ENVIRONMENTAL RESEARCH
ENERGY

CrossMark

OPEN ACCESS

RECEIVED
30 April 2024

REVISED
9 July 2024

ACCEPTED FOR PUBLICATION
9 August 2024

PUBLISHED
4 September 2024

PAPER

Quantifying the impact of energy system model resolution on siting, cost, reliability, and emissions for electricity generation

Anna F Jacobson[1,*], **Denise L Mauzerall**[2,3] and **Jesse D Jenkins**[4,5]

1 Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, United States of America
2 Center for Policy Research on Energy and the Environment, School of Public and International Affairs, Princeton University, Princeton, NJ, United States of America
3 Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, United States of America
4 Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ, United States of America
5 Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ, United States of America
* Author to whom any correspondence should be addressed.

E-mail: anna.f.jacobson@gmail.com

Keywords: energy system models, resolution, structural uncertainty, error, capacity expansion, benchmarks

Abstract

Runtime and memory requirements for typical formulations of energy system models increase non-linearly with resolution, computationally constraining large-scale models despite state-of-the-art solvers and hardware. This scaling paradigm requires omission of detail which can affect key outputs to an unknown degree. Recent algorithmic innovations employing decomposition have enabled linear increases in runtime and memory use as temporal resolution increases. Newly tractable, higher resolution systems can be compared with lower resolution configurations commonly employed today in academic research and industry practice, providing a better understanding of the potential biases or inaccuracies introduced by these abstractions. We employ a state-of-the art electricity system planning model and new high-resolution systems to quantify the impact of varying degrees of spatial, temporal, and operational resolution on results salient to policymakers and planners. We find models with high spatial and temporal resolution result in more realistic siting decisions and improved emissions, reliability, and price outcomes. Errors are generally larger in systems with low spatial resolution, which omit key transmission constraints. We demonstrate that high temporal resolution cannot overcome biases introduced by low spatial resolution, and vice versa. While we see asymptotic improvements to total system cost and reliability with increased resolution, other salient outcomes such as siting accuracy and emissions exhibit continued improvement across the range of model resolutions considered. We conclude that modelers should carefully balance resolution on spatial, temporal, and operational dimensions and that novel computational methods enabling higher resolution modeling are valuable and can further improve the decision support provided by this class of models.

1. Introduction

To address climate change, societies must decarbonize the electricity sector while meeting increasing demands due to electrification of transportation, building heating and cooling, industrial processes, and fuels [16]. Energy system models (ESMs)[6] are frequently used to provide decision support and guidance on decarbonization and energy transitions. These tools can be used to explore optimal configurations of resource investments, capacity siting, reliability, costs, curtailment, and more. Improving the accuracy of these modeling tools is thus critical to improve their decision support capabilities and provide reliable insights to guide cost-effective, resilient energy systems transitions [13, 36].

Most ESMs are formulated as linear programming problems (LP) or mixed-integer linear programming problems (MILP) [14, 40] due to the maturity and relative computational efficiency of solvers for these

[6] This and all other acronyms used in this paper are enumerated in table 1.

Table 1. Acronyms.

Acronym	Definition	Acronym	Definition
CONUS	Contiguous United States	CP	Carbon penalty
ESM	Energy systems model	HR	High resolution
HRS	High resolution baseline	LCOE	Levelized cost of electricity
LP	Linear programming	MILP	Mixed-integer linear programming
MSE	Mean squared error	NG	Natural gas
NGI	No government incentives	NSE	Non-served energy
SCO	Site capacity overlap	UC	Unit commitment
VRE	Variable renewable energy		

classes of problems. However, runtime and memory usage of LPs scale quadratically with model size; MILP fares even worse. A 2023 publication by Jacobson *et al* [26] ran a 6-zone MILP in 2 min for 12 weeks of data, but could not run an analogous 22-week problem within 48 h. Models that do successfully terminate may be extremely slow to do so. A 2016 publication by Frew *et al* [18] showed that increasing the number of days in a model from 168 to 365 increased runtime almost 25-fold. Even with state-of-the-art commercial solvers and computing hardware, large-scale energy systems planning models thus remain computationally constrained.

To improve runtime and memory usage, models commonly employed today in academic research and industry practice commonly decrease model size by representing operations with fewer hours, aggregating geographic regions, or ignoring physical characteristics of systems. Structural uncertainties are created by inaccuracies in how models represent the world [9], including via these simplifications. Decreasing model size by reducing model resolution thus increases errors in salient model outcomes.

Decreasing temporal resolution makes it difficult for models to capture time-dependent variables like wind, solar radiation, and demand as well as the spatiotemporal correlations introduced by larger weather patterns [7, 38]. Decreasing spatial resolution both makes it difficult to capture the full spatiotemporal covariance of weather patterns and omits potential transmission constraints [21]. Omitting physical characteristics of systems, like ramping limits on thermal powerplants, allows models to operate plants with more flexibility than is realistic [34, 37, 46]. Structural uncertainties in ESMs arise from two primary sources: low operational and spatial resolution leading to unexpected operational constraints when systems are implemented, and low temporal and spatial resolution leading to failure to capture the full temporal and spatial covariance structure of weather-dependent time series data.

Unfortunately, it is also impossible to formally bound the magnitude of errors introduced by low model resolution on key outcomes of interest [18], or even to predict the direction of errors in some cases. A model that overestimates wind availability, for example, may underestimate the number of turbines needed to meet demand and recommend unduly low investment. Alternatively, the model could recommend excessive buildout, as wind power seems more valuable. These errors in wind capacity may subsequently induce further errors in optimal storage capacity, thermal retirement, quantity of shed load, etc. The literature confirms that low spatial [4, 10, 21, 22] and temporal [18, 26, 33] resolution impact system cost, reliability, investments, and more. Error may be more likely in systems with high penetrations of variable renewable energy (VRE) [2] or undergoing transmission expansion [21]. Mid-transition systems contain a non-optimized mix of older thermal units and large-scale deployment of VRE [23] and may also be at high risk for error. Errors introduced by structural uncertainty thus have the potential to impede the decision support capabilities provided by these models in ways that are difficult to predict. Because underlying features of models are often complicated and unintuitive with variables that may have unexpected competitive or complementary impacts on each other, we cannot easily predict the results of large-scale ESMs without running them.

Improving the accuracy and quality of recommendations provided by ESMs can be extremely valuable, as energy sector investments and their climate benefits measure in the hundreds of billions of dollars annually [6]. Decarbonizing electricity sector emissions is critical in achieving net zero emissions and reducing public health impacts [30], making ESM accuracy relevant to human welfare as well. While ESMs are highly useful for decision making, planners relying on low resolution models may suggest suboptimal investments or policies. They furthermore may have difficulty justifying proposals should their models be placed under scrutiny, or should different models' uncertainties lead to conflicting recommendations.

Recent algorithmic innovations employing decomposition have improved the computational performance of ESMs [26, 29, 31, 32], enabling higher resolution modeling. Specifically, recent work [26, 35] applies a parallelized Benders decomposition algorithm to enable an electricity system planning model (with a formulation similar to the open-source GenX [27] model) that exhibits *linear* increases in runtime and memory use as temporal resolution (i.e. the number of representative operational time steps modeled) increases. Newly tractable, higher resolution systems can now be compared with lower resolution configurations commonly employed today, providing a better understanding of the potential biases or inaccuracies introduced by low resolution. Here we employ the parallelized decomposition method introduced by [26] and refined in [35] to construct a new high-resolution system (HRS) as a benchmark to better quantify the impact of varying degrees of spatial, temporal, and operational resolution on results salient to policymakers and planners, including costs, emissions, reliability, capacity investment, and local siting decisions. In absence of other methods to strictly bound output errors for a given case study [18, 26], the systematic evaluation of lower resolution systems relative to this HRS is the best way to improve our understanding of the value of improved computational methods and the magnitude of errors likely introduced by commonly employed methods to decrease resolution. Because changes in model resolution on one or more dimensions may interact in unanticipated ways, we examine combinations of model resolution with respect to the following dimensions: *spatial*, the level of geospatial aggregation and number of constrained transmission network paths; *temporal*, the length and number of representative time steps used to represent system operations; and *operational*, the inclusion of more/less realistic physical characteristics of energy systems. We carefully select cases to deeply and simultaneously probe resolution in these three dimensions. We also derive a novel accuracy metric, site capacity overlap (SCO), to quantitatively assess locational inaccuracies related to where resources are sited and built. We employ a two-phase methodology, as described in section 2, in which we downscale investment decisions from varying levels of spatial and temporal aggregation and compare operational results at a consistent spatial and temporal resolution.

