, the differences become apparent, with substantially more investment in solar in southern California, more wind in southern Nevada, and more storage in both northern and southern California for the R8S5H scenario compared to the R4S5H scenario.

Planning with Resiliency Periods

We present the results of two planning regimes: PRM and resiliency day planning. The PRM planning regime corresponds to the approach currently employed by the California Public Utilities Commission and California Energy Commission for their decarbonization planning studies [34]. We examine the performance using two scenarios. The R8S5C scenario is the closest match in terms of yearly load to CAISO's most recent integrated energy policy report (IEPR) load forecasts. The R4S3H is the next closest match, but has slightly lower loads. For reliability considerations, we merge the hotter and cooler pathways together. In other words, to convey the variability of a warming climate, both cooler and hotter perturbations are used for the determination of PRM and selection of resiliency periods. Representative periods, however, are chosen for R8S5C and R4S3H.

We compare the results of these techniques in three ways. The first is the comparison of total cost, as well as direct comparison of fleet composition. The second is the

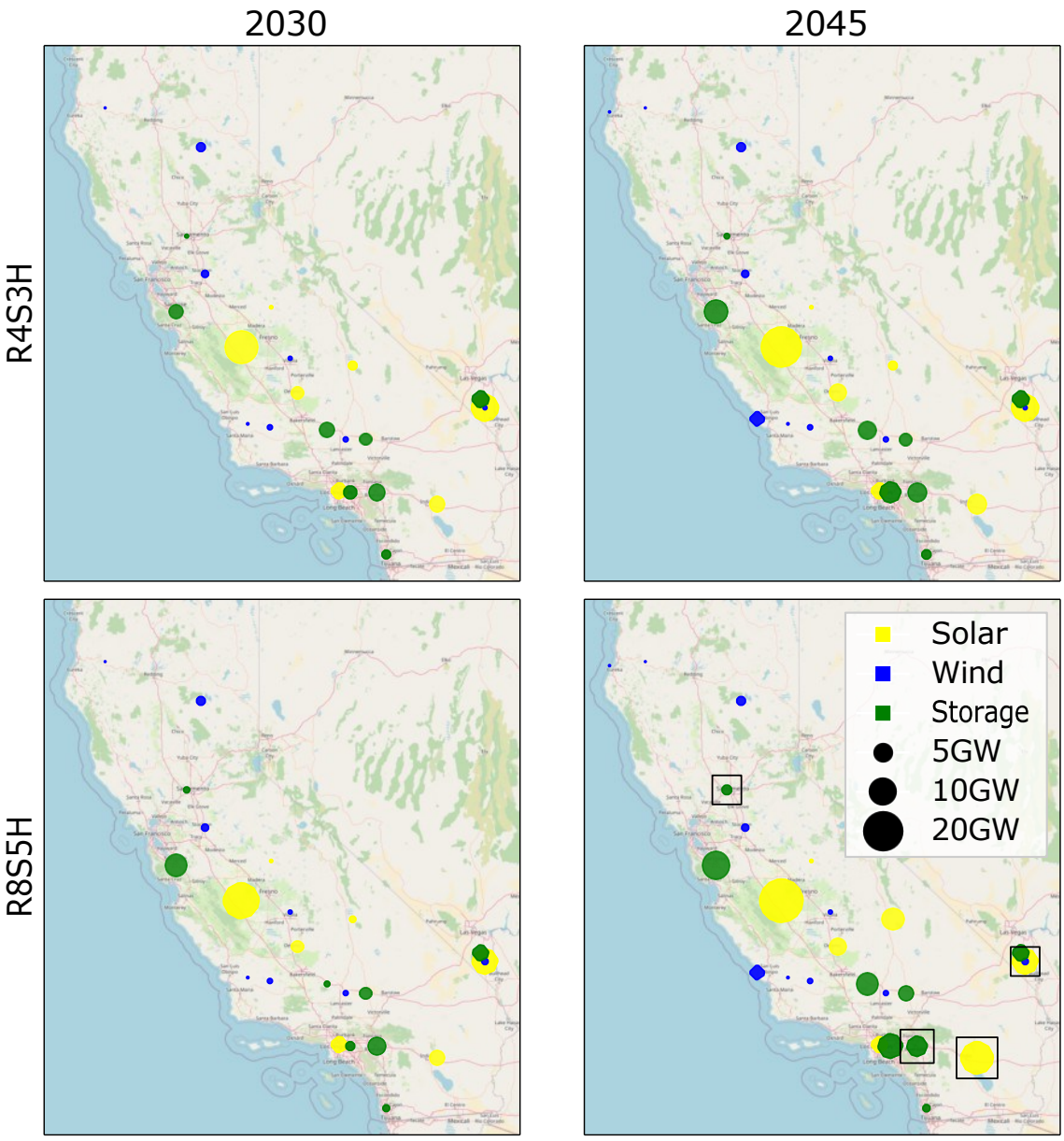

Figure 7.4: Map depicting the installation of storage and renewable projects for 2030 and 2045 across R8S5H and R4S3H, with major differences highlighted.

loss-of-load-expectation of each fleet. The third is a study of what margin above the peak load could be served by each fleet.

Table 7.1: Costs, billions 2025$

	R8S5C		R4S3H	
	PRM	Resil.	PRM	Resil.
Total Cost	365.7	351.2	329.6	322.8
CA Cost	214.9	200.6	179.3	173.3
Maint. Cost	34.8	33.1	29.6	27.6
Inv. Cost	162.3	150.6	131.4	127.8
CA Op. Cost	17.8	16.8	18.3	17.9

As shown in Table 7.1, in the R8S5C scenario, the cost savings for resilient planning over PRM is extremely substantial, over 14 billion or 6.7% in CA costs. This is mostly due to the avoided installation of several GW of thermal capacity. The R4S3H scenario presents moderately lower savings, 6 billion or 3.3%. The savings come primarily from avoided maintenance costs of economically-retired power plants and, to an extent, avoided installation of thermal capacity. This scenario has overall lower loads, so there are overall lower installations necessary to meet reliability needs. As such, the cost savings are considerably lower. In both scenarios, the operating costs are slightly lower using resilient planning as the value of renewable technologies for reliability is higher, resulting in slightly higher overall capacity of these technologies.

