

Task-Aware Spatial and Temporal Aggregation for Capacity Expansion Planning

by

Gabriel Duguey

Master of Science in Mathematics and Physics, École CentraleSupélec, 2025

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

**MASTER OF ENGINEERING IN CIVIL AND ENVIRONMENTAL
ENGINEERING**

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2025

© 2025 Gabriel Duguey. All rights reserved.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by: Gabriel Duguey
Department of Civil and Environmental Engineering
August 8, 2025

Certified by: Saurabh Amin
Professor of Civil and Environmental Engineering,
Co-Director, Operations Research Center, Thesis Supervisor

Accepted by: Heidi Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Task-Aware Spatial and Temporal Aggregation for Capacity Expansion Planning

by

Gabriel Duguey

Submitted to the Department of Civil and Environmental Engineering
on August 8, 2025 in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING IN CIVIL AND ENVIRONMENTAL
ENGINEERING

ABSTRACT

As we plan tomorrow's electricity system, we face fundamental questions: where should new power plants go, which technologies deserve investment, and how much transmission is enough? These decisions are the domain of Capacity Expansion Planning (CEP), a class of optimization models that guide long-term infrastructure investments in power systems. To be realistic, CEP models must capture fine-grained spatial and temporal variations because demand varies by city and climate, while wind and solar output depend on weather patterns that shift hour by hour and location by location. But representing the system with thousands of time steps and hundreds of nodes makes the optimization problem computationally too large to solve.

This thesis addresses the core question: how can spatial and temporal aggregation in CEP models be designed to preserve planning-relevant patterns that drive investment decisions? Existing approaches often treat aggregation as a neutral preprocessing step, relying on heuristics like political boundaries or geographic proximity. In contrast, we propose a task-aware pipeline that treats aggregation as an integral modeling decision, explicitly aligned with planning objectives.

The approach builds a composite similarity metric that blends diverse planning-relevant signals, including, but not limited to, duration curves, ramping behavior, and spatial correlation, and uses k -medoids clustering to define spatial zones. Temporal aggregation is then applied to daily system-wide profiles, selecting representative days that maintain cross-zonal interactions. The result is a reduced spatio-temporal dataset fed into a CEP model. The resulting investment decisions are re-evaluated at full resolution to evaluate their feasibility and real cost.

Experiments on a New England case study show the pipeline consistently outperforms common baselines like political boundaries, geographic proximity, or capacity factor statistics. Among 50 feature weightings, the best design reduces system cost by 13% compared to heuristics. Correlation-based features drive the best results, while raw amplitude and geographic location often degrade performance when used alone.

Thesis supervisor: Saurabh Amin

Title: Professor of Civil and Environmental Engineering,
Co-Director, Operations Research Center

Acknowledgments

I would like to express my deepest gratitude to Professor Saurabh Amin for his guidance, support, and trust throughout this project. His feedback and encouragement have shaped both the direction and the depth of this work. I am also grateful to Rahman Khorramfar for closely following this thesis and for the many valuable discussions along the way.

A special thanks to Aron Brenner, whose generosity with his time, and ideas made a real difference in the development of this project. His contributions went beyond advice, he actively helped strengthen the foundation of the CEP framework and made the process both more rigorous and more enjoyable.

I also want to thank my friends Carlotta Barone-MacDonald, Evrard Constant, and Noé Bertramo. Their presence made this year not only meaningful but also joyful, and their support has quietly but surely contributed to the completion of this thesis.

Finally, I thank my parents for their unwavering support and care throughout this journey, even during times when I may not have been the most restful companion. Their belief in me means more than I can say.

Contents

| | |
|--|-----------|
| <i>List of Figures</i> | 9 |
| <i>List of Tables</i> | 11 |
| 1 Introduction | 13 |
| 1.1 The Energy System Planning Challenge | 13 |
| 1.2 Aggregation as a Design Problem | 14 |
| 1.3 Thesis Objectives and Contributions | 15 |
| 2 Literature Review | 19 |
| 2.1 Temporal Aggregation | 19 |
| 2.1.1 Representative Period Construction | 20 |
| 2.1.2 Linking Periods and Evaluating Aggregation | 23 |
| 2.2 Spatial Aggregation | 24 |
| 2.2.1 From Administrative Boundaries to Data-Driven Clustering | 24 |
| 2.2.2 How Clustering Logic Shapes System Design | 24 |
| 2.2.3 Beyond Similarity: Task-Aware and Hybrid Aggregation | 26 |
| 3 Aggregation Pipeline | 27 |
| 3.1 Feature-Based Similarity for Spatial Clustering | 28 |
| 3.1.1 Feature Construction | 28 |
| 3.1.2 Distance Computation and Aggregation | 29 |
| 3.2 Spatial and Temporal Clustering | 30 |
| 3.2.1 Spatial Aggregation | 30 |
| 3.2.2 Temporal Aggregation | 31 |
| 3.3 Constructing the Aggregated Input Dataset | 32 |
| 4 Evaluation and Experimental Framework | 35 |
| 4.1 Generation and Transmission Expansion Model | 35 |
| 4.2 Validation via Sequential Deaggregation | 39 |
| 4.2.1 Motivation and Methodological Framework | 40 |
| 4.2.2 Spatial Deaggregation Methods | 41 |
| 4.2.3 Temporal Deaggregation Methods | 42 |
| 4.2.4 Integrated Deaggregation Framework | 43 |
| 4.2.5 Our Methodological Choice | 43 |
| 4.3 Experimental Framework | 45 |
| 4.3.1 Hyperparameter Search | 45 |

| | | |
|-------------------|--|-----------|
| 4.3.2 | Benchmarks | 46 |
| 5 | Case Study Data | 49 |
| 5.1 | Construction of Network Data | 49 |
| 5.1.1 | High-Resolution Inputs: 385 Nodes | 49 |
| 5.1.2 | Downscaling for Model Evaluation: 17 Nodes | 51 |
| 5.2 | Planning Model Parameters | 51 |
| 5.2.1 | Modeling Assumptions | 51 |
| 5.2.2 | Data Sources and Parameter Construction | 52 |
| 5.2.3 | Scaling for Numerical Stability | 54 |
| 6 | Results | 57 |
| 6.1 | Feature Weight Analysis | 57 |
| 6.2 | Pareto Frontier | 61 |
| 6.3 | Benchmarks | 63 |
| 7 | Conclusion | 67 |
| 7.1 | Summary and Key Findings | 67 |
| 7.2 | Limitations and Future Directions | 68 |
| A | Values of Planning Model Parameters | 71 |
| B | Grid of Feature Weight Configurations | 75 |
| <i>References</i> | | 79 |

List of Figures

| | | |
|-----|---|----|
| 2.1 | Two-step pipeline from raw time series to representative periods. The original series of length N is divided into K periods of T time steps each. These are grouped and aggregated into a reduced set of \hat{K} representative periods, each of length T (graph from [22]). | 20 |
| 2.2 | Pipeline for clustering-based temporal aggregation, including normalization, assignment, and representation stages (graph from [16]). | 21 |
| 2.3 | Illustration of key modeling decisions in clustering: (a) normalization, (b) distance metrics: Euclidean (ED), Dynamic Time Warping (DTW), and Shape-Based Distance (SBD) (graph from [16]). | 22 |
| 2.4 | Schematic of a DTW-based temporal aggregation pipeline (graph from [27]). | 22 |
| 2.5 | Superposition of intra- and inter-period storage dynamics (graph from [13]). | 23 |
| 2.6 | Comparison of spatial aggregation methods. A power network is aggregated using various clustering criteria (e.g., k -means, capacity-based $f_{\text{cap}}(v)$, time-series-based $f_{\text{time}}(v)$, or a composite quality metric Q), and each aggregation is passed to an optimization model to evaluate impact on siting and system structure (graph from [12]). | 25 |
| 3.1 | Overview of the spatio-temporal aggregation and optimization pipeline. . . . | 27 |
| 5.1 | Construction of fine-resolution network over New England | 50 |
| 6.1 | Top 5 configurations by total deaggregated cost (millions of \\$) for each (C, D) pair. | 58 |
| 6.2 | Correlation between feature weights and a few performance metrics. | 59 |
| 6.3 | Distribution of feature weights for all configurations vs. the top 10 performers. | 60 |
| 6.4 | Pareto frontier in $(C, D, \text{Objective})$ space under fixed feature weights. Red points are non-dominated configurations. | 62 |
| 6.5 | 2D projections of objective vs. aggregation level. Pareto frontier configurations are marked by red crosses. | 62 |
| 6.6 | De-aggregated cost comparison. | 64 |
| 6.7 | Relative objective gain (%) of R25 and others compared to baseline groups. . | 65 |
| 6.8 | Normalized objective score by method and configuration (100% = best). . . | 66 |
| A.1 | Balancing Authority (BA) regions used to extract resource availability from the ReEDS 2.0 dataset [55]. Six zones (p129–p134) corresponding to ISO New England are retained. (Raw figure) | 72 |

| | | |
|-----|---|----|
| A.2 | Baseline transmission line capital cost estimates from WECC Table 2-1 [53]. (Raw figure) | 73 |
| A.3 | Transmission line configurations and thermal ratings from WECC Table 2-8 [53]. These values are used to build representative per-mile cost and capacity metrics. (Raw figure) | 74 |

List of Tables

| | | |
|-----|--|----|
| 3.1 | Summary of feature formats and distance computations used in the composite similarity metric. | 30 |
| 4.1 | Glossary of model variables | 37 |
| 4.2 | Glossary of model sets and parameters | 39 |
| 4.3 | Overview of common approaches for evaluating aggregated CEP solutions. Methods in bold are those implemented in this work. | 44 |
| 5.1 | Summary of key GTEP model parameters (units are after scaling) | 55 |
| 6.1 | Composite performance metrics for ablation configurations E1–E6. | 58 |
| 6.2 | Summary of representative configurations across all (C, D) | 60 |
| 6.3 | Feature composition of best configurations (\checkmark emphasized, \sim moderate). | 60 |
| 6.4 | Configurations on the Pareto frontier under fixed \mathbf{w} , ordered by objective value. | 61 |
| 6.5 | Performance ranking of each method under different aggregation regimes (rank 1 is best). | 63 |
| A.1 | Raw parameter values used in the GTEP model. All values in native units. . | 72 |
| A.2 | Computed parameters used in the GTEP model. All values are expressed in scaled units. | 73 |
| A.3 | Transmission line configuration cost and rating summary based on WECC Tables 2-1 and 2-8 [53]. The final column reports normalized cost per MW-mile. (Built from raw tables A.2 and A.3) | 74 |
| A.4 | Real fuel prices by year for selected fuels from the EIA Annual Energy Outlook 2024, Table 3 [54]. Prices are reported in 2024\$ per MMBtu. (Extracted and filtered for clarity) | 74 |
| B.1 | Summary of feature weight configurations used in the hyperparameter grid search. | 75 |
| B.2 | Full grid of feature weight vectors used for spatial aggregation (Part 1). . . . | 76 |
| B.3 | Full grid of feature weight vectors used for spatial aggregation (Part 2). . . . | 77 |

Chapter 1

Introduction

1.1 The Energy System Planning Challenge

When we think about how our electricity system should evolve over the coming decades, we face an enormously complex puzzle. This puzzle involves deciding where to build new power plants, what types of generation technologies to invest in, how much transmission capacity we need between different regions, and when to make all these investments. This challenge is formally known as Capacity Expansion Planning (CEP), and it represents one of the most important and computationally demanding problems in modern energy system analysis [1, 2].

CEP models are large-scale mathematical optimization programs that compute cost-effective investment portfolios to meet future electricity demand while satisfying technological, environmental, and policy constraints. They represent the long-term evolution of power systems, typically over multi-decade time horizons, and capture both capital expenditures (e.g., building new renewable plants or transmission lines) and operational considerations (e.g., thermal limits or transmission congestion) across thousands of time periods. These models are widely used by researchers, planners, and policymakers to assess renewable integration strategies, evaluate climate mitigation scenarios, and guide infrastructure investments [3].

The decisions we make today about energy infrastructure will shape our electricity system for decades. Power plants often operate for 20–40 years, and transmission lines can last even longer [4, 5]. These long asset lifetimes mean that poor planning decisions could lock us into expensive or carbon-intensive trajectories. At the same time, the rapid growth of wind and solar power is fundamentally changing how we plan power systems. Unlike traditional dispatchable generators, renewable energy sources depend on weather conditions that vary significantly over space and time [6]. This shift toward weather-dependent resources is not a small adjustment—it is a wholesale transformation of how electricity is produced and delivered. Over the coming decades, most new capacity additions are expected to come from wind, solar, and battery storage, reflecting both declining technology costs and growing policy support for decarbonization [4]. CEP models sit at the center of this transition as they provide the primary tool we use to explore siting and sizing decisions for future infrastructure.

To make sound investment decisions, CEP models must capture these spatial and temporal variations in supply and demand. On the spatial side, the power system is modeled as a network of buses and transmission lines, with each node associated with local generation,

load, and storage assets. Resource quality varies widely across geography, a wind farm located in one region may produce significantly more energy than one just 50 kilometers away due to meteorological conditions [7]. Demand also differs by location, influenced by population density, industrial activity, and climate [8]. On the temporal side, the patterns of electricity demand and renewable generation span multiple time scales. Solar output follows a predictable diurnal and seasonal cycle, wind generation can fluctuate hourly, and electricity demand shows distinct daily, weekly, and seasonal variations [9]. Capturing these dynamics with sufficient resolution, often hourly time steps over a full year, is essential to produce robust and actionable investment guidance.

However, achieving this level of detail comes at a steep computational cost. Capturing both high spatial and temporal resolution in CEP models results in extremely large-scale optimization problems. Modeling even a moderately sized system with 19 zones and 8,736 hourly time steps (52 weeks) results in a Mixed Integer Linear Program (MILP) with over 10 billion variables and 30 billion constraints, which cannot be solved using monolithic formulations, that is, treating the entire temporal and spatial scope as a single, undivided optimization problem, on conventional hardware [10]. In practice, models with more than 6 zones or 22 weeks of data are already intractable, requiring over 200 GB of memory and exceeding 48 hours of runtime. Jacobson et al. (2023) [10], find that runtime scales linearly with temporal resolution and quadratically with the number of zones, implying a complexity of roughly $\mathcal{O}(Z^2T)$.

To make these problems tractable, researchers and planners often employ data aggregation techniques that reduce model dimensionality while aiming to preserve its essential structure. Broadly, these fall into two categories: spatial aggregation and temporal aggregation. Spatial aggregation simplifies the network by clustering buses or geographical zones, thereby reducing the number of nodes and branches that must be modeled explicitly. Temporal aggregation reduces the number of modeled time periods, typically by selecting a small set of representative hours or days that capture key system behaviors across the year. Both strategies seek to strike a careful balance between reducing problem size enough to enable computation and retaining enough detail to produce robust and meaningful investment recommendations [11].

1.2 Aggregation as a Design Problem

The growing complexity of long-term power system planning has made dimensionality reduction not just a convenience, but a necessity. To manage the computational scale of CEP models, it is now standard practice to simplify both the spatial and temporal structure of input data. However, while aggregation is widely used, it is often treated as a technical preprocessing step, with relatively little attention paid to how different design choices affect planning outcomes. In reality, aggregation is a modeling decision that shapes the optimization problem itself. The way zones are grouped and time is compressed can materially alter where, when, and how the model decides to invest.

In the spatial domain, aggregation involves clustering buses or geographical units to reduce network size. This is typically done using criteria that are easy to compute or justify like grouping by proximity on the network graph, by geographic coordinates, or by resource or demand averages [12]. Yet these simple heuristics often miss planning-relevant structure.

For instance, two locations that are geographically close might have highly correlated wind profiles, making them poor substitutes from a system reliability standpoint. On the other hand, clustering based on generation or load signals alone may ignore topological constraints, leading to misleading transmission needs or infeasible flows. These trade-offs matter in planning models that rely on spatial diversity for balancing, reliability, and cost optimization.

Temporal aggregation faces analogous challenges. Because CEP models are computationally limited in the number of time steps they can handle, it is common to compress high-resolution time series into a small number of representative periods. Approaches include k -means clustering of days, selection of typical or extreme weeks, or averaging across similar days [13–15]. These methods help reduce problem size, but can obscure ramping events, flatten signals, or exclude rare but critical episodes, each of which can distort how the model values flexibility, storage, or backup generation [13, 16].

Because both spatial and temporal aggregation affect planning behavior, researchers have begun to explore methods that design these reductions more deliberately. This growing body of work recognizes that aggregation is not a neutral operation. It interacts with model objectives, constraints, and technologies in ways that can alter outcomes in significant and sometimes counterintuitive ways. Recent studies advocate for what is increasingly called *task-aware* or *planning-aware* aggregation, in which aggregation choices are explicitly evaluated by their downstream impact on system design and operation [16, 17].

This thesis contributes to this emerging literature by developing a modular and interpretable aggregation framework that integrates these ideas. It enables users to select and combine features they deem important, and to analyze how different aggregation configurations affect the quality of planning outcomes.

1.3 Thesis Objectives and Contributions

This thesis addresses the following question: how can spatial and temporal aggregation in Capacity Expansion Planning (CEP) models be designed to preserve structural patterns that influence long-term investment decisions, while reducing the dimensionality of the optimization problem?

Solving CEP models at high spatial and temporal resolution is computationally intensive, but naive aggregation risks distorting the relationships that drive planning outcomes. This thesis develops a task-aware aggregation pipeline that integrates domain-relevant information into the aggregation process. The goal is to retain the key signals that affect decisions about where to build new generation, how to size transmission, and what technologies to invest in, without overburdening the model with too much detail.

The approach consists of four main components:

1. **Composite similarity metric construction.** Multiple planning-relevant features are computed between each pair of nodes. These include duration curve shape, ramping behavior, spatial correlation of renewable generation, and other features reflecting demand and generation structure. Each feature is normalized, and the set is combined into a weighted distance metric—hereafter referred to as the composite similarity metric—with tunable weights reflecting the planner’s priorities.

2. **Spatial aggregation via clustering.** The composite metric is applied to candidate nodes, and a k -medoids clustering algorithm is used to form spatial zones. Each zone is represented by its medoid node, and its time series is rescaled to preserve total zonal quantities. This step treats aggregation as a modeling decision informed by structural signals, not geographic heuristics.
3. **Temporal aggregation via representative days.** Daily time series are clustered using k -medoids, and selected days are assigned weights to ensure accurate scaling back to annual totals.
4. **CEP model execution and evaluation.** The reduced spatio-temporal dataset is used to solve the CEP problem. Investment decisions are then projected back to the full-resolution system using a deaggregation procedure, allowing operational feasibility and cost to be evaluated without relying on surrogate statistical metrics.

This work contributes both a novel methodological framework and empirical insights for planning-oriented data reduction in CEP models. The key contributions are outlined below:

- **A task-aware, interpretable aggregation pipeline for CEP models.** It introduces a task-aware aggregation framework that coordinates spatial and temporal reduction in a unified and interpretable pipeline. The method allows for fine control and tuning of feature weights, aggregation granularity, and temporal compression.
- **A composite-feature distance metric for spatial clustering.** It proposes a composite-feature distance metric that blends multiple physically meaningful signals into a weighted measure of similarity. Spatial clustering is performed with k -medoids to ensure each zone is represented by a real node. After spatial aggregation, temporal clustering is applied to daily time series using k -medoids, with selected days weighted to preserve annual totals.
- **Downstream, task-focused evaluation framework.** It implements a downstream evaluation framework that reprojects investment decisions to full resolution and evaluates their true cost and feasibility. This ensures that aggregation choices are validated against the planning task, not just data reconstruction metrics.
- **Empirical insights from the New England case study.** It provides empirical insights from a case study of the New England power system. Results show that the pipeline consistently outperforms baseline methods (e.g., political boundaries, pure geographic clustering), with the best configuration achieving a 13% cost reduction. Correlation-based features are found to be most influential, while reliance on amplitude or location alone degrades performance.

The next chapter, Chapter 2, reviews existing methods for aggregation in capacity expansion modeling. It begins with temporal reduction techniques, then covers spatial clustering approaches, and concludes with discussions on recent advances in planning-aware aggregation design. Chapter 3 introduces the full methodology, detailing the construction of composite

feature matrices and the assembly of aggregated datasets. Chapter 4 presents the evaluation framework used for task-awareness through a deaggregation procedure, and outlines the experimental design used to quantify the benefits of different aggregation configurations. Chapter 5 introduces the case study data, including hourly temporal profiles and techno-economic parameters used in the planning model. Chapter 6 analyzes the empirical results of the pipeline, including the trade-offs between resolution and plan quality, the influence of different feature signals on spatial clustering, and the comparative performance of the proposed method against heuristic baselines. Chapter 7 concludes with a synthesis of findings and directions for future research. The full implementation is available at: <https://github.com/gduguey/GridCluster>.

Chapter 2

Literature Review

This chapter reviews how aggregation is used in CEP models. We start with methods that simplify time by choosing representative periods (§2.1), and then look at ways to reduce spatial complexity through network clustering (§2.2). Across all sections, we highlight how different aggregation strategies trade off fidelity, interpretability, and computational performance. The evaluation of how well these aggregation methods preserve planning outcomes is discussed separately in Chapter 4 (§4.2), where a high-resolution validation framework is introduced.

2.1 Temporal Aggregation

Temporal aggregation is one of the most widely used strategies to reduce the computational burden of CEP models. It works by reducing the number of time periods considered in the optimization problem, which directly lowers the number of constraints and decision variables. The motivation is straightforward: a planning horizon for a single target year of hourly data contains 8,760 time steps, but it is common for models to span multiple years and scenarios. Solving such models at full temporal resolution is usually computationally infeasible, especially when coupled with fine spatial granularity. Alternative approaches avoid aggregation altogether by reformulating the optimization problem. Recent work explores decomposition-based techniques that partition high-resolution planning models into linked subproblems to improve tractability while retaining full temporal and spatial detail [10, 18]. These formulations offer a different path to scalability but remain less common and computationally demanding, particularly in large, multi-sector studies. Aggregating time allows researchers and practitioners to make planning models tractable while still attempting to preserve the temporal patterns that influence long-term planning outcomes.

The dominant form of temporal aggregation in the literature is the use of representative periods (most often days, but sometimes weeks or hours) that summarize key patterns in demand and renewable generation. Instead of solving the optimization over the full horizon, planners solve it on a reduced set of representative periods, then rescale the results to approximate yearly or long-term totals [16, 19]. These representative periods may be chosen using heuristics (e.g., days with peak load or peak solar), clustering algorithms based on statistical similarity, or optimization-based techniques that explicitly minimize the deviation of key system metrics [20, 21].

Despite their widespread adoption, temporal aggregation methods involve a cascade of modeling choices, including the selection of periods, similarity measures, and modeling of storage dynamics, all of which substantially affect the fidelity of the results. The rest of this section follows the full aggregation pipeline, from raw time series to selected periods, and unpacks how different methods address these choices and what trade-offs they imply.

