Comparison of temporal resolution selection approaches in energy systems models

Authors: Cara Marcy[1]*, Teagan Goforth[2], Destenie Nock[2]*, Maxwell Brown[3]
1. Office of Air and Radiation, U.S. Environmental Protection Agency, Washington, DC, USA
2. Engineering & Public Policy Department, Carnegie Mellon University, Pittsburgh, PA, USA
3. National Renewable Energy Laboratory, Golden, CO, USA
* Marcy and Nock are co-corresponding authors

Abstract

Capacity expansion models for the power sector are used to project future decisions over the coming decades by simulating investment and operation decisions for the use of electricity. Due to model performance constraints, these models typically do not explicitly simulate every hour within a year, but instead simulate representative time segments (groups of hours). This paper evaluates different approaches for selecting time segments across three methods: sequential, categorical, and clustering, across a wide range of time-segment quantities, for a total of 204 temporal profiles. To measure the performance of each profile's ability to accurately represent data, the root-mean-square-error of each profile's time segments are compared to the data's original hourly data. The temporal alignment across regions is also measured (i.e., how often windy days align across regions). Different spatial resolutions were applied for a subset of the temporal selection methods to investigate the impact spatial resolution has on performance. This paper provides a framework for measuring the value of different temporal selection methods and of adding more granular data to energy system models. Overall, multi-criteria clustering yields the lowest root-mean-square-error across all datasets evaluated and provides a holistic view of the intertwined relationships between renewable generation and electricity demand.

Keywords:

energy system modeling; temporal resolution; electrical load; renewable energy; spatial resolution; capacity expansion planning

Highlights:

- Temporal selection methods analyzed across three commonly used techniques.
- Measured performance of representing model data within representative hours.
- Measured temporal performance across electric load and wind and solar generation.
- Clustering techniques performed better than sequential or categorical techniques.

Abbreviations:

Acronym	Description
BA	Balancing Authority
CEM	Capacity expansion models
EPA	Environmental Protection Agency
IPM	Integrated Planning Model
LDC	Load duration curves
NERC	North American Electric Reliability Corporation
NREL	National Renewable Energy Laboratory
PV	Solar photovoltaic
ReEDs	NERL's Regional Energy Deployment System
RMSE	Root-mean-square-error
VRE	Variable renewable energy

Nomenclature:

Variable	Description
a	Actual value
g	Resource group
h	Time segments
n	Number of observations
p	Predicted value
r	Regions
RMSE	Root-mean-square-error
TTC	Total transfer capability
WF	Weighted frequency
y	Region pairs

1. Introduction

As the energy transition moves forward there is a need to understand how data variability impacts recommendations derived from modeling. A key characteristic of modeling is that it relies on simplifying assumptions to characterize real world behaviors. In this way, models are simplified versions of reality that allow us to observe, understand, and make conjectures about behavior.

Capacity expansion models (CEMs) apply optimization techniques to project power sector investment decisions over a span of decades (IRENA, 2017). These optimization models differ from other types of energy sector models in how they simplify inputs. For example, simulation models simulate prescribed future energy systems and are generally capable of modeling all 8,760 hours within a year (Lund et al., 2017). However, in the case of CEMs, which endogenously evaluate investment decisions across decades, representing every hour of every year over a model's projection period is generally deemed computationally unrealistic. To maintain computational tractability, CEMs leverage simplifications in terms of representative hours within a year (temporal resolution) and representative regions across a set of power producers (spatial resolution). The characteristics of some of the more widely recognized national-scale CEMs are featured in Huntington et al. (2020) and in Chang et al. (2021).

CEMs have become more complex over time due to recent and expected changes within the power sector (Nock & Baker, 2019; Solomon & Krishna, 2011). For example, in the United States, generation from variable renewable energy (VRE), such as wind and solar, recently overtook natural gas as the fastest-growing source of electricity generation (EIA, 2020). One drawback of VRE generation stems from the variability of VRE supply, which makes it a challenge to model (Cole et al., 2017; Mai et al., 2018). Yet incorporating these resource dynamics is important to improving cost estimations of system expansion (Spittler et al. 2020). Looking forward this modeling complexity is compounded by the push to decarbonize the power sector, increase demand response capabilities (Hamidpour et al. 2022), and increase VRE supply (Sepulveda et al., 2018).

In addition to the uncertain nature of VRE, electricity demand varies, influenced by changes in consumption habits like heating and cooling, which adds stochasticity to the grid (Eshraghi et al., 2021). Increasing VRE combined with demand uncertainty creates an inherent risk in the mismatch of supply and demand for energy system planners (Lund, 2018), particularly in deterministic models (Scott et al., 2021). Models of the power sector should aim to have sufficient temporal resolution to incorporate the stochasticity of load and VRE generation while also retaining computational tractability and overall system accuracy.

This paper uses statistical methods to measure the performance of common temporal selection approaches. The key contribution to the literature is benchmarking the performance of temporal section methods across resource and demand types, and identifying best practices. Temporal selection methods dilute information necessary for power-sector investment decisions. Thus, understanding the difference between initial data and simplifying assumptions applied can improve the validity of different energy analyses. Specifically, this paper fills this gap in the literature by evaluating the performance of 204 temporal resolution profiles on load, wind, and solar datasets to identify the discrepancies between temporal selection methods and highlight the benefits of higher temporal resolutions. This is accomplished by quantifying the error between reduced time-segment input (i.e., temporal resolution) and actual variability between load, wind, and solar profiles.

Previous studies have evaluated different temporal resolutions on outcomes of power-sector models

Mallapragada et al., 2018; Reichenberg et al., 2018). Others have presented methods for selecting time-segments based on load *or* VRE availability (Nahmmacher et al., 2016; van der Heijde et al., 2019). Furthermore, Blanford et al. (2018) and Pineda & Morales (2018) have presented methods on selecting time-segments based on correlated load and VRE availability, but do not have comparisons between other temporal selection methods. Here we build on the previous research to systematically compare *across* temporal selection methods, considering both load and VRE availability, by calculating the error of the various methods using yearlong hourly data.

Furthering the complexity, spatial resolution can compound the error produced from temporal resolution. Krishnan & Cole (2016) demonstrate that aggregating regions in their CEM from the 134 balancing authorities to states and North American Electric Reliability Corporation (NERC) regions can mute variability and impact model investment decisions. In addition, the correlation of load and VRE availability between regions may result in the misalignment of time segments in some temporal selection methods. This work fills this void in the literature by investigating three different spatial resolution assumptions (e.g., electricity interconnections, NERC market regions, and native model regions) and then evaluates the temporal selection methods performance at each.

The novelty of this work revolves around quantifying the impact of different temporal modeling assumptions. When selecting how much granularity to incorporate within the models there is an inherent trade-off between solve time and model performance. Too high of a temporal resolution contributes to lengthy solve times which hamper the ability to practically perform analysis. On the other hand, as this paper illustrates, reducing resolution is directly related to decreases in data accuracy. Different temporal selection methods can optimize the representation of the underlying data while meeting the time segment criteria of a given model.

2. Methods

The research approach for measuring the performance of different temporal selection methods involves a mix of energy planning models and temporal selection methods. The data is sourced from Environmental Protection Agency's (EPA) Power Sector Modeling Platform version 6 (EPA Platform v6) using the Integrated Planning Model (IPM) (EPA, 2018). Multiple temporal selection methods are applied to the model regions (i.e., EPA's current approach (EPA, 2018), as well as sequential, categorical, and clustering approaches, as defined in the sections below) across a range of time-segment combinations. Across all approaches, temporal selection methods effectiveness is measured by comparing the root-mean square error (RMSE), where RMSE is defined by the difference between the actual (observed) hourly data and the derived time segments.

In addition, alternative spatial aggregation methods are investigated. Within each dataset different levels of regional aggregation are applied (see Figure 4). Then the RMSE and the temporal-spatial alignment results are calculated. Temporal-spatial alignment measures the frequency at which the hours within a time-segment align with neighboring regions.

2.1 Datasets and Model Regions

EPA Platform v6 (EPA, 2018) 8760-hourly input data was used for evaluating the temporal and spatial resolution of different temporal selection methods. The selection methods were applied to three datasets: electricity load, solar photovoltaic (PV) capacity factors, and onshore wind capacity factors.

For electricity demand, this paper relies on EPA Platform v6 electricity load data (Table 2-2 of EPA, 2018). The data includes a single year (8760 hours) of load data for regions that cover the continental United States. The data is scaled by the peak capacity for each region. EPA Platform v6 has a total of 78 model regions; however, only 63 model regions were included in this analysis, the 11 Canadian regions and 4 US supply-only regions were excluded due to data limitations, and being outside the scope of this analysis.

In terms of resources, only onshore wind and solar PV resource profiles are considered in this analysis due to the wide inclusion of these resources in continental US and capacity expansion models. Wind and solar data from Table 4-39 and 4-43 of EPA (2018) are converted to hourly capacity factors. In addition to the 63 model regions, VRE availability is further subdivided by the state boundaries within each model region. This results in approximately 120 subregions for wind and solar. Each technology and region is then further subdivided by resource class. The data includes up to 10 different resource classes for onshore wind and up to seven for solar PV. In total, the data used contains 621 onshore wind profiles and 245 solar PV profiles.