Prior work has assessed ESM resolution and accuracy. Frysztacki *et al* [21], Brinkerink *et al* [10], and Aryanpur *et al* [2] analyze the impact of spatial resolution on costs and investments. None of these explore the impacts of temporal or operational resolution and their combined interaction with spatial aggregation, nor do they explore the accuracy of resource siting within regions. Frew *et al* [18] examines joint impacts of temporal resolution and spatial extent, but their largest case used is composed of only 10 zones representing the contiguous United States without UC. Poncelet *et al* [37] and Meus *et al* [34] examine the impact of operational resolution via UC representation. The former looks at investments and costs given different capacity and operating reserve scenarios. The latter examines runtime, costs, and fuel shares by UC clustering type. Each of these studies explore monolithic LPs solved with commercial solvers. The studies mentioned above are only a few in a much larger discourse on ESM accuracy (see [7, 14, 22, 36, 39, 40, 42, 46] among others) .

We go beyond previous work by probing three dimensions of resolution in varying combinations, leveraging decomposition to produce an HRS of this scale, examining accuracy of resource siting within regions, and using a two-phase methodology to examine operational outputs in high resolution. Our locational accuracy metrics render this work particularly useful to local planners. By applying locational siting accuracy metrics to cutting-edge systems and investigating resolution in three simultaneous dimensions, this work provides a unique perspective on the magnitude and nature of errors introduced by commonly-employed abstraction methods and helps demonstrate the value of further improvements in model resolution.

Ultimately, we demonstrate that high temporal resolution cannot overcome biases introduced by low spatial resolution, and vice versa. We see asymptotic improvements to total system cost and non-served energy (NSE. i.e. quantity of demand that is unable to be met by resources, an indicator of system reliability,) with increased model resolution. However, other salient outcomes, including errors in local siting and emissions, exhibit continued improvement in accuracy across the range of model resolutions considered herein. We also find some degree of non-monotonicity in accuracy improvements as resolution increases, confirming that trends are hard to extrapolate between cases of different resolutions. Given limited computational power, we recommend that modelers carefully balance spatial, temporal, and operational resolution without neglecting any dimension. We also demonstrate that novel computational methods enabling higher resolution modeling (including the parallelized Benders decomposition algorithm [26, 35] employed herein) are valuable and can further improve the decision support provided by this class of models.

2. Methodology

We use GenX [27], a detailed open-source electricity system capacity expansion model, to examine the impact of spatial, temporal, and operational resolution on outputs for a power system representing the contiguous United States (CONUS) circa 2045 under varying resolution (table 2) and considering two policy environments:

1. *No government incentives (NGI):* Current policies without additional interventions. Incentives introduced via the *Inflation Reduction Act* of 2022 are post-expiry and assumed not renewed.

Table 2. Forms of resolution tested in this study.

Dimension	General description	Resolution implementation
Spatial	Spatial characteristics of systems	Geographic zones, intraregionally homogeneous
Temporal	Time-dependent aspects of the systems (e.g. load, weather)	Days or weeks (timesteps,) reweighted to represent the entire operational year
Operational	Physical rules governing systems	Inclusion or omission of unit commitment (UC) constraints constraining thermal power plants

2. *Carbon penalty (CP):* Carbon emissions are abated exogenously (e.g. via direct air capture or land use management) for net-zero emissions. The cost of abatement is assumed to be $200 per ton [17, 19] and is implemented as a fee for exogenous abatement.

Our highest resolution system (or HRS) used for benchmarking purposes in this study is a 26-zone model of CONUS with 52 weeks of consecutive operational decisions at hourly resolution, yielding a LP with tens of millions of decision variables and constraints.

2.1. A two-phase methodology

To evaluate the impact of resolution on salient model results, we run two ESMs in series per case. We first optimize in a capacity expansion phase at varying resolutions to generate investment decisions. Operational results and siting decisions from capacity expansion model outputs are not directly comparable between cases, as emissions, profits, and generation occur in different zonal resolutions and timesteps with no direct mapping between them. In our second phase, we thus 'downscale' capacity expansion results to a consistent level of spatial resolution (26 zones) and optimize operations across a full year with hourly resolution to generate each system's operational outputs at equivalent resolution to our HRS. In this phase, we prevent further retirements and investments and run GenX in the equivalent of a production cost model. Emissions, profits, and reliability for equally high-resolution outputs can be compared between cases to explore the operational impacts once different investments are implemented. Running two optimizations allows us to examine the impact of resolution on both where resources are built and on how they are run.

Capacity expansion phase simulations are run with regional capacity reserve margins which constrain the model to maintain additional firm capacity beyond modeled demand in all periods. This constraint ensures resulting capacity is likely to be robust to omitted weather conditions and forced outages of generators. After fixing investment and retirement decisions, we omit the capacity reserve constraint in the operations phase and optimize operations to meet modeled demand at least cost.

Our specific methods for converting lower-resolution cases into 26-zone, 52-week systems are included in the supplemental methodology (appendices A.2 and A.3). Our overall methods are summarized in figure 1.

2.2. Computational resources

Simulations for both phases of the optimization are run on the Della high-performance computing cluster at Princeton University with a single node with 200 GB of memory and 53 cores. We employ a customized version of GenX in this study, which implements a parallelized Benders decomposition algorithm to separate investments and parallelize operations according to [26, 35], to enable more computationally efficient performance and higher resolution systems for analysis.

2.3. Case creation

GenX implements generators (e.g. utility scale solar, wind turbines, coal power plants,) as 'resource clusters'. Each represents a number of real world sites (for VRE) or generating units (for existing thermal plants). While sites and units are situated at co-ordinates in space, resource clusters represent an agglomeration of sites within a transmission zone and thus have no explicit location. New-build thermal clusters feature region-specific costs for each transmission zone which are assumed to be uniform within the zone. In each case, we model a consistent number of resource clusters for each resource type per modeled region. As the number of regions (and thus total number of clusters modeled) increases, the spatial extent covered by each cluster shrinks due to a decreasing number of VRE sites or existing thermal units in each cluster, and average costs for thermal new builds reflecting a smaller region. Small regions therefore lead to a better approximation of location for resource clusters, less aggregation or averaging of thermal resource costs or performance characteristics, and more accurate representation of spatio-temporal diversity in wind and solar time series profiles and transmission interconnection costs.

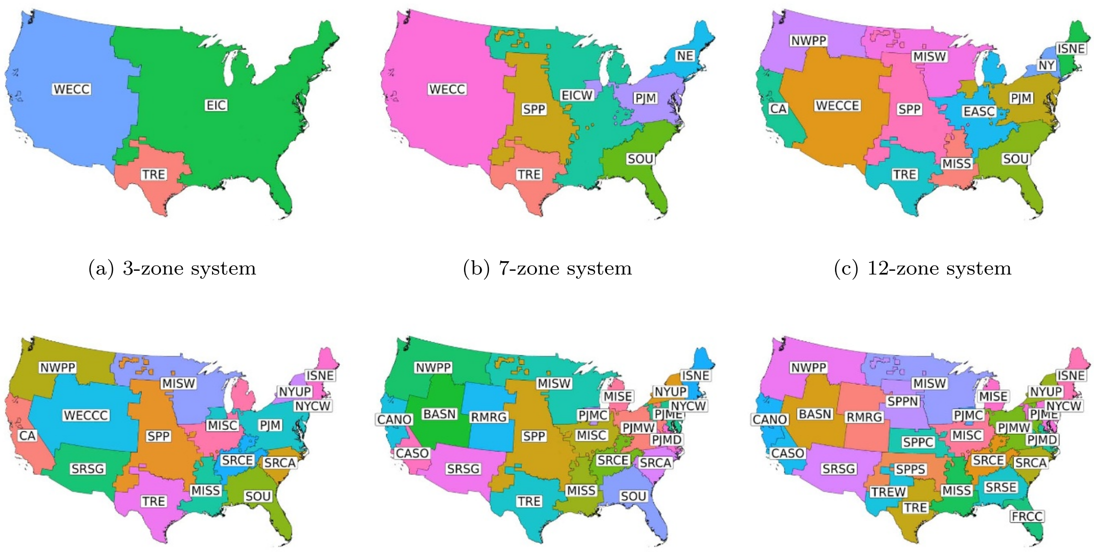
The 64 zones in CONUS and associated transmission constraints defined by the US Environmental Protection Agency's Integrated Planning Model (IPM) ([25]) are the basis for the regional aggregations used