2.1.1 Representative Period Construction

A common starting point for building representative days is to treat the task as a clustering problem. But before clustering can happen, the time series needs to be divided into smaller, comparable chunks [22]. These chunks (often days or weeks) are called periods, and they form the basic units over which similarity is assessed. The choice of period length reflects a fundamental trade-off between capturing relevant temporal variations and maintaining computational tractability. Daily periods (e.g., 24 hourly values) are most commonly used, as they align with diurnal demand and solar generation patterns, and enable realistic modeling of intra-day storage and ramping constraints [14, 23, 24]. Weekly periods can capture longer-term fluctuations such as weather fronts or industrial cycles, but they often degrade computational performance [23] and may not improve optimization outcomes [24]. Autocorrelation analyses have shown clear periodicities at daily and weekly scales, but empirical evaluations favor daily periods for preserving model fidelity without incurring excessive cost [25]. For example, if we are working with hourly data over a year, we might split it into 365 individual days, each with 24 values. These periods are the input to the aggregation pipeline. From this full collection, a much smaller set of representative periods is selected (e.g., 10 or 20 days) that are meant to capture the most important patterns in the original data. This two-step process, where the time series is first sliced into individual periods and then aggregated into a smaller, representative subset, is shown in Figure 2.1. The rest of this section describes how different methods tackle this task, and what trade-offs their choices imply.

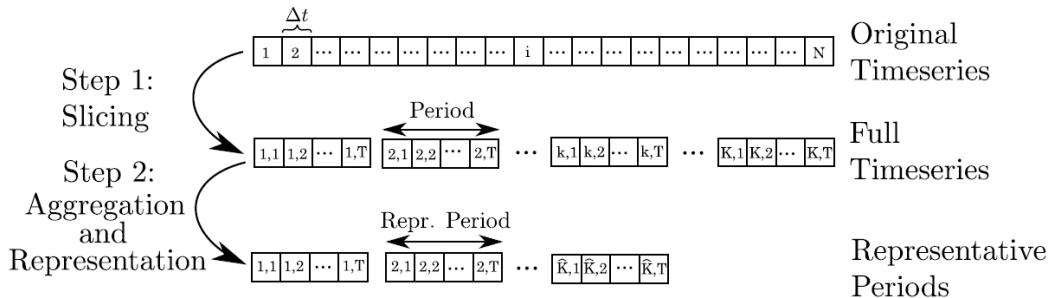


Figure 2.1: Two-step pipeline from raw time series to representative periods. The original series of length N is divided into K periods of T time steps each. These are grouped and aggregated into a reduced set of \hat{K} representative periods, each of length T (graph from [22]).

Once the time series is sliced, Teichgraeber and Brandt (2019) [16] propose a structure for the clustering process shown in Figure 2.2, broken into three main steps: normalizing the input time series, assigning each day to a cluster, and choosing one profile to represent each group. These are not just technical steps, they reflect deeper modeling choices that shape the outcome. For instance, how the time series is normalized affects which features are considered

most important. If each day is normalized on its own (sequence-based normalization), the clustering tends to emphasize the shape of daily profiles (like when peaks occur). But if the entire time series is normalized together (global normalization), then magnitude plays a bigger role. This is illustrated in Figure 2.3a. Similarly, the distance metric used to compare days can steer the clustering in different directions: Euclidean distance focuses on exact value differences, Dynamic Time Warping (DTW) [26] allows time shifts in peaks, and shape-based metrics emphasize structural similarities. Figure 2.3b shows how these metrics differ in the kinds of similarity they detect. Finally, the way representative days are selected, either as synthetic centroids, which average the group, or as real medoid days, which are actual observed days, affects how realistic or smooth the final inputs are. Each of these choices influences not just the clustering outcome, but also which temporal patterns and system behaviors the model is able to retain.

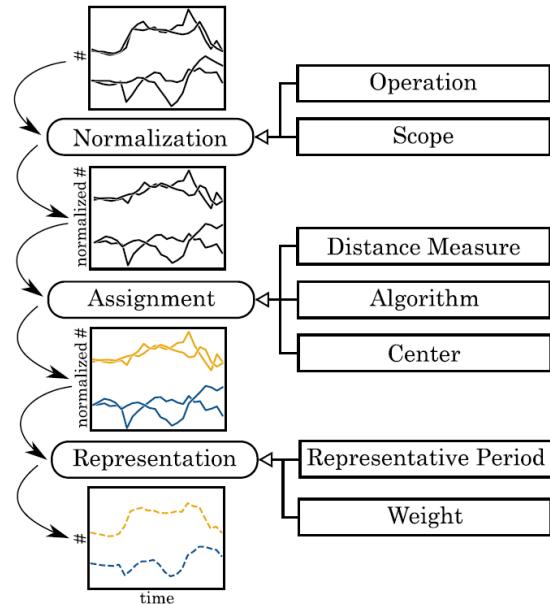


Figure 2.2: Pipeline for clustering-based temporal aggregation, including normalization, assignment, and representation stages (graph from [16]).

In practice, these decisions often reflect trade-offs between interpretability and fidelity. Kotzur et al. (2018) [13] benchmark several clustering algorithms (k -means, k -medoids, hierarchical clustering, and SOMs) and show that while k -means yields smooth centroids, it can produce synthetic profiles that are not physically realizable. k -medoids improves realism by selecting real days, preserving sharp transitions and extreme events, but may perform worse on statistical fit. Reulein et al. (2024) [27] propose a compromise: using Dynamic Time Warping Barycenter Averaging (DBA) to construct synthetic profiles that align time steps nonlinearly, producing smoothed days that better reflect the internal dynamics of wind series. Dynamic Time Warping (DTW) is a distance metric that aligns time series by stretching or compressing them in time, allowing for more flexible comparisons of shape and timing than standard point-wise distances. Figure 2.4 illustrates their DTW-based pipeline, which enhances temporal structure preservation compared to Euclidean-based clustering.

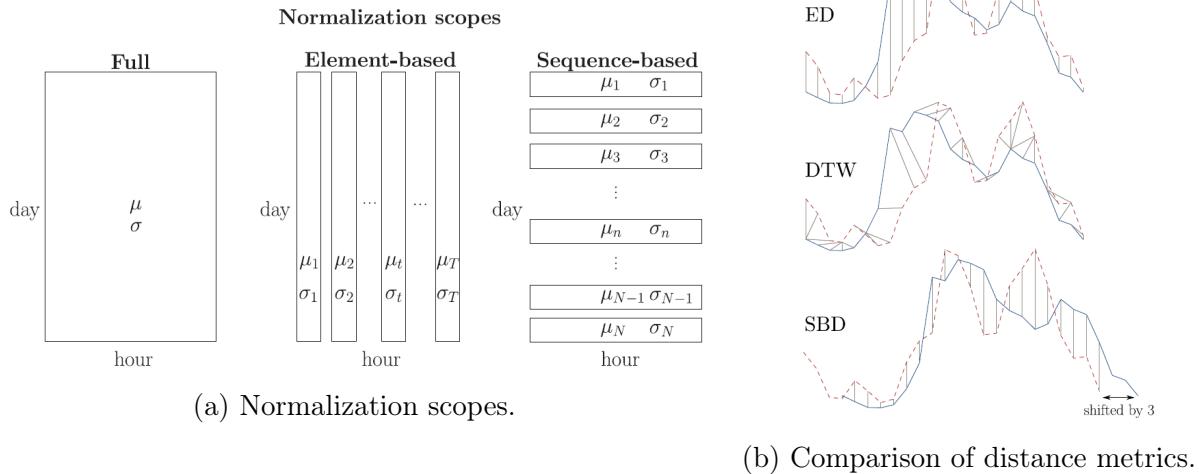


Figure 2.3: Illustration of key modeling decisions in clustering: (a) normalization, (b) distance metrics: Euclidean (ED), Dynamic Time Warping (DTW), and Shape-Based Distance (SBD) (graph from [16]).

This spectrum of methods reveals that the question is not just how to cluster, but what kind of temporal structure one wants to retain. For instance, DBA centroids outperform both medoids and k -means in preserving the variance of wind time series, an important property for models sensitive to short-term intermittency [27]. These improvements are not merely statistical: Reulein et al. (2024) report lower deviations in system cost and generation capacity when using DBA-based profiles in downstream optimization.

Rather than relying on clustering heuristics, some researchers take an optimization approach to representative period selection. Poncelet et al. (2017) [19] formulate an MILP that selects representative days by directly minimizing the error in duration curves across multiple attributes, such as load, residual load, and ramping. This bypasses the statistical biases introduced by clustering, such as overweighting common days, and allows rare but impactful days to be chosen even if they do not form large clusters. Importantly, their method also optimizes the weights assigned to each representative period, treating the full selection process as a constrained optimization problem rather than a statistical summary.

Beyond the selection of representative periods, another key modeling dimension is how these periods are connected, particularly in systems with energy storage. The next subsection

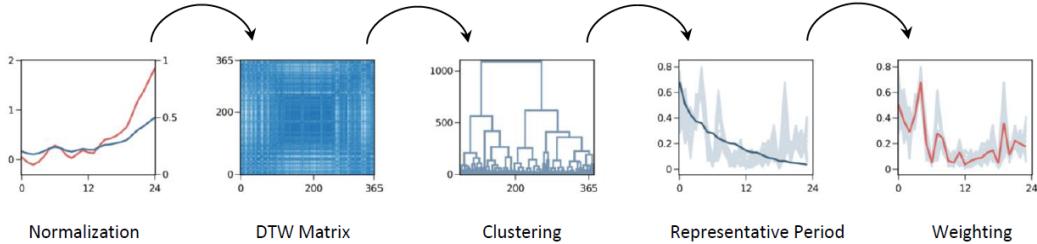


Figure 2.4: Schematic of a DTW-based temporal aggregation pipeline (graph from [27]).

addresses these linkage mechanisms and how aggregation methods are evaluated.

2.1.2 Linking Periods and Evaluating Aggregation

Temporal aggregation methods also vary in how they treat linkages between periods, particularly important when modeling storage and ramping constraints. Most approaches assume each representative day is cyclic, meaning storage must end each day at the same state of charge (SOC) at which it began. This assumption simplifies implementation but distorts inter-day energy flows and biases investment away from long-duration or seasonal storage. Kotzur et al. (2018) [13] address this with a superposition framework that decomposes storage states into intra-day and inter-day components. The former evolves within each representative day, the latter carries energy across periods. This allows aggregated models to emulate seasonal build-up or depletion of storage, as illustrated in Figure 2.5. Without this mechanism, seasonal storage dynamics are fundamentally lost.

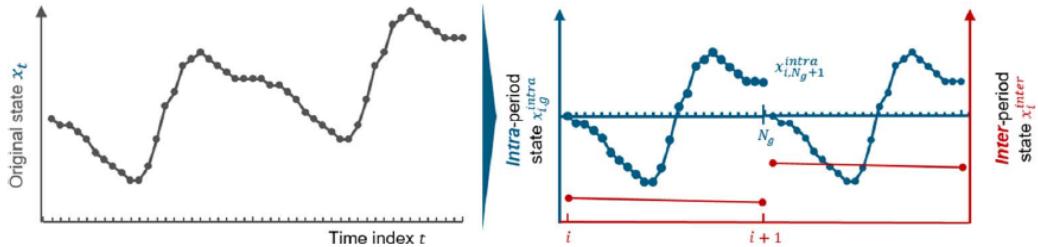


Figure 2.5: Superposition of intra- and inter-period storage dynamics (graph from [13]).

Other authors propose hybrid solutions that mix representative days with chronological sequences, or embed them in rolling horizon simulations. Teichgraeber and Brandt (2022) [22] argue that such hybrids can reduce optimization error while keeping runtime manageable. Reulein et al. (2024) [27] acknowledge the limitation of their current cyclic storage treatment and suggest that incorporating inter-day linkage remains an open direction for shape-based clustering.

Across all these methods, the core challenge is not just how to construct representative periods, but how to know whether they're good enough. Teichgraeber and Brandt (2022) [22] distinguish three kinds of error that can arise when using time-series aggregation. The first is reconstruction error, which measures how well the representative periods approximate the original time series statistically, in terms of their shape, temporal patterns, and temporal structure. The second is input error, which focuses on how the model inputs, such as time-varying load, wind, or solar capacity factors, numerically differ between the full-resolution data and the aggregated version, once weighted and reassembled to cover the full year. This might mean, for instance, that average wind availability is slightly off, or that a key ramping event is smoothed out. The third is optimization error, which quantifies how much the model's outputs, like investment decisions or system costs, change as a result of using the aggregated data. These three error types don't always align: a close statistical match to the time series doesn't always yield a good system design. That's why evaluation must combine both statistical diagnostics (e.g. RMSE, covered variance) and model-based metrics (e.g. cost

gaps, changes in installed capacity). Beyond the selection of representative periods, another crucial modeling dimension is how these periods are connected, particularly in systems with energy storage. The next subsection addresses these linkage mechanisms and how aggregation methods are evaluated.

2.2 Spatial Aggregation

Spatial aggregation is the process of simplifying the network by merging buses, generators, and transmission links into larger zones. It is often essential for computational feasibility in CEP models. But unlike temporal aggregation, which mainly reduces input size, spatial aggregation alters the physical and economic topology of the grid itself. As a result, it can have significant consequences for power flows, renewable siting, and investment decisions.

2.2.1 From Administrative Boundaries to Data-Driven Clustering

A common starting point for spatial aggregation is the use of administrative or political boundaries. Models like ReEDS in the U.S. [28] and LIMES-EU in Europe [29] use predefined zones such as states or balancing authorities to reduce geographic complexity. This approach is convenient, data are standardized and policy boundaries are respected, but it is poorly matched to the actual structure of electricity systems. Wind, solar, and load patterns follow natural rather than institutional gradients [9], and rigid political aggregation risks masking both high-quality resource sites and intra-regional transmission constraints.

To overcome these limitations, many studies apply unsupervised clustering methods that group nodes based on engineered similarity features. Hörsch and Brown (2017) [30] introduce a k -means-based spatial clustering over node coordinates and show that investment patterns in European CEP models begin to converge at a resolution of roughly 200–300 zones. However, their method is grid-agnostic and does not account for heterogeneity in generation or demand.

2.2.2 How Clustering Logic Shapes System Design

Frysztacki et al. (2022) [12] offer a more systematic exploration of how different clustering strategies affect model outcomes. They compare four methods, geographic (k -means), statistical (Ward’s clustering on mean capacity factors $f_{\text{cap}}(v)$), temporal (Ward’s clustering on time-series profiles $f_{\text{time}}(v)$), and electrical (modularity-based clustering using admittance-weighted Q), in a large-scale European test case (see Figure 2.6). Each method is evaluated against a high-resolution benchmark with 1250 nodes.

Their results show that clustering logic fundamentally shapes system architecture. For example, $f_{\text{time}}(v)$ clustering produces zones with more coherent renewable profiles and better siting accuracy under 100% decarbonization, while modularity-based clustering excels at preserving power flow patterns under moderate targets. In contrast, k -means clustering performs poorly across all dimensions, its lack of grid or resource awareness yields high siting error and unrealistic flow paths. Crucially, the authors demonstrate that different methods lead to different technology mixes: e.g., modularity clustering increases hydrogen storage due to poor spatial alignment of renewables, while $f_{\text{time}}(v)$ better captures coastal wind

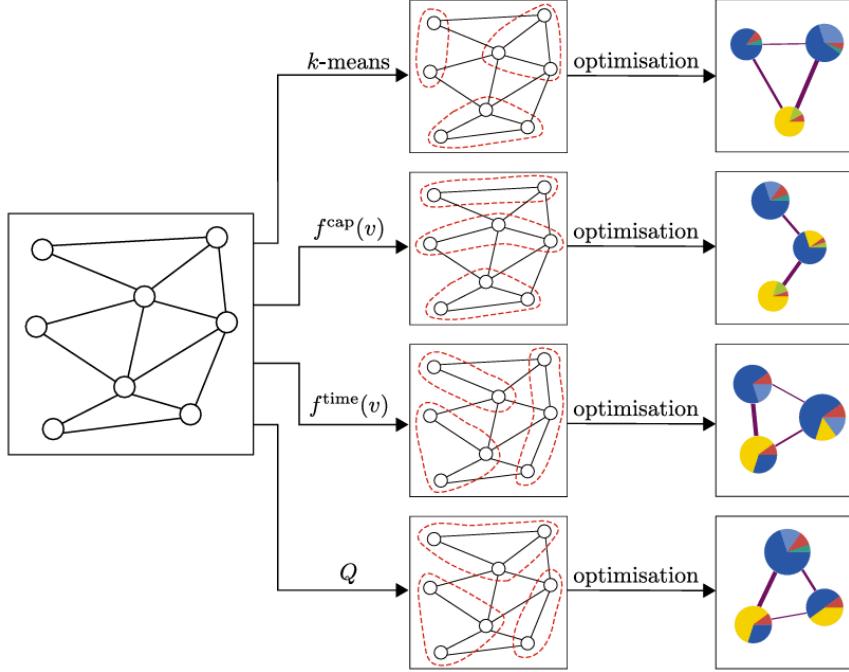


Figure 2.6: Comparison of spatial aggregation methods. A power network is aggregated using various clustering criteria (e.g., k -means, capacity-based $f_{\text{cap}}(v)$, time-series-based $f_{\text{time}}(v)$, or a composite quality metric Q), and each aggregation is passed to an optimization model to evaluate impact on siting and system structure (graph from [12]).

zones and transmission needs. This reinforces a key modeling insight: aggregation is not neutral. It embeds assumptions about what features matter and directly influences optimal infrastructure design.

These findings echo broader concerns about how aggregation affects both power flows and resource representation. When nodes are merged, internal lines are often removed or replaced by a copper-plate approximation, implicitly assuming unlimited intra-zonal transmission. Brinkerink et al. (2023) [31] and Frysztacki and Brown (2020) [32] show that this can overstate flexibility and distort dispatch, especially in zones with internal congestion. Similarly, averaging capacity factor profiles across aggregated nodes can dilute high-quality renewable sites. Brown et al. (2018) [33] find that coarse models shift wind investments inland, away from optimal coastal zones, due to spatial averaging.

Several studies quantify these distortions. Staffell and Pfenninger (2018) [34] vary resolution in a European model and show that coarser aggregation underestimates curtailment and transmission needs, particularly under high-VRE futures. Frysztacki et al. (2022) define spatial fidelity metrics (e.g., siting error, cost gap) to directly compare reduced models to full-resolution baselines, concluding that preserving both resource and load heterogeneity is essential to avoid bias.

2.2.3 Beyond Similarity: Task-Aware and Hybrid Aggregation

While most methods rely on clustering over hand-engineered features, some studies pursue more task-aware approaches that explicitly account for the modeling objective. Radu et al. (2021) [35] propose a bilevel optimization framework that selects a reduced set of buses to preserve investment fidelity. Unlike clustering, which groups similar nodes, this method retains those most critical to replicating the siting decisions of a full-resolution model. Their LP-relaxation-based heuristic achieves >90% node reduction with minimal error in siting and system cost, highlighting that strategic retention of important nodes can outperform similarity-based aggregation.

Patil et al. (2022) [36] advance this task-aware perspective by combining spatial and technological aggregation in a hybrid two-stage framework. They enforce geographic contiguity via constrained k -medoids clustering, then reduce technological complexity by clustering time series of VRE generation within each region. This preserves both physical topology and temporal dynamics. Their analysis reveals a trade-off frontier: with 33 spatial zones and 38 VRE types per zone, the model retains accuracy within 5% of baseline while reducing runtime by over 90%. Their key insight is that effective aggregation must reflect not just proximity, but purpose, in this case, optimizing over spatial, temporal, and technological diversity.

Taken together, these studies converge on a common point: spatial aggregation must be fit for purpose. Grid planning under deep decarbonization requires resolution that respects resource and transmission constraints, not just computational convenience. Generic clustering may preserve statistical similarity but distort outcomes; task-aware strategies, while more complex, better preserve economic and operational realism.

The implications are clear. For broad policy screening, aggregation by administrative zones may suffice. But for infrastructure design under high-VRE futures, spatial reduction must reflect both grid physics and investment logic. As the next section will show, these challenges are compounded when spatial aggregation interacts with temporal resolution choices.

Chapter 3

Aggregation Pipeline

This chapter introduces the methodology developed to implement spatio-temporal aggregation for long-term power system planning. The pipeline has three main stages, as described in Figure 3.1. First, it extracts a set of interpretable features from demand and renewable generation data, chosen to reflect operational and planning-relevant temporal patterns. These features are used to define a composite similarity metric over system nodes, that is, a weighted combination of multiple feature distances that captures different aspects of spatial and temporal structure relevant to planning. Second, the network is partitioned into spatial clusters using k -medoids, guided by the composite distance matrix. For each cluster, a single representative node (the medoid) is selected, and its time series is rescaled to match the aggregate behavior of the cluster. Third, the temporal dimension is compressed by selecting a small number of representative days. Importantly, this temporal clustering is performed over the already-aggregated network, so that it reflects spatial diversity in daily profiles. The resulting spatio-temporal dataset is then used as input to a reduced CEP model. The codebase implementing the full aggregation pipeline, CEP model, and evaluation framework is publicly available at: <https://github.com/gduguey/GridCluster>.

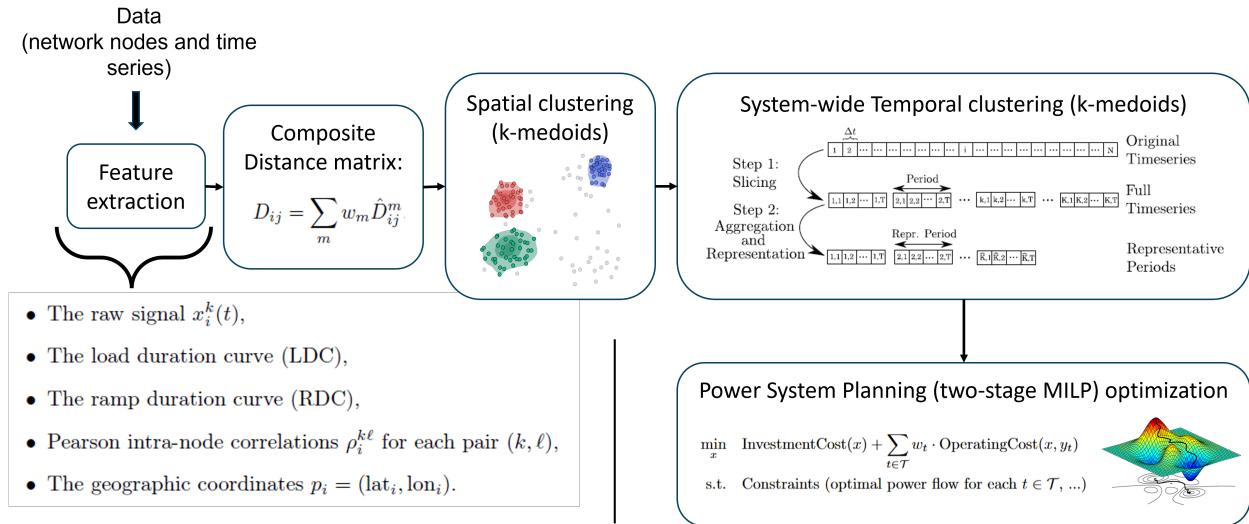


Figure 3.1: Overview of the spatio-temporal aggregation and optimization pipeline.