2.2 Temporal Selection Methods

Temporal selection methods were analyzed across three categories: sequential, categorical, and clustering. These individual methods reflect commonly used approaches in CEMs (Connolly et al., 2009; Huntington et al., 2020). We also include the EPA Platform v6 selection method (EPA, 2018), which applies a mixed approach (combining both categorical and clustering techniques).

2.2.1 Sequential Method

The sequential method averages every set of hours within a specified interval. For example, an interval of two hours would average the data in hours one and two together, then hours three and four together, etc., resulting in half as many time segments. A two-hour interval results in 12 time-segments per day, 4,380 per year. Intervals can also span across multiple days, such as a 120-hour interval averaging the hours across five days, which yields 73 time-segments. However, load and solar data have diurnal patterns (i.e., solar generation follows the rise and fall of the sun), meaning that averaging values across five days yields poor results. Figure 1(a) shows a six-hour interval, four time-segments per day. Figure 1(b) illustrates how the six-hour interval aggregation method spans across a period of 6 days. The sequential approach presents results for every multiple of 8,760, yielding 32 profiles total.

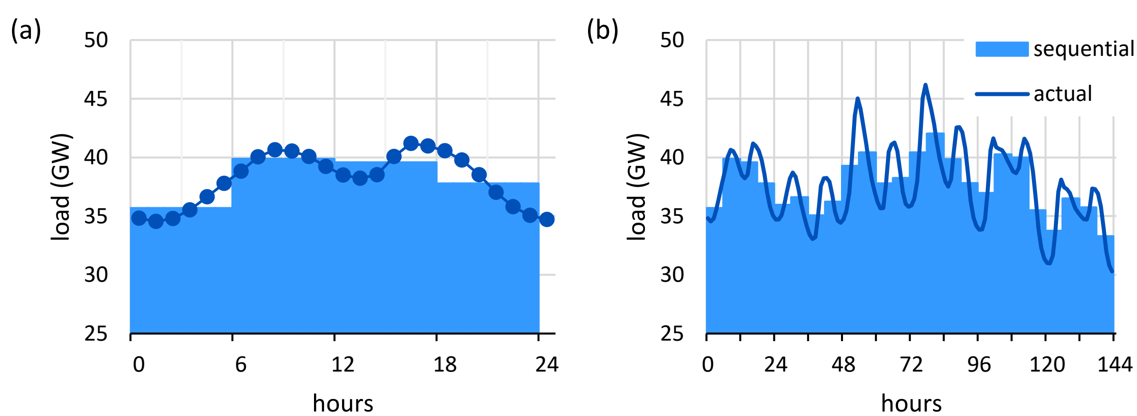


Figure 1: Sequential Temporal Selection Method for Example Load Dataset

2.2.2 Categorical Method

The categorical method better respects diurnal load and VRE availability compared to the sequential method by selecting representative days throughout the year to define a time-segment set. Day-types are selected categorically, for example, by month of the year or day of the week. Additionally, one can reduce the number of time segments within a representative day by applying categorical hour-types. For example, representing all 24 hours within a single representative day-type for each month of the year results in 288 time-segments; however, this can be reduced further to 72 time-segments by using 4-hour intervals within each day-type.

Figure 2(a) walks through creating a representative day. Here each hour of the day is averaged across 6 days to form a representative day. Next, Figure 2(b) maps the representative day across the original 6-days of data.

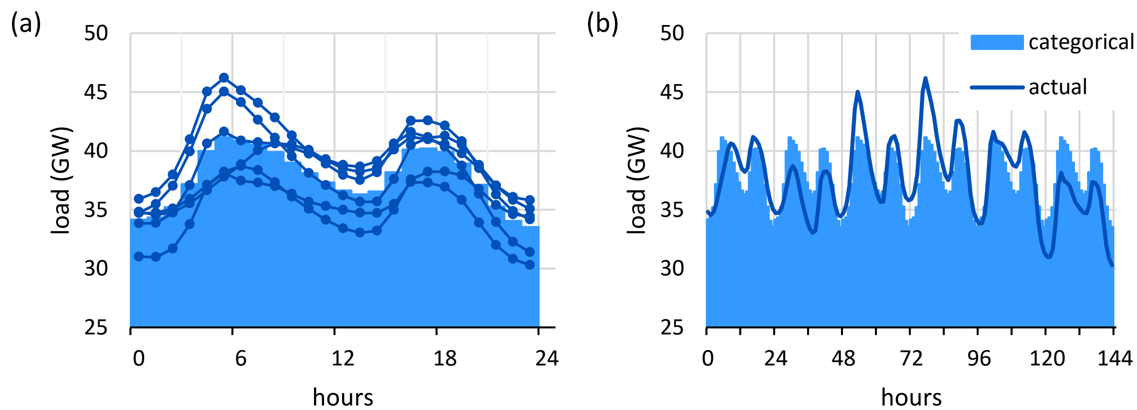


Figure 2: Categorical Temporal Selection Method for Example Load Dataset

The three most common categories (month-, day-, and hour-type) are applied. Month-types explored include using all months, grouping into pairs of consecutive months, and grouping months by seasonal attributes (where spring/fall are combined). Day-types explored were all weekdays, weekend/weekday groupings, and a third category that identified the peak day within each month. Weekday/weekend groupings were selected because load profiles differ across the two day-types, as electricity consumption shifts from commercial/industrial loads on the weekdays to residential loads on the weekends, and a peak day was included to ensure peak characteristics of load could also be captured. Hour-types explored included all hours within the day and then 4-hour intervals within a day-type.

Every combination of these categories was investigated, yielding 18 categorical day-type profiles. The profile time-segment count ranged between 18 and 864 time-segments. A final category was added that included a single representative day for each week out of the year, which resulted in 1,248 time-segments. Table 1 summarizes the profiles used for the categorical method.

Table 1: Categorical Profiles

PROFILE	MONTH-TYPE		DAY-TYPE		HOUR-TYPE		SEGMENT #
W52-D1-H24	Weekly	52	Single	1	24 Hours	24	1248
M12-D3-H24	Monthly	12	W-day/W-end/Peak	3	24 Hours	24	864
M06-D3-H24	Bi-Monthly	6	W-day/W-end/Peak	3	24 Hours	24	432
M03-D3-H24	Seasons	3	W-day/W-end/Peak	3	24 Hours	24	216
M12-D2-H24	Monthly	12	W-day/W-end	2	24 Hours	24	576
M06-D2-H24	Bi-Monthly	6	W-day/W-end	2	24 Hours	24	288
M03-D2-H24	Seasons	3	W-day/W-end	2	24 Hours	24	144
M12-D1-H24	Monthly	12	Single	1	24 Hours	24	288
M06-D1-H24	Bi-Monthly	6	Single	1	24 Hours	24	144
M03-D1-H24	Seasons	3	Single	1	24 Hours	24	72
M12-D3-H06	Monthly	12	W-day/W-end/Peak	3	4-Hour Int	6	216
M06-D3-H06	Bi-Monthly	6	W-day/W-end/Peak	3	4-Hour Int	6	108
M03-D3-H06	Seasons	3	W-day/W-end/Peak	3	4-Hour Int	6	54
M12-D2-H06	Monthly	12	W-day/W-end	2	4-Hour Int	6	144
M06-D2-H06	Bi-Monthly	6	W-day/W-end	2	4-Hour Int	6	72
M03-D2-H06	Seasons	3	W-day/W-end	2	4-Hour Int	6	36
M12-D1-H06	Monthly	12	Single	1	4-Hour Int	6	72
M06-D1-H06	Bi-Monthly	6	Single	1	4-Hour Int	6	36
M03-D1-H06	Seasons	3	Single	1	4-Hour Int	6	18

Note: For this paper, seasons were based on the EPA Platform v6 definition of seasons (EPA, 2018), where fall and spring are combined into one shoulder season.

2.2.3 Clustering Method

The clustering method identifies hours of the year with common characteristics, and groups them together. Load duration curves (LDCs) are a commonly applied example of this approach. Here hours within a year are sorted from highest to lowest load value and then grouped (or clustered) together. For example, one could sort all the load hours and then use the top 1% of hours as one time-segment, followed by the next 4% and so on. Figure 3 presents a breakdown of this approach. In Figure 3(a) the hours are sorted across 6 days and averaged them by their sort order. Figure 3(b) shows how the segments look across the 6 days. For this paper agglomerative hierarchical clustering is applied to group hours together. The agglomerative hierarchical clustering method starts with each point as a separate cluster, then measures the distance between points, or a set of points, and merges sets with the shortest Euclidean distances until a user-specified number of clusters is achieved.