Step		Description	Output
1		Use PowerGenome [41] to create 24 cases of GenX input with varying levels of spatial, temporal, and operational resolution for a no government incentives and carbon penalty case, Table 3.	GenX inputs, varying resolution
2		Run GenX [27], a capacity expansion model, to optimize each case. Discard operations and return optimal investments for generators, storage, and transmission.	GenX outputs, varying resolution
3		Translate investments of lower resolution cases into individual investment sites and transmission lines within the United States.	Individual resource sites
4		Turn investment sites into high resolution (HR) GenX inputs, generating 48 directly comparable investment portfolios whose resources have identical load curves and VRE profiles, though their capacity may differ due to differing levels of investment.	GenX inputs, HR
5		Prevent investments and retirements and optimize operations on the HR investment portfolios derived above.	GenX outputs, HR
6		Analyze the impact and wider implications of resolution; how do results differ case by case?	Results and discussion

Figure 1. Graphical methodology. High resolution (HR) cases are run at 26-zone, 52-week resolution with unit commitment (UC) constraints active.

in this study (figure 2) and determine resource clusters' characteristics via methods included in the appendix, section A. We used the open-source data compilation software tool PowerGenome [41] to create these cases. To create multiple temporal resolutions, we subset a full operational year into reweighted representative days or weeks with hourly resolution. We use k-means clustering as built into GenX [27] to group timesteps based on their demand and VRE profiles, and then select a representative timestep per group. We also forcibly include 'extreme' timesteps (one minimum each for solar and wind, one maximum for load) in all cases except three spatial variations of the 15-day case with representative timesteps only. We test three spatial variations of cases that omit thermal plant unit commitment (UC) decisions and constraints, while all other cases apply a linear relaxation of these UC decisions and constraints. The full suite of cases sampled is in table 3.

Because the 26-zone, 52-week case with UC is the largest run, it serves as our highest-resolution system or HRS. We note that 26-zones is still relatively low resolution for a real-world transmission network with tens of thousands of nodes, and we consider only a single year of weather-related time series data in this study. Computational and algorithmic advances may unlock higher resolution modeling; once feasible, users may wish to include more weather years or higher spatial resolution. Our HRS should not be considered 'full' resolution outside the context of this report. Still, the HRS is larger than many models actively in use for planning, policy analysis and decision support (see section 4) and is thus a useful referent to better understand the magnitude and directionality of errors introduced by commonly employed abstraction methods.



(a) 3-zone system (b) 7-zone system (c) 12-zone system

(d) 16-zone system (e) 22-zone system (f) 26-zone system

Figure 2. Region aggregations used. Name abbreviations are listed in table A1 in the appendix.

Table 3. A list of all systems tested. Cases correspond to rows, and show (left to right) temporal resolution with an indicator for inclusion of extreme timesteps, operational resolution via inclusion of unit commitment, and how many spatial resolutions are sampled. 6 spatial resolutions are simulated for cases at 52-week resolution with UC; 3 spatial resolutions are simulated for systems with lower operational or temporal resolution (figure 2). In total, 24 cases are simulated for both the the no government incentives and carbon penalty cases, with hourly resolution per-timestep. * This case serves as our highest resolution system (HRS).

	Temporal		Operational	Spatial
	Time Slices	Extreme Timesteps?	Unit Commitment?	Number of Zones Simulated
	52 Weeks	NA	Yes	3, 7, 12, 16, 22, and 26*
	52 Weeks	NA	No	3, 16, and 26
	30 Weeks	Yes	Yes	3, 16, and 26
Cases	10 Weeks	Yes	Yes	3, 16, and 26
	30 Days	Yes	Yes	3, 16, and 26
	15 Days	Yes	Yes	3, 16, and 26
	15 Days	No	Yes	3, 16, and 26

3. Results

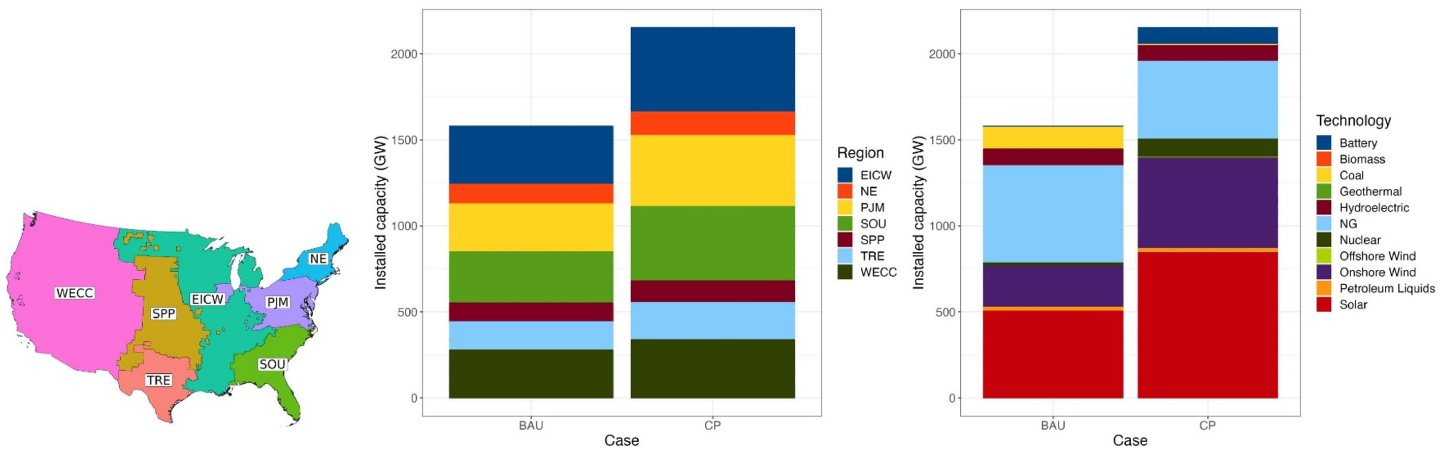
Results compare investments and operations for the NGI and CP cases by spatial, temporal, and operational resolution. Recommendations from lower resolution systems are more costly, less reliable, and have higher CO_2 emissions. Systems with low resolution cannot predict optimal VRE placement or locations of transmission bottlenecks. High resolution models benefit human welfare in multiple ways when their recommendations are implemented in the real world. Effects are more significant locally than nationally.

In section 3.1, we look at the impacts of varying resolution on system-wide metrics (cost, installed capacity). In section 3.2, we analyze differences in recommended VRE and transmission siting locations. In section 3.3, we look at how operations results vary with capacity expansion model resolution. In section 3.4, we compare the two phases of optimization to determine what information was missed by low resolution simulations in absence of high resolution spatial and temporal constraints.

3.1. Aggregate results

We see a shift in types of VRE built (e.g. solar to wind) and an overall decrease in VRE installation in lower resolution simulations. Implementing recommendations from lower-resolution systems also changes recommendations on where to invest. Investments recommended by low resolution simulations are more expensive to operate and less cost-effective.

Optimal capacity as recommended by the HRS for the CP and NGI cases is shown in figure 3, both (figure 3(a)) by region and (figure 3(b)) by technology. Introduction of a carbon penalty in the CP case



(a) Total installed capacity by region

(b) Total installed capacity by technology

Figure 3. Optimal installed capacity by region (a) and technology (b) for the highest resolution system (HRS) which has 26-zone, 52-week resolution with unit commitment (UC). Results are included for both the no government incentives (NGI) and carbon penalty (CP) cases. The carbon penalty increases total capacity 36.3%, from 1580 to 2154 GW, increases solar capacity 67.7%, and onshore wind capacity 124.2%. The carbon penalty increases fraction of buildout on the east coast. Aggregations listed in (a) are copied here from figure 2(b) and are listed in appendix table A1.

increases solar and wind buildout, coal retirement, and the proportion of installed capacity on the east coast relative to NGI case.

Systems with lower spatial resolution recommend varying VRE investment relative to the HRS (figures 4(a) and (c)) with per-technology impacts up to 140 GW in the 3-zone case. Large zones combine weather patterns across greater spaces, underestimating VRE variability. By omitting intraregional transmission constraints, large regions also fail to predict transmission bottlenecks over wider areas, leading to overestimation of resource deliverability. Both effects lead poorly spatially resolved models to overestimate VRE performance, thereby underestimating the amount of capacity needed to meet demand.