The cumulative investments in 2045 are shown in Fig. 7.5 for both scenarios. We also visualize the fleet resulting from planning with no reliability requirement ("No Rel.") to help compare the planning regimes. Both scenarios show generally similar levels of renewable and storage buildout between the PRM and resilient planning regimes. In the R8S5C scenario, there is a slight increase in renewable and slight decrease in storage build

for resilient planning. In the R4S3H scenario, the renewable buildout is nearly the same, but the storage buildout is slightly higher. The key difference between the three regimes is the level of retirement and construction of thermal technology. Without any reliability, there are considerable retirements. PRM encourages heavy buildout of thermal units and nearly no retirements. Resilient planning is a middle ground between the two results. The

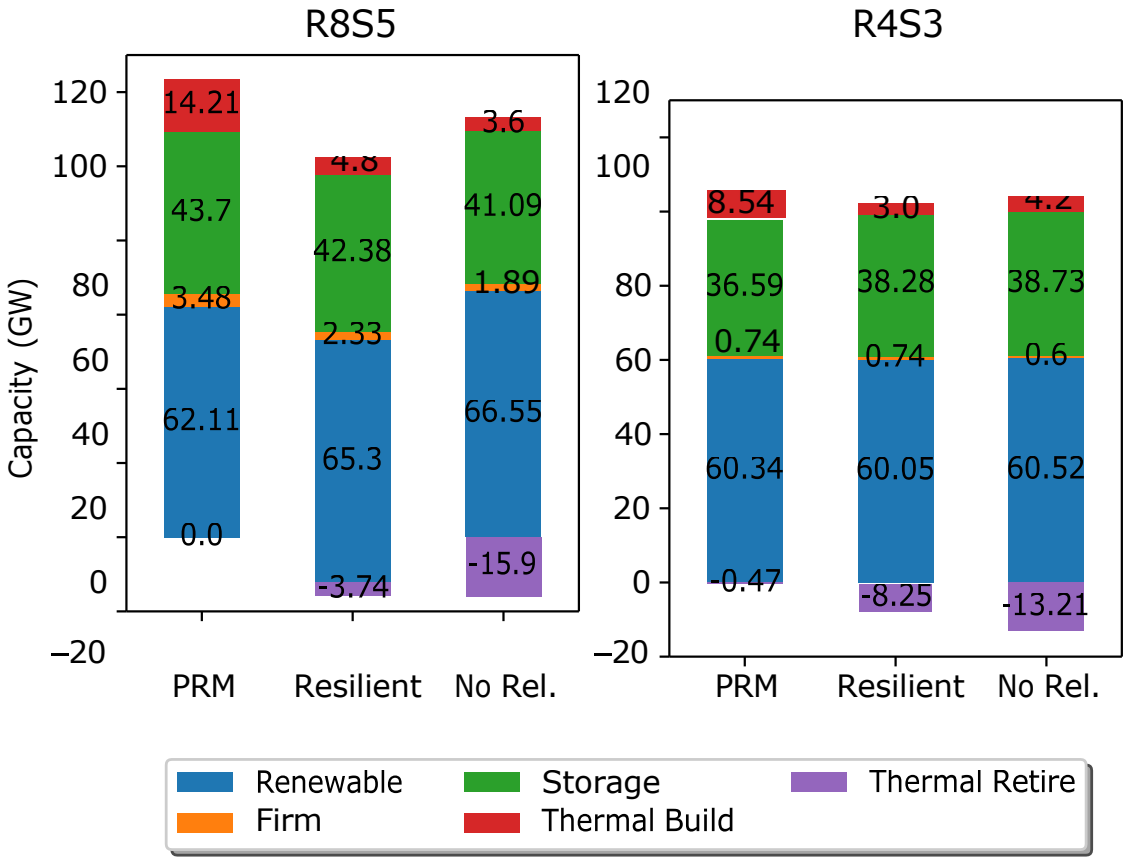

Figure 7.5: Comparison of fleets resulting from planning with PRM, the proposed resiliency days modeling, and no reliability requirement. The proposed method is less dependent on thermal units for reliability.

overall lower reliance on thermal units in the two scenarios manifests in different ways. The total thermal fleets in 2045 are shown in Fig. 7.6, including both existing units, planned retirements, economic retirements, and new construction. Not shown are combustion tur-

bine and reciprocating engine capacities which have negligible installed capacities in all scenarios. Planned retirements are minor, consisting of 2 peakers in all scenarios. Generally, combined-cycle gas turbines (CCGTs) are leveraged more heavily in 2045. In the PRM scenarios, CCGTs are the main candidate for new construction. While two smaller candidate units (steam turbine, reciprocating engine) are available, there is nearly no construction of these resources as the CCGT has a slightly lower $/MW capital cost. These resources are almost exclusively used to satisfy the reliability constraints, so their operating characteristics are less important than their capital cost in this context. These units are not visualized here due to that lack of utilization. Peakers see more retirements in general; these resources are particularly expensive to run, so pose little benefit to normal operation. However, the maintenance costs are low compared to capital costs of new construction, so peakers are retained when they serve to satisfy reliability constraints. In the lower-load R4S3H scenario, both CCGT and peakers see retirement. We have demonstrated that the fleets produced by the resilient planning regime are more economical than the PRM fleets. It is critical to demonstrate that the fleet meets the required level of reliability; otherwise, these cost savings are worthless. We seek to demonstrate that our planning approach leads to more economical, sufficiently-reliable fleets by more appropriately accounting for the load-serving potential of all resources during extreme load events. In order to demonstrate this, we directly look at the loss of load expectation. Then, we examine the level of load above these events that could be actually served by each fleet.

For each of the investment years, we run 10 years of unit dispatch, with 5 each corresponding to the cooler and hotter warming scenarios, and look for any load shedding.

Neither planning regime experiences any load shedding in this test. This should come as no surprise; dispatch of extreme load periods was directly modeled in the resilient planning regime, and the PRM regime has even higher capacities. This study thus satisfies the 1-in-10 loss of load expectation which informs CAISO's reliability margin.

Although this analysis is not shown, it is worth noting that if only the 1-hour gross load peak is selected for resilient planning, load shedding actually does occur in this dispatch stage. This highlights the importance of looking at the correlation between renewable generation and load rather than load alone for reliability considerations, as well as selecting net load peaks of several time scales as the resiliency periods.

We have shown that each fleet meets the reliability requirements; now we seek to demonstrate the degree to which each fleet exceeds the reliability requirements. We conduct a study showing the level of load above the peak load that could be served by each fleet. We start by selecting the 1-in-10 load day for each investment year. As is, this represents a reliability threshold. Then, we increasingly scale the load of this day and look to both the number of hours with load shedding. The results of this experiment are shown in Fig. 7.7. The resilient planning regime typically starts requiring load shed after a 5% or 10% increase in the load. The PRM regime starts shedding load at significantly higher percentages in every case. It also experiences fewer hours of load shedding. In the R8S5C scenario, the PRM regime can meet demand 25% higher than the 1-in-10 load in 3 out of the 5 investment years. This demonstrates that the PRM scenario is overbuilt.

The key result of this analysis is that the existing method, rather than meet the prescribed reliability standard, far exceeds it. On the other hand, the proposed method more

accurately tailors the portfolio to meet the reliability standard, at a lower overall cost. While higher levels of reliability are not adverse, significantly exceeding reliability standards by maintaining or building excess capacity is not the most efficient investment. In this sense, the proposed method is an improvement, as it meets reliability targets at significantly reduced costs. Finally, the proposed method is flexible, and could readily be applied to meet more or less stringent reliability targets by adjusting the specified probabilistic set of days D.

As previously discussed, the PRM constraint is artificial in the sense that it attempts to distill complex energy resource dispatch information into a single constraint. In the theoretically most rigorous planning model, thousands of days would be simulated. Periods of extreme weather, across a variety of patterns, would be directly simulated, and thus the resiliency requirements would be met. However, this type of analysis is currently infeasible due to computational limitations. The proposed method essentially occupies a subspace of that theoretical planning model. PRM-based models, however, transform into an entirely separate model by the addition of this constraint, and so the decisions made are of unknown optimality compared to real operations.