This approach differs from black-box aggregation schemes by preserving full interpretability of each step and enabling control over key design parameters. For instance, users can tune the relative weights of feature types to explore different planning priorities, such as geographic diversity, ramping capability, or renewable complementarity. Each aggregation choice is evaluated downstream through its impact on model fidelity and planning outcomes.

The remainder of the chapter is structured as follows. Section 3.1 describes the construction of the composite distance matrix process and explains how each feature contributes to the task of spatial clustering. Section 3.2 details the spatial and temporal clustering steps, including the construction of cluster-level time series and the selection of representative days. Section 3.3 explains how the clustered inputs are rescaled and assembled into a reduced spatio-temporal dataset. This aggregated input will serve as the basis for the CEP model introduced in Chapter 4, where we also describe the evaluation framework used to assess the effectiveness of different aggregation configurations.

3.1 Feature-Based Similarity for Spatial Clustering

To support aggregation that reflects the operational structure of the power system, we begin by constructing a set of features that characterize each node’s temporal behavior, spatial context, and planning relevance. These features are inspired by the approach of Poncelet et al. [19], and are based on their ability to represent aspects of system behavior that matter for long-term investment decisions, such as temporal fluctuations, flexibility, complementarity, and peak intensity. Unlike task-agnostic approaches that optimize for data reconstruction alone, our goal is to extract features that help preserve the structural drivers of capacity expansion and transmission planning.

3.1.1 Feature Construction

Each feature is computed from the raw hourly time series $x_i^k(t)$, where $i \in \mathcal{N}$ is a node of the network \mathcal{N} , $k \in \mathcal{K} = \{\text{demand, wind, solar}\}$ is a time series type, and $t \in \mathcal{T}$, with $|\mathcal{T}| = 8760$ (spanning a year hourly). The first feature is the raw time series itself, normalized by its maximum value at each node and for each variable type, so that shape and magnitude are both preserved: $\tilde{x}_i^k(t) = x_i^k(t) / \max_t |x_i^k(t)|$. This retains differences in peak demand or resource intensity across nodes, which are essential for generation and siting decisions. Next, we extract two sorted-profile features. The Duration Curve (DC) is obtained by sorting $x_i^k(t)$ in decreasing order, and the Ramp Duration Curve (RDC) is built from the sorted values of the time differences $\Delta x_i^k(t) = x_i^k(t+1) - x_i^k(t)$. Both are min-max normalized to preserve shape while removing scale effects. These curves summarize long-run system characteristics relevant to capacity sizing (DC) and intra-day operational flexibility (RDC). The DC captures how often high signals occur, informing how much firm capacity is needed to ensure reliability. The RDC highlights the magnitude and frequency of short-term ramps, which drive the need for flexible resources such as peaking plants, batteries, and demand response in expansion plans.

We then compute the intra-node correlation vector for each node, defined as the Pearson correlation coefficients $\rho_i^{k\ell}$ between all pairs of variable types $(k, \ell) \in \mathcal{K} \times \mathcal{K}$. These values

capture how demand and generation co-vary at a single location, for instance, whether solar generation aligns with peak load or offsets it. Including these correlations in the feature set helps identify nodes with high self-balancing potential or those requiring external support. To incorporate spatial structure, we use the geographic coordinates $p_i = (\text{lat}_i, \text{lon}_i) \in \mathbb{R}^2$. This enables the clustering to preserve spatial contiguity and recognize climate-driven gradients. The final feature is the only one computed directly between node pairs: the inter-node correlation matrix $D_{ij}^{\text{inter-corr}}$. For each node pair $(i, j) \in \mathcal{N} \times \mathcal{N}$ and time series type pair $(k, \ell) \in \mathcal{K} \times \mathcal{K}$, we compute the Pearson correlation between the max-normalized time series $\tilde{x}_i^k(t)$ and $\tilde{x}_j^\ell(t)$, transform it into a distance using an exponential decay $d_{ij}^{k\ell} = \exp(-\rho_{ij}^{k\ell})$, and average across all variable pairs:

$$D_{ij}^{\text{inter-corr}} = \frac{1}{|\mathcal{K}|^2} \sum_{k,\ell} d_{ij}^{k\ell}$$

The exponential form ensures that strongly correlated nodes are tightly grouped, while anticorrelated or noisy pairs are sharply penalized. As a result, the metric highlights spatial complementarity and shared temporal fluctuations, helping preserve cross-zonal balancing relationships in the aggregated system.

3.1.2 Distance Computation and Aggregation

Each of the five per-node feature vectors is used to compute a pairwise distance matrix $D^{(m)} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$, where the index $m \in \{1, 2, 3, 4, 5, 6\}$ refers to the six feature types. The sixth feature corresponds to the inter-node correlation matrix, which is computed directly between node pairs rather than from per-node vectors. Additional features could easily be incorporated, such as electrical distance or network-theoretic measures, to further align the clustering with power system topology. For real-valued features (raw time series, DC, RDC, intra-node correlations), we use Euclidean distance:

$$D_{ij}^{(m)} = \|f_i^{(m)} - f_j^{(m)}\|_2$$

For spatial coordinates, we compute great-circle distances using the Haversine formula:

$$D_{ij}^{\text{geo}} = 2R \cdot \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta\phi_{ij}}{2} \right) + \cos(\phi_i) \cos(\phi_j) \sin^2 \left(\frac{\Delta\lambda_{ij}}{2} \right)} \right)$$

where $\Delta\phi_{ij}$ and $\Delta\lambda_{ij}$ are the latitude and longitude differences (in radians), and R is Earth's radius. Each resulting matrix $D^{(m)}$, including the inter-node correlation matrix $D^{\text{inter-corr}}$, is then min-max normalized to produce a scaled matrix $\tilde{D}^{(m)} \in [0, 1]^{|\mathcal{N}| \times |\mathcal{N}|}$.

The final composite distance matrix is constructed as a convex combination of the normalized matrices:

$$D_{ij} = \sum_{m=1}^6 w_m \cdot \tilde{D}_{ij}^{(m)} \quad \text{subject to} \quad \sum_{m=1}^6 w_m = 1, \quad w_m \geq 0$$

The weight vector $\mathbf{w} = (w_1, \dots, w_6)$ governs the contribution of each feature to the overall similarity structure. By adjusting the weights, planners can tune the clustering to emphasize

| Feature | Form | Distance Metric | Normalization |
|---|-----------------|-----------------|---------------|
| Raw time series $\tilde{x}_i^k(t)$ | Node vector | Euclidean | Max |
| DC | Node vector | Euclidean | Min-max |
| RDC | Node vector | Euclidean | Min-max |
| Intra-node correlation $\rho_i^{k\ell}$ | Node vector | Euclidean | Min-max |
| Geographic coordinates p_i | Node vector | Haversine | None |
| Inter-node correlation $D_{ij}^{\text{inter-corr}}$ | Pairwise matrix | — | — |

Table 3.1: Summary of feature formats and distance computations used in the composite similarity metric.

particular system aspects, such as renewable co-variation, demand profile shape, ramping dynamics, or spatial coherence. These weights are treated as hyperparameters and selected through downstream evaluation, as described in Chapter 4, Section 4.3.1. A summary of feature forms and distance metrics is provided in Table 3.1.

This structured yet interpretable design stands in contrast to black-box approaches such as autoencoders or learned embeddings, which often obscure the relationship between input features and resulting clusters. By building the similarity metric from explicit, physically meaningful components, this approach preserves transparency and control, both of which are essential for ensuring that the aggregation reflects the priorities and logic of the planning model it supports.

3.2 Spatial and Temporal Clustering

This section explains how the composite distance matrix defined in Section 3.1 is used to perform spatial clustering of network nodes, and how the resulting zonal structure is used to construct a temporally aggregated representation of system behavior.

3.2.1 Spatial Aggregation

We begin by partitioning the full node set \mathcal{N} into $C \in \mathbb{N}$ spatial zones using the composite distance matrix $D \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$ constructed from the weighted combination of planning-relevant features. Clustering is performed using the k -medoids algorithm, which selects actual data points as cluster centers, unlike k -means, which clusters in a continuous space and can choose a centroid not corresponding to any of the discrete cluster points. This is important in the context of power systems, since each spatial zone must be represented by a real node whose time series can be used in the aggregated model. The implementation uses the PAM (Partitioning Around Medoids) method with heuristic initialization and multiple random restarts to improve robustness and avoid poor local minima. Let $\mathcal{R} = \{i_1, \dots, i_C\} \subset \mathcal{N}$ denote the selected medoid set. Each node $i \in \mathcal{N}$ is assigned to its closest medoid according to the distance matrix: $c(i) = \arg \min_{j \in \mathcal{R}} D_{ij}$. For each cluster c , the representative node $i_c \in \mathcal{R}$ serves as the spatial proxy, and its time series form the raw input to the temporal

aggregation step.

While the primary goal of spatial clustering is to reduce the number of modeled locations and simplify the network topology, care is taken to preserve heterogeneity that matters for investment and dispatch. The use of features such as ramping behavior, correlation structure, and geographic continuity ensures that each zone groups nodes with similar planning-relevant characteristics. As a result, the aggregated topology can retain information about where capacity should be sited, which zones are likely to be flexible or volatile, and how cross-zonal interactions may play out. The cluster definition also implies a modeling simplification: intra-zonal transmission is treated as unconstrained (i.e., copperplate assumption), and only inter-zonal links are retained in the reduced grid. While this assumption facilitates tractability, it may obscure internal bottlenecks within large zones, its limitations are assessed in Chapter 4, Section 4.2.

Once clusters are defined, we construct representative time series for each zone. For each time series type $k \in \mathcal{K}$, we extract the max-normalized series from the medoid node i_c . However, to ensure that this representative series reflects the average behavior of the cluster in both shape and magnitude, we apply a rescaling step. Specifically, the medoid series is multiplied by the ratio of the cluster-wide average intensity to the medoid average intensity. This yields a rescaled time series $\bar{x}_c^k(t)$ that captures the shape of the medoid but reflects the mean intensity of the entire zone. These rescaled series $\bar{x}_c^k(t)$ across all $c \in \{1, \dots, C\}$ and $k \in \mathcal{K}$ form a compact, zonal description of system dynamics and serve as the input to temporal clustering.

3.2.2 Temporal Aggregation

The goal of temporal clustering is to reduce the 365-day hourly input data into a small number of representative daily profiles that capture the diversity of system operations across all spatial zones. To do so, we treat each day as a multi-dimensional snapshot of the full system. For each day $d \in \{1, \dots, 365\}$, we construct a three-dimensional tensor $\mathbf{T}_d \in \mathbb{R}^{K \times H \times C}$, where $K = |\mathcal{K}|$ is the number of time series types (e.g., demand, wind, solar), $H = 24$ is the number of hours per day, and C is the number of spatial clusters. Each entry $\mathbf{T}_d[k, h, c]$ represents the value of the time series of type k at hour h of day d in cluster c . This tensor encodes both inter-zonal interactions and intra-day operational patterns.

To enable clustering, we flatten each tensor into a vector $\mathbf{v}_d = \text{vec}(\mathbf{T}_d) \in \mathbb{R}^{KHC}$. Before computing distances, we normalize each vector using its ℓ_1 -norm: $\hat{\mathbf{v}}_d = \mathbf{v}_d / (\sum_i \mathbf{v}_d[i] + \varepsilon)$. This normalization ensures that clustering emphasizes the relative shape and spatial structure of daily profiles rather than absolute magnitudes. It avoids overemphasis on peak days and allows the selection process to focus on typical operational patterns and ramping behavior. In practice, this is often complemented by the explicit inclusion of extreme days (e.g., peak demand or low renewable days) to ensure system adequacy is not underestimated [13, 14]. Clustering is then performed using k -medoids on the euclidean-based distance matrix induced by these normalized vectors, yielding a reduced set $\mathcal{D} = \{d_1, \dots, d_D\} \subset \{1, \dots, 365\}$ of representative days. Each day is shared across all clusters, ensuring a temporally coherent basis for dispatch and investment decisions. That is, the same set of representative days is used system-wide, so that temporal patterns remain aligned across zones. This avoids inconsistencies that could arise if each zone had its own independent set of representative

periods, such as mismatched peaks or ramping events that distort inter-zonal coordination. By enforcing a common temporal structure, the model preserves the simultaneity of events across the network.

To maintain fidelity with the original time distribution, each selected day $d \in \mathcal{D}$ is assigned a weight $\omega_d = |\pi(d)|$, where $\pi(d)$ is the set of days mapped to medoid d . These weights are used in the CEP model to scale cost and emissions over the year. Crucially, the interaction between spatial and temporal aggregation is preserved: temporal clustering is performed on the zonal time series obtained from spatial medoids, and daily profiles reflect diversity across both space and time. This coordinated approach avoids the common pitfall of applying separate clustering steps that overlook how spatial resolution constrains temporal patterns. It also ensures that key phenomena, such as solar complementarity between east and west zones, or ramping diversity between coastal and inland regions, are captured in the reduced dataset.

The result is a compact yet expressive spatio-temporal input representation, tailored to support planning decisions in the CEP model. The final step, presented in the next section, assembles this reduced dataset into a structured input format and integrates it into the planning model.

3.3 Constructing the Aggregated Input Dataset

The final step of the pipeline transforms the clustered and rescaled time series into a reduced spatio-temporal dataset suitable for optimization. This section describes how the output of the spatial and temporal clustering steps, namely the medoid-based zones and representative days, is used to build the input data for a compressed capacity expansion planning model. The construction ensures three essential properties: conservation of annual energy quantities, temporal representativeness via weighted sampling, and full consistency across spatial zones. These design features enable the reduced model to approximate long-run investment outcomes without introducing incoherence or systematic bias.

The aggregated system is defined over a reduced node set $\mathcal{R} = \{i_1, \dots, i_C\} \subset \mathcal{N}$, and a reduced temporal horizon composed of D representative days. Each day $d \in \mathcal{D} = \{d_1, \dots, d_D\} \subset \{1, \dots, 365\}$ is assigned a weight $\omega_d \geq 0$, such that $\sum_d \omega_d = 365$, indicating the frequency with which the day occurs in the original dataset. These two sets, spatial zones and weighted representative days, define the index structure of the reduced model: nodes \mathcal{R} , time $\mathcal{T}_{\text{agg}} = \mathcal{D} \times \{1, \dots, 24\}$, and time series types $\mathcal{K} = \{\text{demand}, \text{wind}, \text{solar}\}$.

Although medoid nodes serve as proxies for their clusters, the raw time series $x_{i_c}^k(t)$ cannot be used directly in all cases: doing so would misrepresent the total energy consumption of the cluster. To address this, each medoid profile is post-processed to ensure consistency with the aggregate cluster behavior. For demand, where energy quantities are additive, we apply a rescaling step that adjusts the medoid profile to match the cluster's total annual consumption. For each cluster c and time series type k , we define a scaling factor:

$$s_c^k = \frac{\sum_{i \in \mathcal{C}_c} \sum_{t \in \mathcal{T}} x_i^k(t)}{\sum_{t \in \mathcal{T}} x_{i_c}^k(t)}$$

This factor guarantees that the rescaled medoid demand matches the total annual energy

of its cluster. It is applied to demand profiles. For renewable profiles such as wind and solar, which represent normalized capacity factors, no scaling is applied. The aggregated time series are then defined as:

$$\bar{x}_c^k(d, h) = \begin{cases} s_c^k \cdot x_{i_c}^k(d, h), & \text{if } k = \text{demand} \\ x_{i_c}^k(d, h), & \text{if } k \in \{\text{wind, solar}\} \end{cases}$$

for all $d \in \mathcal{D}, h \in \{1, \dots, 24\}$, where $x_{i_c}^k(d, h)$ is the value of time series k at hour h of day d in the original full-resolution dataset. This results in a set of scaled, cluster-level time series $\bar{x}_c^k : \mathcal{D} \times \{1, \dots, 24\} \rightarrow \mathbb{R}_{\geq 0}$ for each $c \in \{1, \dots, C\}$ and $k \in \mathcal{K}$. These time series retain the temporal shape and inter-cluster complementarity of the medoids while conserving the total energy across the original system.

To ensure temporal fidelity in the optimization, each representative day is weighted according to its empirical frequency. These weights are used throughout the model to scale objective terms and annualized constraints. Let $y(d, h)$ be any cost- or energy-related quantity. Then its contribution to the annual objective is computed as:

$$\sum_{d \in \mathcal{D}} \omega_d \sum_{h=1}^{24} \text{Cost}(y(d, h))$$

This weighted summation ensures that daily dispatch patterns are properly scaled to reflect full-year behavior. Constraints related to total generation, storage cycles, or emissions can also be formulated using the same weighting structure, preserving annual consistency despite the reduced temporal sample.

The final outcome of the aggregation process is a compressed dataset with the following elements: (i) a set of representative nodes \mathcal{R} , each associated with a spatial cluster, (ii) a set of representative days \mathcal{D} with weights ω_d , (iii) scaled, daily time series $\bar{x}_c^k(d, h)$ for each cluster, and (iv) scaling factors s_c^k that link the medoid profiles to the full cluster behavior. These elements form the input to the CEP model introduced in Chapter 4, Section 4.1. Together, they ensure that the reduced model remains tractable, coherent, and reflective of the full-resolution system's operational characteristics.

Chapter 4

Evaluation and Experimental Framework

Chapter 3 ended with the construction of aggregated input data for a power system planning model. These inputs result from a specific aggregation configuration, defined by the number of spatial clusters, the number of representative days, and the feature weights used for clustering. This chapter builds the framework needed to assess how these design choices influence the quality and feasibility of planning outcomes.

We start by presenting the CEP model used in this study (§4.1), which takes the form of a Generation and Transmission Expansion Planning (GTEP) model and serves as the base optimization problem on which all aggregation configurations are tested. We then review existing approaches for validating solutions obtained on aggregated inputs and explain the deaggregation procedure we adopt (§4.2), where we start from fixed investment decisions and re-simulate operations at increasing levels of resolution to test the feasibility and realism of each solution. Finally, we describe how this evaluation process is used experimentally to tune hyperparameters and compare our pipeline to a set of benchmarks (§4.3). This full framework provides the foundation for the results in Chapter 6.

4.1 Generation and Transmission Expansion Model

We describe the GTEP model used to evaluate the datasets produced by the aggregation pipeline. The model is a transport-based MILP that co-optimizes investment and dispatch over the reduced network, accounting for load, storage dynamics, resource profiles, and operational constraints. The model operates on a set of spatial clusters \mathcal{C} and a time horizon $\mathcal{T}_{\text{agg}} = \{1, \dots, T\}$ indexed by representative hours, each with weight $\omega_t \in [0, 1]$. Each zone $c \in \mathcal{C}$ has demand $D_c(t)$, obtained from the scaled and aggregated profiles of Chapter 3. For reference, all sets, parameters, and decision variables used in the model are summarized in Tables 4.2 and 4.1.

Transmission is modeled using a simplified transport flow formulation, which enforces nodal power balance and imposes line-wise flow capacity limits, but omits the underlying electrical physics. Unlike DCOPF models, which include voltage phase angle variables and derive flows from linearized power flow equations subject to Kirchhoff's laws and line admittances, the transport model treats flows as abstract quantities constrained only by capacity and conservation. There are no phase angles, no impedance parameters, and no

cycle constraints. This abstraction significantly reduces computational burden and avoids numerical instability, making it well-suited for running a large number of experiments across different spatial and temporal aggregation configurations. It is commonly used in regional CEP tools such as NREL's Resource Planning Model [37]. However, the simplicity comes at the cost of physical fidelity. In meshed networks, actual power flows obey Kirchhoff's laws and distribute according to network reactances; transport models cannot represent loop flows or internal congestion and may underestimate line loading. As such, this approach is best viewed as a screening tool that supports approximate investment decisions, with the understanding that final layouts must be validated under more realistic power flow conditions.

The set of technologies is denoted \mathcal{G} , and is partitioned into dispatchable units $\mathcal{G}_{\text{disp}}$ (e.g., gas turbines) and variable renewables \mathcal{G}_{var} (e.g., wind, solar). For each technology $g \in \mathcal{G}$, the model distinguishes between marginal cost φ_g , per-unit capacity \bar{P}_g , ramp rate r_g , and installable potential Γ_g^{\max} . The time-varying availability of variable renewables is given by $a_c^g(t) \in [0, 1]$, which reflects the cluster-specific capacity factor profile for technology g at time t . The set of interzonal transmission links is denoted $\mathcal{E} \subseteq \mathcal{C} \times \mathcal{C}$, where each pair $(c, c') \in \mathcal{E}$ represents an admissible directional connection between zones. For each link $(c, c') \in \mathcal{E}$, $\bar{T}_{cc'}$ denote the maximum installable transmission capacity (in MW), representing technical or economic limits on line sizing.

Investment variables include generation capacity $G_c^g \in \mathbb{Z}_+$ for dispatchable technologies and $G_c^g \in \mathbb{R}_+$ for variable renewables, as well as storage power capacity $S_c \in \mathbb{R}_+$ and transmission capacity $T_{cc'} \in \mathbb{R}_+$ for each interzonal link. Operational variables comprise generation dispatch $P_c^g(t)$, curtailment $q_c^g(t)$, load shedding $\ell_c(t)$, power flows $f_{cc'}(t)$, storage charging $C_c(t)$, discharging $D_c(t)$, and energy level $e_c(t)$. Storage dynamics are further characterized by fixed charging and discharging efficiencies, $\eta^{\text{ch}} \in (0, 1]$ and $\eta^{\text{dis}} \in (0, 1]$, and a duration parameter $\tau \in \mathbb{R}_+$ defining the maximum energy-to-power ratio.