Systems with lower temporal resolution recommend less onshore wind and more solar capacity, though trends are non-monotonic (figures 4(d) and (b)). Using few timesteps misrepresents wind availability, solar availability, demand, and the temporal linkages between them. Our clustering method, when used on our data, has technology-dependent biases: appendix figure A2(a) shows that increasing temporal resolution inflates the overall availability of wind and decreases that of solar, explaining the increased installed wind capacity in capacity expansion models with lower temporal resolution. Excluding extreme timesteps (a minimal day for wind and solar, a maximum day for load) when clustering makes wind appear more available and solar less so relative to the 15-day case with extreme timesteps (appendix figure A2(a)). In the 26-zone, 15-day cases, excluding extreme timesteps encourages natural gas (NG) retirement in the NGI case due to underestimation of intermittency that must be overcome using thermal power (figure 4(b)). In the 15-day CP cases, excluding extreme timesteps increases onshore wind (due to increased availability) and battery buildout (figure 4(d)).

Omission of UC constraints discourages retirement of NG in the 26-zone 52-week NGI case, as thermal plants can mobilize all capacity at once in the absence of ramping limits, performing better than is realistically feasible. In the CP case, the ability to leverage thermal power to meet sudden spikes in demand leads the model to build less onshore wind capacity.

Investments suggested by cases with low spatial resolution in the capacity expansion phase are more costly to operate in the operations phase (figures 5(a) and (c)). Most added cost is variable (figures 5(a) and (c)). An increase in fuel costs in particular indicates that operational models, provided suboptimal investments, cannot meet demand cheaply when subjected to realistic operational constraints and must mobilize expensive resources to meet local demand. In section 3.4, we confirm that low resolution capacity expansion models have unrealistically high expectations for VRE performance. Trends are weaker for systems with low temporal resolution (figures 5(b) and (d)). While temporally low resolution systems misrepresent weather patterns, they do not fail to see transmission constraints like spatially low resolution systems do. This may explain the relative impact seen here.

3.2. Siting accuracy

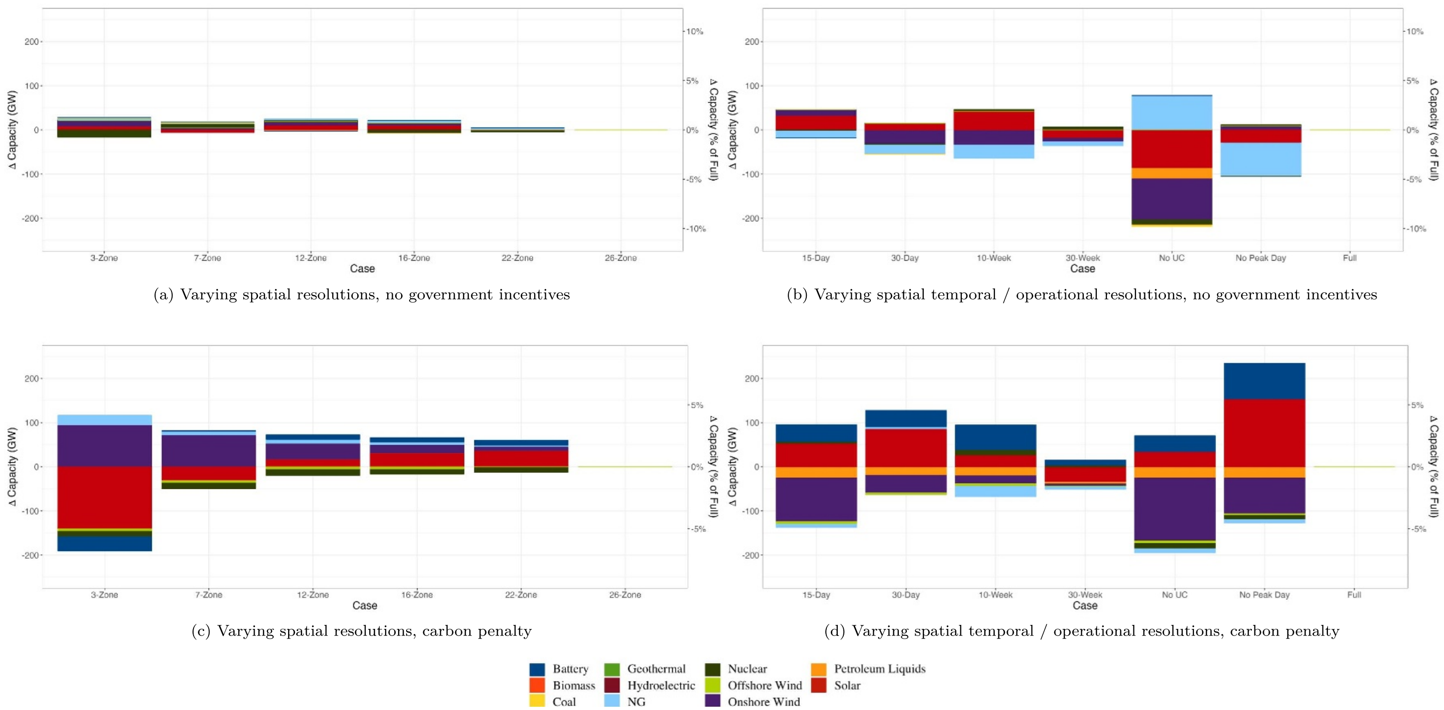
Accuracy of resource siting is strongly impacted by capacity expansion phase model resolution and is most impacted by the lowest level of resolution in the system. A model with low spatial resolution, for example, cannot recover accuracy using high temporal resolution. We measure locational accuracy for VRE using the percentage of investment sites, weighted by capacity, that are selected by both the HRS and the lower resolution case. We call this metric 'SCO' and implement it using the formula below

$$\mathrm{SCO}_{\mathrm{tech}} = 100\% \cdot \frac{\mathrm{sites}_{\mathrm{HRS,tech}} \cap \mathrm{sites}_{\mathrm{case,tech}}}{\mathrm{sites}_{\mathrm{HRS,tech}} \cup \mathrm{sites}_{\mathrm{case,tech}}}.$$

High SCO implies strong agreement between a case and the HRS in terms of where investment should occur. To any local planner, SCO is a more relevant metric than system-wide in investments. While change in total capacity is less than 10% in all cases (figure 4) SCO can be as low as 7%. Overall installed capacity, when used as an accuracy metric, significantly underestimates the impact of uncertainties on local planners.

Total transmission capacity varies little with resolution, indicating that our methods for estimating transmission routing and costs at varying degrees of resolution (including intraregional backbone networks) are reasonably accurate (figure 6). Still, transmission tends to be underbuilt near dense VRE investment (figures A4 and 7). Decreased transmission capacity in spatially low resolution regions with heavy VRE investment is indicative that spurlines as incorporated (see appendix section A.2) are insufficient in capturing the dynamics of intraregional networks and the amount of transmission needed in a system with heavy investment in renewables.

VRE buildout is densely concentrated in spatially low resolution models. Figures A4 and 7 show the sites selected in the CP case. Geographically disperse VRE is closer to multiple sources of demand and can take advantage of non-homogeneous national weather patterns. Spatially low resolution models cannot tease out these benefits; 3-zone systems have 9x fewer degrees of freedom than the HRS when selecting sites due to a more limited number of resource clusters. Models with poor spatial resolution cannot recapture local



(a) Varying spatial resolutions, no government incentives

(b) Varying spatial temporal / operational resolutions, no government incentives

(c) Varying spatial resolutions, carbon penalty

(d) Varying spatial temporal / operational resolutions, carbon penalty

Figure 4. Difference in installed capacity by technology relative to the high resolution system (HRS) with unit commitment (UC) and 26-zone, 52-week resolution. Varying zonal resolutions are shown in (c) and (a), varying operational and temporal resolutions are in (d) and (b). (a) and (c) have full temporal and operational resolution (52-weeks with UC). (b) and (d) are run at 26-zone resolution. Increasing spatial resolution generally increases solar buildout and decreases onshore wind. Trends are strongest for the 3-zone cases, though generally aggregate capacity is most impacted by temporal and operational resolution. Impacts are universally stronger for the the carbon penalty than no government incentives cases. Omission of UC decreases renewable energy buildout.