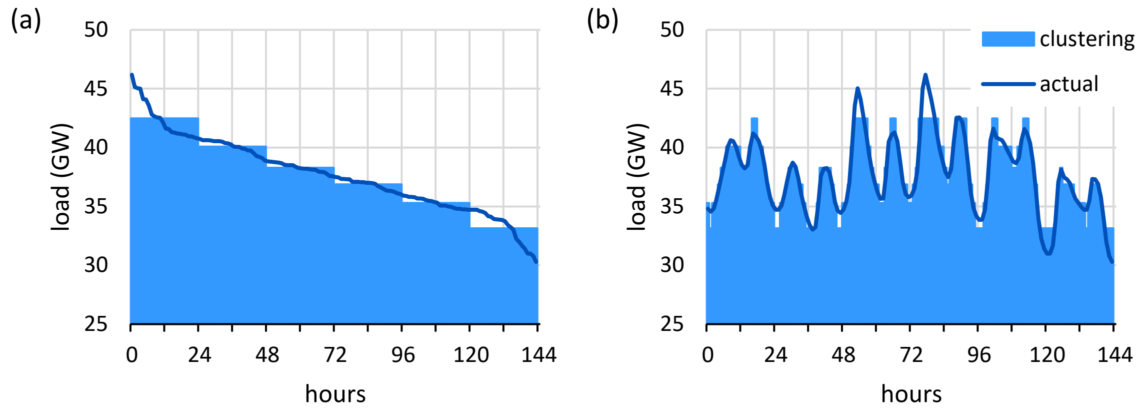


Figure 3: Clustering Temporal Selection Method for Example Load Dataset

Here the clustering approaches are examined across four different data groupings: three include clustering the data on each of the three datasets in this analysis (load, solar, and wind) separately, and the fourth approach involves clustering data across all three datasets at once (referred to as 3-way or multi-criteria clustering approach throughout this paper). Next, two methods for clustering were applied: an hourly method and a day-type method. The hourly clustering method clusters the data across two dimensions: the 8,760-hours of the year and the number of profiles (e.g., wind, solar, and load). The day-type clustering method clusters the data across three dimensions: the 365 days of the year, the 24-hours of the day, and the number of profiles. The day-type method yields best-fit representative day-types, while the hourly method clusters data regardless of the hour of day.

Under the day-type method, the four clustering approaches were applied to 18 specified time-segment numbers, ranging from one to 50 days or 24 to 1,200 (50×24) time-segments, and resulting in 72 profiles. Under the hourly method, the four clustering approaches were applied to 20 different specified time-segment numbers, matching the time segment numbers that aligned with the day-type method and adding two additional profiles at 6 and 12 time-segments, and resulting in 80 profiles.

2.2.3 Summary of Temporal Selection Methods

Table 2 summarizes the temporal selection methods evaluated in this paper and the number of time segments considered for each method. A total of 204 temporal selection profiles were tested.

Table 2: Summary of Profiles

Methods Considered	**Methods Description**	**Time Segments Considered**	**Time Segments Description**
Sequential	The sequential approach averages hours across a set interval.	1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 24, 30, 40, 60, 73, 120, 146, 219, 292, 365, 438, 584, 730, 876, 1095, 1460, 1752, 2190, 2920, 4380, 8760	every factor of 8760; total of 32 profiles
Categorical	The categorical approach groups hours based on a set of attributes associated with that hour.	18, 36, 36, 54, 72, 72, 72, 108, 144, 144, 144, 216, 216, 288, 288, 432, 576, 864, 1248	see Table 1 for day-type combinations applied; total of 19 profiles
Clustering	*Hourly Approach:* The hourly clustering approach clusters hours together that are closest in value to one another.	6, 12, 24, 48, 72, 96, 120, 144, 168, 192, 216, 240, 360, 480, 600, 720, 840, 960, 1080, 1200	hour list = 6, 12, 24, 48, 72, 96, 120, 144, 168, 192, 216, 240, 360, 480, 600, 720, 840, 960, 1080, 1200; 20 profiles per dataset (load, wind, solar, and all 3 combined), total of 80 profiles
	Day-Type Approach: The day-type clustering approach clusters days together where the values within each hour of the day are closest to one another.	24, 48, 72, 96, 120, 144, 168, 192, 216, 240, 360, 480, 600, 720, 840, 960, 1080, 1200	day list = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50; 18 profiles per dataset (load, wind, solar, and all 3 combined), total of 72 profiles
EPA Platform v6 (IPM)	Defines time-segments by a three by 24-step LDC. The year is divided into three seasons, which are sorted into LDCs and clustered into six groups. Each group is then separated into four time-of-day categories.	72	total of 1 profile

2.3 Relative Root Mean Square Error Measure

The RMSE is a frequently used measure of the differences between values predicted by a model and true values that have been observed. To measure the performance of each profile's ability to represent data accurately, the RMSE across each dataset's original spatial resolution is calculated using Equation 1. There are multiple statistical ways to measure the error of modeling efforts. The RMSE is used in this analysis because it returns the error as a single value that is easily comparable between profiles and penalizes large errors more than smaller errors. One advantage of using RMSE is that the RMSE gives a higher weight to larger errors, thereby potentially identifying approaches that are more successful at representing critical load hours, like peak load.

Explicitly, the RMSE can be computed as a function of the summed differences across regions (*r*) and time segments (*h*) from predicted (*p*) values the actual (*a*) values at each resource group (*g*) for a number of observations (*n*):

$$RMSE = \sqrt{\frac{\sum_{r,h,g}\left(p_{r,h,g} - a_{r,h,g}\right)^2}{n}}$$

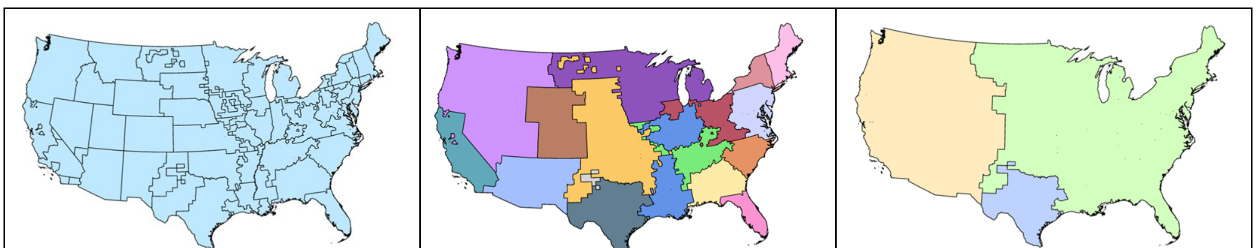
Equation 1

Here the predicted value (p) is the calculated average value for each time-segment derived from the different methods summarized in Table 2. The actual value (*a*) is the corresponding hourly value of the resource. For the results, a relative RMSE for each profile is calculated by scaling the RMSE for a given profile by the max RMSE value for each dataset. The max RMSE for each dataset is defined as the RMSE from applying a single time segment, an annual average value.

One challenge with using a statistical method to evaluate input assumptions is that it does not consider the accuracy of the model outcomes. It is important to note that the level of accuracy of the temporal resolution only matters to the extent that it impacts model results. For instance, high temporal resolution may not be necessary in a system predominantly served by dispatchable resources now and in the future. But since these models are used to evaluate a range of scenarios, ensuring the model is well equipped to evaluate all alternative futures holds merit. Additionally, the statistical approach allows for a more efficient means of evaluating different temporal selection methods compared with the time associated with setting up and running different CEMs for each individual profile presented here.

2.4 Spatial Aggregation Methods

For a subset of temporal selection methods, the impact that different spatial resolutions have on the performance metrics is evaluated. The spatial resolutions evaluated in this paper includes the EPA Platform v6 model regions (63 regions), NERC market regions (16 regions), and interconnect regions (3 regions), as seen in Figure 4. The datasets were aggregated up to each region by summing load data and averaging wind and solar capacity factors at each hour.



(a) EPA Platform v6 model	(b) NERC	(b) US Interconnect

Figure 4: Spatial Aggregation Regions for (a) *EPA Platform v6 model regions, (b) NERC market regions, and (c) Interconnect regions.*

2.5 Temporal-Spatial Alignment Measure

Temporal-spatial alignment is important for the instances where information is transferred from one region to another. In CEMs, this transfer plays out in the form of electricity trading. Clustering time segment selection can occur at different spatial resolutions. Clustering time-segments at individual model regions, may yield lower RMSE as compared to NERC or interconnect regions, but may lead to mismatches of information across regions. This information is critical in evaluating interregional trade, where mismatches of information could lead to unrealistic interregional transfers in model results.

EPA Platform v6 defines 140 model region pairs that allow for interregional electricity trade (Table 3-21 of EPA, 2018), referred to as total energy transfer capabilities. Total energy transfer capabilities define the upper limit of what can be transferred on an hourly basis given existing transmission infrastructure. The maximum megawatt value of electricity trade between each pair is used as the basis for evaluating the temporal-spatial alignment, as summarized in Figure 5.