The objective function minimizes the sum of investment and operational costs across all technologies and time periods:

$$\min \quad \sum_{c,g} C_g^{\text{inv}} G_c^g + \sum_c C_S^{\text{inv}} S_c + \sum_{(c,c')} C_T^{\text{inv}} T_{cc'} + \sum_{t \in \mathcal{T}_{\text{agg}}} \omega_t \sum_c \left(\sum_{g \in \mathcal{G}_{\text{disp}}} \varphi_g P_c^g(t) + \psi \ell_c(t) \right)$$

Generation constraints ensure that dispatchable units operate within their installed capacity limits (Eq.(4.1)), while variable renewables are constrained by their availability factors and may curtail excess production when needed (Eq.(4.2)). These fundamental operational limits capture the physical characteristics of different generation technologies and their temporal patterns:

$$P_c^g(t) \leq G_c^g \cdot \bar{P}_g \quad \forall g \in \mathcal{G}_{\text{disp}}, c \in \mathcal{C}, t \in \mathcal{T}_{\text{agg}} \quad (4.1)$$

$$P_c^g(t) + q_c^g(t) = G_c^g \cdot a_c^g(t) \quad \forall g \in \mathcal{G}_{\text{var}}, c \in \mathcal{C}, t \in \mathcal{T}_{\text{agg}} \quad (4.2)$$

Ramping constraints for dispatchable technologies limit the rate at which generation can change between consecutive time periods (Eq. (4.3)), reflecting the physical inertia and operational limitations of conventional power plants. This constraint becomes particularly

Table 4.1: Glossary of model variables

| Symbol | Type | Description |
|---|----------------|---|
| Investment Variables | | |
| G_c^g | integer / real | Installed capacity of generation technology g in cluster c (integer if $g \in \mathcal{G}_{\text{disp}}$, real if $g \in \mathcal{G}_{\text{var}}$) |
| S_c | real | Installed storage power capacity in cluster c |
| $T_{cc'}$ | real | Installed transmission capacity between clusters c and c' |
| Operational Variables (per time step t) | | |
| $P_c^g(t)$ | real | Power generated by technology g in cluster c |
| $q_c^g(t)$ | real | Curtailment of renewable g in cluster c |
| $\ell_c(t)$ | real | Load shedding in cluster c |
| $f_{cc'}(t)$ | real | Power flow from cluster c to c' |
| $C_c(t)$ | real | Storage charging power in cluster c |
| $D_c(t)$ | real | Storage discharging power in cluster c |
| $e_c(t)$ | real | Storage energy level in cluster c |

important when evaluating the impact of temporal aggregation on system flexibility requirements:

$$|P_c^g(t) - P_c^g(t-1)| \leq G_c^g \cdot \bar{P}_g \cdot r_g \quad \forall g \in \mathcal{G}_{\text{disp}}, c \in \mathcal{C}, t > 1 \quad (4.3)$$

The load balance constraint ensures power system equilibrium at each node and time period by requiring that total generation plus storage discharge and transmission inflows equal demand plus storage charging, transmission outflows, and any necessary load shedding (Eq. (??)). This fundamental power system constraint maintains the physical feasibility of the solution across the aggregated network:

$$\sum_g P_c^g(t) + D_c(t) - C_c(t) + \sum_{c'} f_{c'c}(t) - \sum_{c'} f_{cc'}(t) + \ell_c(t) = D_c(t) \quad \forall c \in \mathcal{C}, t \in \mathcal{T}_{\text{agg}} \quad (4.4)$$

Storage dynamics are captured through a comprehensive set of constraints that model energy state evolution, charging and discharging limits, and intertemporal consistency. The energy balance equations track storage levels over time while accounting for charging and discharging efficiencies (Eqs. (4.5)–(4.6)). Equation (4.5) sets the initial state of charge $e_c(1)$ based on an exogenous parameter e_c^0 . In long-term planning models, this value is typically chosen within the feasible energy range to avoid artificially constraining early-period operations. Since the objective is to evaluate structural investment decisions rather than short-term operational trajectories, the specific value of e_c^0 is not critical, provided it does not bias the ability to charge or discharge at the start of the horizon. Power capacity limits constrain the rate of

energy transfer (Eq. (4.7)), and energy capacity limits define the maximum storable energy (Eq. (4.8)). No terminal or cyclical constraint is imposed on the state of charge, allowing storage to end the horizon at any feasible level:

$$e_c(1) = e_c^0 + \eta^{\text{ch}} C_c(1) - \frac{D_c(1)}{\eta^{\text{dis}}} \quad \forall c \in \mathcal{C} \quad (4.5)$$

$$e_c(t) = e_c(t-1) + \eta^{\text{ch}} C_c(t) - \frac{D_c(t)}{\eta^{\text{dis}}} \quad \forall c \in \mathcal{C}, t > 1 \quad (4.6)$$

$$0 \leq C_c(t), D_c(t) \leq S_c \quad \forall c \in \mathcal{C}, t \in \mathcal{T}_{\text{agg}} \quad (4.7)$$

$$0 \leq e_c(t) \leq \tau S_c \quad \forall c \in \mathcal{C}, t \in \mathcal{T}_{\text{agg}} \quad (4.8)$$

Transmission constraints limit power flows between adjacent zones according to the installed transmission capacity, allowing bidirectional flows while respecting infrastructure limits (Eq. (4.9)). This transport model approximation enables efficient computation while capturing the essential trade-offs between transmission investment and generation dispatch:

$$-T_{cc'} \leq f_{cc'}(t) \leq T_{cc'} \quad \forall (c, c') \in \mathcal{E}, t \in \mathcal{T}_{\text{agg}} \quad (4.9)$$

Resource potential constraints ensure that the total installed capacity of variable renewable technologies does not exceed the available resource potential in each region, reflecting land use limitations, resource quality constraints, and other factors that limit renewable energy deployment (Eq. (4.10)):

$$\sum_c G_c^g \leq \Gamma_g^{\max} \quad \forall g \in \mathcal{G}_{\text{var}} \quad (4.10)$$

To reflect physical or policy-imposed siting constraints, we impose upper bounds $\bar{T}_{cc'}$ on the amount of transmission capacity that can be built along each link (c, c') (Eq. (4.11)). This allows the model to trade off line reinforcement against local generation or storage investment while respecting regional build limits.

$$0 \leq T_{cc'} \leq \bar{T}_{cc'} \quad \forall (c, c') \in \mathcal{E} \quad (4.11)$$

This MILP formulation uses integer variables only for dispatchable capacity to mirror real-world indivisibility constraints such as discrete power plant units, while treating variable renewables and other technologies as continuous to maintain computational tractability. While this approach enables efficient exploration of many configurations, it omits certain operational features such as unit commitment decisions (startup costs, minimum up and down times, or binary on/off constraints), which may affect dispatch realism in systems with high thermal generation. These elements could be incorporated in an extended formulation but are excluded here for simplicity. Emissions are also not modeled explicitly, so the investment outcomes produced by the model cannot be directly linked to carbon targets or used for policy evaluation. The next section details how this model is used for evaluation by reconstructing full-resolution outputs from fixed investment plans.

Table 4.2: Glossary of model sets and parameters

| Symbol | Description |
|--|--|
| Sets | |
| \mathcal{C} | Set of spatial clusters (zones) |
| \mathcal{T}_{agg} | Set of representative time steps (indexed by t) |
| \mathcal{G} | Set of generation technologies |
| $\mathcal{G}_{\text{disp}} \subseteq \mathcal{G}$ | Set of dispatchable generation technologies |
| $\mathcal{G}_{\text{var}} \subseteq \mathcal{G}$ | Set of variable renewable generation technologies |
| $\mathcal{E} \subseteq \mathcal{C} \times \mathcal{C}$ | Set of directional transmission links between clusters |
| Parameters | |
| ω_t | Weight of representative time step t |
| $D_c(t)$ | Demand in cluster c at time t |
| φ_g | Marginal cost of generation technology g |
| \bar{P}_g | Per-unit power capacity of generation technology g |
| r_g | Ramping rate of dispatchable generator g |
| Γ_g^{\max} | Maximum installable potential of variable renewable g |
| $a_c^g(t)$ | Availability factor of renewable g at cluster c , time t |
| C_g^{inv} | Investment cost per unit of generation technology g |
| C_S^{inv} | Investment cost per unit of storage power capacity |
| C_T^{inv} | Investment cost per unit of transmission capacity |
| $\bar{T}_{cc'}$ | Maximum allowable transmission capacity between c and c' |
| $\eta^{\text{ch}}, \eta^{\text{dis}}$ | Charging and discharging efficiency of storage |
| τ | Maximum storage duration (energy-to-power ratio) |
| ψ | Cost of unserved energy (load shedding penalty) |
| e_c^0 | Initial state of charge in cluster c |

4.2 Validation via Sequential Deaggregation

When a CEP model is simplified in space or time, the resulting plan may look cost-efficient in the aggregated model but fail to operate reliably once tested on the full dataset. Understanding how researchers evaluate this risk is the focus of this section. We first outline, in Section 4.2.1, why this question matters and how different studies have tried to measure the fidelity of aggregated solutions. We then review, in Sections 4.2.2 and 4.2.3, common approaches to downscaling results in space and time, before looking at integrated workflows in Section 4.2.4. Finally, Section 4.2.5 explains the procedure used in this work, applying these general ideas to the GTEP model from Section 4.1 to compare alternative aggregation configurations.

4.2.1 Motivation and Methodological Framework

Evaluating the reliability of aggregated capacity expansion model (CEP) solutions remains a central and unresolved challenge in energy planning. Aggregation compresses a high-dimensional problem, with decision space \mathbb{R}^m , into a reduced form \mathbb{R}^n where $n \ll m$. The resulting mapping $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is surjective but non-injective as many fine-resolution solutions correspond to the same aggregated representation, and the inverse mapping f^{-1} is inherently non-unique [38]. This loss of information complicates the evaluation of aggregated results, because a single low-resolution solution can translate into multiple possible high-resolution system designs, each with potentially different feasibility and performance.

This difficulty is further compounded in multi-sector models, where the inclusion of coupled infrastructures such as electricity, hydrogen, or heating increases dimensionality and tightens interdependencies across space and time. To keep such models computationally tractable, strong simplifications are often introduced, such as coarse spatial zoning, aggressive temporal aggregation, or budget-based relaxations of inter-sectoral flows [18]. While these abstractions are necessary for solvability, they increase the risk that aggregated investment decisions cannot be implemented feasibly when tested against the original, fully resolved system representation.

The literature describes three broad paradigms for evaluating aggregated CEP results [39]. The first is **direct benchmarking** against a high-resolution solution of the same model. This approach provides a clean measure of accuracy—differences in total system cost, investment patterns, or dual values such as marginal energy prices [23], but it requires solving a reference model that is often computationally infeasible for realistic systems. Consequently, direct accuracy metrics are typically available only on small test cases or highly simplified formulations.

The second paradigm is **iterative operational validation**, where aggregated planning models are linked to detailed production cost models (PCM) (e.g., PLEXOS), which simulate system dispatch and reliability over full temporal resolution under fixed investment decisions. This process adds two layers of detail absent from the planning solve. It restores hourly or sub-hourly time steps, revealing flexibility shortages such as ramping or reserve violations, and it introduces finer operational constraints like unit commitment and reserve requirements [39–44]. The workflow is typically iterative. When the PCM reveals frequent load shedding, reserve shortfalls, or high redispatch costs, these results are fed back into the planning model by tightening flexibility constraints or adding new investment options, and the process is repeated until operational feasibility is achieved. However, it preserves the spatial aggregation of the original CEP solution, effectively treating each zone as a copperplate network. As a result, intra-zonal transmission congestion or nodal feasibility issues remain undetected.

The third and increasingly prominent paradigm is **disaggregation-based evaluation** [38]. This approach maps aggregated investment decisions back to the full spatial and temporal resolution of the original dataset and tests their operational feasibility. A fixed investment plan is redispatched on the fine-resolution network, and any infeasibilities (load shedding, excessive curtailment, violations of ramping or network constraints) are recorded, often translated into penalty costs to quantify the "repair" needed for feasibility. This enables evaluation of model fidelity without solving the full-scale high-resolution planning problem, making it applicable to larger instances. Yet, it introduces its own methodological ambiguity:

the mapping f^{-1} from aggregated zones to detailed network nodes is not unique and typically relies on heuristics, which can bias results.

Taken together, these approaches illustrate both what researchers measure and the limitations they face. Accuracy metrics are insightful but hinge on solving intractable reference models, operational feasibility metrics capture spatial and temporal inconsistencies but depend on non-systematic disaggregation choices. These challenges motivate the structured deaggregation framework developed in this chapter. By defining a consistent, reproducible mapping and a high-resolution feasibility check, it provides a scalable and rigorous method to validate aggregated solutions and quantify the trade-offs between model compression and fidelity.

4.2.2 Spatial Deaggregation Methods

The literature on spatial deaggregation mostly describes two established families of approaches, with a recent third option proposed by Frysztacki et al. (2023) [38]. Existing methods can be broadly categorized as either heuristic downscaling of regional capacities or full high-resolution re-optimization, with Frysztacki et al. introducing an intermediate optimization heuristic. These options, along with the broader evaluation paradigms discussed above, are summarized in Table 4.3.

- **Heuristic Allocation and Siting Tools:** The simplest approach allocates the same share of the aggregated cluster capacity G_c^g to each node i within the cluster c , subject to local land-use constraints. If the cluster contains $|c|$ nodes, the allocated capacity at node i is:

$$G_i^g = \frac{G_c^g}{|c|} \quad \forall i \in c, \forall g \in \mathcal{G} \quad (4.12)$$

This uniform split is computationally trivial but ignores heterogeneity in resource quality, demand, and transmission access. To improve realism, recent workflows employ spatial siting tools such as CERF [45], Net Zero America [46], or The Nature Conservancy’s siting datasets (2023) [47]. These tools downscale regional capacities by ranking candidate sites based on high-resolution data on land availability, environmental exclusions, interconnection costs, and empirical siting patterns. Investments are assigned to the lowest-cost or most feasible sites within each region. While this captures heterogeneity absent from naive uniform allocation, it does not perform a full high-resolution optimization: spatial placement is exogenous, relying on data quality and static cost curves. Operational feasibility within clusters remains untested, and no new intra-regional transmission is planned.

- **Regional Re-optimization:** This approach solves the full capacity expansion formulation independently for each cluster while constraining total capacity to match aggregated results:

$$\sum_{i \in c} G_i^g = G_c^g \quad \forall g \in \mathcal{G} \quad (4.13)$$

While computationally intensive, this method preserves the optimization structure and can account for intra-cluster transmission constraints and resource heterogeneity.

- **Minimal Excess Electricity:** Developed by Frysztacki et al. (2023) [38], this method provides a compromise between uniform allocation and full re-optimization. It allocates variable renewable capacities G_i^g (technology $g \in \mathcal{G}_{\text{var}}$ at node i) within a cluster by minimizing the total excess energy that cannot be locally consumed or exported:

$$\min_{G_i^g} \sum_{\substack{i \in c \\ g \in \mathcal{G}_{\text{var}} \\ t \in \mathcal{T}_{\text{agg}}}} \left[a_i^g(t) G_i^g - D_i(t) - \alpha \sum_{j \in \mathcal{N}(i)} F_{ij} \right]^+ \quad (4.14)$$

Here $a_i^g(t)$ denotes the availability factor of renewable technology g at node i and time t , $D_i(t)$ the local electricity demand, and F_{ij} the existing transfer capacity between node i and its neighbor $j \in \mathcal{N}(i)$. The parameter $\alpha = 0.7$ is a heuristic scaling factor proposed in [38] to approximate that only a fraction of the nominal transmission capacity can reliably be used for exporting surplus power, reflecting competing flows and network inefficiencies. The objective penalizes renewable generation that cannot be consumed locally or feasibly exported, pushing capacity allocations toward nodes with better demand matching or access to transmission links. This provides a computationally light alternative to full high-resolution re-optimization, which would explicitly model operational dispatch, cost trade-offs, and network constraints but is significantly more expensive to solve. However, MinEx remains a heuristic as it does not explicitly coordinate dispatchable generation or storage siting, and cannot design new intra-cluster transmission capacity. It only reallocates variable renewable capacities based on static approximations of demand and transfer potential. As a result, when aggregated plans are disaggregated using this method, residual feasibility issues often persist. Frysztacki et al. report that even highly resolved European models (100–200 nodes) still exhibit 3–7% unmet demand under full-resolution operational testing.

4.2.3 Temporal Deaggregation Methods

Temporal deaggregation restores the full chronological resolution of demand, renewable profiles, and storage dynamics to test whether investment decisions from a temporally aggregated CEP model remain feasible when applied to the original dataset. The standard approach fixes the aggregated investments and resolves a detailed operational dispatch problem over all time steps, typically using a high-resolution PCM. Several recent studies, such as the LA100 study (NREL, 2021) [43], Yates et al. (2024) [40], Wessel et al. (2022) [44], Dyreson et al. (2022) [41], and Cohen et al. (2022) [42], have adopted this approach to reveal errors introduced by representative period selection, including underestimated storage requirements and overestimated system flexibility. When both spatial and temporal aggregation are applied, spatial disaggregation must generally precede this step so that investments are first mapped to individual nodes before evaluating their temporal feasibility.

4.2.4 Integrated Deaggregation Framework

While spatial and temporal deaggregation have each been studied in isolation, the literature contains little evidence of workflows that combine both dimensions in a coordinated, iterative manner. Mongird and Rice (2024) [39] review existing practice and note that prior studies typically link capacity expansion models either to high-resolution PCMs [40–44] for temporal feasibility checks or to spatial siting models for land-use and interconnection analysis [45–47], but rarely both. They explicitly state that "*we could not identify a previously published workflow that has included all of them simultaneously*". Their own contribution is to propose, at a conceptual level, an integrated multiscale framework where long-term planning is followed by spatial siting refinement, chronological dispatch validation, and feedback to adjust investment decisions if infeasibilities are found. This is presented as a design recommendation rather than a fully implemented, demonstrated workflow.

In line with this conceptual approach, a practical integrated deaggregation framework can be structured as a sequential process:

1. **Spatial Deaggregation:** Map cluster-level capacities G_i^c to node-level capacities G_i^g using one of the spatial allocation methods described above.
2. **Temporal Deaggregation:** Fix the spatially disaggregated investments and resolve operational dispatch over the full chronological horizon to test temporal feasibility.
3. **Feasibility Assessment:** Quantify any load shedding, curtailment, or other constraint violations that arise when the aggregated solution is expressed at full spatio-temporal resolution.

This sequential approach does not fully capture the feedback loops or co-dependencies between space and time envisioned in [39], but it provides a tractable method to approximate an integrated evaluation while maintaining clear separation of steps.

4.2.5 Our Methodological Choice

Among the options in Table 4.3, this work adopts a disaggregation-based evaluation combined with regional re-optimization for spatial allocation. We adopt a two-stage, optimization-based deaggregation approach to evaluate solutions produced by aggregated GTEP models. Unlike heuristic redistribution formulas, our procedure re-solves constrained MILPs at full spatial and temporal resolution. This ensures that feasibility is assessed under the same physical and operational rules as the planning model.

In the first stage, we perform spatial deaggregation by solving a full-resolution GTEP problem while constraining total investment per cluster to match the aggregated solution. For each cluster c and generation technology g , node-level capacities G_i^g must satisfy $\sum_{i \in c} G_i^g = G_c^g$, and storage capacities obey $\sum_{i \in c} S_i = S_c$. Transmission investments are disaggregated similarly: for each aggregated link (c, c') connecting two clusters, we consider the set of physical lines $\mathcal{E}_{cc'} = \{(i, j) : i \in c, j \in c'\}$ and enforce $\sum_{(i,j) \in \mathcal{E}_{cc'}} T_{ij} = T_{cc'}$, where T_{ij} is the built capacity on line (i, j) and $T_{cc'}$ is the aggregated investment between clusters. These constraints preserve the total aggregated decisions while allowing the optimizer to

Table 4.3: Overview of common approaches for evaluating aggregated CEP solutions. Methods in bold are those implemented in this work.

| Method | How it works | Key implication |
|--|---|---|
| <i>Evaluation Paradigms</i> | | |
| Direct benchmarking | Solve a high-resolution reference model and compare costs and investments. | Clear accuracy metric but rarely tractable for realistic systems. |
| Operational validation | Test aggregated plans in a detailed PCM, iteratively adjusting constraints. | Captures flexibility issues but ignores intra-zonal congestion. |
| Disaggregation-based evaluation | Map aggregated investments to fine resolution and re-dispatch. | Scalable, results depend on chosen disaggregation rule. |
| <i>Spatial Deaggregation Methods</i> | | |
| Heuristic allocation | Split capacity evenly or by static siting rules. | Simple but may give infeasible or unrealistic layouts. |
| Regional re-optimization | Optimize within clusters under fixed totals. | Feasible and network-aware but computationally heavy. |
| Minimal excess electricity | Allocate renewables to reduce local oversupply. | Fast compromise, ignores dispatchables and new lines. |

determine their distribution across individual nodes and lines. The model otherwise includes the same operational constraints as the aggregated GTEP formulation, ensuring that the spatial allocation is feasible with respect to power balance, ramping, storage dynamics, and line limits.

In the second stage, we fix all node- and line-level investments obtained from Stage 1 and solve the operational MILP over the full chronological horizon $\mathcal{T} = \{1, \dots, 8760\}$. The objective minimizes operating costs and penalties for unserved energy, subject to all dispatch, storage, and flow constraints with fixed investments. This step tests whether the spatially feasible investment plan remains operable under real chronological conditions and flexibility requirements. Any residual load shedding $\ell_v(t)$ or high curtailment $q_v^g(t)$ in this full-resolution solution indicates feasibility challenges introduced by the aggregation process.

By deaggregating every candidate plan, we explicitly reconstruct a feasible high-resolution implementation that respects operational and network constraints over the full year. Each aggregated solution is thus translated into a valid upper bound to the true high-resolution optimum. This eliminates ambiguity in comparing results: solutions can be directly ordered by their realized cost and feasibility penalties in the detailed model, providing a consistent and rigorous basis for evaluation. As detailed in the next section, this ordering replaces the iterative feedback loops commonly used to tune aggregation and allows systematic selection

of the most reliable configurations.

4.3 Experimental Framework

With the GTEP model and the deaggregation procedure in place, we now turn to how the aggregation pipeline is tested in practice. This section introduces the experimental setup used to evaluate different aggregation configurations and to benchmark our approach against methods from the literature.

4.3.1 Hyperparameter Search

The aggregation pipeline described in Chapter 3 relies on three key choices, hereafter referred to as hyperparameters, that govern how the network is clustered and temporally compressed. As introduced in Chapter 1, we use the term *task-aware* to describe the approach of tuning these hyperparameters based on the deaggregated outputs of the GTEP model, so as to align aggregation with downstream modeling needs. This is precisely the approach taken in this chapter.

The full set of hyperparameters includes eight values: six feature weights that define the spatial similarity metric, and two structural parameters that control aggregation granularity:

- The weight vector $\mathbf{w} \in \mathbb{R}^6$ determines how strongly each feature type (raw time series, load duration curves, ramp duration curves, intra-node correlations, geographic coordinates, and inter-node correlations) contributes to the spatial similarity metric.
- The number of spatial clusters C sets how many zones are kept
- The number of representative days D sets the temporal resolution of the model.

To understand how feature design influences the performance of spatial aggregation, we perform a grid search over the feature weight vector \mathbf{w} . By testing multiple weight combinations, we aim to assess the role of each signal type and identify configurations that perform better than others.

However, the quality of a given \mathbf{w} may depend strongly on the aggregation level, namely, the number of spatial clusters C and representative days D used. A weight vector that performs well at one resolution may underperform at another. To account for this, we evaluate each \mathbf{w} over a small set $\mathcal{S} = \{(C, D)\}$ of spatial and temporal aggregation levels. For each combination (\mathbf{w}, C, D) , we generate aggregated input data, solve the corresponding GTEP model, and evaluate the full-resolution cost after deaggregation. The cost includes both investment and operational components.