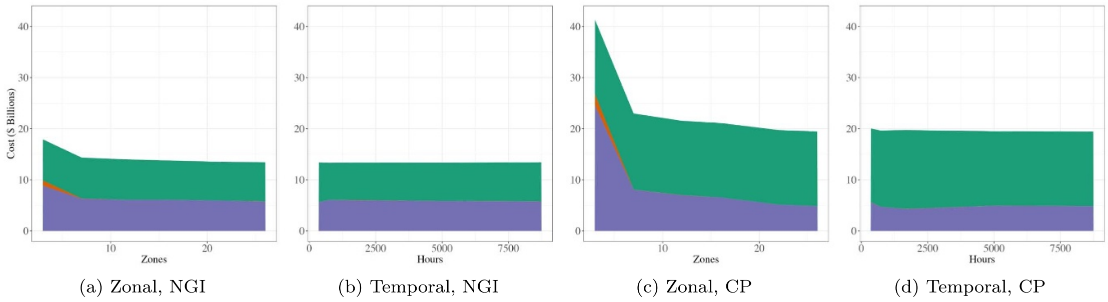


Figure 5. Total cost by spatial and temporal resolution. Cost is subdivided into fixed (e.g. investment,) variable (e.g. fuel) and non-served energy (NSE). Total cost is \$19.5 B for the carbon penalty (CP) case and \$13.4 B for the no government incentives (NGI) case. The highest resolution system (HRS) is at the extreme right of each subfigure. In the HRS, fixed costs are 75% and 56.8% of total costs for the CP and NGI cases, respectively due to a shift to low-emitting resources with heavy upfront investment and low variable costs. For the CP case, lowest spatial resolution cases have 212.5% of the cost of the highest-resolution model.

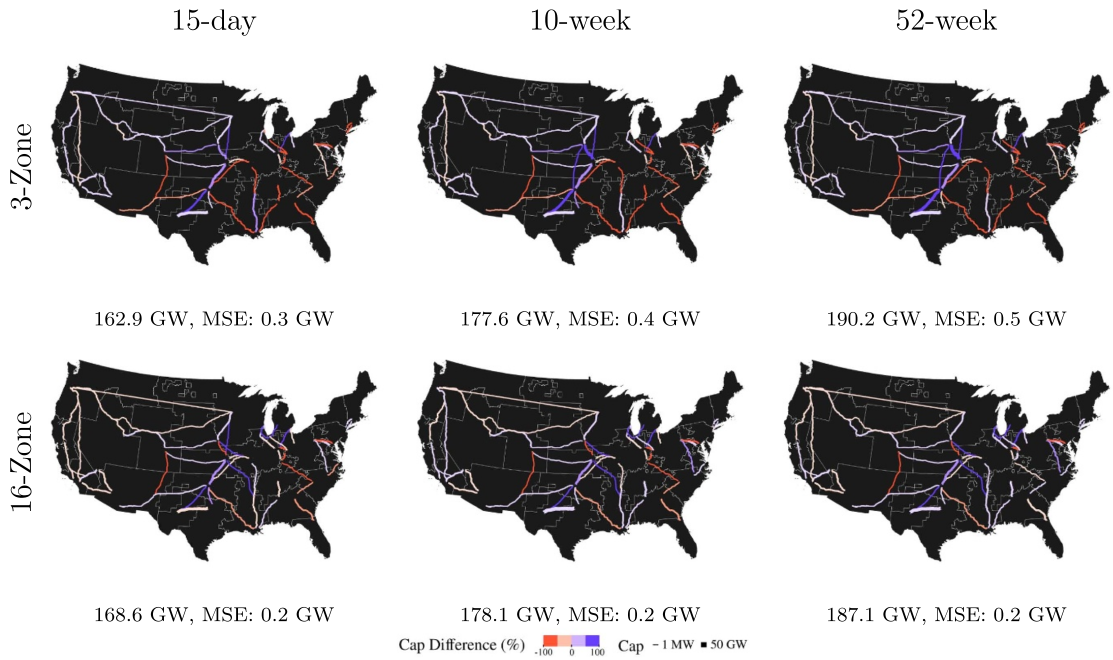


Figure 6. Differential transmission capacity by resolution, Carbon penalty case. 26-zone cases are omitted, as transmission is identical to the highest resolution system (HRS). Color of line shows the difference (%) in capacity between the cases and the HRS: $\mathrm{cap}_{\mathrm{GW}} - \mathrm{cap}_{\mathrm{GW,HRS}}$. Blue indicates the low-resolution case overinvested in the given line, while red indicates underinvestment. Line thickness shows the capacity of the line in the given case. Metrics under given cases indicate the total capacity and the mean squared error (MSE) of capacity per line. $\mathrm{MSE}_{\mathrm{cap}} = \sqrt{\sum (\mathrm{cap}_{\mathrm{GW}} - \mathrm{cap}_{\mathrm{GW,HRS}})^2} \div (\text{number of lines})$.

weather patterns and, as demand is modeled on the regional level, cannot parse the benefit of having VRE near multiple urban areas. Cheapest sites, often geographically close, receive investment per resource cluster; binning sites into fewer clusters leads to system-wide aggregation of selected sites. This explains the heavy onshore wind investment near Oklahoma (figure 7) in the 3-zone case. Adding more VRE resource clusters per region would enable models to better parse weather patterns intraregionally. However, models would still be unable to predict transmission bottlenecks, and runtime impacts may occur. The ideal number of regions and clusters per region given system characteristics should be explored experimentally before any large scale study.

While VRE buildout shifts in temporally low resolution systems, it is not geographically aggregated as it is in spatially low resolution ones. Locational inaccuracies in lower temporal resolutions are due to an inability to recapture weather patterns and correlations between VRE and demand. These systems do not have fewer degrees of freedom when selecting sites and do not ignore transmission constraints. As a result, cases with

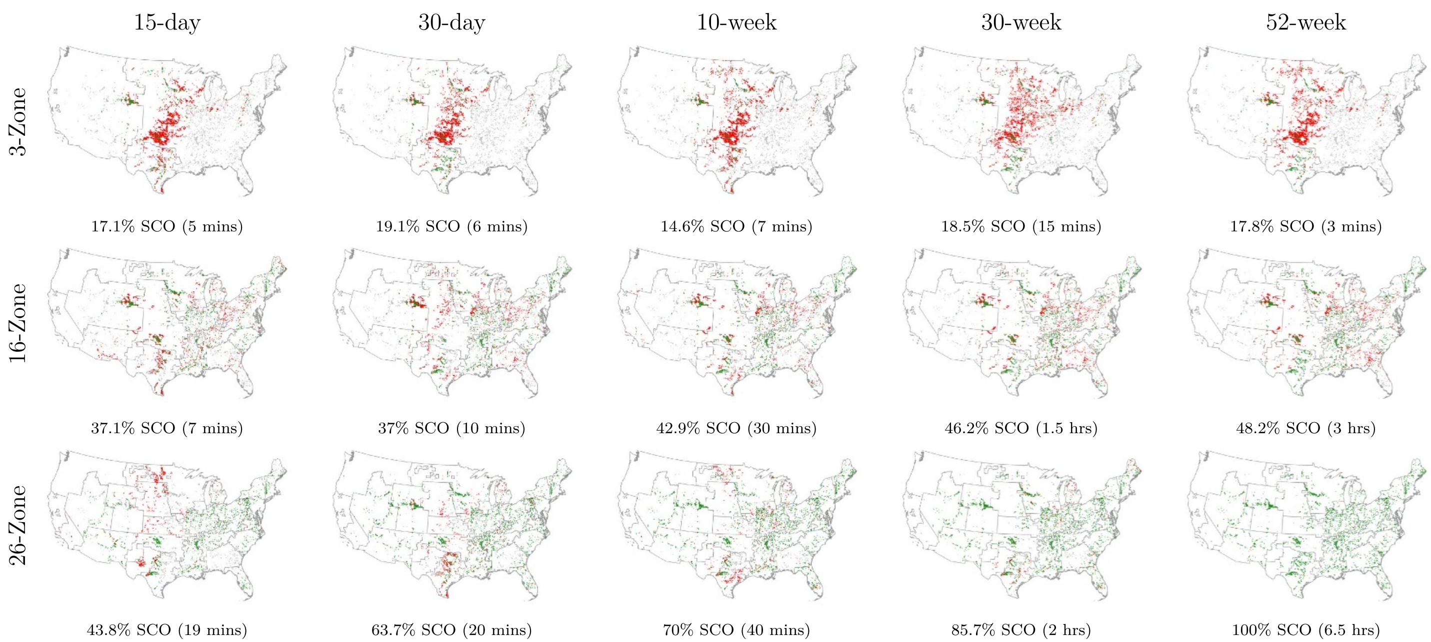


Figure 7. Difference plot for onshore wind capacity by case. Carbon penalty case. Cases with higher spatial resolution are lower on the grid; cases with higher temporal resolution are towards the right. We consider sites selected by the highest resolution system (HRS) to be optimal; site capacity overlap (SCO) is the percentage of onshore wind sites, weighted by capacity, that is invested in by both the HRS and a given low resolution case. $\mathrm{SCO} = 100\% \cdot \frac{\mathrm{sites}_{\mathrm{HRS}} \cap \mathrm{sites}}{\mathrm{sites}_{\mathrm{HRS}} \cup \mathrm{sites}}$. Sites selected by low resolution cases are shown in color: those also selected by the HRS (optimal) are in green, while those not chosen by the HRS (suboptimal) are in red. (HRS-only investments are in gray). Cases with higher resolution are better able to recapitulate siting results, selecting fewer sites that were suboptimal in (i.e. not selected by) the HRS. Next to SCO data is the runtime of the relevant optimization.