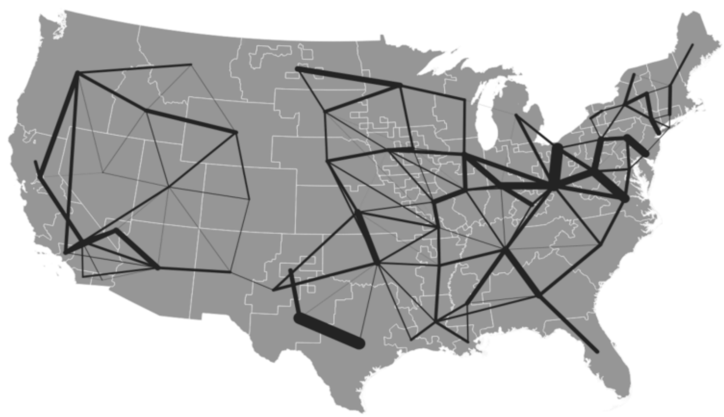


Figure 5: Maximum Energy Total Transfer Capabilities Between Model Regions

Each region (r) contains a set of data that matches all 8,760 hours to a given number of time-segments. For each model region pair (y) identified, the frequency at which the hours within a given time-segment align are measured. The weighted frequency (WF), as defined in Equation 2, is weighted by the total transfer capability between the regional pair (TTC_y).

$$WF = \frac{\sum_y \left(TTC_y \times \left| r_y^o \cap r_y' \right| \right)}{\sum_y \left(TTC_y \right) \times 8{,}760}$$

Equation 2

Equation 2 is applied across all regional pairs within the clustering results. The sequential and categorical approaches were excluded from this section of the analysis because, although their RMSE results may vary with different spatial resolutions, the temporal-spatial alignment results would not. They result in 100% frequency match regardless of the spatial resolution assumed. This is due to the fact that the data in these approaches is grouped based on information like interval hours or the months of the year – information which does not change from one region the next – whereas the clustering approach could have hours grouped in different time segments from one region to the next based on differences in load or VRE availability. See supplemental materials for addition details on Equation 2.

3. Results and Discussion

This section first presents the RMSE results for the sequential, categorical, and clustering approaches, and then a comparison across all three temporal selection methods. Then, the clustering approach is used to examine temporal-spatial alignment.

3.1 Sequential Method

The sequential approach averages hours across a set interval. One of the main advantaged to a sequential approach is its ability to maintain chronology. This can be beneficial for modeling technologies like energy storage, which require chronology to account for charging/discharging. The main disadvantage to this approach is that it is not an effective tool for reducing model complexity down to the size typically needed in CEMs.

As the datasets are hourly, an interval of 1-hour (i.e., 8,760 time-segments) results in zero error. Conversely, choosing an interval of 8,760-hours (i.e., an average value for the entire year) for each dataset defines the max error for each dataset.

The RMSE, shown in Figure 6(a), for load (blue) is less than that of solar (yellow) and wind (green). One reason for this stems from a larger set of profiles for solar and wind (245 and 621 respectively) compared to load (63 total) as there are multiple resource groups for wind and solar within each model region (see Section 2.1 Datasets and Model Regions). Secondly, the capacity factors for wind and solar have higher variability than load. For example, the capacity factor for solar may reach as high as 100% during the midday in the summer and will drop to zero overnight. The load hours are scaled to reach 100% at the peak hour of demand, but never drop to zero, as there is always demand on the system. To account for these differences, the rest of the results in the paper report the relative RMSE, as shown in Figure 6(b). The relative RMSE adjusts the profiles for load, solar, and wind scaling them by their max RMSE (the annual average).

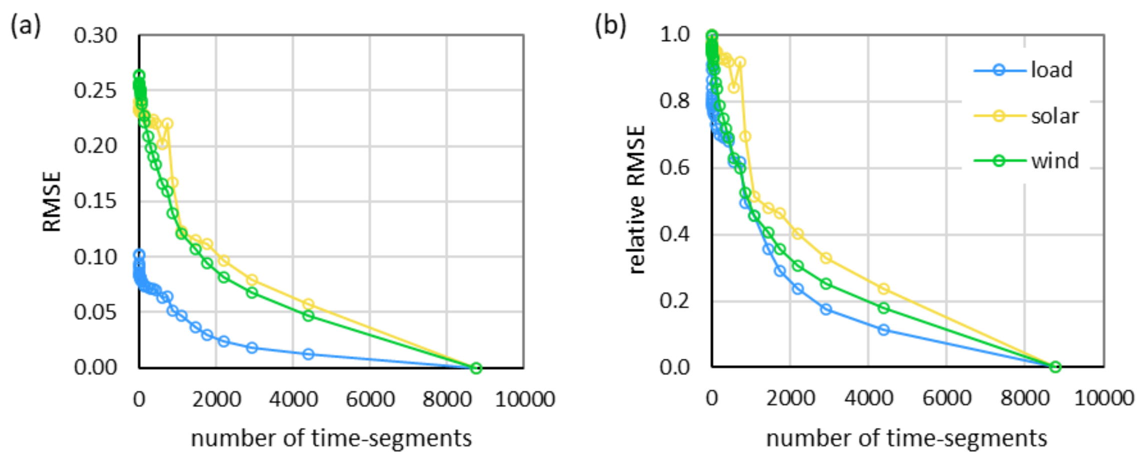


Figure 6: Sequential Approach RMSE (a) and Relative RMSE (b) by Dataset

Figure 6 illustrates there is a distinct tradeoff between computational complexity (number of time segments) and the associated error. Using just a two-hour interval instead of all 8,760 hours reduces the number of time segments in half but only results in 0.11-0.24 increase in the relative RMSE. Unsurprisingly, this relationship is not linear and instead the RMSE increases exponentially as the number of time-segments decreases.

While the sequential results have illustrated the expected tradeoff between accuracy and complexity across the entire 8,760 landscape, the rest of the results in this paper focus on practical reductions in model size, examining time-segment numbers at or below 1,300 and then at or below 130.

3.2 Categorical Method

The categorical approach groups hours based on a set of attributes associated with that hour, like month of the year or day of the week. This approach allows for representative day-types that can be used to map data across the year. Representative days can account for chronology like the sequential approach and have the added benefit of lower error results at lower time-segment numbers.

Figure 7 shows the relative RMSE results for the 19 categorical profiles evaluated for each of the three datasets (load (a), solar (b), and wind (c)) and then the average (d) relative RMSE across all three. The top panel row (Figure 7 a-d) shows the profiles between zero and 1,300 time-segments and the bottom panel row (Figure 7 e-h) hows the RMSE of highly reduced time-segments more commonly seen in today's CEMs (i.e., up to 130).

Load and solar data perform better under the categorical approach compared to wind data. In Figure 7(e–h), the relative RMSE drops by 0.13 for load, 0.10 for solar, but only 0.02 for wind, between the profile with 18 time-segments and the one with 120 time-segments. The difference in the categorical day-type approaches relative RMSE results between load and solar versus wind illustrates the strong diurnal alignment of load and solar data. This suggests that alternative metrics beyond day-types that focus on wind availability may be needed to improve representation of wind data in CEMs.

It is also important to note that there is a diminishing incremental improvement on the relative RMSE for the categories explored here, especially for load and solar data. As shown in Figure 7(a-d), there seems to be little incremental benefit between the three day-types shown with time segments beyond 500. These profiles include: M12-D2-H24, M12-D3-H24, and W52-D1-H24 (see Table 1 for definition). Increasing from 576 time-segments to 1,248 only yields and additional reduction in the relative RMSE of 0.03 for load and solar and 0.06 for wind.