To compare configurations across different aggregation levels, we define a set of relative performance metrics that capture both quality and consistency. Let $z_i^{(C,D)}$ be the deaggregated cost for weight configuration i under aggregation level (C, D) , and let $z_{(C,D)}^*$ denote the best cost achieved in that group. We then compute the following three metrics for each configuration in each group:

- (i) Its **relative rank**, based on the ordering of cost values across all configurations. The best-performing configuration receives rank 1, the second-best receives rank 2, and so on. This gives a direct comparison of how each design fares within each aggregation setting (C, D) .
- (ii) A **discrete score** $s_i^{(C,D)} \in \{1.0, 0.8, 0.6, 0.4, 0.2, 0.0\}$ assigned according to this rank. Only the top five configurations in each group receive a non-zero score, with the top configuration scoring 1.0. A score of 0.0 indicates performance outside the top five and thus no contribution to overall quality.
- (iii) Its **relative objective gap** $g_i^{(C,D)} = (z_i^{(C,D)} - z_{(C,D)}^*) / z_{(C,D)}^*$, which measures the normalized distance to the best cost in the group and provides a scale-independent way of comparing how far a configuration is from the group's optimum.

These quantities are then aggregated across all $(C, D) \in \mathcal{S}$ to yield the following global metrics for each configuration:

| Metric | Formula |
|-----------------|--|
| Average rank | $\frac{1}{ \mathcal{S} } \sum_{(C,D) \in \mathcal{S}} \text{rank}_i^{(C,D)}$ |
| Mean score | $\frac{1}{ \mathcal{S} } \sum_{(C,D) \in \mathcal{S}} s_i^{(C,D)}$ |
| Total score | $\sum_{(C,D) \in \mathcal{S}} s_i^{(C,D)}$ |
| Mean gap | $\frac{1}{ \mathcal{S} } \sum_{(C,D) \in \mathcal{S}} g_i^{(C,D)}$ |
| Gap variability | Standard deviation of $\{g_i^{(C,D)}\}_{(C,D) \in \mathcal{S}}$ |

These quantities help identify feature weightings that perform well not only in a single configuration, but consistently across a range of model resolutions. High scores and low cost gaps reflect strong performance, while low variability signals robustness to the level of aggregation.

After this search, we can select a high-performing weight vector \mathbf{w}^* and, with it fixed, investigate the trade-off between spatial and temporal aggregation. By varying (C, D) while keeping \mathbf{w}^* constant, we construct a Pareto frontier that reveals how each aggregation dimension affects solution quality under a fixed runtime or model complexity budget.

4.3.2 Benchmarks

To assess the added value of our aggregation pipeline, we compare it against a set of benchmark methods that reflect both naive heuristics and state-of-the-art alternatives in the literature. All methods use the same temporal aggregation procedure, only the spatial clustering strategy is varied. Our goal is to evaluate whether our pipeline offers measurable improvements in planning quality, as determined by full-resolution investment and dispatch outcomes.

We distinguish between two groups of benchmarks: simple spatial baselines and temporally informed spatial clustering methods. The first group includes three naive spatial aggregation strategies that do not rely on signal-based similarity:

Random cluster assignment. This method assigns each bus in the system to one of C clusters uniformly at random, using a fixed seed for reproducibility. Within each cluster, the representative node is the one with the smallest index. This approach serves as a lower bound for aggregation quality, as it ignores all physical, geographic, or signal-based structure. Despite its simplicity, it provides a useful sanity check for how harmful uninformed clustering can be.

Administrative boundaries. This method defines spatial clusters based on political divisions, in our case, the states of the New England region. The number of clusters C is thus fixed by the number of states, and temporal clustering proceeds as usual. This approach is common in applied energy planning contexts, where jurisdictional boundaries often guide model structure. While interpretable and convenient, such boundaries may not align with network structure or operational behavior, leading to distortions in aggregated models.

K -medoids on node positions. Here, clustering is performed using the standard k -medoids algorithm, with Euclidean distance between geographic coordinates. This is equivalent to the "spatial k -means" baseline studied in [12]. It assumes that spatial proximity implies similarity in load, generation, and flexibility patterns. While intuitive, this assumption can fail in practice, especially in systems with high renewable penetration.

In addition to these naive methods, we benchmark against two more sophisticated approaches drawn from recent literature:

K -medoids on raw time series. This method clusters nodes based on their raw hourly profiles (e.g., demand or capacity factors), using Euclidean distance as the similarity metric. It mirrors one of the leading methods in [12], where Ward's method was applied to raw profiles with an electrical topology constraint. Our version uses k -medoids instead of hierarchical clustering and does not enforce network-aware constraints, due to data limitations described below.

K -medoids on average capacity factors. Similar to the previous method, but using aggregated features, specifically, average capacity factors over time, as the clustering signal. This smooths temporal fluctuations and focuses on resource availability patterns. It approximates the setup used in [12] when clustering based on mean resource characteristics.

The paper by Frysztacki et al. [12] introduces a particularly influential spatial clustering strategy based on Ward's method, adapted to the context of electricity networks. Ward's method is a hierarchical agglomerative algorithm that greedily merges the pair of clusters whose union causes the smallest increase in within-cluster variance. At each step, it minimizes the following objective:

$$\Delta(A, B) = \frac{|A| \cdot |B|}{|A| + |B|} \cdot \|\mu_A - \mu_B\|^2$$

where $|A|$ and $|B|$ denote the sizes of the clusters, and μ_A , μ_B their centroids. This expression quantifies the increase in total variance resulting from the merge. To adapt this method to power systems, the authors introduce a connectivity constraint. Two clusters may only be merged if they are connected by at least one transmission line. This ensures that aggregated zones are physically meaningful and maintain consistency with the underlying network topology. It is worth noting, even though not done in their implementation, that the approach could also be extended to incorporate topology-aware distance metrics, such as electrical admittance or impedance, rather than relying purely on Euclidean distances in signal space. Such extensions would allow the clustering to capture both resource behavior and the structural properties of the grid.

However, our case study operates under a greenfield assumption. As explained in Chapter 5, we do not assume access to a predefined transmission network or electrical topology. This means the key constraint used in topology-aware Ward clustering, merging only connected nodes, cannot be applied. Our benchmark implementations thus rely on standard k -medoids without topological constraints. Nonetheless, our methods remain comparable in spirit to those in [12], as both rely on clustering node-level features derived from operational or resource profiles.

Altogether, these five benchmarks provide a representative spectrum of aggregation strategies, ranging from uninformed spatial heuristics to temporally driven spatial clustering. By comparing our pipeline against this range of methods across multiple aggregation regimes, we can assess the robustness and relative performance of our approach under realistic modeling conditions.

Chapter 5

Case Study Data

5.1 Construction of Network Data

This section walks through how we built our spatial and temporal inputs. We begin from a fine-grained dataset that includes wind and solar capacity factors across 385 locations in New England, based on high-resolution generation modeling from [48]. We pair the generation data with demand, disaggregated from 67 county-level records. However, this full-resolution dataset is too large for our evaluation framework, which requires repeated deaggregation experiments. We therefore construct a coarse 17-node version of the network that enables tractable experimentation.

5.1.1 High-Resolution Inputs: 385 Nodes

We begin by describing the renewable generation dataset, which provides hourly capacity factors at 385 nodes. The starting point for the network data construction is high-resolution renewable generation data developed in [48], which provides hourly wind and solar capacity factors across 385 spatial nodes in New England over several years. These nodes, defined on a $0.25^\circ \times 0.25^\circ$ grid, each correspond to a square cell that spans 0.25° in both latitude and longitude, offering a detailed view of renewable potential across the region. For each node, the wind and solar profiles are encoded as hourly time series $\{w_t^i\}_{t=1}^T$ and $\{s_t^i\}_{t=1}^T$.

To enable system-level modeling, we pair this with hourly demand data from 67 counties in New England, scaled to 2050 projections. We obtain hourly electricity demand from historical county-level data in [49], originally reported for the year 2013 and covering 67 counties¹. To reflect projected electrification by 2050, the demand is uniformly scaled by a factor of 1.6. This multiplier is motivated by long-term growth trends observed in ISO New England's forecasts. The 2024 Capacity, Energy, Loads, and Transmission (CELT) report [50] projects an increase in electricity use from 119,179 GWh in 2024 to 140,001 GWh by 2033, representing a cumulative growth of 17.5% over 9 years. Assuming compound growth, this corresponds to an average annual increase of approximately 1.8%, since $(1.018)^9 \approx 1.175$. Extending this trajectory to 2050 yields a cumulative growth of roughly 60%, as $(1.018)^{26} \approx 1.6$.

¹In 2025, New England comprises 68 counties due to administrative changes.

This introduces a spatial mismatch: while generation is given at the grid level (385 nodes), demand is reported at the county level (67 nodes). To bridge this gap, we develop a disaggregation method that distributes county-level demand down to generation nodes based on population weights. Population data is provided as a dense point-based dataset, extracted from the 2023 Kontur Population Dataset [51], where each point corresponds to a 400m-resolution hexagon with associated population and geographic coordinates. Each population point is first assigned to a county using spatial joins. The point is then snapped to its nearest grid node using Haversine distance. This results is that a single node may receive population contributions from multiple counties, depending on spatial proximity. We compute the total population falling into each county-node pair, and define a weight matrix by normalizing these values within each county. These weights are then used to allocate the county-level demand to the corresponding nodes. As a result, the demand assigned to each node may reflect contributions from multiple counties. This approach ensures that total demand is conserved at each timestep, while respecting both geographic and demographic distribution.

This yields a unified dataset with wind, solar, and demand time series at each of the 385 nodes, used as the basis for fine-resolution analysis. Figure 5.1 illustrates the spatial layout of the 385 grid nodes and the corresponding distribution of disaggregated demand.

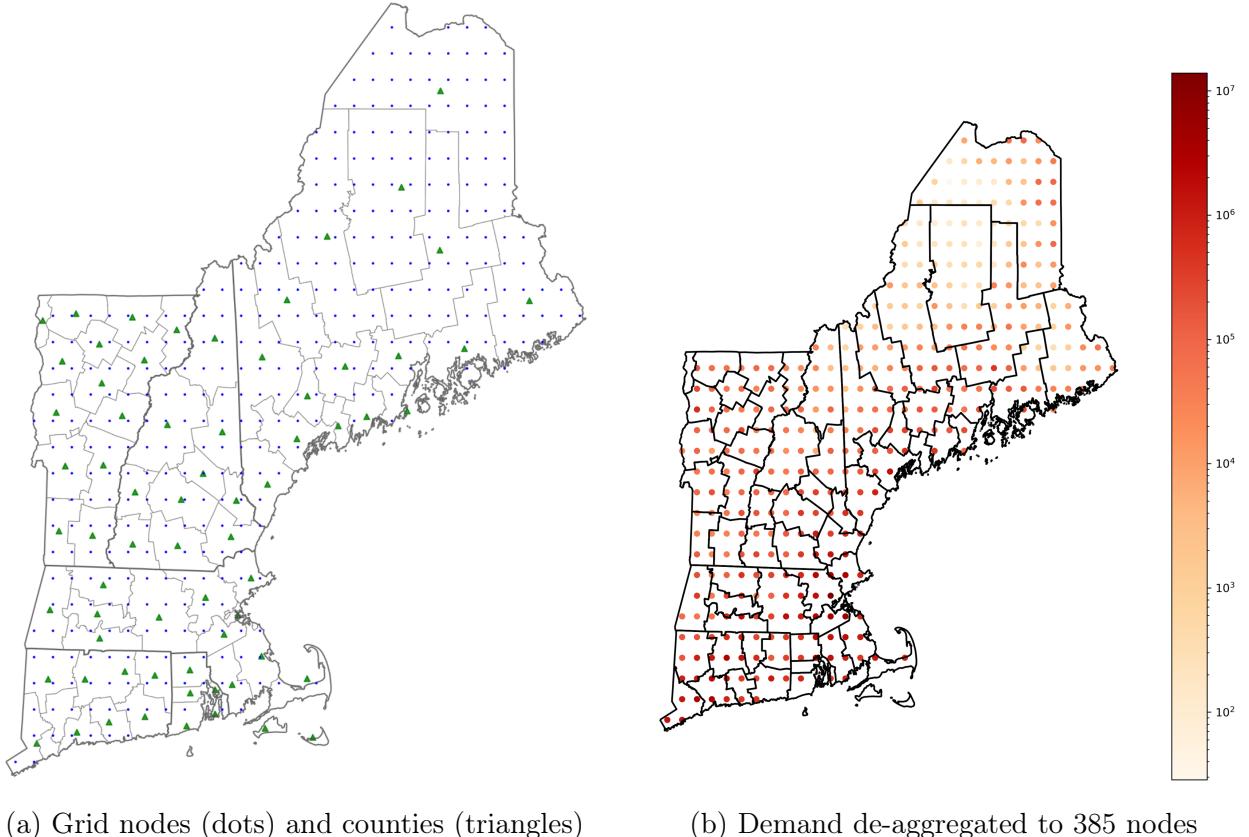


Figure 5.1: Construction of fine-resolution network over New England

5.1.2 Downscaling for Model Evaluation: 17 Nodes

The full 385-node dataset is too large for repeated experiments. Our evaluation framework relies on spatial deaggregation through regional re-optimization (Chapter 4), which solves small capacity expansion problems within each cluster to distribute investments. This process becomes too heavy to run hundreds of times. An alternative could have been the Minimal Excess Electricity method proposed by [38], discussed in Chapter 4 and Section 4.2.2, which offers a lighter-weight heuristic for spatial allocation by minimizing local renewable oversupply. However, that method requires an existing transmission topology to estimate inter-nodal transfer potential, something not available in our greenfield setting. We therefore adopt a coarser 17-node network that allows tractable, repeatable deaggregation validation workflows. This reduced representation, consistent with the aggregation scheme used in [49], serves as the basis for all planning and evaluation models used in this work.

The mapping from fine to coarse resolution is based on geographic proximity. Each fine-resolution node is assigned to its nearest coarse node using a k -nearest neighbor search. Once this mapping is in place, we aggregate the time series data. For demand, aggregation is additive, all demand at fine nodes belonging to a coarse zone is summed. For wind and solar, where averaging might obscure important temporal features, we instead select a representative medoid time series from within each zone. This medoid is the time series that minimizes the total distance to others in its zone and better preserves local temporal patterns.

This results in a compact 17-node dataset, where each node has an associated demand profile and representative wind and solar time series. To define the transmission investment space for GTEP, we assume a fully connected greenfield topology. That is, any pair of coarse nodes may be connected, giving a total of $|\mathcal{L}| = \binom{17}{2} = 136$ potential transmission links. Since our model uses a transport formulation, no electrical parameters such as line susceptance are required. Nonetheless, the dense network connectivity significantly increases the complexity of optimization.

5.2 Planning Model Parameters

This section describes the modeling assumptions and parameter inputs used in the GTEP model. We begin by clarifying key modeling choices and system design assumptions. We then detail the parameter values used for technologies, costs, and constraints, along with their sources and derivation methods. Finally, we describe a rescaling step applied to all data for improved numerical stability during optimization.

5.2.1 Modeling Assumptions

The GTEP model used in this study includes a minimal but representative set of technologies: utility-scale photovoltaic (PV), onshore wind, natural gas combined-cycle turbines (CCGT), battery storage, and transmission. No hydroelectric or nuclear technologies are included, reflecting a testbed-style design where only a few major asset classes are used to explore aggregation effects. All assets are sited and sized in a greenfield context, with no pre-existing infrastructure or topology. As a result, we do not model electrical quantities such

as susceptance or admittance. Instead, the network is modeled as a transport graph, with each of the 17 nodes connected to all others by potential transmission links. This fully connected topology enables long-range transfers but comes at a cost as the number of candidate lines scales quadratically, creating a computational burden. In total, $\binom{17}{2} = 136$ links are considered.

5.2.2 Data Sources and Parameter Construction

We extract all techno-economic assumptions from established planning datasets and processing pipelines. The complete set of parameter values is reported in Appendix A.

Generation Technology Selection. We adopt representative technologies from the National Renewable Energy Laboratory (NREL) 2024 Annual Technology Baseline (ATB) [4]: NG 2-on-1 Combined Cycle (F-Frame) 95 % CCS for CCGT, Utility PV - Class 5 for PV, Land-Based Wind - Class 4 - Technology 1 for wind, and Utility-Scale Battery Storage - 4Hr for storage, with a maximum storage duration (energy-to-power ratio) of $\tau = 4$ hours. These selections follow the ATB's *Representative Value* guideline, representing typical near-term deployments based on commonly built plant configurations. To derive techno-economic input parameters (e.g., CAPEX, fixed and variable O&M, capacity factor), we filter the ATB dataset for the *Moderate* scenario, the *Market+Policies* financial assumptions case, a projection year of 2050, and a cost recovery period of 30 years. The *Moderate* scenario represents mid-range expectations for technology cost reductions, grounded in bottom-up engineering models from 2022 to 2030 and moderate learning rates out to 2050. The *Market+Policies* case reflects average financing terms and the impacts of current federal policies (e.g., tax credits and regulatory treatment), contrasting with the more limited *R&D-only* financial sensitivity case. The resulting filtered values provide a consistent baseline reflective of plausible mid-range cost and performance evolution.

Generation CAPEX. For each selected technology, we extract capital expenditure (CAPEX) values from the filtered ATB dataset [4] and compute annualized investment costs for use in the planning model. The ATB supports two approaches for this computation depending on the technology: either it provides a precomputed fixed charge rate (FCR), or we can apply a capital recovery factor (CRF) manually. Both approaches yield an annualized CAPEX, but differ in what financial components are explicitly modeled. The FCR encapsulates financing assumptions into a single multiplier applied to CAPEX. It accounts for the weighted average cost of capital (WACC), the debt-equity structure, the cost recovery period, and relevant tax treatments, producing a fully loaded annualized cost without requiring the modeler to specify these details individually. The annualized investment cost is given by:

$$\text{Annualized CAPEX} = \text{CAPEX} \times \text{FCR}$$

When no FCR is provided, we compute annualized CAPEX using a CRF factor, which amortizes investment over N years at a fixed discount rate r , without incorporating tax effects or financing structures:

$$\text{CRF}(r, N) = \frac{r}{1 - (1 + r)^{-N}} \quad \Rightarrow \quad \text{Annualized CAPEX} = \text{CAPEX} \times \text{CRF}$$

In our case, we assume $r = 0.06$ and $N = 30$. For wind and solar, the ATB provides an FCR, allowing direct computation of annualized CAPEX. For CCGT and storage, which lack an FCR entry under our selected filters, we apply the CRF method described above.

Other Generation Parameters. Additional properties such as ramp rates r_g , nameplate capacity limits \bar{P}_g , and heat rate levels for CCGT are sourced from the ATB [4].

Transmission Line CAPEX. The ATB does not provide cost estimates for transmission infrastructure, and line-level capital costs are generally difficult to source consistently across voltage levels and configurations. To construct distance-based estimates, we begin by analyzing typical buildout patterns observed in recent ISO-NE Regional System Plans [52], which suggest that most planned transmission expansions involve 230kV and 345kV HVAC lines, with an approximate mix of 90% 345kV and 10% 230kV.

We use capital cost estimates from the Western Electricity Coordinating Council (WECC) report [53], which provides \$959,700 per mile for 230kV single-circuit (SC) lines rated at 400MW, and \$1,343,800 per mile for 345kV SC lines rated at 750MW. To obtain a distance- and capacity-normalized value, we compute a weighted average cost per MW-mile using Table 2-1 (costs) and Table 2-8 (ratings) of the WECC report:

$$\text{Cost}_{\text{avg}} = \frac{0.9 \times 1,343,800}{750} + \frac{0.1 \times 959,700}{400} \approx \$1,853/\text{mi/MW}$$

These values are expressed in 2014 dollars. To align with our 2050 cost basis, we escalate using an assumed 2% annual inflation rate over 36 years $\$1,853 \times (1.02)^{36} \approx \$3,780/\text{mi/MW}$. We then annualize capital costs using the CRF method, assuming a 40-year recovery period and an 8% discount rate, following common practice in long-term transmission planning studies such as SPP's ITP20 [5] ($N = 40$, $r = 0.08$).

Transmission Line Rating \bar{T}_ℓ . We adopt a simplified, length-dependent rule to assign thermal limits to each possible line. This approach reflects a common simplification in planning models where transmission capacity scales linearly with physical length. Specifically, we assume a fixed transfer capacity per mile and define the thermal limit of each line as:

$$\bar{T}_\ell = \gamma \cdot d_\ell$$

where \bar{T}_ℓ is the transfer limit (in MW), d_ℓ is the line length (in miles), and γ is a fixed capacity-per-mile coefficient (in MW/mi). To set γ , we draw from the WECC transmission study [53]. We use a representative HVAC line capacity of 715MW, computed as a weighted average of typical line ratings based on the voltage mix described in the previous paragraph, and we divide it by the median line length across all candidate transmission links in our coarse network, denoted $\text{median}(d_\ell)$: $\gamma = 715/\text{median}(d_\ell)$.

Fuel Prices φ_g . Delivered natural gas prices are taken from the U.S. Energy Information Administration's Annual Energy Outlook 2025 [54], using 2050 projections for the electric power sector from Table 3 (Energy Prices by Sector and Source, Reference case). Prices are expressed in real 2024 dollars per MMBtu and converted to \$/MWh using a heat rate from the NREL ATB.

Storage Efficiency $\eta^{\text{ch}}, \eta^{\text{dis}}$ and SOC e_c^0 . We assume symmetric round-trip efficiency RTE = 0.85, as reported in the NREL ATB, yielding $\eta^{\text{ch}} = \eta^{\text{dis}} = \sqrt{0.85} \approx 0.922$. The initial state of charge is set to $e_c^0 = 0.5$, a common assumption in long-term energy storage planning that provides a neutral baseline, allowing both charging and discharging flexibility, without biasing the model in either direction at the start.

Renewable Resource Limits Γ_g^{\max} . To constrain investment decisions for variable renewable generation, we impose resource availability limits reflecting the maximum buildable capacity of wind and solar technologies in ISO New England. These limits are derived from supply curves provided by the ReEDS 2.0 model developed by NREL² [55]. A supply curve in this context quantifies the total developable capacity (in MW) available at varying levels of resource quality, after applying land-use exclusions such as protected areas, steep slopes, or dense forests. ReEDS classifies resource quality into discrete bins (e.g., Classes 1–5 for solar), and aggregates potential by balancing authority (BA) region.

For our analysis, we select the "reference" siting scenario, which applies standard exclusion rules and is most consistent with long-term planning assumptions. We retain only the six BA regions comprising ISO New England (codes p129–p134) and restrict capacity estimates to high-quality classes (1–6 for wind and 1–3 for solar). Summing across these filters yields total buildable potentials in ISO-NE. These values serve as hard upper bounds on technology-specific capacity investment in the GTEP model.

Penalty for Shedding ψ . Unserved demand is discouraged via a penalty of $\psi = \$10,000/\text{MWh}$, in line with previous studies (e.g., [49]).

A summary of all key parameter values used in the model is provided in Table 5.1, with all units expressed after applying the scaling procedure detailed in the next subsection.