7.3.3 Policy Implications

The key takeaway of this study is that PRM-based reliability requirements are too inflexible. Feasibly, PRM-based constraints could suffer from the opposite problem as well, failing to account for severe energy droughts. This problem is demonstrated by the fact resilient planning using only the peak load day produces a fleet that has to shed load on days with lower gross load, but also less renewable generation.

In this case, the use of PRM results in chronic underestimation of renewables' contributions towards reliability. This in turn creates an artificial demand for new power plants, and a lack of retirement of existing power plants. The lack of retirement has considerable social implications. The resiliency regime retires roughly 60 peakers more than the PRM regime in both scenarios. Peakers are often located in densely-inhabitated areas [27], so reducing the reliance on these resources may have an even greater social impact than reducing system-wide carbon emissions. Retiring these resources may have further social benefits if the land can be converted to other uses, such as housing or greenspaces. If the state moves forward under the assumption that no existing power plants will be retired, these potential benefits will be left untapped.

On the other hand, the construction of new power plants is associated with considerable embodied carbon emissions. Even if these resources never turn on, and thus never emit, in the context of normal operation, there is a notable cost associated with their construction. We follow two estimates of the emissions associated with power plant construction to obtain estimates of 450 and 1280 tons per MW [125, 106]. For the R8S5C scenario, the PRM regime is associated with an additional 3.2-9.2 million tons of emissions. Simultaneously, the 2045 emissions target is approximately 12 million tons. Although embodied emissions are not considered as part of the energy sector emissions targets, the scale of emissions associated with this capacity that is built but not required to meet reliability needs is extremely relevant. Thus, it is crucial that resources not be overbuilt, not just for cost reduction, but for a holistic view of satisfying California's climate goals.

Long-term planning models like the present model have a separate objective from short-term planning models. The goal of long-term planning models is not to determine exactly how much capacity of each resource will be purchased for the next 20 years. The goal is to determine long-term trends in capacity. Especially from the perspective of state agencies, understanding these trends is critical as they inform at a high level the implementation of various programs. The use of PRM can affect this in several ways. Thermal capacity becomes more favorable and renewable capacity is undervalued, leading to severely overbuilt fleets. The excess of thermal capacity has both cost and social ramifications. PRM also undervalues the reliability contributions of renewable and storage capacity to the extend that it reduces the amount of investment in these technologies.

It is necessary to note that reliability studies, in particular in CAISO, encompass a large number of contingencies, some of which are not suitable to be modeled here. We are not advocating for the replacement of near-term resource adequacy studies by the proposed resiliency period modeling. However, we suggest that for the task of long-term planning, the current planning reserve margin studies may be overly rigid, and undervalue the combined value of storage, wind, and solar. This is of interest to policymakers, because these long-term planning models are not directly informing utilities which resources they should buy, but informing state agencies where regulatory and funding efforts should be directed over the next decades.

We would also like to point out that the proposed resiliency days method is ultimately a data driven method. The method is effectively driven by the available projections and statistical analysis of weather-generation-load patterns. More or less frequent events

could be chosen as the resiliency days, in line with the desired loss-of-load expectation. The resiliency days could also incorporate modeling for generator outages, line outages, and other climate-related risks. The key advantage of the proposed method is direct simulation of low-likelihood events. The specific implementation is flexible.

Simultaneously, the resiliency days method is enabled by the Lagrangian relaxation-based solution methodology, which provides substantial improvements in CPU time over traditional optimization solvers used alone. This technique allows us to solve a detailed dispatch model without sacrificing the temporal scope.

A caveat of these studies is that perfect foresight to which climate pathway will occur is impossible. This highlights the need for study of a range of pathways, as well as the importance of repeating these studies every few years to ensure that the trajectories take advantage of the best available science. Conversely, the advantage of the proposed resiliency days method is that it better represents the capacity needed to ensure reliability. The excess capacity held in the PRM scenarios, although unintentional, could be advantageous if the future loads are greater than the projection. However, we suggest that this type of uncertainty can be more robustly handled through scenario analysis, rather than unintended effects of PRM. Finally, we analyze a range of scenarios concerning climate and emissions, but do not consider various other uncertainties, including the future price of storage and renewable technologies, adoption of hydrogen fuels, increased energy efficiency, and so on. These factors are all relevant to a holistic understanding of future power system operations and investment in California.

7.4 Conclusion

In this chapter, we investigate the impacts of climate and socioeconomic scenario uncertainty on California's decarbonization pathways. First, we formulate decarbonization planning as a MILP problem. We propose a novel method for ensuring reliability, referred to as resiliency period planning. Then, we discuss the surrogate Lagrangian relaxation-based technique that enabled computational tractability of this large MILP model. We present the results of planning under a range of climate and socioeconomic scenarios and show that the range of required new capacity of renewables and storage varies on the order of 52GW. We then compare the traditional approach to reliability to the proposed method and find that the proposed method meets the required reliability threshold at costs savings of 6.7% by more rigorously valuing the contribution of renewable technologies. This study, however, presents a limited version of the reliability studies performed by regulators and utilities. Future work will expand upon the reliability studies to further investigate the comparative performance of the proposed method and planning reserve margin-based approaches. It is also acknowledged that the direct adoption of this proposed technique has significant hurdles for the risk-adverse utility industry. Instead, we hope that this study elucidates the shortcomings of the existing methods.