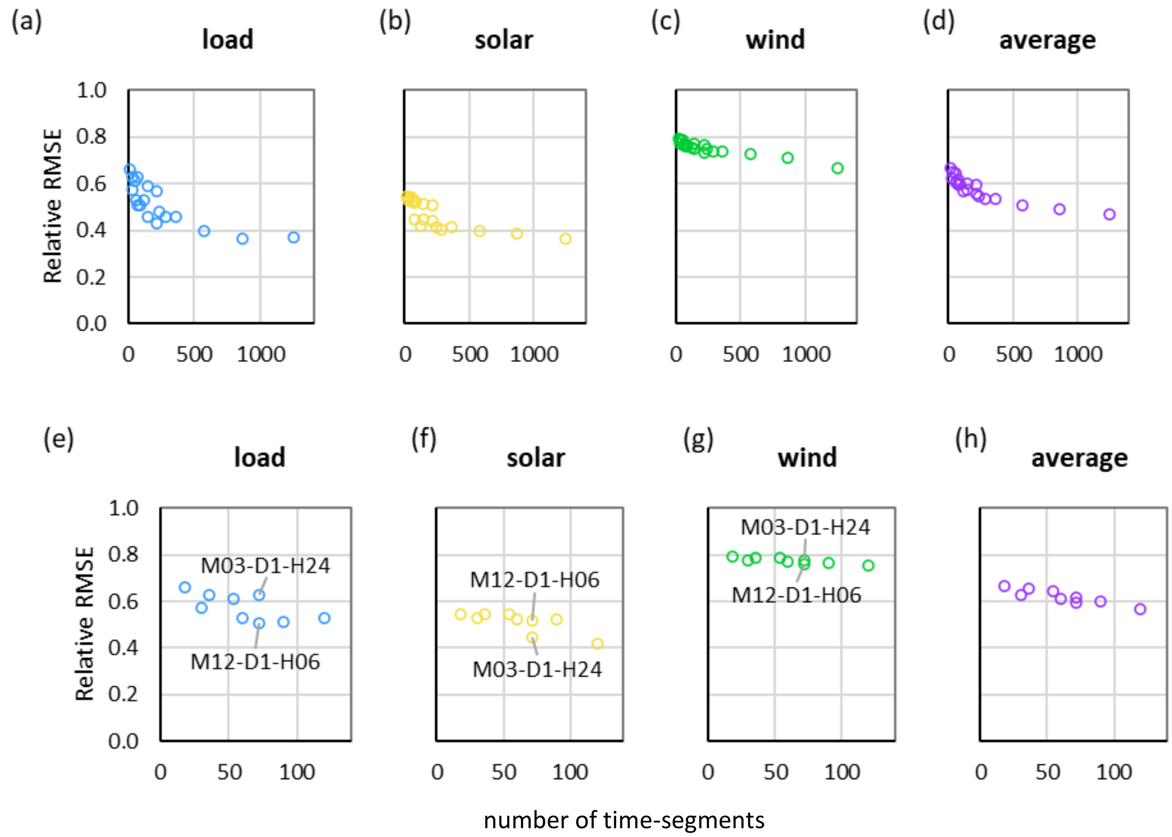


Figure 7: Categorical Approach Relative RMSE for Each Dataset Load (a, e), Solar (b, f), and Wind (c, g), and their Average (d, h). Note the x-axis differences between a-d and e-h. The second row (e-h) presents a zoomed in view of the results.

Figure 7(e-h) also highlights the tradeoffs between representative hours verses representative months. In these panels, there are two categorical day-type approach combinations that result in 72 time-segments, M03-D1-H24 and M12-D1-H06. For load, the M12-D1-H06 day-type (more representative months, fewer representative hours) results in the lower relative RMSE, while the M03-D1-H24 (fewer representative months, more representative hours) results in the lower for solar (with wind they are nearly the same). This suggests that the daily hourly intervals are more important to reflect for solar, whereas monthly trends may be more important for load.

3.3 Clustering Method

The clustering approach groups hours based on their data characteristics. This approach can allow for the lowest measures of error achievable for a given number of time segments, particularly for hourly clustering approach. In exchange for this reduced error, this approach loses its ability to maintain chronology, which limits its ability for advanced representation of technologies like storage. The day-type clustering approach allows for the development of representative days, which allows for some limited representation of chronology and still achieves relatively low error results. In addition, both hourly and day-type clustering introduces challenges with temporal-spatial alignment, as discussed in Section 3.5 Temporal-Spatial Alignment.

For each number of segments identified, the clustering approach was applied across four cluster approaches: clustering the data on each of the three datasets (load, solar, and wind) separately, and then a fourth multi-criteria clustering approach, which clustered data across all three datasets at once (referred to as 3-way). Two clustering methods were applied: first clustering across all 8,760 hours (hourly approach) and second across 365 days (day-type approach).

Figure 8 shows the relative RMSE results from the clustering method applied across 8,760 hours (hourly approach) and Figure 9 across 365 days (day-type approach). Each series in the figures shows a different clustering approach applied: clustering data on load (blue), solar (yellow), wind (green), or clustering on all three (purple). Both figures show the relative RMSE for each dataset: load (a), solar (b), wind (c) and then the average value for each profile (d), this time only for results with less than 130 time-segments.

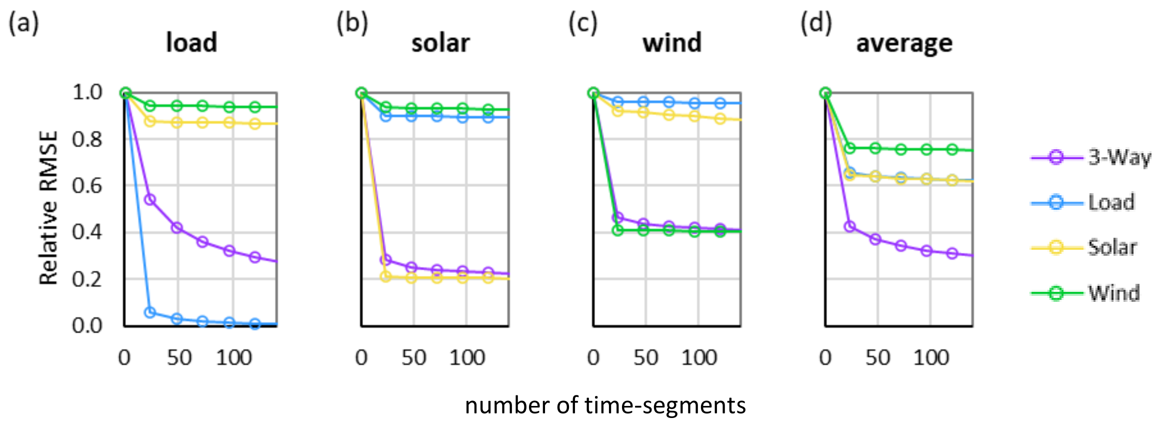


Figure 8: Hourly Clustering Approaches Relative RMSE for Each Dataset Load (a), Solar (b), and Wind (c), and their Average (d)

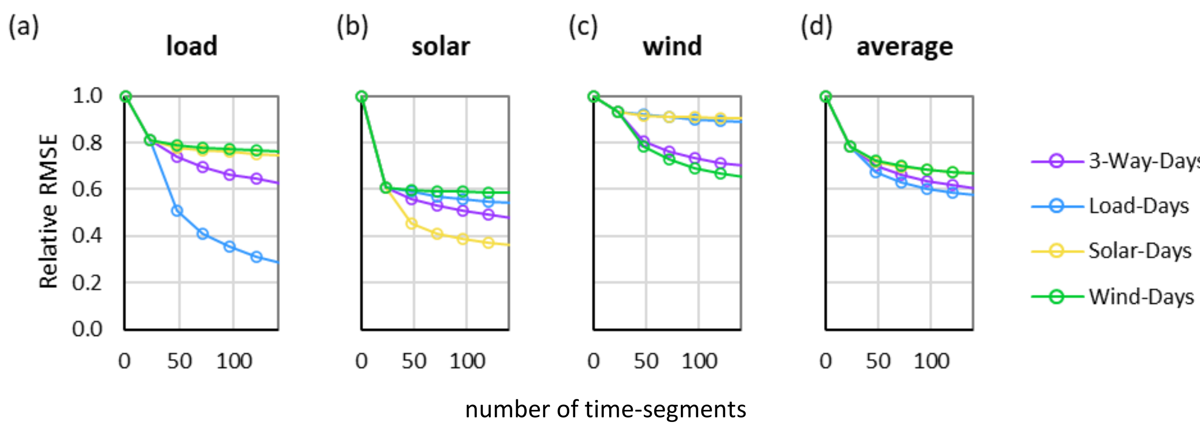


Figure 9: Day-Type Clustering Approaches Relative RMSE for Each Dataset Load (a), Solar (b), and Wind (c), and their Average (d)

The first observation is that fitting time-segments to reduce the error for one dataset results in a increasing the relative RMSE in other datasets. For hourly clustering (Figure 8), the relative RMSE increases by 0.88 on average for load clustering load data, as compared to solar or wind clustering load data. We see similar trends for the solar and wind data (0.72 and 0.53 respectively).

In almost all cases, the next lowest relative RMSE to the matching approaches (e.g., load clustering load data, etc.) is the multi-criteria clustering approach, which clusters data on load, solar, and wind simultaneously. When comparing Figure 8 and Figure 9, the hourly sets have a lower relative RMSE compared to the day-type results for the matching and multi-criteria cluster approaches and higher relative RMSE for the non-matching approaches. Interestingly, when looking at the average for the day-type clustering approach, Figure 9(d), load clustering outperforms multi-criteria clustering, mainly due to the notable difference seen in the relative RMSE results for load Figure 9(a).

3.4 Comparison Across All Approaches

This section compares the relative RMSE results across the three temporal selection methods: sequential, categorical, and clustering. Figure 10 shows the relative RMSE for profiles up to 1,300 time-segments. The relative RMSE is measured for load (a), solar (b), and wind (c) and then the average (d). The EPA platform v6 results are shown in red, sequential green, categorical purple, and clustering blue.

The clustering results are broken out into hourly (light blue) and day-type (dark blue). Both hourly and day-type clustering approaches are further broken out into multi-criteria clustering (o marker) and each dataset's matching or native clustering approach (x markers). Only the multi-criteria approach is included in the average (d).