5.2.3 Scaling for Numerical Stability

To ensure numerical stability and improve solver performance in Gurobi, all power values are scaled from MW to GW ($\times 10^{-3}$) and all cost values from \$ to M\$ ($\times 10^{-6}$). Derived parameters such as cost per line and fuel cost per GWh are recomputed using scaled units. Transmission capacities are also updated. This preprocessing reduces the spread of coefficients in the MILP formulation and helps avoid numerical instability during optimization.

²See <https://github.com/NREL/ReEDS-2.0>, directory /inputs/supply_curve/, files wind-ons_supply_curve-reference_ba.csv and upv_supply_curve-reference_ba.csv, accessed August 2025.

Table 5.1: Summary of key GTEP model parameters (units are after scaling)

| Symbol | Unit | Description |
|---------------------|----------------|--|
| c_g | M\$/GW/yr | Annualized CAPEX for generation technology g |
| c_c | M\$/GW/yr | Annualized CAPEX for storage (power capacity) |
| τ | hour | Storage duration (energy-to-power ratio) |
| η^{ch} | [0–1] | Charging efficiency of storage |
| η^{dis} | [0–1] | Discharging efficiency of storage |
| e_c^0 | [0–1] | Initial state of charge (SOC) of storage |
| r_g | [0–1] | Ramp rate for dispatchable generator g |
| \bar{P}_g | GW | Nameplate capacity of generator g |
| Γ_g^{\max} | GW | Resource potential limit for renewable g |
| \bar{T}_ℓ | GW | Thermal limit (line rating) of line ℓ |
| d_ℓ | miles | Length of line ℓ |
| γ | GW/mile | Capacity-per-mile transfer coefficient |
| c_ℓ | M\$/GW/mile/yr | Annualized CAPEX of transmission line ℓ |
| φ_g | M\$/GWh | Fuel cost for generator g (e.g. CCGT) |
| ψ | M\$/GWh | Penalty for unserved demand (shedding) |

Chapter 6

Results

This chapter presents the main experimental results of our study. We aim to understand whether our aggregation pipeline can produce reliable and efficient input data for CEPs, and under which conditions it performs best. We address this question through three complementary investigations. First, we analyze how the choice of feature weights in the spatial clustering process affects solution quality (§6.1). This allows us to isolate which structural signals matter most for downstream planning. Second, we examine the trade-off between compression and fidelity by exploring the Pareto frontier over spatial and temporal aggregation levels (§6.2). Finally, we compare our pipeline to a set of benchmark methods from the literature to assess its practical performance against standard heuristics and data-driven baselines (§6.3).

All experiments are conducted on the Engaging Cluster at MIT, a large-scale computing platform equipped with heterogeneous CPU and GPU nodes. Our workloads are submitted via SLURM using dedicated batch scripts, and all GTEP instances are solved using Gurobi version 10.0.2. Jobs are run on nodes equipped with 2x48-core AMD EPYC 9474F processors and 376 GB of memory.

6.1 Feature Weight Analysis

We begin by exploring how the choice of feature weights in the spatial clustering step impacts downstream model performance. As described in Section 4.3.1, we evaluate the impact of spatial feature design by testing multiple weight vectors \mathbf{w} across several aggregation levels (C, D) drawn from a fixed set \mathcal{S} . In this chapter, we consider 50 distinct weight configurations (Appendix B) and evaluate each over the set $\mathcal{S} = \{(5, 7), (5, 14), (5, 21), (10, 7), (10, 14), (10, 21)\}$. The weight grid includes equal-weight baselines, single-feature tests, structured combinations, targeted sweeps, and randomized samples to ensure broad coverage of the feature space. The full weight grid is detailed in Appendix B.

To understand which weightings lead to better performance, we compute the deaggregated objective value for each (\mathbf{w}, C, D) tuple and apply the performance metrics introduced in Section 4.3.1. These include total score, mean score, mean objective gap, and standard deviation of the gap.

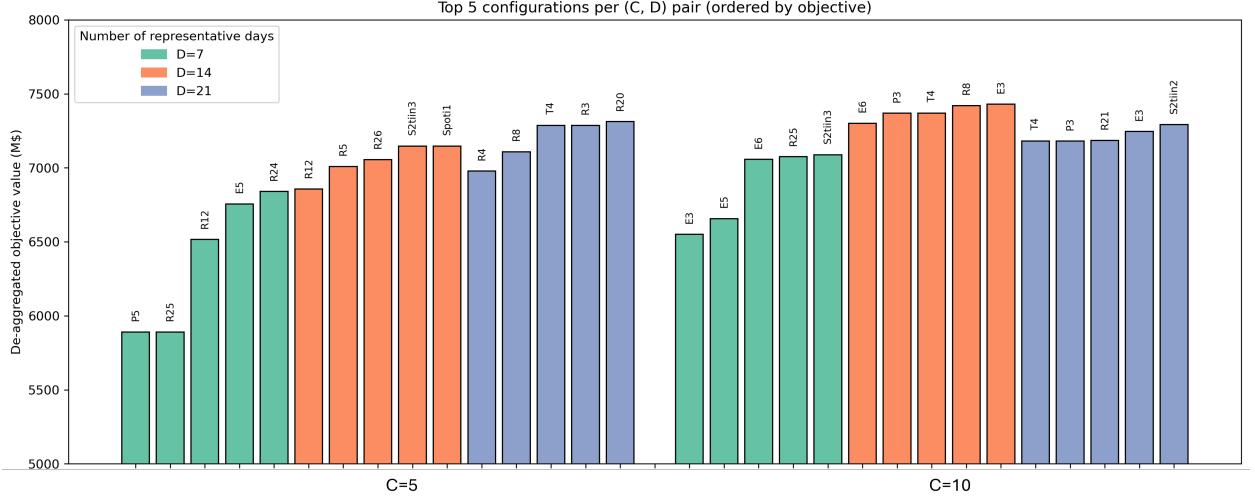


Figure 6.1: Top 5 configurations by total deaggregated cost (millions of \\$) for each (C, D) pair.

A first global view is given in Figure 6.1, which displays the top five performing configurations for each aggregation level. Bars are grouped by (C, D) pair, with C on the x-axis and D represented by color. Within each group, configurations are ordered by their deaggregated total cost, defined as the sum of investment and operational costs, with lower values indicating better performance. Each bar is annotated with the name of the corresponding weight configuration, which can be cross-referenced with its full vector in Appendix B. We observe that while some configurations appear in the top five across multiple (C, D) pairs, the identities of the best-performing weightings vary significantly from one setting to another. Ideally, we would expect certain feature combinations to generalize well across aggregation levels, indicating strong and robust signals for guiding spatial clustering. The observed variability suggests that no single design dominates consistently. This motivates a more systematic analysis to determine whether some configurations perform reliably well across the full set \mathcal{S} , a question that will be addressed using the performance metrics introduced in Section 4.3.1.

To better understand what drives strong performance, we isolate the contribution of individual features. Table 6.1 shows how the pipeline performs when only one feature class is used (ablation analysis) using the performance metrics. Duration curves and correlation features emerge as the most effective: configurations E3, E5, and E6 all achieve good scores. Features like raw time series and spatial coordinates perform poorly in isolation.

Table 6.1: Composite performance metrics for ablation configurations E1–E6.

| ID | Mean Rank | Mean Score | Total Score | Mean Gap | Std. Gap |
|--------------------|--------------|--------------|-------------|----------|--------------|
| E3 (duration only) | 10.17 | 0.267 | 1.6 | 0.059 | 0.066 |
| E6 (inter-corr.) | 22.00 | 0.267 | 1.6 | 0.095 | 0.078 |
| E5 (intra-corr.) | 18.33 | 0.200 | 1.2 | 0.066 | 0.044 |
| E2 (time series) | 25.83 | 0.000 | 0.0 | 0.101 | 0.070 |
| E4 (ramp duration) | 21.83 | 0.000 | 0.0 | 0.095 | 0.079 |
| E1 (position only) | 39.17 | 0.000 | 0.0 | 0.116 | 0.089 |

This pattern is reinforced by a correlation analysis. Figure 6.2 quantifies the relationship between each feature's weight and global performance metrics. We aim to maximize correlation with mean score and total score, both indicators of strong performance, and minimize correlation with mean objective gap, which measures deviation from the best-known cost. Among all features, inter-cluster and intra-cluster correlation weights stand out with strong positive correlation to the score metrics and negative correlation to the gap, indicating that emphasizing structural relationships across nodes tends to yield better outcomes. In contrast, geographic position and raw time series amplitude show the opposite trend as they are negatively correlated with scores and positively correlated with gap, suggesting that these features degrade performance when overemphasized. Duration curves and ramping duration curves occupy a middle ground, with weaker and less consistent correlations that may reflect case-specific utility.



Figure 6.2: Correlation between feature weights and a few performance metrics.

Figure 6.3 further illustrates these trends by comparing the distribution of feature weights in the top 10 configurations against those across the entire set of 50. For both position and time series amplitude, the median weights in the top 10 are significantly lower than in the full population, confirming their limited usefulness. Intra-cluster correlation, by contrast, shows both a higher median and greater spread among top performers, reinforcing its importance. Inter-cluster correlation exhibits higher mean weight in the top group but not a higher median, suggesting it plays a helpful, though perhaps more variable, role. This distributional comparison supports the idea that robust configurations consistently avoid over-reliance on raw or locational signals and instead prioritize features that reflect structural alignment.

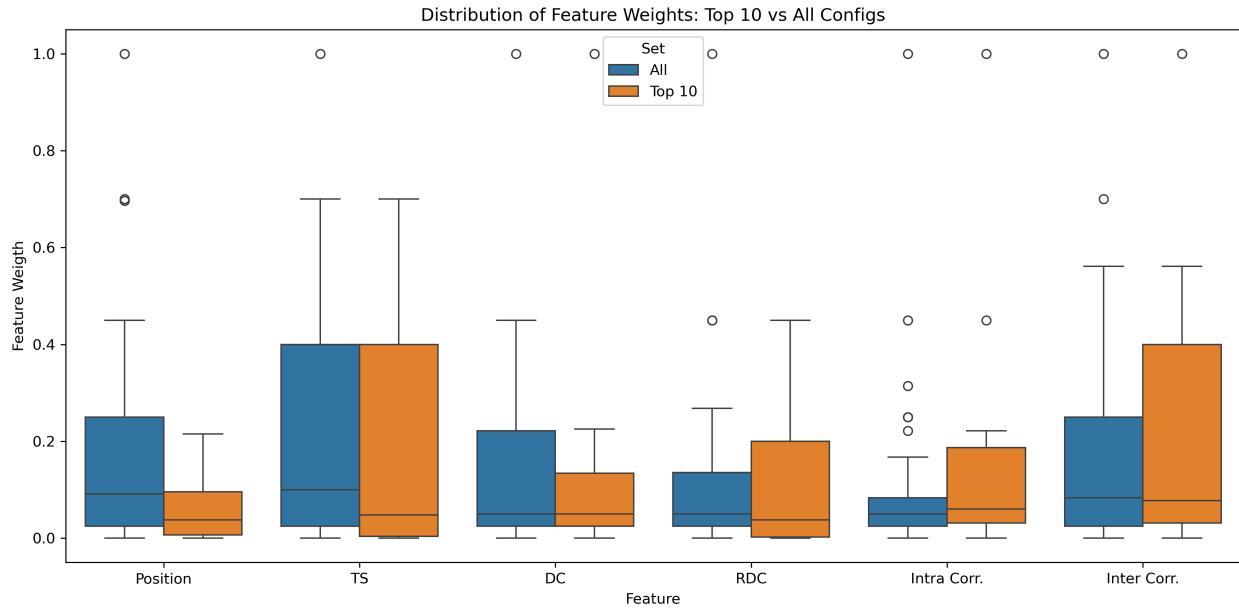


Figure 6.3: Distribution of feature weights for all configurations vs. the top 10 performers.

Table 6.2: Summary of representative configurations across all (C, D) .

| ID | Mean Rank | Mean Score | Total Score | Mean Gap | Std. Gap | Verdict |
|-----|--------------|--------------|-------------|----------|--------------|-------------|
| T4 | 18.33 | 0.400 | 2.4 | 0.090 | 0.098 | Best score |
| R12 | 13.00 | 0.267 | 1.6 | 0.057 | 0.047 | Good |
| R25 | 13.00 | 0.233 | 1.4 | 0.047 | 0.030 | Most stable |
| E3 | 10.17 | 0.267 | 1.6 | 0.059 | 0.066 | Best Rank |
| P3 | 28.83 | 0.300 | 1.8 | 0.111 | 0.114 | Unstable |
| E1 | 39.17 | 0.000 | 0.0 | 0.116 | 0.089 | Worst Rank |
| R2 | 30.50 | 0.000 | 0.0 | 0.114 | 0.082 | Bad |

Table 6.3: Feature composition of best configurations (\checkmark emphasized, \sim moderate).

| ID | Position | TS | DC | RDC | Intra | Inter | Verdict |
|-----|----------|--------------|--------|--------------|--------------|--------------|-------------|
| T4 | \sim | \checkmark | \sim | \checkmark | \sim | \checkmark | Best score |
| R12 | \sim | | | \checkmark | \checkmark | \sim | Good |
| R25 | | | | | \checkmark | \checkmark | Most stable |

The previous analyses focused on individual feature effects. We now turn to feature combinations to understand how they interact in high-performing configurations. Tables 6.2 and 6.3 summarize the key patterns. While several top configurations include ramping duration curves (RDC), their presence alone does not distinguish strong performers. What stands out instead is that every top configuration assigns substantial weight to either inter-cluster correlation, intra-cluster correlation, or both. This confirms the central role of structural coherence in driving good outcomes. Among these, R25 is the most structurally focused design, assigning nearly all its weight to correlation-based features. Its reliable performance across aggregation levels further supports the value of this emphasis.

In summary, our analysis shows that structural signals such as node correlations are critical for building reliable aggregates. Naively intuitive features like geography or raw time series can degrade performance unless used in combination with more informative signals.

6.2 Pareto Frontier

We now fix a weight vector \mathbf{w} and explore how aggregation level affects model performance. This analysis isolates the impact of spatial and temporal resolution by holding feature emphasis constant. We select configuration P4, while it is not the top performer overall, it offers consistent behavior and serves as a clean baseline to study the role of C and D .

We run the full pipeline on all (C, D) pairs and evaluate the deaggregated cost. Figure 6.4 visualizes the resulting Pareto frontier, and Table 6.4 lists the non-dominated configurations. The results reveal a strong asymmetry: increasing C (spatial clusters) dramatically improves performance, while increasing D (representative days) has weaker effects unless C is already high. This structure is made clearer in Figure 6.5, which projects the results onto 2D plots. The steep slope along C reflects the dominant role of spatial detail. Temporal refinement offers marginal gains unless spatial granularity is sufficient.

Table 6.4: Configurations on the Pareto frontier under fixed \mathbf{w} , ordered by objective value.

| Objective | C (clusters) | D (days) |
|-----------|----------------|------------|
| 5926.10 | 10 | 3 |
| 6521.22 | 8 | 3 |
| 7213.51 | 3 | 30 |
| 7283.13 | 5 | 7 |
| 7311.99 | 3 | 20 |
| 7314.13 | 3 | 15 |
| 7323.55 | 3 | 7 |
| 7433.99 | 3 | 3 |

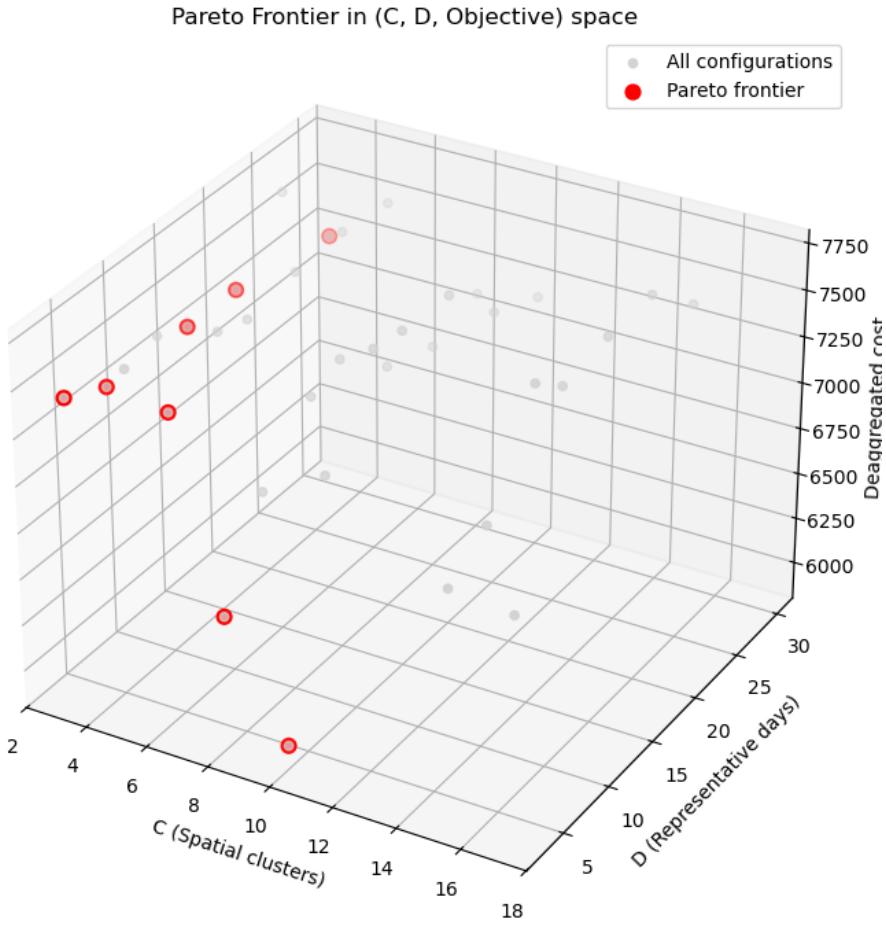


Figure 6.4: Pareto frontier in $(C, D, \text{Objective})$ space under fixed feature weights. Red points are non-dominated configurations.

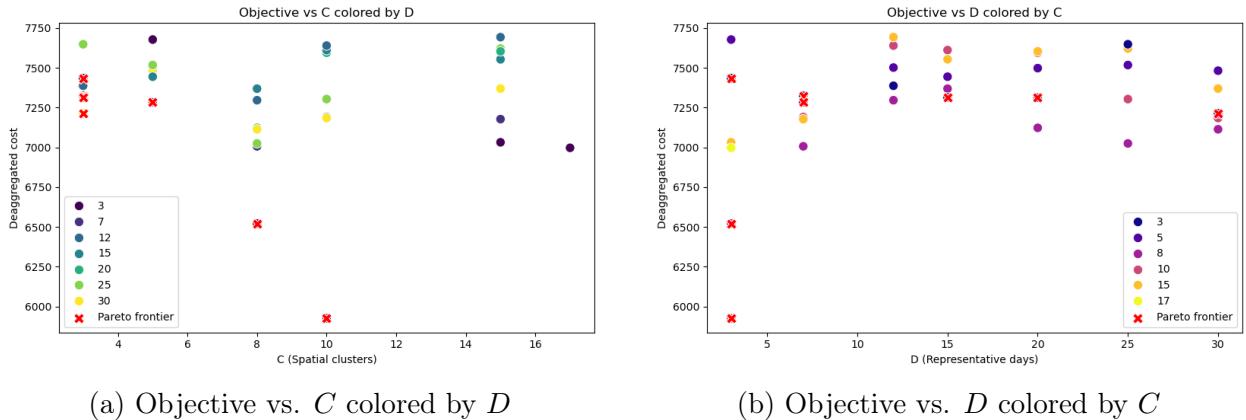


Figure 6.5: 2D projections of objective vs. aggregation level. Pareto frontier configurations are marked by red crosses.

These results show that spatial compression introduces more severe distortion than temporal compression. Carefully selecting C can yield compact yet accurate models, while over-simplification along either axis leads to degraded performance.

6.3 Benchmarks

Finally, we evaluate our pipeline against five benchmark methods described in Section 4.3.2. These include three naive heuristics (random, position-based, and administrative boundaries) and two signal-driven approaches (clustering based on raw series or average capacity factors). All methods are tested under three aggregation regimes: a low spatial, high temporal setting ($C = 3, D = 30$), a balanced setting with moderate spatial and temporal resolution ($C = 6, D = 12$), and a high spatial, low temporal setting ($C = 10, D = 3$). The political boundaries benchmark is only evaluated at $C = 6$, corresponding to the six states in the New England region.

We select configuration R25 as a representative of our pipeline due to its stable performance in Section 6.1. As shown in Table 6.5, our pipeline ranks first or second in all regimes, outperforming all other methods in both robustness and peak quality. This highlights its consistent ability to generalize across different compression tradeoffs. By contrast, competing methods such as average CF clustering and raw time series clustering show fluctuating ranks, occasionally performing well but lacking stability. Simpler methods like random clustering or geometry-based k -medoids suffer from even greater variability, with no clear dominance across settings. Note that the random method is excluded from the high spatial, low temporal configuration ($C = 10, D = 3$) because the deaggregation process did not converge within a 12-hour runtime limit. Results for this setting are therefore shown without it.

Table 6.5: Performance ranking of each method under different aggregation regimes (rank 1 is best).

| Method | Low spatial, high temp. | Medium resolution | High spatial, low temp. |
|------------------------|-------------------------|-------------------|-------------------------|
| Our pipeline | 1 | 1 | 2 |
| Avg CF (k -medoids) | 2 | 2 | 4 |
| States | — | 5 | — |
| Geo (k -medoids) | 5 | 6 | 3 |
| Random | 3 | 3 | — |
| Raw profiles | 4 | 4 | 1 |

Figure 6.6 complements these rankings by showing the absolute deaggregated objective values achieved by each method in each configuration. The bars are sorted by performance within each setting, allowing us to compare the relative effectiveness of all methods. In the ($C = 3, D = 30$) configuration, all methods yield higher costs due to coarse spatial resolution, and the differences between them remain small. As spatial resolution increases in ($C = 6, D = 12$) and ($C = 10, D = 3$), performance diverges more clearly. Our pipeline achieves some of the lowest total costs in both of these settings, showing that its advantage becomes more pronounced when sufficient spatial granularity is available.