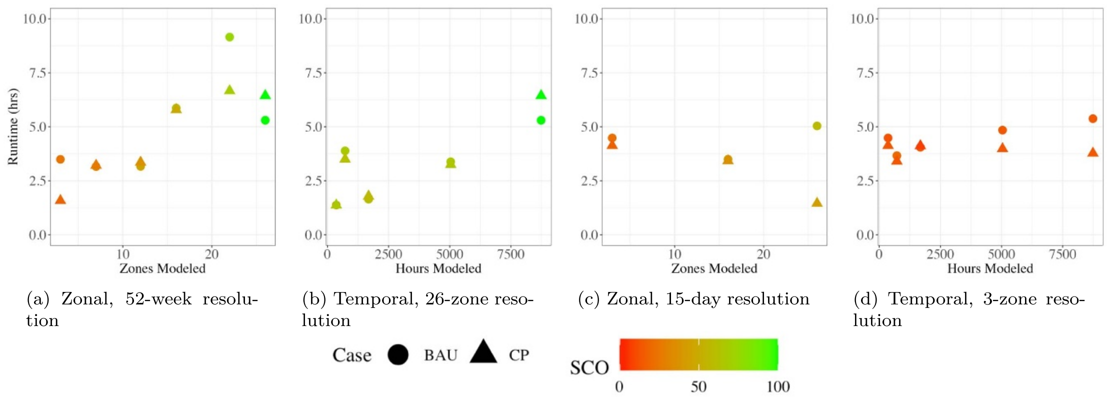


Figure 8. Resolution, accuracy, and runtime. *X*-axis is resolution, *y*-axis is runtime (hrs), and color is site capacity overlap of onshore wind, $SCO_{onshore}$. An analagous figure using SCO_{solar} is figure A5 in the appendix. Figures include both the carbon penalty (CP) and no government incentives (NGI) cases. (a) and (b) have highest resolution for the dimension not on the *x*-axis. (c) and (d) have lowest resolution for the excluded dimension. Increasing resolution improves accuracy, though this trend is masked in cases where the non-graphed dimension of resolution is low.

low temporal resolution shift capacity around per cluster within regions but still result in relatively geographically dispersed capacity.

Omission of UC in the 26-zone case leads to 75.1% SCO_{solar}, 31.0% $SCO_{onshore}$ for NGI and 72.8% SCO_{solar}, 60.5% $SCO_{onshore}$ for CP (appendix figure A3). These inaccuracies are due partially to total shifts in installation (figure 4(d)). In addition to total shifts in capacity, cases without UC decrease investment near areas with large thermal capacity; in the NGI case, VRE capacity decreases in the same regions of the mid-east with the highest coal generation in the capacity expansion phase. This is because these resources are operating better than is feasible and less VRE is needed near them to meet spikes in demand.

There are three forms of error reflected in SCO: Changes in total installed capacity by resource type or zone, changes in which resource clusters are selected within a zone, or changes from sites being reassigned to new clusters due to different levels of spatial aggregation. This last source of error is not relevant in specific zones that do not change the degree of aggregation between cases (e.g. NWPP, which is the same in both the 16- and 26-zone cases, figure 2). For the 26-zone case, SCO thus only reflects the first two sources of error listed above. It is important to note that SCO does not differentiate between capacity that has relocated across the country vs capacity that has only shifted within a county. For the most part, however, SCO in the cases tested displays minimal arbitrary short-range relocations: Changes due to selection of different clusters tends to be larger-scale, as clusters generally represent large, spatially adjacent groupings of project sites. We visually affirm minimal relocation between proximate sites by noting the rare intermixing of gray and red points on figures 7 and A4 when viewed up close, indicating few regions where the HRS and the given low-resolution case select different sites within the same local area. Still, decision makers and stakeholders should always consult a map of investment sites in tandem with the quantitative SCO metric to better gauge how far investment has moved and whether this amounts to a salient difference in geospatial siting decisions.

If one dimension of resolution is low, high resolution in other dimensions cannot overcome it; a model is able to increase its accuracy by increasing resolution in the spatial (temporal) dimension only if its resolution in the temporal (spatial) dimension is already high. Poorly spatially resolved CP cases with full temporal resolution only reach 17.8% $SCO_{onshore}$. Poorly temporally resolved models at 26-zone resolution only reach 43.8% $SCO_{onshore}$ (figure 7). Figure 8 shows that $SCO_{onshore}$ improves with a given dimension of resolution only when the run is otherwise high resolution. Lowering (figure 8(c)) spatial or (figure 8(d)) temporal resolution decreases both SCO and the overall corrolation between resolution and SCO. Similar trends persist for SCO_{solar} (appendix figures A4 and A5).

Offshore wind has high capital costs which may be prohibitive [28]; a primary benefit of offshore is its consistency relative to onshore with high availability off the east coast. The HRS CP case invested in 4 GW of offshore wind near New Jersey. The 30-week, 26-zone CP case invested in 4 GW near New Jersey, 30 MW near Maine. No other cases invested in offshore wind. Lower temporal resolution, by underrepresenting onshore wind variability, sees less benefit to offshore resources. Spatially low resolution cases overestimate the ease of transmitting VRE power to the northeast, decreasing need for local offshore wind. The trend that increasing spatial resolution leads to replacement of off- with onshore wind has been demonstrated previously [22]. Because virtually all cases had $SCO_{offshore}$ of 0% in the CP case, (100% for NGI,) a map for offshore wind is omitted here.

3.3. Operational impacts

Poor investments may lead to higher costs and emissions. Systems that have insufficient supply to meet demand in any given timestep may lead to NSE if operators must enforce rolling blackouts due to an inability to produce energy to meet the given load. When investments from low resolution capacity expansion phase systems are tested in operations models, they lead to higher costs, generator profits, NSE, and CO_2 emissions. Implementing recommendations from these low resolution systems threatens the prices and supply of electricity faced by consumers, as well as air quality across CONUS.

Investments from low resolution systems are frequently unable to meet demand. Investments from poorly spatially resolved systems incur higher NSE when operated (figure 9), indicating lower reliability. Investments from the 3-zone CP cases have up to 6.3% NSE with an average of 0.2% of demand unmet across the entire timeseries. Much of the NSE in these cases occurs in the northeast, lining up with under-investment in VRE (figures 7 and A4) in the zone ISONE and decreased transmission capacity (figure 6).

Investments from spatially aggregated models lead to higher CO_2 emissions: going from 16- to 26-zones at full temporal resolution decreases CO_2 emissions in the CP case by 30% (figure 10). According to our cost of exogenous carbon abatement, we expect to save \$12B in carbon mitigation from the carbon fee due to this improvement. When poorly sited resources from lower resolution cases cannot meet demand, models must either rely on high-emitting local resources or enforce blackouts. In our systems, NSE had costs of up to \$2,000 / MWh, leading models to prioritize demand over preventing emissions, explaining the higher emissions seen here.

Low resolution cases have higher generator profits and electricity prices. Appendix figure A8 shows resource profits grouped by region for CP cases. When demand exceeds available generation, power generators are able to inflate their prices while still finding markets for expensive electricity. This practice is called 'scarcity pricing.' High NSE and profits are highly spatially correlated across figure 9 and appendix figure A8; appendix figure A9 demonstrates high electricity prices in low-resolution cases. These trends confirm scarcity pricing at low resolutions.

3.4. Optimization phase comparison

Inaccuracies occur largely because low resolution capacity expansion models cannot predict their resources' optimal operations when high resolution operational constraints are enforced. Here we compare predicted capacity expansion phase operations with final operations from the operations phase to explore why low resolution models proposed the resource mixes that they did. Results indicate that low resolution models expect their resources to over-perform; decision makers relying on lower resolution systems are making recommendations that based on models that underestimate operational costs and overestimate VRE performance.