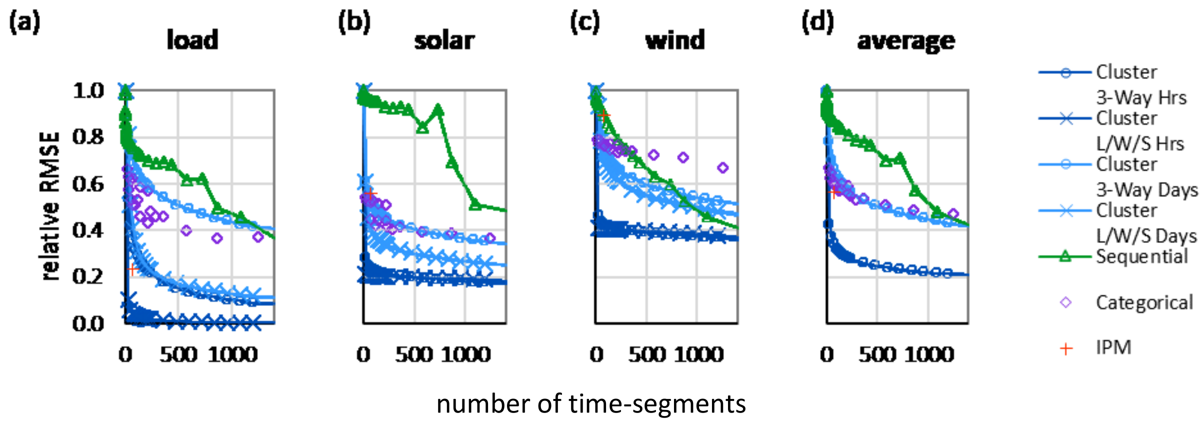


Figure 10: All Approaches Relative RMSE for Each Dataset Load (a), Solar (b), and Wind (c), and the Average (d)

One observation is the decrease in the incremental relative RMSE reduction as the temporal resolution increases. In most cases, the sequential profiles (green) act as an upper bound for relative RMSE; except for some wind day-type profiles. This is because wind data doesn't have the same diurnal patterns of load and solar.

The clustering approaches have the least error. The lower bound for the relative RMSE for all profiles tested is the hourly clustering approach. However, as observed in the previous section, the matching clustering approaches only perform well for the dataset in which they match and perform poorly when they do not match. When looking at the average (d) the multi-criteria hourly clustering approach shows the lowest relative RMSE across all time segments.

Another takeaway from Figure 10 is that the wind dataset has the highest relative RMSE for all profiles, which is partly a function of the number of wind profiles represented, but also indicates that representing wind in temporal resolution approaches is more difficult than the other datasets due to the variable nature of the data.

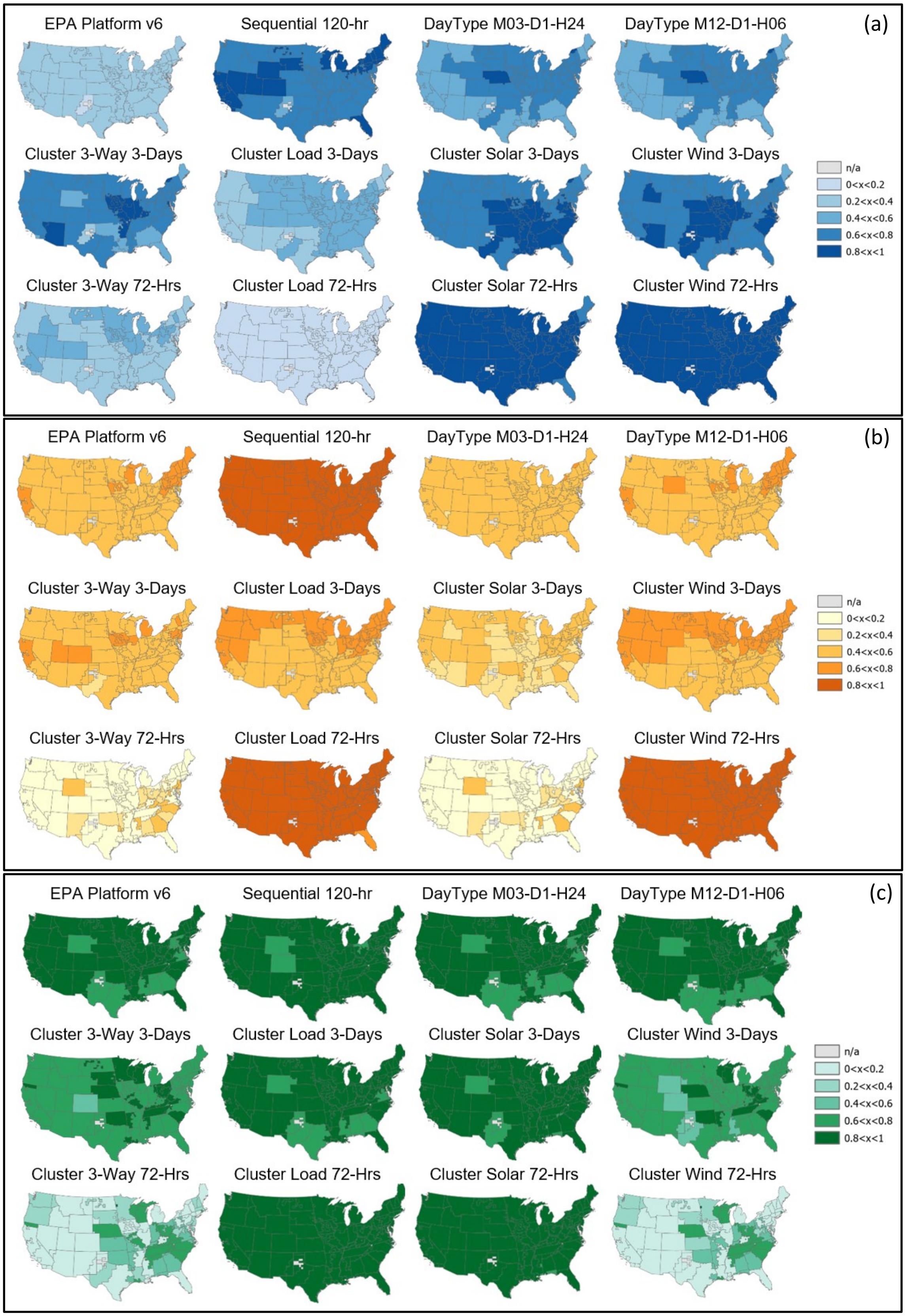


Figure 11: Select Profiles Regional Relative RMSE for Each Dataset Load (a), Solar (b), and Wind (c)

Figure 11 displays regional differences in relative RMSE for select profiles (all with near 72 time-segments). The first key observation is that not all regions necessarily have the same level of error, meaning results will vary and what is the best fit for one region may not be the best fit for another.

For the load dataset in Figure 11(a), day clustering and day-type approaches perform better in the south and west compared to the central and east, perhaps due to less seasonal variation in those regions. For the solar dataset (b), the best performing profile across all regions is the Cluster Solar 72-Hrs profile. In those cases, Wyoming and regions in the Southeast have higher relative RMSEs compared to the rest of the regions. This suggests that the solar data is more variable in those regions. Across the wind dataset (c), Wyoming, Colorado, and parts of Texas has a lower relative RMSE than other regions. One reason for this may be related to the strength of the resource in these regions. There tends to be more availability of wind in these regions, which results in more consistent availability of the resource.

Overall, it is more advantageous to use a three-way clustering approach as opposed to a single approach if interested in appropriately characterizing all three sets. It is also important to be aware of differences in regional performance of temporal resolution to identify which regions a model may be less or more accurate in modeling load, solar, and wind resources.

3.5 Temporal-Spatial Alignment

The results have thus far utilized temporal selection methods applied at the model region level (63 regions); however, temporal selection approaches can also be applied at different levels of spatial aggregation. In this section the clustering approaches are applied at the NERC (16 regions) and interconnect level (3 regions) and the impact on the results is measured. Aligning time segments across broader regions improves the evaluation of trade outcomes, but often at the expense of the RMSE.

Figure 12 shows the results from changing the spatial resolution for a subset of profiles, the clustering profiles highlighted in Figure 11 (e.g., 72-hour and 3-day clustering approaches). In Figure 12, clustering (x marker) is more sensitive to spatial aggregation compared to day-type clustering (o marker).

It appears that for most of the profiles considered, decreasing the spatial resolution from 63 regions to 3 regions results in an increase in the error. This is especially true when considering multi-criteria clustering and each dataset's matching or native clustering approach. Specifically, in Figure 12, going from 63 to 3 regions results in large increases (greater than 0.3) in the relative RMSE for the matching hourly cluster approaches for all datasets and for the multi-criteria clustering approach for the solar and wind datasets. Its only in the cases where the relative RMSE is already relatively high at 63 regions where the value would decrease when the method is applied to fewer regions.