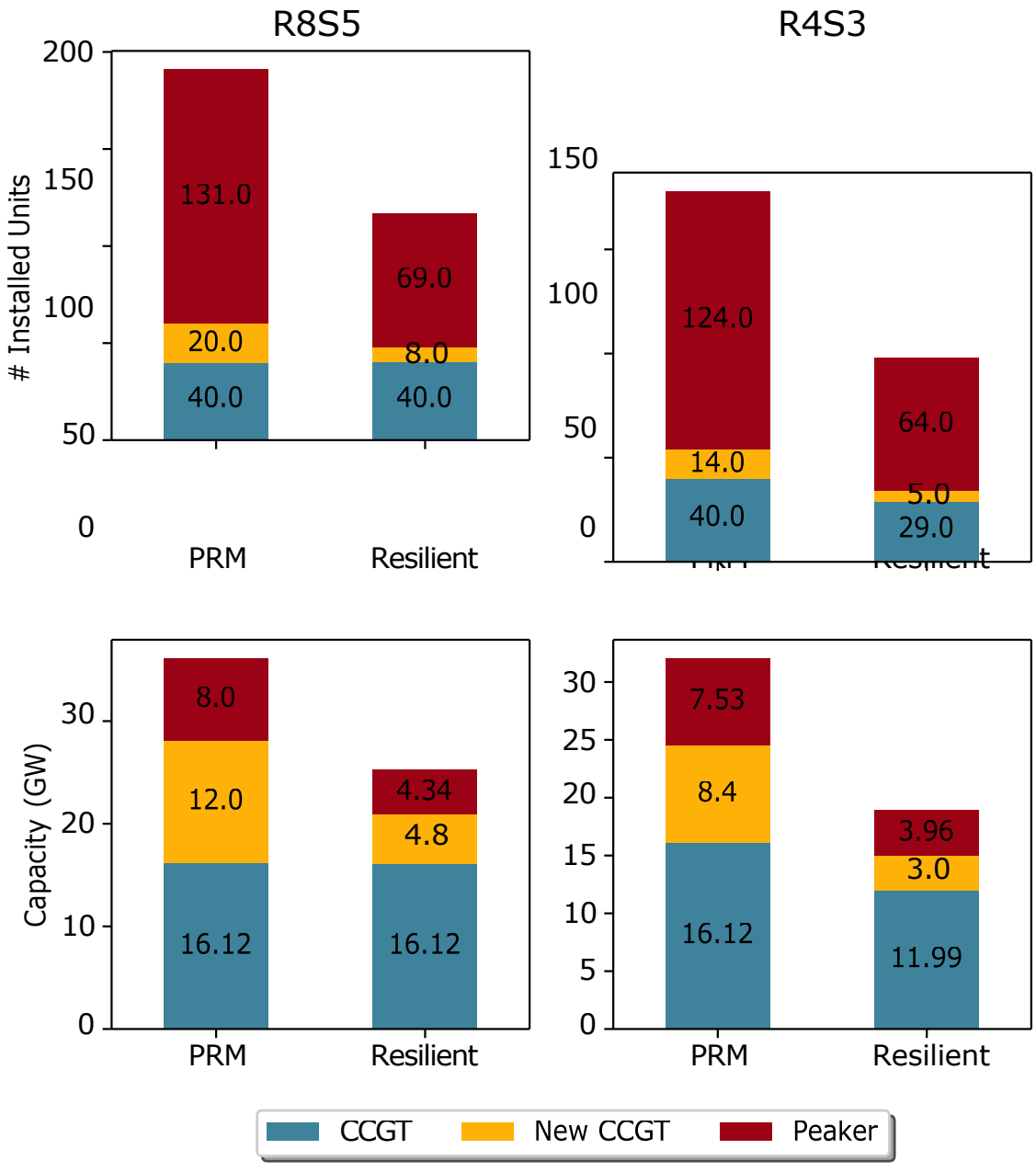

Figure 7.6: Comparison of thermal fleet components under PRM and resiliency days modeling.

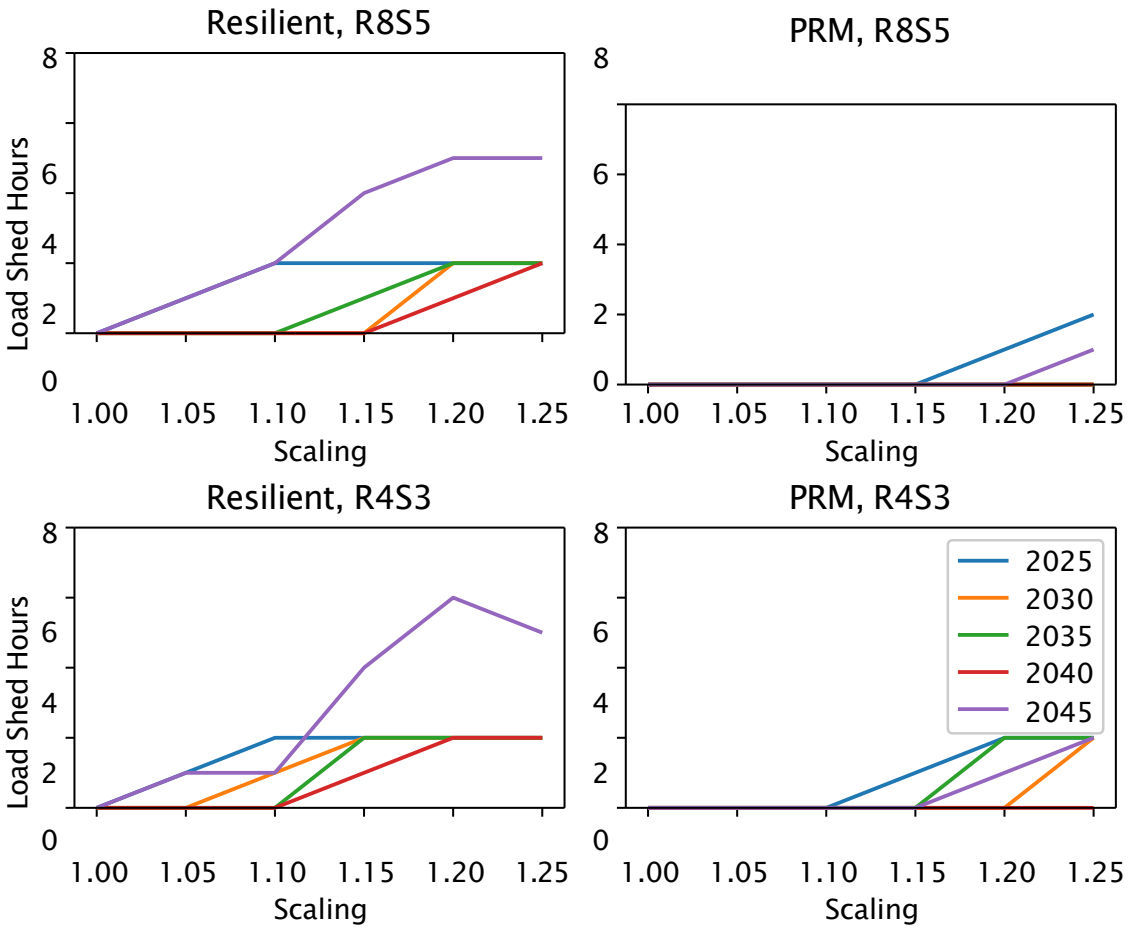

Figure 7.7: Study showing the load shedding behavior of fleets resulting from PRM and the proposed method as the peak load day is scaled higher.

Chapter 8

Conclusions

This dissertation studied the issue of planning the decarbonization of California's electrical grid through a number of lenses. In Chapter 2, a detailed decarbonization planning model was developed and formulated as a MILP optimization problem. Chapter 3 discussed the surrogate Lagrangian relaxation-based solution methodology. Chapter 4 presented results from the proposed model and method, and compared them with the status quo. In Chapter 5, two novel representative period selection algorithms were proposed. A case study considering the integration of smart vehicle charging for MHD BEVs was presented in Chapter 6. In Chapter 7, a case study of the impact of climate change pathways on decarbonization planning was shown, and a novel method for resilient planning was proposed.

The methods proposed in this dissertation push the boundaries of what has been considered computationally feasible in capacity expansion planning. The case studies consider novel, concrete applications of the methods. Still, the area of capacity expansion

planning is diverse and in urgent need. In the US, the grid will see more rapid change over the next decades than ever before. Comprehensive planning is necessary to ensure this change happens with the best possible cost, reliability, and environmental efficacy.

Planning models like the ones developed in this dissertation have a wide range of practical applications. For instance, they can help understand the long-term impacts of decisions like decommissioning of nuclear facilities or removal of hydroelectric dams. They can also help understand the long-term impacts of climate change, and ensure that the power system is resilient to extreme weather events.

The proposed studies assumed perfect foresight. To account for the uncertainties in future loads, costs, and so on, the proposed model could be converted to a stochastic optimization model. Further, these planning studies need to be repeated periodically as time passes and projections of future load, cost, and weather conditions are improved and updated.

Bibliography

[1] H. Abdulkareem Saleh Abushamah, M.R. Haghifam, and T. Ghanizadeh Bolandi. A novel approach for distributed generation expansion planning considering its added value compared with centralized generation expansion. *Sustain. Energy, Grids Netw.*, 25:100417, 2021.