Spatially low resolution models overestimate VRE mobilization (figure 11(a)). These systems poorly capture weather and demand patterns, leading to altered VRE availability when modeled at high resolution. Furthermore, they ignore intraregional transmission across large areas, failing to predict bottlenecks. As a result of poor siting with concentrated areas of VRE, low resolution systems provide recommendations with low VRE consistency. Due to failure to predict bottlenecks, these systems overestimate deliverability. Impacts decrease by 16-zones.

Omission of UC leads capacity expansion models to overestimate NG generation (figure 11(a)). In GenX, thermal resources with minimum run constraints can power down fully with UC. Without UC, they are constrained to run at their minimum output value across the timeseries. This is the cause of the VRE curtailment seen in the first phase non-UC optimization: NG under minimum run constraints is unable to shut down and power up intermittently to accommodate times of high VRE generation. We confirm that NG with carbon capture (with minimum run constraints) replaces VRE in times of curtailment. We also confirm that plants with nonzero generation in the non-UC case always have capacity factor (CF) greater than the minimum output, and that this does not hold in the UC optimization.

Temporally low resolution models have marginally higher VRE operations in their investment phase (figure 11(b)). This is because they fail to predict weather patterns, leading to suboptimal siting within regions. Consistent with the rest of our results, temporally low resolution models show smaller error than spatially low resolution ones, with smaller discrepancies between optimization phases. Impacts are largely masked by 30 weeks.

Low resolution models consistently underestimate variable costs, appendix figure A7. High resolution in the operations phase of optimization is critical for predicting cost trends; low resolution models should not be trusted for cost measurements due to their misrepresentation of operational constraints.

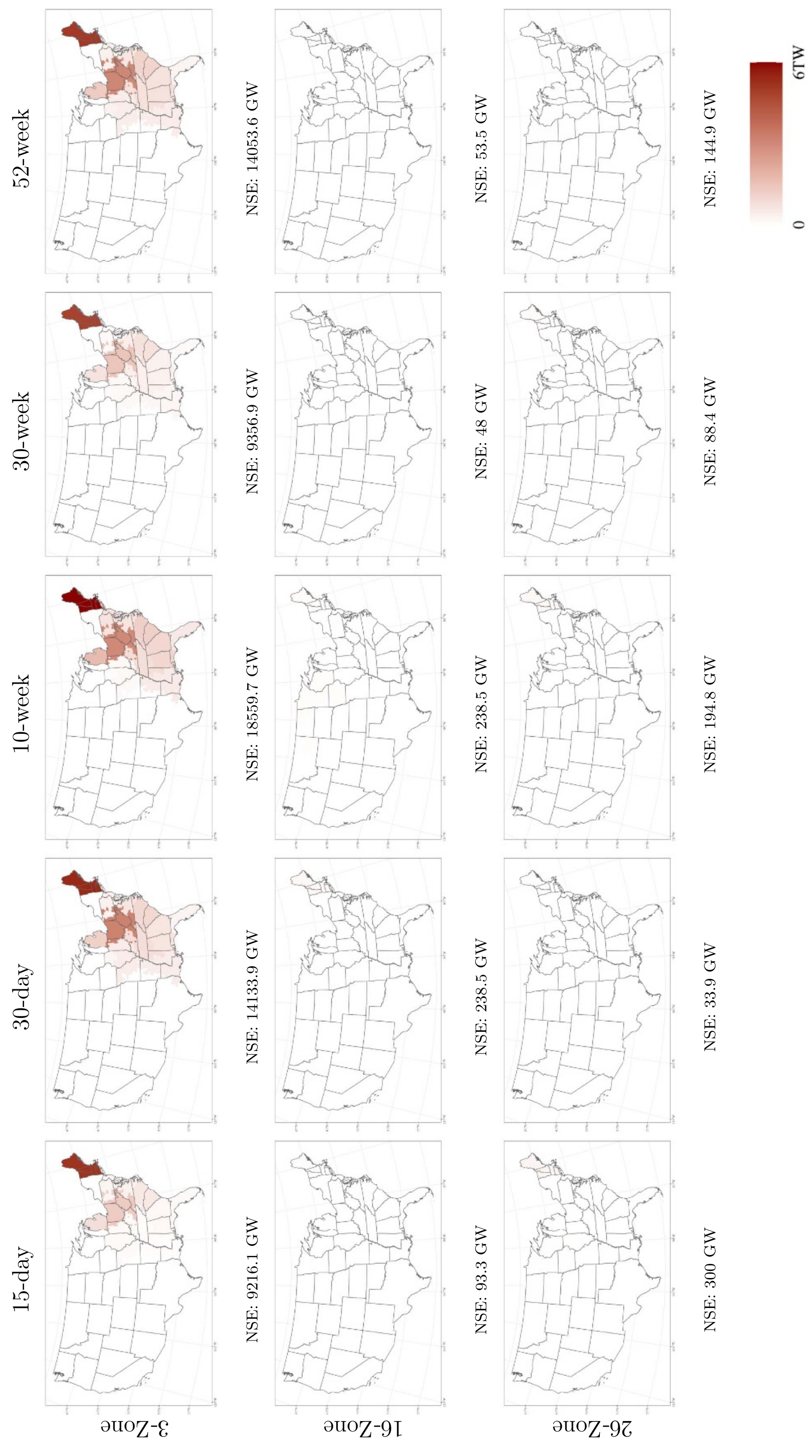


Figure 9. Systemwide non-served energy (NSE) by spatial and temporal resolution. Excessive NSE in specific regions is likely due to suboptimal local renewable resource investment and underbuilt transmission capacity, causing low generation which may need to be transmitted long distances.

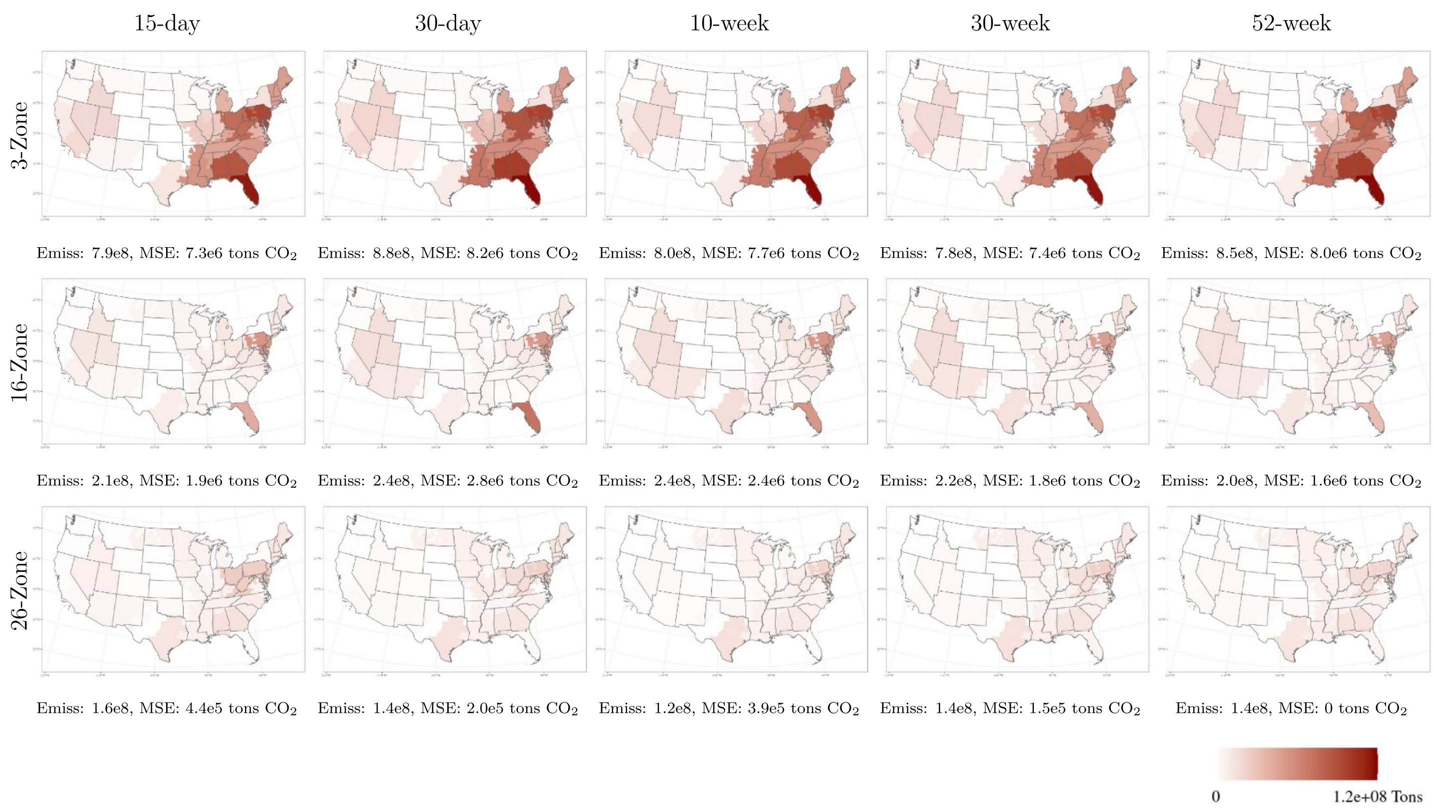


Figure 10. Systemwide CO_2 emissions by spatial and temporal resolution. Carbon penalty case. Results show decreasing total emissions as spatial and temporal resolution increases. Mean squared error (MSE) of location of emissions correlates inversely with model resolution. $\mathrm{MSE_{emiss}} = \sqrt{\sum(\mathrm{emiss} - \mathrm{emiss_{HRS}})^2} \div 26$. All emissions shown here must be abated at \$200 per ton.