Interestingly, as seen in Figure 12(a), the largest decrease (-0.18) in the relative RMSE from 63 to 3 regions is for the day-type multi-criteria cluster approach applied to load data (purple o markers). This is likely due the fact that larger regions more effectively smooth out extremes in variability seen in wind and solar day-type data while still capturing daily load patterns. This is consistent with observations made in Section 3.3 Clustering Method.

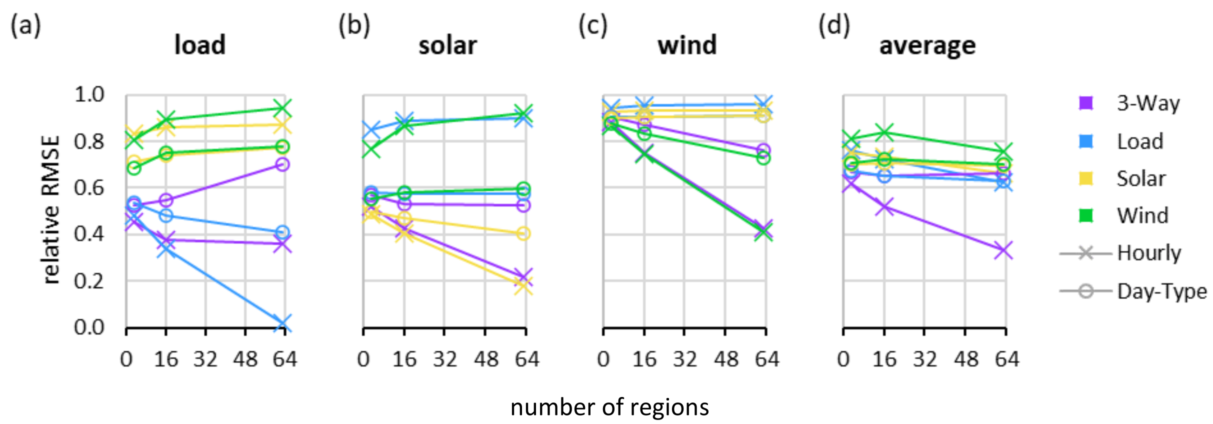


Figure 12: Clustering Approaches Relative RMSE for Each Dataset Load (a), Solar (b), and Wind (c), and their Average (d) at Different Spatial Resolutions

Another consideration with temporal resolution is the extent the selected hours align from one region to the next. Figure 13 shows the impact of applying different spatial resolutions to the temporal selection method has on hourly alignment. The y-axis shows the percentage of hours that fall within like time-segments weighted across all trade regions. Figure 13(a,c) shows the day-type clustering approach and (b,d) the hourly clustering approach. Each of the series shows the weighted frequency for the different spatial resolutions across the four profiles within each clustering approach (load, solar, wind, and 3-way). The square markers show the frequency at 3 regions, the circles at 16 regions, and the triangles at 63 regions.

Selecting the three interconnections as the point of spatial resolution applied to the temporal selection method yields near perfect alignment of hours across trade regions, although, as observed previously in Figure 12, at the expense of higher relative RMSE results. This near perfect alignment of hours is due to the facts that little to no trade of electricity is occurs across the three interconnections. Of the 306 GW of energy total transfer capabilities across regions evaluated in this analysis, 98% (or 300 GW) of that energy total transfer capabilities occurs within interconnection boundaries. For NERC regions, 49% (or 155 GW) of the energy total transfer capabilities evaluated occurs within NERC boundaries.

For the three spatial resolutions evaluated, at higher time-segments (a,b) the clustering approach essentially reflects the share of the total energy transfer capabilities that occurs within its respective boundaries. Fewer time-segments (c,d) lead to a higher probability of more aligned hours. Figure 13(c,d) show the day-type clustering approach performs better than the hourly clustering approach at aligning hours because hours of the day within the day-type are already aligned, ensuring greater alignment. Although hourly clustering yields lower relative RMSE results, this comes at a tradeoff between interregional hourly alignment as compared to the day-type clustering approach.

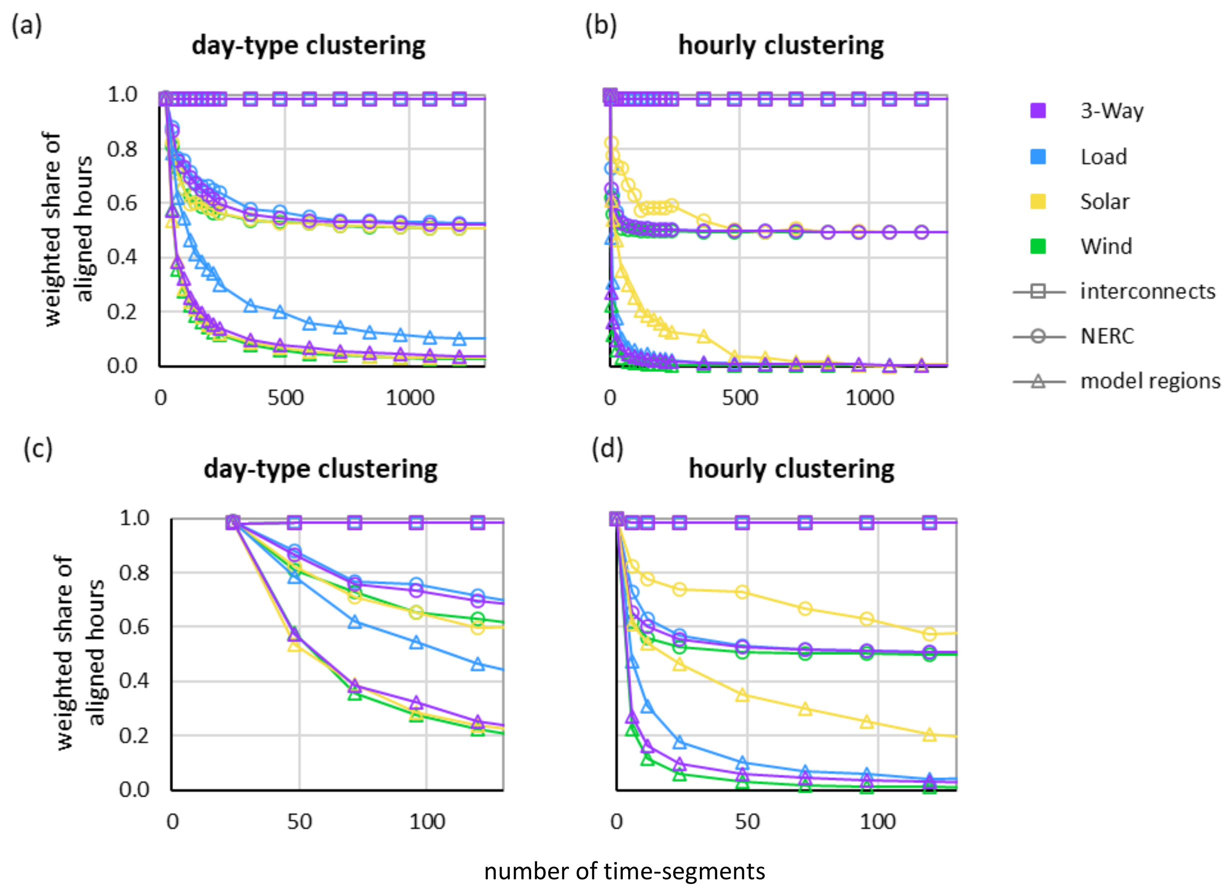


Figure 13: Interregional Alignment for Day-Type (a, c) and Hourly (b, d) Clustering Profiles at Different Spatial Resolutions

Across the four profiles (load, solar, wind, and 3-way) and within each clustering approach and spatial resolution evaluated there is little difference between alignment results. The exceptions to this are the load-day-type clustering profiles in (a,c) and the solar-hourly clustering profiles in (b,d). Weather conditions on a day-to-day basis likely assist in the greater alignment between neighboring regions relative to the other profiles in the clustering approach.

Conclusions

Our work has highlighted differences in time-segment representations across three temporal selection methods, and three subnational region groupings. The purpose of this paper is to provide guidance for modelers and assess the most-commonly applied temporal selection methods. This work is also useful for considering the tradeoffs between model resolution and fidelity to underlying data.

Through this research, there are a few salient insights into temporal selection methods. First is that the sequential approach had the highest error. In certain instances, for example where modelers want to retain a high number of time segments and properly represent technologies like energy storage, the method could still prove viable but, in general, there are better methods for selecting data for CEMs.

One of the approaches that particularly excelled is the hourly multi-criteria clustering of simultaneous wind, solar, and load datasets. Although it could be bested in terms of reduced error by other approaches for each individual dataset, across all datasets it performed particularly well. This could prove to be a viable approach for modelers that are not particularly concerned with a specific technology or load patterns and instead want a holistic view of the intertwined relationships of technology availability and electricity demand. One challenge with the hourly multi-criteria clustering approach is the lack of chronology, which can be important for modeling certain technologies, like storage.