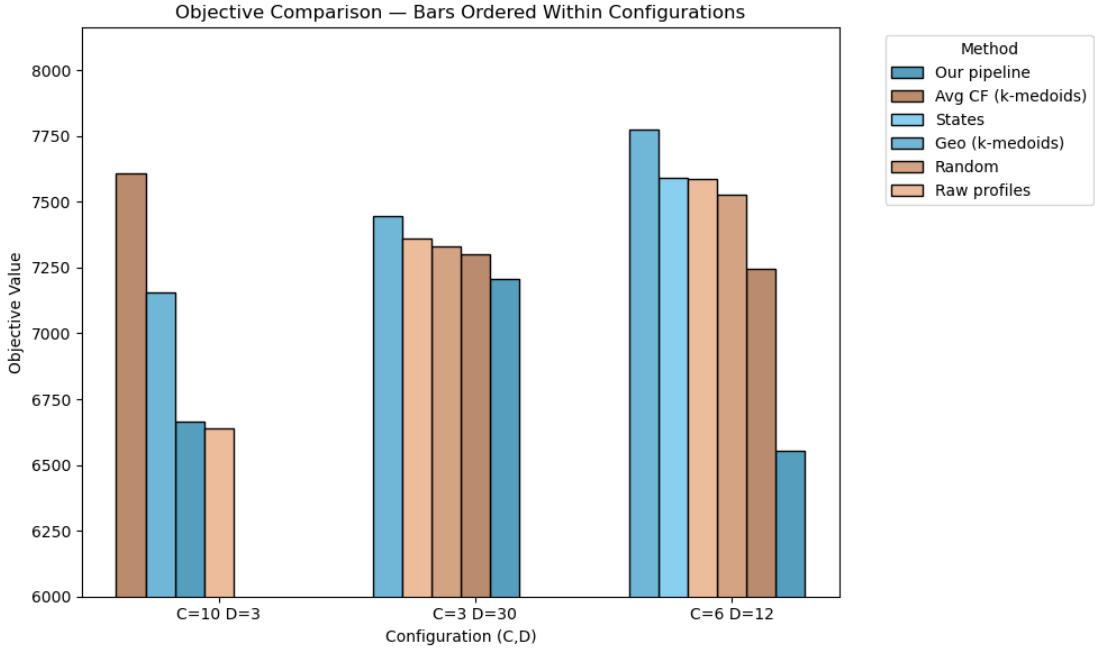
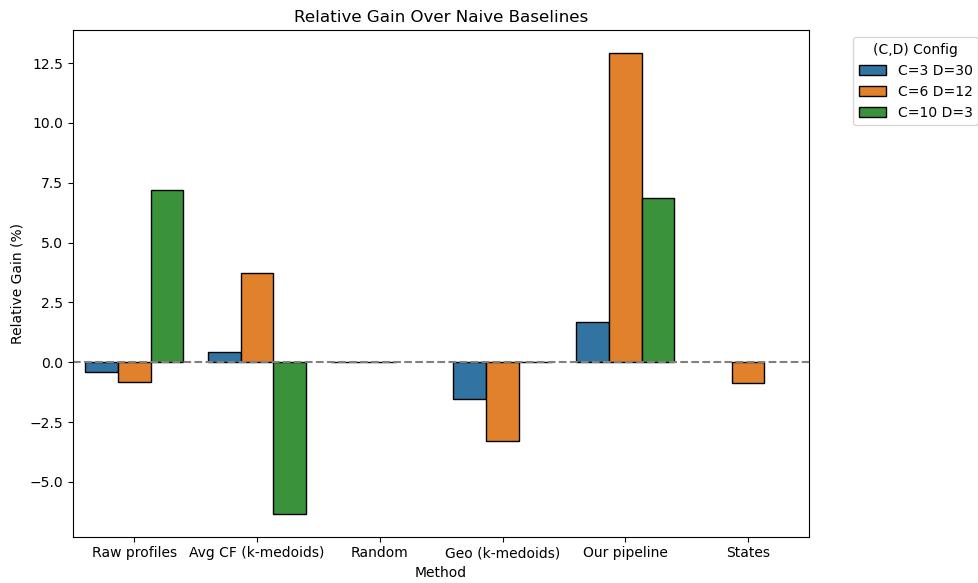


Figure 6.6: De-aggregated cost comparison.

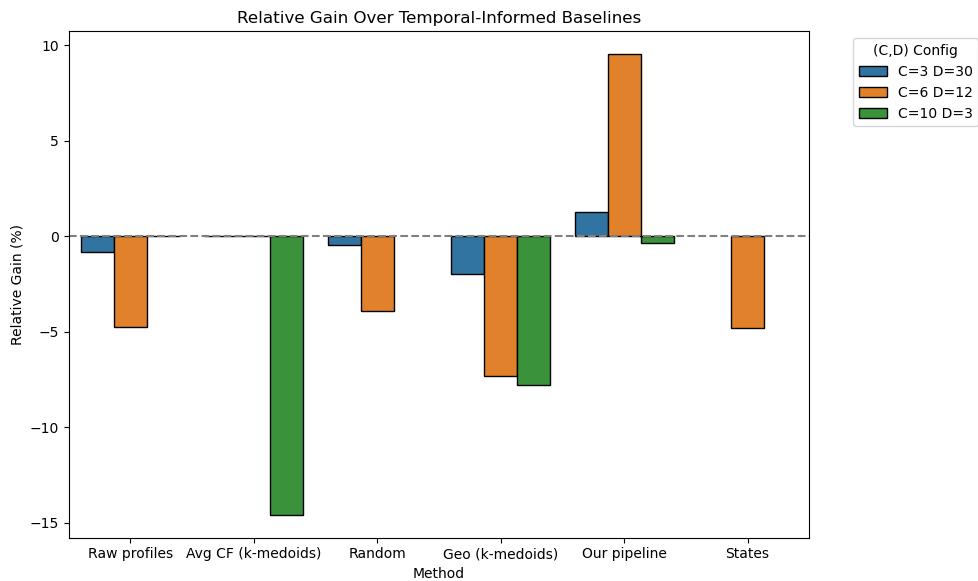
To make these differences more visible, Figure 6.7 quantifies the relative improvement of our pipeline over two groups of baselines. The top panel compares its performance to naive spatial aggregation methods, while the bottom panel compares it to the signal-based clustering approaches. The results are consistent. In the low-resolution regime ($C = 3, D = 30$), where all methods perform poorly, the relative gain of our pipeline is modest, about 1% compared to both baseline groups. However, in the more realistic mid-resolution setting ($C = 6, D = 12$), our method outperforms naive baselines by roughly 13% and signal-based ones by nearly 10%. Even in the high-spatial regime ($C = 10, D = 3$), where the raw time series baseline performs unusually well, our pipeline still maintains a 7% advantage over naive methods and near parity with the temporal benchmark.

Finally, Figure 6.8 synthesizes these results across all three configurations using a normalized radar plot. Each axis corresponds to one (C, D) setting, and scores are scaled such that 100% marks the best method in each case. The profile for our pipeline is smooth and close to the outer edge in all directions, reflecting strong and balanced performance. In contrast, other methods show jagged profiles with noticeable dips, often performing well in one setting but poorly in another. This visual reinforces the robustness of our pipeline compared to more volatile alternatives.

Altogether, these experiments demonstrate that our method, when equipped with the R25 configuration, consistently delivers strong performance across a range of spatial and temporal settings. It achieves top rankings, maintains low objective costs, and outperforms both heuristic and data-driven clustering baselines. These results underscore the value of combining structural signal emphasis with moderate aggregation resolution, and highlight our pipeline’s practical utility in guiding model simplification without sacrificing downstream feasibility or optimality.



(a) Relative gain over naive baselines



(b) Relative gain over temporal baselines

Figure 6.7: Relative objective gain (%) of R25 and others compared to baseline groups.

Normalized Objective Score by Method and Configuration

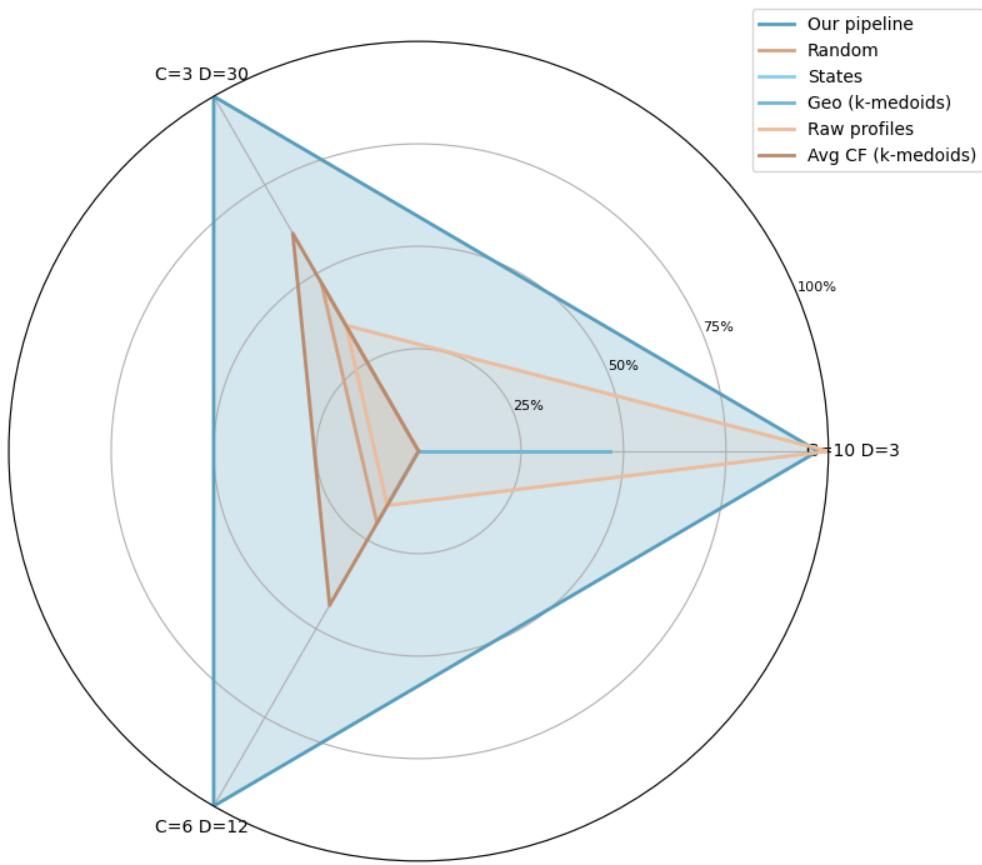


Figure 6.8: Normalized objective score by method and configuration (100% = best).

Chapter 7

Conclusion

7.1 Summary and Key Findings

This thesis developed and evaluated a task-aware framework for spatio-temporal aggregation in CEP models, with a focus on interpretability, configurability, and downstream performance. The framework integrates a composite-feature spatial clustering approach with representative-day temporal reduction, enabling dimensionality reduction that preserves key planning-relevant signals. Each step of the pipeline is explicitly designed to balance computational tractability with fidelity to the structural characteristics that influence investment outcomes in long-term power system planning.

We began by motivating the need for aggregation as a modeling decision rather than a neutral preprocessing step. Both spatial and temporal aggregation fundamentally reshape the input data and thereby influence siting, technology mix, and system costs. This motivates the need for planning-aware methods that prioritize model outcomes over statistical reconstruction. To this end, the aggregation pipeline presented here (Chapter 3) combines diverse system features, such as load duration curves, renewable profiles, ramping behavior, and spatial correlations, into a modular and interpretable similarity metric. Spatial clustering is performed using a k -medoids algorithm over this composite metric, while temporal aggregation proceeds via clustering of daily system profiles. These aggregated inputs form the basis of a reduced CEP model.

To evaluate the performance of each aggregation configuration, we proposed a sequential deaggregation framework (Chapter 4) that projects investment decisions back to higher-resolution simulations and assesses their feasibility and cost. This approach enables aggregation hyperparameters to be selected based on their planning performance rather than input-level similarity.

Applied to the New England power system, the aggregation pipeline led to the following key findings (Chapter 6):

- **Performance gains over heuristics.** The task-aware aggregation consistently outperforms standard baselines such as political boundaries, geographic proximity, and capacity-factor-based grouping. In the New England case study, the best feature weightings deliver up to a 13% reduction in total system cost compared to these heuristics.

- **Importance of correlation-based features.** Features that capture temporal correlation between zones—particularly renewable generation correlations—are the most effective in producing high-quality investment outcomes. In contrast, feature sets that rely heavily on raw amplitude measures or geographic coordinates tend to degrade performance.
- **Operational validation of aggregated solutions.** Investment plans obtained from the aggregated models remain operationally feasible when re-evaluated at full resolution and capture most of the benefits predicted by the reduced model. The integrated approach to spatial clustering and representative day selection thus preserves key ramping events and cross-zonal interactions, leading to investment decisions that remain cost-efficient under full chronological simulation.
- **Spatial granularity is a key driver of planning accuracy.** The number and quality of spatial clusters have a decisive impact on model performance. Reducing the number of zones too aggressively erases important geographic and correlation structure, leading to misplaced generation, underestimated transmission needs, and higher operational costs when tested in full resolution.

These results reinforce the central argument of this thesis: aggregation should be treated as a design problem, with methods tailored to the structural properties and planning goals of the underlying optimization model. By framing aggregation as a tunable component of the modeling pipeline, rather than a fixed preprocessor, we enable more flexible and outcome-aligned simplifications of large-scale CEP models.

7.2 Limitations and Future Directions

Despite the contribution of this work, several limitations remain, both in model design and empirical scope. Below we summarize key constraints and possible extensions.

Temporal aggregation applied after spatial clustering. In this thesis, temporal aggregation is applied after spatial clustering, using the reduced network as input. This ensures that the selected representative days match the structure seen by the planning model. However, it creates two issues. First, important temporal patterns may be lost if they were only visible in the original high-resolution data. Second, because temporal aggregation is redone for each spatial configuration, the representative days change when the number of spatial clusters or the spatial clustering method changes. This makes it hard to isolate the impact of spatial aggregation, as temporal input varies at the same time. A cleaner approach would perform temporal aggregation once on the full dataset, then reuse the same days across all spatial variants. Moreover, because system-wide daily profiles are built from the clustered zones, poorly chosen spatial clusters can distort temporal signals—masking ramping diversity or misaligning peaks between zones—thereby reducing the quality of representative days.

No real transmission network. Due to lack of access to realistic network data (e.g., shift factors, thermal ratings), this study uses a greenfield grid and a simplified transport model instead of a DCOPF. While this is acceptable for comparing aggregation strategies, it limits post-hoc evaluation. In particular, advanced techniques like the *min-excess method* [38], which assess feasibility using electrical distances, could not be used. The alternative is full re-optimization at high resolution after deaggregation, a method that scales poorly. To make the experiments feasible, we downscaled the network to 17 nodes instead of using the full 385-node dataset. The lack of network data also limits the clustering features: physical metrics like impedance or connectivity could not be included.

Beyond these two structural constraints, a few additional limitations affect the modeling framework, experimental design, and generalizability of the results:

- **Weight tuning and generalization.** One goal was to find a robust feature weighting across all aggregation levels. However, results show that no fixed set of weights performs best across all (C, D) configurations. This suggests that weights should be tuned jointly with the number of clusters and days for each run. Although the reduced models are smaller, tuning feature weights and aggregation levels still requires solving many CEP instances across configurations, which remains computationally intensive for large systems.
- **Standard temporal clustering.** The pipeline uses k -medoids with Euclidean distance over normalized daily profiles, a standard but limited method. It does not align peaks across days or ensure inclusion of extreme events. Temporal aggregation is based on a fixed set of representative days, which preserves diversity in daily profiles but fails to model chronological linkages beyond the day level. This limits the representation of seasonal storage dynamics and extended weather events. Alternatives like Dynamic Time Warping (DTW) or DTW-Barycenter Averaging (DBA) could better capture system stress days and shape variation. Future pipelines could also explicitly include peak and drought days using heuristic or optimization-based selection.
- **Simplified GTEP model.** The model includes core features like investment, dispatch, ramping, and storage dynamics. However, it omits unit commitment constraints (e.g., binary on/off variables, minimum up/down times), carbon caps, and policy constraints. These are not needed for testing aggregation, but matter for realism and could influence how aggregation affects results.
- **Single-region case study.** All experiments use data from New England. This provides a focused testbed but limits generalization. Aggregation performance can depend on geographic diversity and system structure. Applying the method to other regions (e.g., ERCOT, MISO, Europe) would test whether key design choices and feature signals transfer across systems.
- **Post-hoc analysis.** While the pipeline evaluates objective quality and feasibility, further insights could be gained by analyzing siting patterns and operational outcomes such as renewable curtailment or load shedding. These metrics would help assess whether aggregation biases planning decisions and where those effects manifest spatially.

In short, the two most important limitations are (1) the sequencing of temporal aggregation after spatial clustering, and (2) the absence of detailed network data. These choices were necessary for tractability, but they limit the realism and consistency of the results. Addressing them, by improving temporal design, adding physical grid structure, and expanding the modeling framework, offers a clear path for future work.

Appendix A

Values of Planning Model Parameters

This appendix presents the full set of parameter values used in the planning model described in Chapter 5. It supports the discussion in Section 5.2, which outlines how each parameter enters the formulation of the GTEP model. We separate the information into two categories: (i) raw parameters, which are directly extracted from data sources such as the NREL ATB [4], WECC reports [53], or EIA projections [54], and (ii) computed parameters, which are derived from raw values using standard financial or engineering formulas. For completeness, we also display key data visualizations and tabulated values drawn from the literature that support our modeling assumptions.

Table A.1 reports all raw parameters used in the model, including technology-specific capital costs, financial multipliers (e.g., FCR), physical performance metrics (e.g., heat rate, round-trip efficiency), resource limits, and transmission line specifications. These values are directly taken from standardized datasets such as the NREL ATB, WECC cost reports, and EIA projections. Table A.2 presents all derived parameters used in the GTEP formulation, including annualized CAPEX, marginal production costs, transmission cost-per-mile expressions, and network flow limits. The formulas used to compute each parameter are included in the final column. All computed values are expressed in scaled units consistent with the modeling framework. When relevant, inflation adjustments and amortization are applied using compound growth and capital recovery factors.

Figure A.1 shows the six ReEDS balancing areas covering ISO New England, from which renewable resource availability are extracted. Figures A.3 and A.2 summarize thermal ratings and capital cost estimates for standard transmission line configurations published by WECC. These values are consolidated in Table A.3, which reports normalized costs per MW-mile based on nameplate ratings. Together, they support the construction of representative transmission parameters, including the per-mile annualized CAPEX cost C_{Tran} and the capacity-per-distance coefficient γ used in Table A.2. Finally, Table A.4 displays real fuel price projections from the EIA AEO (2024), with values indexed in 2024 dollars. The 2050 natural gas price from this table underpins the marginal cost φ_{CCGT} used in the model.

Table A.1: Raw parameter values used in the GTEP model. All values in native units.

| Parameter | Value | Source |
|--|-------------------|-----------------------------|
| CAPEX _{PV} (2050) | 682.90 \$/kW | NREL ATB (2024) [4] |
| FCR _{PV} (2050) | 0.0697 | NREL ATB (2024) [4] |
| Γ_{PV}^{\max} | 2,360.33 GW | ReEDS 2.0, Classes 1–3 [55] |
| CAPEX _{Wind} (2050) | 1,114.91 \$/kW | NREL ATB (2024) [4] |
| FCR _{Wind} (2050) | 0.0739 | NREL ATB (2024) [4] |
| Γ_{Wind}^{\max} | 57.90 GW | ReEDS 2.0, Classes 1–6 [55] |
| CAPEX _{CCGT} (2050) | 1,955.80 \$/kW | NREL ATB (2024) [4] |
| Heat Rate h_{CCGT} (2050) | 6.85 MMBtu/MWh | NREL ATB (2024) [4] |
| Ramping r_{CCGT} | 5% | NREL ATB (2024) [4] |
| Nameplate Capacity \bar{P}_{CCGT} | 640 MW | NREL ATB (2024) [4] |
| NG Fuel Price (2050) | 4.46 \$/MMBtu | EIA AEO (2025) [54] |
| CAPEX _{Storage} | 1,035.77 \$/kW | NREL ATB (2024) [4] |
| Storage Duration | 4 hours | NREL ATB (2024) [4] |
| Storage Round-Trip Efficiency (RTE) | 85% | NREL ATB (2024) [4] |
| Initial SOC e_c^0 | 50% | Assumed |
| Transmission Cost $C_{230\text{kV}}$ (230 kV SC) | 959,700 \$/mile | WECC [53] |
| Transmission Cost $C_{345\text{kV}}$ (345 kV SC) | 1,343,800 \$/mile | WECC [53] |
| Transmission Capacity $\bar{T}_{230\text{kV}}$ (230 kV SC) | 400 MW | WECC [53] |
| Transmission Capacity $\bar{T}_{345\text{kV}}$ (345 kV SC) | 750 MW | WECC [53] |
| Load Shedding Cost ψ | 10,000 \$/MWh | [49] |

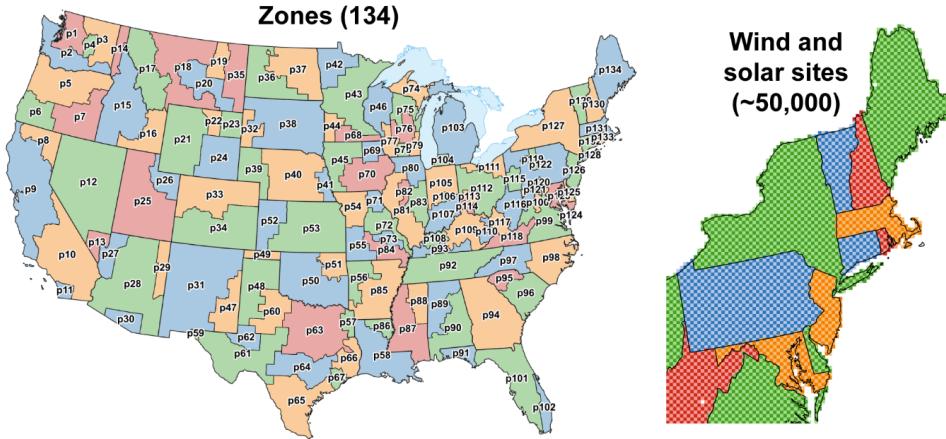


Figure A.1: Balancing Authority (BA) regions used to extract resource availability from the ReEDS 2.0 dataset [55]. Six zones (p129–p134) corresponding to ISO New England are retained. (Raw figure)

Table A.2: Computed parameters used in the GTEP model. All values are expressed in scaled units.

| Parameter | Value | Unit | Formula |
|---------------------------------------|-------------------------------|----------------|---|
| C_{PV} | 47.62 | M\$/GW/yr | $\text{CAPEX}_{\text{PV}} \times \text{FCR}_{\text{PV}}$ |
| C_{Wind} | 82.41 | M\$/GW/yr | $\text{CAPEX}_{\text{Wind}} \times \text{FCR}_{\text{Wind}}$ |
| C_{CCGT} | 142.09 | M\$/GW/yr | $\text{CAPEX}_{\text{CCGT}} \times \text{CRF}(0.06, 30)$ |
| C_{Storage} | 75.25 | M\$/GW/yr | $\text{CAPEX}_{\text{Storage}} \times \text{CRF}(0.06, 30)$ |
| C_{Tran} | 0.3170 | M\$/GW/mile/yr | $\text{CAPEX}_{\text{Tran}} \times \text{CRF}(0.08, 40)$ |
| $\text{CAPEX}_{\text{Tran}}$ | 3.78 | \$/mile/GW | $\left(\frac{0.9 \times C_{345\text{kV}}}{T_{345\text{kV}}} + \frac{0.1 \times C_{230\text{kV}}}{T_{230\text{kV}}} \right) \times (1.02)^{36}$ |
| Line capacity γ | $0.715/d_{\ell}^{\text{med}}$ | GW/mile | $(0.9 \times \bar{T}_{345\text{kV}} + 0.1 \times \bar{T}_{230\text{kV}}) / d_{\ell}^{\text{med}}$ |
| Marginal cost φ_{CCGT} | 0.0306 | M\$/GWh | Fuel price \times heat rate |
| $\eta^{\text{ch}}, \eta^{\text{dis}}$ | 0.922 | — | $\sqrt{\text{RTE}}$ |