[2] Behnam Alizadeh and Shahram Jadid. A dynamic model for coordination of generation and transmission expansion planning in power systems. *Int. J. Electr. Power Energy Syst.*, 65:408–418, 2015.

[3] Osten Anderson, Nanpeng Yu, and Mikhail Bragin. Optimize deep decarbonization pathways in california with power system planning using surrogate level-based lagrangian relaxation, 2023.

[4] Osten Anderson, Nanpeng Yu, Konstantinos Oikonomou, Patrick Maloney, and Di Wu. Representative period selection for robust capacity expansion planning in low-carbon grids. In *2024 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, pages 1–5, 2024.

[5] Osten Anderson, Nanpeng Yu, Konstantinos Oikonomou, and Di Wu. On the selection of intermediate length representative periods for capacity expansion. *arXiv Preprint*, 2024.

[6] Russell W Bent, Alan Berscheid, and G Loren Toole. Generation and transmission expansion planning for renewable energy integration. *Power Systems Computation Conference 2011*, 11 2010.

[7] Atri Bera, Andrew Benson, and Tu Nguyen. Reliability-based sizing of energy storage for systems with very high renewable penetration. In *2023 IEEE PESGM*, pages 1–5, 2023.

[8] Bijan Bibak and Hatice Tekiner-Mogulkoc. Influences of vehicle to grid (V2G) on power grid: An analysis by considering associated stochastic parameters explicitly. *Sustain. Energy, Grids Netw.*, 26:100429, 2021.

[9] Geoffrey Blanford, James Merrick, John Bistline, and David Young. Simulating annual variation in load, wind, and solar by representative hour selection. *Energy J.*, 39, 07 2018.

[10] Brennan Borlaug, Matteo Muratori, Madeline Gilleran, David Woody, William Muston, Thomas Canada, Andrew Ingram, Hal Gresham, and Charlie McQueen. Heavy-duty truck electrification and the impacts of depot charging on electricity distribution systems. *Nat. Energy*, 6:673 – 682, 2021.

[11] Cameron Bracken, Travis Thurber, and Nathalie Voisin. Hourly Wind and Solar Generation Profiles at 1/8th Degree Resolution. https://doi.org/10.5281/zenodo.10214348, November 2023.

[12] Cameron Bracken, Nathalie Voisin, Casey D. Burleyson, Allison M. Campbell, Z. Jason Hou, and Daniel Broman. Standardized benchmark of historical compound wind and solar energy droughts across the continental united states. *Renewable Energy*, 220:119550, 2024.

[13] Mikhail A Bragin. Lagrangian relaxation for mixed-integer linear programming: Importance, challenges, recent advancements, and opportunities. *arXiv preprint arXiv:2301.00573*, 2023.

[14] Mikhail A Bragin, Peter B Luh, Bing Yan, and Xiaorong Sun. A scalable solution methodology for mixed-integer linear programming problems arising in automation. *IEEE Trans. Autom. Sci. Eng*, 16(2):531–541, 2018.

[15] Mikhail A. Bragin, Peter B. Luh, Bing Yan, and Xiaorong Sun. A scalable solution methodology for mixed-integer linear programming problems arising in automation. *IEEE Trans. Autom. Sci. Eng*, 16(2):531–541, 2019.

[16] Mikhail A Bragin, Peter B Luh, Joseph H Yan, Nanpeng Yu, and Gary A Stern. Convergence of the surrogate Lagrangian relaxation method. *J. Optim. Theory Appl.*, 164(1):173–201, 2015.

[17] Mikhail A Bragin and Emily L Tucker. Surrogate "level-based" Lagrangian relaxation for mixed-integer linear programming. *Sci. Rep.*, 22(1):1–12, 2022.

[18] Mikhail A Bragin, Bing Yan, and Peter B Luh. Distributed and asynchronous coordination of a mixed-integer linear system via surrogate Lagrangian relaxation. *IEEE Trans. Autom. Sci. Eng*, 18(3):1191–1205, 2020.

[19] Mikhail A Bragin, Zuzhao Ye, and Nanpeng Yu. Toward efficient transportation electrification of heavy-duty trucks: Joint scheduling of truck routing and charging. *Transportation Research Part C: Emerging Technologies*, 2024.

[20] Simone Brant, Eric Dupré, Michele Kito, and Judith Iklé. The 2018 resource adequacy report. Technical report, California Public Utilities Commission, 2019.

[21] Casey Burleyson, Zarrar Khan, Misha Kulshresta, , Nathalie Voisin, and Jennie Rice. When do different scenarios of projected electricity demand start to meaningfully diverge? *Under review in Applied Energy*, 2024.

[22] Implementation of ab 2127 electric vehicle charging infrastructure assess- ments. https://efiling.energy.ca.gov/GetDocument.aspx?tn=251866&DocumentContentId=86859, 2023.

[23] Kevin Carden, Trevor Bellon, Aaron Burdick, Charles Gulian, and Arne Olson. Incremental elcc study for mid-term reliability procurement. Technical report, California Public Utilities Commission, 2023.

[24] Francesco Careri, Camillo Genesi, Paolo Marannino, Mario Montagna, Stefano Rossi, and Ilaria Siviero. Generation expansion planning in the age of green economy. *IEEE Trans. Power Syst.*, 26(4):2214–2223, 2011.

[25] Lily Chow and Simone Brant. The 2017 resource adequacy report. Technical report, California Public Utilities Commission, 2018.

[26] 2021 winter storm Uri after-action review findings report, 2021.

[27] Clean Energy Group. Phase out peakers: Replacing polluting urban power plants with renewables and battery storage. https://www.cleanegroup.org/publication/phase-out-peakers/, 2020.

[28] Wesley Cole, Daniel Greer, Jonathan Ho, and Robert Margolis. Considerations for maintaining resource adequacy of electricity systems with high penetrations of pv and storage. *Applied Energy*, 279:115795, 2020.

[29] Wesley J Cole and Akash Karmakar. Cost projections for utility-scale battery storage: 2023 update. Technical report, National Renewable Energy Laboratory, June 2023.

[30] Assess the battery-recharging and hydrogen-refueling infrastructure needs, costs and timelines required to support regulatory requirements for light-, medium-, and heavy-duty zero-emission vehicles, September 2023.

[31] Gustavo Henrique de Rosa, Douglas Rodrigues, and J o ã o Paulo Papa. Opytimizer: A nature-inspired python optimizer, November 2020.

[32] Jonathan Donadee, Robbie Shaw, Oliver Garnett, Eric Cutter, and Liang Min. Potential benefits of vehicle-to-grid technology in California: High value for capabilities beyond one-way managed charging. *IEEE Electrific. Mag.*, 7(2):40–45, 2019.

[33] Edison International. Countdown to 2045: Realizing california's pathway to net zero. https://www.edison.com/our-perspective/countdown-to-2045, 2023.

[34] Energy + Environmental Economics. Inputs & Assumptions: 2019-2020 Integrated Resource Planning . Technical report, California Public Utilities Commission, 2019.

[35] Gerald Gamrath and Marco E. Lübbecke. Experiments with a generic dantzig-wolfe decomposition for integer programs. In *The Sea*, 2010.