Additionally, this analysis highlights the tradeoffs between temporal and spatial resolution for time selection methods. Applying broader regional definitions to clustering approaches improves the alignment of hours across neighboring model regions. Alignment of hours is important for modeling trade between regions; however, the improvement of aligned hours often came at the expense of the higher error. In particular, hourly clustering profiles, which often performed the best at reducing error performed poorly at aligning hours across regions. This suggests that modelers should use caution when interpreting results that have large quantities of trade across regions unless care is taken to ensure those hours are aligned. However, more generally, improving temporal-spatial alignment should not come at the expense of the key performance metric, the relative RMSE, as the majority of power sector investment decisions within CEMs are based on intra-regional rather than inter-regional model decisions.

For modelers concerned with striking a balance between representing the underlying input data, modeling chronology, and aligning interregional trade, an alternative approach that excelled well across all of these concerns was the day-type load clustering approach, which yielded the next lowest relative RMSE after the hourly clustering approach across all three datasets. In addition, with certain care, the categorical approaches could be designed in a way to achieve similar results to the day-type load clustering approach and avoid concerns regarding interregional trade, but as the results show, this is sensitive to the categories selected. Both of approaches also tend to perform significantly worse at representing wind data, which does not reflect the same level of diurnal pattern compared to load and solar.

A fruitful direction for future work would be to test the temporal resolution approaches identified within a model to measure the impact on model results and computational time. It is important to note that the level of accuracy of the temporal resolution only matters to the extent that it impacts model outcomes and changes the dispatch or level of investment for capacity expansion. The value of this work has been quantifying the error reduction in including high temporal resolution at different spatial scales. This work can help other energy modelers understand how the temporal resolution can impact the accuracy of their energy analyses, and lead to better representation of resource profiles in energy system models.

Acknowledgements

This work is funded in part by the National Science Foundation [grant number 2017789]. Nock acknowledges support from the Scott Institute for Energy Innovation, where she is an energy fellow.

Data availability

Analysis model is publicly available [See: https://github.com/cmarcy/temporal] and can be used without any software licenses.

Disclaimer

The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Author contributions: Cara Mary: Conceptualization, Methodology, Validation, Writing – original draft, Visualization; Teagan Goforth: Data Curation, Formal analysis, Writing – original draft; Destenie Nock: Writing – review and editing; Maxwell Brown: Writing - review and editing.

References

Blanford, G. J., Merrick, J. H., Bistline, J. E. T., & Young, D. T. (2018). Simulating Annual Variation in Load, Wind, and Solar by Representative Hour Selection. *The Energy Journal*, *39*(3). https://doi.org/10.5547/01956574.39.3.gbla

Chang, M., Thellufsen, J. Z., Zakeri, B., Pickering, B., Pfenninger, S., Lund, H., & Østergaard, P. A. (2021). Trends in tools and approaches for modelling the energy transition. *Applied Energy*, *290*. https://doi.org/10.1016/j.apenergy.2021.116731

Cole, W., Frew, B., Mai, T., Sun, Y., Bistline, J., Blanford, G., Young, D., Marcy, C., Namovicz, C., Edelman, R., Meroney, B., Sims, R., Stenhouse, J., & Onohoo-vallett, P. (2017). *Variable Renewable Energy in Long-Term Planning Models : A Multi-Model Perspective Variable Renewable Energy in Long-term Planning Models : A Multi-model Perspective*. *November 2017*, 35.

Connolly, D., Lund, H., Mathiesen, B. V, & Leahy, M. (2009). *A review of computer tools for analysing the integration of renewable energy into various energy systems Keywords: Energy models Energy tools Energy systems Renewable energy Sustainable energy Energy systems modelling*. https://doi.org/10.1016/j.apenergy.2009.09.026

EIA. (2020). *Electricity generation, capacity, and sales in the United States - U.S. Energy Information Administration (EIA)*. https://www.eia.gov/energyexplained/electricity/electricity-in-the-us-generation-capacity-and-sales.php

EPA. (2018). *Documentation for EPA's Power Sector Modeling Platform v6 Using the Integrated Planning Model*. *November*.

Eshraghi, H., Rodrigo de Queiroz, A., Sankarasubramanian, A., & DeCarolis, J. F. (2021). Quantification of climate-induced interannual variability in residential U.S. electricity demand. *Energy*, *236*, 121273. https://doi.org/10.1016/j.energy.2021.121273

Hamidpour, H., Aghaei, J., Pirouzi, S., Niknam, T., Nikoobakht, A., Lehtonen, M., ... & Catalão, J. P. (2022). Coordinated expansion planning problem considering wind farms, energy storage systems and demand response. Energy, 239, 122321. https://doi.org/10.1016/j.energy.2021.122321

Huntington, H. G., Bhargava, A., Daniels, D., Weyant, J. P., Avraam, C., Bistline, J., Edmonds, J. A., Giarola, S., Hawkes, A., Hansen, M., Johnston, P., Molar-Cruz, A., Nadew, M., Siddiqui, S., Vaillancourt, K., & Victor, N. (2020). Key findings from the core North American scenarios in the EMF34 intermodel comparison. *Energy Policy*, *144*, 111599. https://doi.org/10.1016/j.enpol.2020.111599

IRENA. (2017). Planning for the Renewable Future: Long-term modelling and tools to expand variable renewable power in emerging economices. In */publications/2017/Jan/Planning-for-the-renewable-future-Long-term-modelling-and-tools-to-expand-variable-renewable-power*.

Krishnan, V., & Cole, W. (2016). Evaluating the value of high spatial resolution in national capacity expansion models using ReEDS. *IEEE Power and Energy Society General Meeting*, *2016-Novem*(July). https://doi.org/10.1109/PESGM.2016.7741996

Lund, H. (2018). Renewable heating strategies and their consequences for storage and grid infrastructures comparing a smart grid to a smart energy systems approach. *Energy*, *151*, 94–102. https://doi.org/10.1016/j.energy.2018.03.010

Lund, H., Arler, F., Østergaard, P. A., Hvelplund, F., Connolly, D., Mathiesen, B. V., & Karnøe, P. (2017). Simulation versus optimisation: Theoretical positions in energy system modelling. *Energies*, *10*(7), 1–17. https://doi.org/10.3390/en10070840

Mai, T., Bistline, J., Sun, Y., Cole, W., Marcy, C., Namovicz, C., & Young, D. (2018). The role of input assumptions and model structures in projections of variable renewable energy: A multi-model perspective of the U.S. electricity system. *Energy Economics*, *76*. https://doi.org/10.1016/j.eneco.2018.10.019

Mallapragada, D. S., Papageorgiou, D. J., Venkatesh, A., Lara, C. L., & Grossmann, I. E. (2018). Impact of model resolution on scenario outcomes for electricity sector system expansion. *Energy*, *163*, 1231–1244. https://doi.org/10.1016/j.energy.2018.08.015

Nahmmacher, P., Schmid, E., Hirth, L., & Knopf, B. (2016). Carpe diem: A novel approach to select representative days for long-term power system modeling. *Energy*, *112*, 430–442. https://doi.org/10.1016/j.energy.2016.06.081

Nock, D., & Baker, E. (2019). Holistic multi-criteria decision analysis evaluation of sustainable electric generation portfolios: New England case study. *Applied Energy*, *242*(December 2018), 655–673. https://doi.org/10.1016/j.apenergy.2019.03.019

Pineda, S., & Morales, J. M. (2018). Chronological time-period clustering for optimal capacity expansion planning with storage. *IEEE Transactions on Power Systems*, *33*(6), 7162–7170. https://doi.org/10.1109/TPWRS.2018.2842093

Reichenberg, L., Siddiqui, A. S., & Wogrin, S. (2018). Policy implications of downscaling the time dimension in power system planning models to represent variability in renewable output. *Energy*, *159*, 870–877. https://doi.org/10.1016/j.energy.2018.06.160

Scott, I. J., Carvalho, P. M., Botterud, A., & Silva, C. A. (2021). Long-term uncertainties in generation expansion planning: Implications for electricity market modelling and policy. Energy, 227, 120371. https://doi.org/10.1016/j.energy.2021.120371

Sepulveda, N. A., Jenkins, J. D., de Sisternes, F. J., & Lester, R. K. (2018). The Role of Firm Low-Carbon Electricity Resources in Deep Decarbonization of Power Generation. *Joule*, *2*(11), 2403–2420. https://doi.org/10.1016/j.joule.2018.08.006

Solomon, B. D., & Krishna, K. (2011). The coming sustainable energy transition: History, strategies, and outlook. *Energy Policy*. https://doi.org/10.1016/j.enpol.2011.09.009

Spittler, N., Shafiei, E., Davidsdottir, B., & Juliusson, E. (2020). Modelling geothermal resource utilization by incorporating resource dynamics, capacity expansion, and development costs. Energy, 190, 116407. https://doi.org/10.1016/j.energy.2019.116407 van der Heijde, B., Vandermeulen, A., Salenbien, R., & Helsen, L. (2019). Representative days selection for district energy system optimisation: a solar district heating system with seasonal storage. *Applied Energy*, *248*, 79–94. https://doi.org/10.1016/j.apenergy.2019.04.030