Table 2-1 Baseline Transmission Costs

| LINE DESCRIPTION | NEW LINE COST 2014 (\$/MILE) |
|-----------------------|------------------------------|
| 230 kV Single Circuit | \$959,700 |
| 230 kV Double Circuit | \$1,536,400 |
| 345 kV Single Circuit | \$1,343,800 |
| 345 kV Double Circuit | \$2,150,300 |
| 500 kV Single Circuit | \$1,919,450 |
| 500 kV Double Circuit | \$3,071,750 |
| 500 kV HVDC Bi-pole | \$1,536,400 |
| 600 kV HVDC Bi-pole | \$1,613,200 |

Assumptions: Aluminum Conductor Steel Reinforced (ACSR), Tubular (230 kV)/ Lattice (345 kV – 600 kV), > 10 miles

Figure A.2: Baseline transmission line capital cost estimates from WECC Table 2-1 [53]. (Raw figure)

Table 2-8 Transmission Line Configuration Adopted from WREZ

| | 230 KV SINGLE CIRCUIT | 230 KV DOUBLE CIRCUIT | 345 KV SINGLE CIRCUIT | 345 KV DOUBLE CIRCUIT | 500 KV SINGLE CIRCUIT | 500 KV DOUBLE CIRCUIT | 500 KV HVDC BI-POLE | 600 KV HVDC BI-POLE |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|---------------------------|---------------------------|
| Capacity (MW) | 400 | 800 | 750 | 1500 | 1500 | 3000 | 3000 | 3000 |
| No. of Conductors Per Phase | 1 | 1 | 2 | 2 | 3 | 3 | 3 | 3 |
| No. of Circuits Per Line | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 |

Figure A.3: Transmission line configurations and thermal ratings from WECC Table 2-8 [53]. These values are used to build representative per-mile cost and capacity metrics. (Raw figure)

Table A.3: Transmission line configuration cost and rating summary based on WECC Tables 2-1 and 2-8 [53]. The final column reports normalized cost per MW-mile. (Built from raw tables A.2 and A.3)

| Configuration | Cost [\$/mile] | Rating [MW] | Cost per MW [\$/MW/mile] |
|--------------------|----------------|-------------|--------------------------|
| 230 kV SC | 959,700 | 400 | 2,399 |
| 230 kV DC | 1,536,400 | 800 | 1,921 |
| 345 kV SC | 1,343,800 | 750 | 1,792 |
| 345 kV DC | 2,150,300 | 1,500 | 1,434 |
| 500 kV SC | 1,919,450 | 1,500 | 1,280 |
| 500 kV DC | 3,071,750 | 3,000 | 1,024 |
| 500 kV HVDC bipole | 1,536,400 | 3,000 | 512 |
| 600 kV HVDC bipole | 1,613,200 | 3,000 | 538 |

Table A.4: Real fuel prices by year for selected fuels from the EIA Annual Energy Outlook 2024, Table 3 [54]. Prices are reported in 2024\$ per MMBtu. (Extracted and filtered for clarity)

| Fuel Type | 2024 | 2030 | 2040 | 2050 |
|---------------------|-------|-------|-------|-------|
| Natural Gas | 2.72 | 3.43 | 4.31 | 4.46 |
| Distillate Fuel Oil | 26.14 | 22.16 | 23.85 | 25.74 |
| Residual Fuel Oil | 17.39 | 16.44 | 17.46 | 17.85 |
| Steam Coal | 2.49 | 2.24 | 2.06 | 2.36 |
| Uranium | 0.75 | 0.75 | 0.75 | 0.75 |
| Hydrogen | 9.59 | 10.93 | 12.47 | 12.29 |

Appendix B

Grid of Feature Weight Configurations

Table B.1 summarizes the set of feature weight configurations \mathbf{w} explored in our experiments. Each configuration assigns normalized weights to six interpretable features: spatial position, raw time series (TS), duration curve (DC), ramp duration curve (RDC), intra-cluster correlation, and inter-cluster correlation. The grid includes equal-weight baselines, single-feature tests, structured combinations (pairs and triplets), targeted sweeps across selected pairs, and random samples from a Dirichlet distribution to cover a broad portion of the weight space. The full list of numerical vectors is detailed in Tables B.2 and B.3.

Table B.1: Summary of feature weight configurations used in the hyperparameter grid search.

| ID | Description | Values |
|-----------|--------------------------|---|
| B0 | Equal weights | All weights = 0.167 |
| E1–E6 | Single-feature | One weight = 1, rest = 0 |
| P1–P6 | Pair-dominant combos | Two weights = 0.45, rest = 0.025 |
| T1–T5 | Triplet emphasis sets | Three weights = 0.25, rest = 0.083 |
| Spoti1–3 | Position–TS sweep | (0.1, 0.7), (0.4, 0.4), (0.7, 0.1), rest = 0.05 |
| S2tiin1–3 | TS–Inter-Corr. sweep | (0.1, 0.7), (0.4, 0.4), (0.7, 0.1), rest = 0.05 |
| R1–R26 | Random Dirichlet samples | Randomized, normalized weights |

Table B.2: Full grid of feature weight vectors used for spatial aggregation (Part 1).

| ID | Position | Time Series | Duration Curves | Ramp Duration Curves | Intra Correlation | Inter Correlation |
|---------|----------|-------------|-----------------|----------------------|-------------------|-------------------|
| B0 | 0.167 | 0.167 | 0.167 | 0.167 | 0.167 | 0.167 |
| E1 | 1 | 0 | 0 | 0 | 0 | 0 |
| E2 | 0 | 1 | 0 | 0 | 0 | 0 |
| E3 | 0 | 0 | 1 | 0 | 0 | 0 |
| E4 | 0 | 0 | 0 | 1 | 0 | 0 |
| E5 | 0 | 0 | 0 | 0 | 1 | 0 |
| E6 | 0 | 0 | 0 | 0 | 0 | 1 |
| P1 | 0.45 | 0.45 | 0.025 | 0.025 | 0.025 | 0.025 |
| P2 | 0.025 | 0.45 | 0.45 | 0.025 | 0.025 | 0.025 |
| P3 | 0.025 | 0.45 | 0.025 | 0.45 | 0.025 | 0.025 |
| P4 | 0.45 | 0.025 | 0.025 | 0.025 | 0.025 | 0.45 |
| P5 | 0.025 | 0.025 | 0.025 | 0.025 | 0.45 | 0.45 |
| P6 | 0.025 | 0.025 | 0.45 | 0.45 | 0.025 | 0.025 |
| T1 | 0.083 | 0.25 | 0.25 | 0.25 | 0.083 | 0.083 |
| T2 | 0.083 | 0.25 | 0.083 | 0.083 | 0.25 | 0.25 |
| T3 | 0.25 | 0.083 | 0.25 | 0.083 | 0.083 | 0.25 |
| T4 | 0.083 | 0.25 | 0.083 | 0.25 | 0.083 | 0.25 |
| T5 | 0.25 | 0.083 | 0.25 | 0.083 | 0.25 | 0.083 |
| Spoti1 | 0.1 | 0.7 | 0.05 | 0.05 | 0.05 | 0.05 |
| Spoti2 | 0.400 | 0.400 | 0.05 | 0.05 | 0.05 | 0.05 |
| Spoti3 | 0.7 | 0.1 | 0.05 | 0.05 | 0.05 | 0.05 |
| S2tiin1 | 0.05 | 0.1 | 0.05 | 0.05 | 0.05 | 0.7 |
| S2tiin2 | 0.05 | 0.400 | 0.05 | 0.05 | 0.05 | 0.400 |
| S2tiin3 | 0.05 | 0.7 | 0.05 | 0.05 | 0.05 | 0.1 |
| R1 | 0.269 | 0.261 | 0.267 | 0.031 | 0.01 | 0.162 |
| R2 | 0.207 | 0.458 | 0.012 | 0.153 | 0.01 | 0.16 |
| R3 | 0.154 | 0.071 | 0.213 | 0.038 | 0.314 | 0.21 |
| R4 | 0.215 | 0.014 | 0.225 | 0.268 | 0.222 | 0.055 |
| R5 | 0.137 | 0.071 | 0.151 | 0.009 | 0.071 | 0.561 |

Table B.3: Full grid of feature weight vectors used for spatial aggregation (Part 2).

| ID | Position | Time Series | Duration Curves | Ramp Curves | Duration | Intra Correlation | Inter Correlation |
|-----|----------|-------------|-----------------|-------------|----------|-------------------|-------------------|
| R6 | 0.697 | 0.098 | 0.05 | | 0.069 | 0.063 | 0.023 |
| R7 | 0.078 | 0.498 | 0.218 | | 0.151 | 0.028 | 0.028 |
| R8 | 0.008 | 0.26 | 0.119 | | 0.159 | 0.003 | 0.452 |
| R9 | 0.505 | 0.068 | 0.057 | | 0.057 | 0.103 | 0.210 |
| R10 | 0.201 | 0.123 | 0.337 | | 0.053 | 0.123 | 0.162 |
| R11 | 0.151 | 0.381 | 0.055 | | 0.179 | 0.222 | 0.012 |
| R12 | 0.102 | 0.02 | 0.007 | | 0.324 | 0.367 | 0.18 |
| R13 | 0.121 | 0.034 | 0.383 | | 0.193 | 0.043 | 0.227 |
| R14 | 0.007 | 0.487 | 0.061 | | 0.22 | 0.076 | 0.149 |
| R15 | 0.072 | 0.019 | 0.316 | | 0.135 | 0.254 | 0.204 |
| R16 | 0.216 | 0.605 | 0.022 | | 0.052 | 0.011 | 0.093 |
| R17 | 0.119 | 0.077 | 0.428 | | 0.107 | 0.08 | 0.190 |
| R18 | 0.019 | 0.206 | 0.01 | | 0.55 | 0.188 | 0.028 |
| R19 | 0.001 | 0.292 | 0.212 | | 0.226 | 0.255 | 0.013 |
| R20 | 0.111 | 0.031 | 0.497 | | 0.244 | 0.1 | 0.016 |
| R21 | 0.063 | 0.067 | 0.221 | | 0.172 | 0.369 | 0.108 |
| R22 | 0.022 | 0.216 | 0.247 | | 0.142 | 0.255 | 0.118 |
| R23 | 0.298 | 0.225 | 0.01 | | 0.046 | 0.013 | 0.408 |
| R24 | 0.066 | 0.125 | 0.418 | | 0.05 | 0.093 | 0.247 |
| R25 | 0.05 | 0.016 | 0.066 | | 0.034 | 0.514 | 0.32 |
| R26 | 0.127 | 0.26 | 0.206 | | 0.026 | 0.283 | 0.098 |

References

- [1] B. F. Hobbs. “Optimization methods for electric utility resource planning”. In: *European Journal of Operational Research* 83.1 (1995), pp. 1–20.
- [2] M. V. Pereira and L. M. Pinto. “Multi-stage stochastic optimization applied to energy planning”. In: *Mathematical programming* 52.1 (1991), pp. 359–375.
- [3] N. E. Koltsaklis and A. S. Dagoumas. “State-of-the-art generation expansion planning: A review”. In: *Applied energy* 230 (2018), pp. 563–589.
- [4] National Renewable Energy Laboratory. *2024 Annual Technology Baseline*. Accessed July 2025. Golden, CO: National Renewable Energy Laboratory, 2024. URL: <https://atb.nrel.gov/>.
- [5] Southwest Power Pool, Inc. *The Value of Transmission*. Evaluates transmission projects placed in service from 2012–2014; Benefit-Cost Ratio 3.5 over a 40-year horizon. Little Rock, AR: Southwest Power Pool, Jan. 2016. URL: <https://www.spp.org/documents/35297/the%20value%20of%20transmission%20report.pdf>.
- [6] J. Cochran, M. Miller, O. Zinaman, M. Milligan, D. Arent, B. Palmintier, M. O’Malley, S. Mueller, E. Lannoye, A. Tuohy, et al. *Flexibility in 21st Century Power Systems* (No. NREL/TP-6A20-61721, 1130630). 2014.
- [7] J. Kiviluoma, H. Holttinen, D. Weir, R. Scharff, L. Söder, N. Menemenlis, N. A. Cutululis, I. Danti Lopez, E. Lannoye, A. Estanqueiro, et al. “Variability in large-scale wind power generation”. In: *Wind Energy* 19.9 (2016), pp. 1649–1665.
- [8] L. G. Swan and V. I. Ugursal. “Modeling of end-use energy consumption in the residential sector: A review of modeling techniques”. In: *Renewable and sustainable energy reviews* 13.8 (2009), pp. 1819–1835.
- [9] S. Pfenninger and I. Staffell. “Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data”. In: *Energy* 114 (2016), pp. 1251–1265.
- [10] A. Jacobson, F. Pecci, N. Sepulveda, Q. Xu, and J. Jenkins. “A computationally efficient benders decomposition for energy systems planning problems with detailed operations and time-coupling constraints”. In: *INFORMS Journal on Optimization* 6.1 (2024), pp. 32–45.
- [11] A. F. Jacobson, D. L. Mauzerall, and J. D. Jenkins. “Quantifying the impact of energy system model resolution on siting, cost, reliability, and emissions for electricity generation”. In: *Environmental Research: Energy* 1.3 (2024), p. 035009.

- [12] M. M. Frysztacki, G. Recht, and T. Brown. “A comparison of clustering methods for the spatial reduction of renewable electricity optimisation models of Europe”. In: *Energy Informatics* 5.1 (2022), p. 4.
- [13] L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten. “Time series aggregation for energy system design: Modeling seasonal storage”. In: *Applied Energy* 213 (2018), pp. 123–135.
- [14] P. Nahmmacher, E. Schmid, L. Hirth, and B. Knopf. “Carpe diem: A novel approach to select representative days for long-term power system modeling”. In: *Energy* 112 (2016), pp. 430–442.
- [15] T. Schütz, M. H. Schraven, H. Harb, M. Fuchs, and D. Müller. “Clustering algorithms for the selection of typical demand days for the optimal design of building energy systems”. In: *Efficiency, cost, optimization, simulation and environmental impact of energy systems* (2016).
- [16] H. Teichgraeber and A. R. Brandt. “Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison”. In: *Applied energy* 239 (2019), pp. 1283–1293.
- [17] A. Brenner, R. Khorramfar, and S. Amin. “Learning spatio-temporal aggregations for large-scale capacity expansion problems”. In: *Proceedings of the ACM/IEEE 14th International Conference on Cyber-Physical Systems (with CPS-IoT Week 2023)*. 2023, pp. 68–77.
- [18] F. Parolin, Y. Weng, P. Colbertaldo, and R. Macdonald. “Sectoral and spatial decomposition methods for multi-sector capacity expansion models”. In: *arXiv preprint arXiv:2504.08503* (2025).
- [19] K. Poncelet, H. Höschle, E. Delarue, A. Virag, and W. D’haeseleer. “Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems”. In: *IEEE Transactions on Power Systems* 32.3 (2016), pp. 1936–1948.
- [20] A. Belderbos and E. Delarue. “Accounting for flexibility in power system planning with renewables”. In: *International Journal of Electrical Power & Energy Systems* 71 (2015), pp. 33–41.
- [21] S. 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”. In: *Applied energy* 197 (2017), pp. 1–13.
- [22] H. Teichgraeber and A. R. Brandt. “Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities”. In: *Renewable and Sustainable Energy Reviews* 157 (2022), p. 111984.
- [23] L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten. “Impact of different time series aggregation methods on optimal energy system design”. In: *Renewable energy* 117 (2018), pp. 474–487.

- [24] S. Pineda and J. M. Morales. “Chronological time-period clustering for optimal capacity expansion planning with storage”. In: *IEEE Transactions on Power Systems* 33.6 (2018), pp. 7162–7170.
- [25] B. Bahl, T. Söhler, M. Hennen, and A. Bardow. “Typical periods for two-stage synthesis by time-series aggregation with bounded error in objective function”. In: *Frontiers in Energy Research* 5 (2018), p. 35.
- [26] F. Petitjean, A. Ketterlin, and P. Gançarski. “A global averaging method for dynamic time warping, with applications to clustering”. In: *Pattern recognition* 44.3 (2011), pp. 678–693.
- [27] D. Reulein, C. Li, D. Pinel, C. A. Andresen, H. Farahmand, and M. Parvania. “The Aggregation of Wind Time-Series in Capacity Expansion Models”. In: (2025).
- [28] J. Ho, J. Becker, M. Brown, P. Brown, I. Chernyakhovskiy, S. Cohen, W. Cole, S. Corcoran, K. Eurek, W. Frazier, et al. *Regional energy deployment system (reeds) model documentation (version 2020)*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2021.
- [29] M. Haller, S. Ludig, and N. Bauer. “Decarbonization scenarios for the EU and MENA power system: Considering spatial distribution and short term dynamics of renewable generation”. In: *Energy policy* 47 (2012), pp. 282–290.
- [30] J. Hörsch and T. Brown. “The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios”. In: *2017 14th international conference on the European Energy Market (EEM)*. IEEE. 2017, pp. 1–7.
- [31] M. Brinkerink, E. Mayfield, and P. Deane. “The role of spatial resolution in global electricity systems modelling”. In: *Energy Strategy Reviews* 53 (2024), p. 101370.
- [32] M. Frysztacki and T. Brown. “Modeling curtailment in germany: how spatial resolution impacts line congestion”. In: *2020 17th international conference on the European energy market (EEM)*. IEEE. 2020, pp. 1–7.
- [33] T. Brown, D. Schlachtberger, A. Kies, S. Schramm, and M. Greiner. “Synergies of sector coupling and transmission reinforcement in a cost-optimised, highly renewable European energy system”. In: *Energy* 160 (2018), pp. 720–739.
- [34] I. Staffell and S. Pfenninger. “The increasing impact of weather on electricity supply and demand”. In: *Energy* 145 (2018), pp. 65–78.
- [35] D. Radu, A. Dubois, M. Berger, and D. Ernst. “Model reduction in capacity expansion planning problems via renewable generation site selection”. In: *2021 IEEE Madrid PowerTech*. IEEE. 2021, pp. 1–6.
- [36] S. Patil, L. Kotzur, and D. Stolten. “Advanced spatial and technological aggregation scheme for energy system models”. In: *Energies* 15.24 (2022), p. 9517.
- [37] T. Mai, E. Drury, K. Eurek, N. Bodington, A. Lopez, and A. Perry. *Resource planning model: an integrated resource planning and dispatch tool for regional electric systems*. Tech. rep. National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2013.

- [38] M. M. Frysztacki, V. Hagenmeyer, and T. Brown. “Inverse methods: How feasible are spatially low-resolved capacity expansion modelling results when disaggregated at high spatial resolution?” In: *Energy* 281 (2023), p. 128133.
- [39] K. Mongird and J. Rice. “An Integrated and Iterative Multiscale Modeling Framework for Robust Capacity Expansion Planning”. In: *Current Sustainable/Renewable Energy Reports* 11.4 (2024), pp. 77–84.
- [40] D. Yates, J. Szinai, and A. Jones. “Modeling the Water Systems of the Western US to Support Climate-Resilient Electricity System Planning”. In: *Earth’s Future* 12.1 (2024), e2022EF003220.
- [41] A. Dyreson, N. Devineni, S. W. Turner, T. De Silva M, A. Miara, N. Voisin, S. Cohen, and J. Macknick. “The role of regional connections in planning for future power system operations under climate extremes”. In: *Earth’s Future* 10.6 (2022), e2021EF002554.
- [42] S. M. Cohen, A. Dyreson, S. Turner, V. Tidwell, N. Voisin, and A. Miara. “A multi-model framework for assessing long-and short-term climate influences on the electric grid”. In: *Applied Energy* 317 (2022), p. 119193.
- [43] J. Cochran, P. Denholm, M. Mooney, D. Steinberg, E. Hale, G. Heath, B. Palmintier, B. Sigrin, D. Keyser, D. McCamey, et al. “LA100: The Los Angeles 100% Renewable Energy Study—Executive Summary”. In: *National Renewable Energy Laboratory* (2021).
- [44] J. Wessel, J. D. Kern, N. Voisin, K. Oikonomou, and J. Haas. “Technology pathways could help drive the US West Coast grid’s exposure to hydrometeorological uncertainty”. In: *Earth’s Future* 10.1 (2022), e2021EF002187.
- [45] C. R. Vernon, J. S. Rice, N. Zuljevic, K. Mongird, K. D. Nelson, G. C. Iyer, N. Voisin, and M. T. Binsted. “cerf: A Python package to evaluate the feasibility and costs of power plant siting for alternative futures”. In: *Journal of Open Source Software* 6.PNNL-SA-164611 (2021).
- [46] E. Larson. *Net-zero America: Potential pathways, infrastructure, and impacts*. Princeton University, 2020.
- [47] N. Conservancy. *Power of place: national*. 2023.
- [48] L. Qiu, R. Khorramfar, S. Amin, and M. F. Howland. “Decarbonized energy system planning with high-resolution spatial representation of renewables lowers cost”. In: *Cell Reports Sustainability* 1.12 (2024).
- [49] R. Khorramfar, D. Mallapragada, and S. Amin. “Electric-gas infrastructure planning for deep decarbonization of energy systems”. In: *Applied Energy* 354 (2024), p. 122176.
- [50] ISO New England. *CELT 2024: Expanded economic model informs long-term forecast*. Accessed: 2025-08-06. ISO New England, May 2024. URL: <https://isoneswire.com/2024/05/23/celt-2024-expanded-economic-model-informs-long-term-forecast/>.
- [51] Kontur Inc. *Kontur Population Dataset (2023)*. <https://www.kontur.io/datasets/population-dataset/>. Accessed August 6, 2025. High-resolution global population dataset at 400m H3 hexagon scale. 2023.

- [52] ISO New England Inc. *Final RSP Project List and Asset Condition List – June 2025*. Accessed: 2025-08-06. July 2025. URL: <https://www.iso-ne.com/system-planning/system-plans-studies/rsp-project-list-asset-condition-list>.
- [53] R. Pletka, J. Khangura, A. Rawlins, E. Waldren, and D. Wilson. “Capital costs for transmission and substations: updated recommendations for WECC transmission expansion planning”. In: *Black and Veatch PROJECT* 181374 (2014).
- [54] U.S. Energy Information Administration. *Annual Energy Outlook 2025*. Report. Accessed: 2025-08-06. U.S. Department of Energy, Apr. 2025. URL: <https://www.eia.gov/outlooks/aeo/>.
- [55] National Renewable Energy Laboratory. *Model Documentation — ReEDS 2.0*. Accessed: 2025-08-06. Aug. 2025. URL: https://nrel.github.io/ReEDS-2.0/model_documentation.html.