[36] Timo Gerres, José Pablo Chaves Ávila, Francisco Martín Martínez, Michel Rivier Abbad, Rafael Cossent Arín, and Álvaro Sánchez Miralles. Rethinking the electricity market design: Remuneration mechanisms to reach high res shares. results from a spanish case study. *Energy Policy*, 129:1320–1330, 2019.

[37] Shaghayegh Gharghabi, Shima Imani, Anthony Bagnall, Amirali Darvishzadeh, and Eamonn Keogh. Matrix profile XII: Mpdist: A novel time series distance measure to allow data mining in more challenging scenarios. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 965–970, 2018.

[38] Liz Gill, Aleecia Gutierrez, and Terra Weeks. 2021 SB 100 joint agency report: Achieving 100 percent clean electricity in California: An initial assessment. Technical report, California Energy Commission, 2021.

[39] Logan Goldie-Scot. A behind the scenes take on lithium-ion battery prices. https://about.bnef.com/blog/behind-scenes-take-lithium-ion-battery-prices/, 2019.

[40] Yang Gu, James D. McCalley, and Ming Ni. Coordinating large-scale wind integration and transmission planning. *IEEE Trans. Sustain. Energy*, 3(4):652–659, 2012.

[41] Omar J. Guerra, Diego A. Tejada, and Gintaras V. Reklaitis. An optimization framework for the integrated planning of generation and transmission expansion in interconnected power systems. *Appl. Energy*, 170:1–21, 2016.

[42] Philipp Andreas Gunkel, Claire Bergaentzlé, Ida Græsted Jensen, and Fabian Scheller. From passive to active: Flexibility from electric vehicles in the context of transmission system development. *Appl. Energy*, 277:115526, 2020.

[43] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual. "https://www.gurobi.com", 2023.

[44] Ali Hajebrahimi, Innocent Kamwa, Erick Delage, and Morad Mohamed Abdelmageed Abdelaziz. Adaptive distributionally robust optimization for electricity and electrified transportation planning. *IEEE Trans. Smart Grid*, 11(5):4278–4289, 2020.

[45] Elaine Hart and Ana Mileva. Advancing resource adequacy analysis with the gridpath ra toolkit: A case study of the western us. Technical report, Gridlab, 2022.

[46] Yuan Hu, Zhaohong Bie, Tao Ding, and Yanling Lin. An NSGA-II based multi-objective optimization for combined gas and electricity network expansion planning. *Appl. Energy*, 167:280–293, 2016.

[47] Bowen Hua, Ross Baldick, and Jianhui Wang. Representing operational flexibility in generation expansion planning through convex relaxation of unit commitment. *IEEE Trans. Power Syst.*, 33(2):2272–2281, 2018.

[48] Sanderson Hull, Arne Olson, Charlie Duff, Mengyao Yuan, Patrick O'Neill, and Joe Hooke. Least cost carbon reduction policies in PJM. https://www.ethree.com/wp-content/uploads/2020/10/E3-Least_Cost_Carbon_Reduction_Policies_in_PJM-1.pdf, 2020.

[49] IBM, LLC. ILOG CPLEX. "http://www.ilog.com/products/cplex", 2023.

[50] Shima Imani, Frank Madrid, Wei Ding, Scott Crouter, and Eamonn Keogh. Matrix profile XIII: Time series snippets: A new primitive for time series data mining. In *2018 IEEE International Conference on Big Knowledge (ICBK)*, pages 382–389, 2018.

[51] R. A. Jabr. Robust transmission network expansion planning with uncertain renewable generation and loads. *IEEE Trans. Power Syst.*, 28(4):4558–4567, 2013.

[52] Andrew Jones, Deeksha Rastogi, Pouya Vahmani, Alyssa Stansfield, Kevin Reed, Travis Thurber, Paul Ullrich, and Jennie Rice. Continental united states climate projections based on thermodynamic modification of historical weather. *Scientific Data*, 10, 09 2023.

[53] S. Kannan, S. Baskar, James D. McCalley, and P. Murugan. Application of NSGA-II algorithm to generation expansion planning. *IEEE Trans. Power Syst.*, 24(1):454–461, 2009.

[54] E. Keogh, J. Lin, and W. Truppel. Clustering of time series subsequences is meaningless: implications for previous and future research. In *Third IEEE International Conference on Data Mining*, pages 115–122, 2003.

[55] Amin Khodaei, Mohammad Shahidehpour, Lei Wu, and Zuyi Li. Coordination of short-term operation constraints in multi-area expansion planning. *IEEE Trans. Power Syst.*, 27(4):2242–2250, 2012.

[56] Leander Kotzur, Peter Markewitz, Martin Robinius, and Detlef Stolten. Impact of different time series aggregation methods on optimal energy system design. *Renewable Energy*, 117:474–487, 2018.

[57] Kevin De León. Clean energy and pollution reduction act of 2015. *California Statutes of 2015*, 2015.

[58] Kevin De León. 100 percent clean energy act of 2018. *California Statutes of 2018*, 2018.

[59] Can Li, Antonio Conejo, John Siirola, and Ignacio Grossmann. On representative day selection for capacity expansion planning of power systems under extreme operating conditions. *Int j. Electr. Power Energy Syst.*, 137:107697, 11 2021.

[60] Can Li, Antonio J. Conejo, John D. Siirola, and Ignacio E. Grossmann. On representative day selection for capacity expansion planning of power systems under extreme operating conditions. *Int. J. Electr. Power Energy Syst.*, 137:107697, 2022.

[61] Xinzhou Li, Yitong Tan, Xinxin Liu, Qiangqiang Liao, Bo Sun, Guangyu Cao, Cheng Li, Xiu Yang, and Zhiqin Wang. A cost-benefit analysis of V2G electric vehicles supporting peak shaving in Shanghai. *Electr. Power Syst. Res.*, 179:106058, 2020.

[62] Anbang Liu, Mikhail A Bragin, Xi Chen, and Xiaohong Guan. Accelerating level-value adjustment for the polyak stepsize. *arXiv preprint arXiv:2311.18255*, 2023.

[63] Anbang Liu, Peter B Luh, Bing Yan, and Mikhail A Bragin. A novel integer linear programming formulation for job-shop scheduling problems. *IEEE Robot. Autom. Lett.*, 6(3):5937–5944, 2021.

[64] Yixian Liu, Ramteen Sioshansi, and Antonio J. Conejo. Hierarchical clustering to find representative operating periods for capacity-expansion modeling. *IEEE Trans. Power Syst.*, 33(3):3029–3039, 2018.

[65] Patrick Maloney, Ping Liu, Qingyu Xu, James Mccalley, Benjamin Hobbs, Sara Daubenberger, Anders Johnson, and Stan Williams. Wind capacity growth in the Northwest United States: Cooptimized versus sequential generation and transmission planning. *Wind Eng.*, 43:0309524X1881496, 01 2019.

[66] James H. Merrick. On representation of temporal variability in electricity capacity planning models. *Energy Econ.*, 59:261–274, 2016.

[67] S.M. Moghddas-Tafreshi, H.A. Shayanfar, A. Saliminia Lahiji, A. Rabiee, and J. Aghaei. Generation expansion planning in pool market: A hybrid modified game theory and particle swarm optimization. *Energy Convers. Manag.*, 52(2):1512–1519, 2011.

[68] Amir Motamedi, Hamidreza Zareipour, Majid Oloomi Buygi, and William D. Rosehart. A transmission planning framework considering future generation expansions in electricity markets. *IEEE Trans. Power Syst.*, 25(4):1987–1995, 2010.

[69] Salman Nazir, Hisham Othman, Khoi Vu, Shiyuan Wang, Dipayan Banik, Atri Bera, Cody Newlun, Andrew Benson, and Jim Ellison. pirp: A probabilistic tool for long-term integrated resource planning of power systems. In *2022 IEEE EESAT*, pages 1–5, 2022.

[70] Najmeh Neshat and M.R. Amin-Naseri. Cleaner power generation through market-driven generation expansion planning: an agent-based hybrid framework of game theory and particle swarm optimization. *J. Clean. Prod.*, 105:206–217, 2015.

[71] Riccardo Novo, Paolo Marocco, Giuseppe Giorgi, Andrea Lanzini, Massimo Santarelli, and Giuliana Mattiazzo. Planning the decarbonisation of energy systems: The importance of applying time series clustering to long-term models. *Energy Convers. Manag.*, 15:100274, 2022.

[72] George A. Orfanos, Pavlos S. Georgilakis, and Nikos D. Hatziargyriou. Transmission expansion planning of systems with increasing wind power integration. *IEEE Trans. Power Syst.*, 28(2):1355–1362, 2013.

[73] Bryan Palmintier and Mort Webster. Impact of unit commitment constraints on generation expansion planning with renewables. In *2011 IEEE PESGM*, pages 1–7, 2011.

[74] F. Pedregosa et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 12:2825–2830, 2011.

[75] Stefan Pfenninger. Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability. *Appl. Energy*, 197:1–13, 2017.

[76] Salvador Pineda and Juan M. Morales. Chronological time-period clustering for optimal capacity expansion planning with storage. *IEEE Trans. Power Syst.*, 33(6):7162–7170, 2018.

[77] G. Pleßmann and P. Blechinger. How to meet EU GHG emission reduction targets? A model based decarbonization pathway for Europe's electricity supply system until 2050. *Energy Strategy Rev.*, 15:19–32, 2017.

[78] Boris Teodorovich Polyak. Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*, 9(3):14–29, 1969.

[79] 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. *Appl. Energy*, 162:631–643, 2016.

[80] Kris Poncelet, Hanspeter Höschle, Erik Delarue, Ana Virag, and William D'haeseleer. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Transactions on Power Systems*, 32(3):1936–1948, 2017.

[81] Jingtao Qin, Yuanqi Gao, Mikhail Bragin, and Nanpeng Yu. An optimization method-assisted ensemble deep reinforcement learning algorithm to solve unit commitment problems. *IEEE Access*, 11:100125–100136, 2023.

[82] Niranjan Raghunathan, Mikhail A Bragin, Bing Yan, Peter B Luh, Khosrow Moslehi, Xiaoming Feng, Yaowen Yu, Chien-Ning Yu, and Chia-Chun Tsai. Exploiting soft constraints within decomposition and coordination methods for sub-hourly unit commitment. *Int. J. Electr. Power Energy Syst.*, 139:108023, 2022.

[83] Vignesh Ramasamy, Jarett Zuboy, Michael Woodhouse, Eric O'Shaughnessy, David Feldman, Jal Desai, Andy Walker, Robert Margolis, and Paul Basore. U.S. solar photovoltaic system and energy storage cost benchmarks, with minimum sustainable price analysis: Q1 2023. Technical report, National Renewable Energy Laboratory, 2023.

[84] Pedro J. Ram´ırez, Dimitrios Papadaskalopoulos, and Goran Strbac. Co-optimization of generation expansion planning and electric vehicles flexibility. *IEEE Trans. Smart Grid*, 7(3):1609–1619, 2016.

[85] Seyyed A. Rashidaee, Turaj Amraee, and Mahmud Fotuhi-Firuzabad. A linear model for dynamic generation expansion planning considering loss of load probability. *IEEE Trans. Power Syst.*, 33(6):6924–6934, 2018.

[86] Andrew Reimers, Wesley Cole, and Bethany Frew. The impact of planning reserve margins in long-term planning models of the electricity sector. *Energy Policy*, 125, 10 2018.

[87] Blue Marble Analytics. Gridpath. https://github.com/blue-marble/gridpath.

[88] California Air Resources Board. Advanced clean fleets regulation - drayage truck requirements. https://ww2.arb.ca.gov/resources/fact-sheets/advanced-clean-fleets-regulation-drayage-truck-requirements, 2023.

[89] California Air Resources Board. Current california ghg emission inventory data. https://ww2.arb.ca.gov/ghg-inventory-data, 2023.

[90] California Energy Commission. Statewide vehicle stock forecast and scenarios. https://www.energy.ca.gov/data-reports/california-energy-planning-library/forecasts-and-system-planning/demand-side-modeling, 2022.

[91] California Energy Commission. Data show clean power increasing, fossil fuel decreasing in california. https://www.energy.ca.gov/news/2023-08/data-show-clean-power-increasing-fossil-fuel-decreasing-california, 2023.

[92] Department of Energy. Electric vehicle battery pack costs in 2022 are nearly 90% lower than in 2008, according to DOE estimates. https://www.energy.gov/eere/vehicles/articles/fotw-1272-january-9-2023-electric-vehicle-battery-pack-costs-2022-are-nearly, 2023.

[93] Energy + Environmental Economics. E3 RESOLVE model. https://github.com/AaronHolm/RESOLVE, 2019.

[94] Federal Energy Regulatory Commission. Open access transmission tariff (OATT) reform. https://www.ferc.gov/power-sales-and-markets/open-access-transmission-tariff-oatt-reform, 2023.

[95] MIT Energy Initiative and Princeton University ZERO lab. Genx: a configurable power system capacity expansion model for studying low-carbon energy futures. https://github.com/GenXProject/GenX.

[96] National Renewable Energy Laboratory. Regional energy deployment system model 2.0 (reeds 2.0). https://www.nrel.gov/analysis/reeds/index.html.

[97] National Renewable Energy Laboratory. BLAST-Lite. https://www.nrel.gov/transportation/blast.html, 2023.

[98] National Renewable Energy Laboratory. Fleet DNA project data. www.nrel.gov/fleetdna, 2023.

[99] United Nations. Paris agreement. *United Nations Treaty Collection*, 2015.

[100] U.S. Energy Information Administration. Prices for california's emissions credits increase in early 2022 auction. https://www.eia.gov/todayinenergy/detail.php?id=51918, 2022.

[101] U.S. Energy Information Agency. Us electricity profile 2022. https://www.eia.gov/electricity/state/, 2023.

[102] Office of Fossil Energy and Carbon Management. Carbon negative shot. https://www.energy.gov/fecm/carbon-negative-shot, 2023.

[103] Nick Schlag, Zach Ming, Arne Olson, Lakshmi Alagappan, Ben Carron, Kevin Steinberger, and Huai Jiang. Capacity and reliability planning in the era of decarbonization: Practical application of effective load carrying capability in resource adequacy. Technical report, Energy + Environmental Economics, 2020.

[104] Ian J. Scott, Pedro M.S. Carvalho, Audun Botterud, and Carlos A. Silva. Clustering representative days for power systems generation expansion planning: Capturing the effects of variable renewables and energy storage. *Appl. Energy*, 253:113603, 2019.

[105] Jiraporn Sirikum, Anulark Techanitisawad, and Voratas Kachitvichyanukul. A new efficient GA-Benders' decomposition method: For power generation expansion planning with emission controls. *IEEE Trans. Power Syst.*, 22(3):1092–1100, 2007.

[106] Pamela Spath and Margaret Mann. Life cycle assessment of a natural gas combined cycle power generation system. Technical report, National Renewable Energy Laboratory, 12 2000.

[107] Jethro Ssengonzi, Jeremiah X. Johnson, and Joseph F. DeCarolis. An efficient method to estimate renewable energy capacity credit at increasing regional grid penetration levels. *Renewable and Sustainable Energy Transition*, 2:100033, 2022.

[108] Xiaorong Sun, Peter B Luh, Mikhail A Bragin, Yonghong Chen, Jie Wan, and Fengyu Wang. A novel decomposition and coordination approach for large day-ahead unit commitment with combined cycle units. *IEEE Trans. Power Syst.*, 33(5):5297–5308, 2018.

[109] Adam Suski, Tom Remy, Debabrata Chattopadhyay, Chong Suk Song, Ivan Jaques, Tarek Keskes, and Yanchao Li. Analyzing electric vehicle load impact on power systems: Modeling analysis and a case study for Maldives. *IEEE Access*, 9:125640–125657, 2021.

[110] Holger Teichgraeber and Adam R. Brandt. Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. *Appl. Energy*, 239:1283–1293, 2019.

[111] Holger Teichgraeber, Lucas Elias Küpper, and Adam R. Brandt. Designing reliable future energy systems by iteratively including extreme periods in time-series aggregation. *Applied Energy*, 304:117696, 2021.

[112] Diego A. Tejada-Arango, Maya Domeshek, Sonja Wogrin, and Efraim Centeno. Enhanced representative days and system states modeling for energy storage investment analysis. *IEEE Trans. Power Syst.*, 33(6):6534–6544, 2018.

[113] Diego A. Tejada-Arango, Germán Morales-España, Sonja Wogrin, and Efraim Centeno. Power-based generation expansion planning for flexibility requirements. *IEEE Trans. Power Syst.*, 35(3):2012–2023, 2020.

[114] Tjark Thien, Tobias Blank, Benedikt Lunz, and Dirk Uwe Sauer. Life cycle cost calculation and comparison for different reference cases and market segments. In *Electrochemical Energy Storage for Renewable Sources and Grid Balancing*, pages 437–452. Elsevier, Amsterdam, 2015.

[115] Andreas Thingvad, Lisa Calearo, Peter B. Andersen, Mattia Marinelli, Myriam Neaimeh, Kenta Suzuki, and Kensuke Murai. Value of V2G frequency regulation in Great Britain considering real driving data. In *2019 IEEE ISGT Europe*, pages 1–5, 2019.

[116] Andrew W. Thompson and Yannick Perez. Vehicle-to-everything (V2X) energy services, value streams, and regulatory policy implications. *Energy Policy*, 137:111136, 2020.

[117] William W. Tso, C. Doga Demirhan, Clara F. Heuberger, Joseph B. Powell, and Efstratios N. Pistikopoulos. A hierarchical clustering decomposition algorithm for optimizing renewable power systems with storage. *Appl. Energy*, 270:115190, 2020.

[118] Felipe Verástegui, Álvaro Lorca, Daniel Olivares, and Matias Negrete-Pincetic. Optimization-based analysis of decarbonization pathways and flexibility requirements in highly renewable power systems. *Energy*, 234:121242, 2021.

[119] Bin Wang and Cong Zhang. Medium- & heavy-duty electric vehicle infrastructure load, operations and deployment tool (HEVI-LOAD). https://www.energy.ca.gov/sites/default/files/2021-09/5%20LBNL-FTD-EAD-HEVI-LOAD%20Medium-%20and%20Heavy-Duty%20Load%20Shapes_ADA.pdf, 2021.

[120] Bo Wang, Min Zhou, Bo Xin, Xin Zhao, and Junzo Watada. Analysis of operation cost and wind curtailment using multi-objective unit commitment with battery energy storage. *Energy*, 178:101–114, 2019.

[121] Meiye Wang and Michael T. Craig. The value of vehicle-to-grid in a decarbonizing California grid. *J. Power Sources*, 513:230472, 2021.

[122] Michal Wierzbowski, Wojciech Lyzwa, and Izabela Musial. MILP model for long-term energy mix planning with consideration of power system reserves. *Appl. Energy*, 169:93–111, 2016.

[123] Jianghua Wu, Peter B Luh, Yonghong Chen, Mikhail A Bragin, and Bing Yan. A novel optimization approach for sub-hourly unit commitment with large numbers of units and virtual transactions. *IEEE Trans. Power Syst.*, 37(5):3716–3725, 2021.

[124] Jianghua Wu, Peter B Luh, Yonghong Chen, Bing Yan, and Mikhail A Bragin. Synergistic integration of machine learning and mathematical optimization for unit commitment. *IEEE Trans. Power Syst.*, 2023.

[125] X.D. Wu, J.L. Guo, and G.Q. Chen. The striking amount of carbon emissions by the construction stage of coal-fired power generation system in china. *Energy Policy*, 117:358–369, 2018.

[126] Xpress optimization, LLC. Xpress optimization Reference Manual. "https://www.fico.com/en/products/fico-xpress-optimization", 2023.

[127] Lei Xu, Hasan Ü mitcan Yilmaz, Zongfei Wang, Witold-Roger Poganietz, and Patrick Jochem. Greenhouse gas emissions of electric vehicles in Europe considering different charging strategies. *Transp. Res. D*, 87:102534, 2020.

[128] Xing Yao, Ying Fan, Fan Zhao, and Shao-Chao Ma. Economic and climate benefits of vehicle-to-grid for low-carbon transitions of power systems: A case study of China's 2030 renewable energy target. *J. Clean. Prod.*, 330:129833, 2022.

[129] Ali Yeganefar, Mohammad Reza Amin-Naseri, and Mohammad Kazem Sheikh-El-Eslami. Improvement of representative days selection in power system planning by incorporating the extreme days of the net load to take account of the variability and intermittency of renewable resources. *Applied Energy*, 272:115224, 2020.

[130] Chin-Chia Michael Yeh, Nickolas Kavantzas, and Eamonn Keogh. Matrix profile VI: Meaningful multidimensional motif discovery. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 565–574, 2017.

[131] Chin-Chia Michael Yeh, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen, and Eamonn Keogh. Matrix profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 1317–1322, 2016.

[132] Zhenhai Zhang, Jie Shi, Yuanqi Gao, and Nanpeng Yu. Degradation-aware valuation and sizing of behind-the-meter battery energy storage systems for commercial customers. In *2019 IEEE PES GTD Gd. Int. Conf. Expo. Asia*, pages 895–900, 2019.

[133] Xing Zhao, Peter B Luh, and Jihua Wang. Surrogate gradient algorithm for Lagrangian relaxation. *J. Optim. Theory Appl.*, 100(3):699–712, 1999.