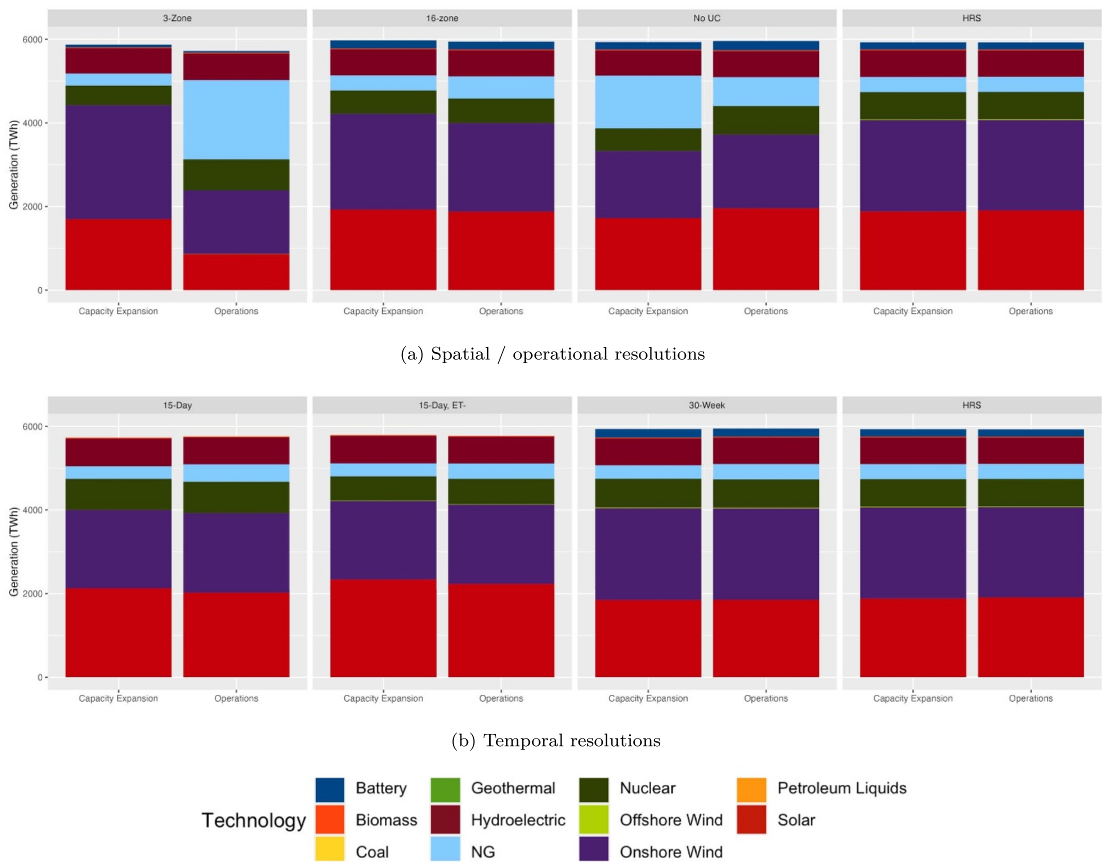


Figure 11. Generation by technology. Carbon penalty case. Figure compares operations from the capacity expansion and operations phases. For low resolution cases, first phase optimizations are misinformed about how much they can mobilize resources. Results here are for the carbon penalty case. Results for the no government incentives case show similar trends and are included as figure A10 in the appendix. HRS: High resolution system. UC: Unit commitment.

4. Discussion

Mathematical modeling is key in creating energy portfolios that improve human welfare whilst mitigating emissions. While low resolution ESMs are known to be less accurate [10, 18, 21, 22, 26], it is impossible to assert an upper-bound on degree of error [18, 26] on costs, investments, emissions, and more without running full systems. We leverage modern mathematical techniques to create new, high resolution models to better quantify errors resulting from low resolution. We prove that lower resolution systems' results are suboptimal with impacts to consumer welfare if recommendations are implemented.

Higher resolution increases accuracy and improves outcomes for prices, CO_2 emissions, reliability, and siting for renewables. Improvement for reliability (figure 9) tapers asymptotically with resolution, as our systems incur high penalties from lost load and prioritize reducing it. For other metrics, like siting accuracy or emissions, improvements are non-asymptotic due to features of the system: 26-zones is still low resolution relative to a real-world transmission network with thousands of nodes, such that even our higher resolution systems saw benefits from adding regions. Asymptotes in improvements for these other metrics may appear once larger systems are enabled by algorithmic advances. Still, because improvements persist at all levels we examine (appendix table A4,) we conclude that modelers should always increase resolution where possible, confirming the utility of recent algorithmic advances.

While accuracy improves continually as resolution is added, improvements are often non-monotonic. In some cases (figure 4) trends on capacity buildout reverse in direction as resolution is added. Runtime (figure A6) or accuracy metrics (figure 7) may decrease at intermediate resolutions before increasing again. These trends are further confirmation that quirks of LP and emergent characteristics of complex systems may lead to unexpected behavior. Our inability to precisely extrapolate to upper-bound error on variables or objective value is further reason to increase resolution to the greatest extent possible; when the direction or magnitude of biases are unknowable ad hoc, mitigating them to the greatest extent possible is the only way to increase confidence in results.

Spatial resolution is more impactful than temporal resolution at the scales tested here. In spatially low resolution models, systems are missing information on the structure of the transmission system and may propose infeasible operations as a result. Temporally low resolution models with high spatial resolution have trouble recapitulating weather patterns but do not misrepresent the feasibility of delivering power; as a result, they are more likely to result in feasible operations. That said, our systems' HRS include only one representative weather year, leading to relatively simple demands on the temporal downscaling methodologies. A study with more data may see more impact of temporal resolution, as more information would be needed to recapture the full gamut of VRE availability and load curves across multiple futures. We therefore do not conclude temporal resolution to be strictly less consequential than spatial outside the scope of this study.

We show models to be impacted by their lowest level of resolution between spatial and temporal (figures 7, 8, 10, appendix table A4). A model that has lost accuracy due to low resolution in one dimension cannot recover it using another; computational resources put into increasing resolution may be wasted if any aspect of the model is left unduly low resolution.

We note three sources of error that can diminish SCO: Changes in total installed capacity by resource type or zone, changes in which clusters are selected within a zone, and changes due to sites being reassigned to new clusters due to different levels of aggregation. The first is relevant to system planners at all scales. The second is relevant to local planners determining where to invest, as sites within a given cluster tend to be spatially proximate and thus errors in SCO due to shifts in capacity between clusters imply salient differences in associated impacts at the local level. The third source of spatial error is an artifact of changing spatial resolution; we observe that in some circumstances can result in shifts between relatively proximate locations with less salience. In general, figures 7 and A4 indicate that most error in SCO reflect longer-distance shifts of interest to local planners.

Due to computational limitations, virtually all studies are forced to decrease resolution. In appendix figure A6, we confirm the runtime trends expected from the decomposition described in Jacobson *et al* [26]: linear runtime scaling for the first phase model with temporal resolution, quadratic scaling with spatial resolution. In appendix table A3, we show that memory usage for selected cases also increases dramatically at larger scales. While our linear scaling with temporal resolution is extremely promising, the runtime and memory usage of our largest cases confirm that decreasing resolution is still necessary for the largest models.

We include some sample studies across the sector and their resolutions employed in table 4, but note that downscaling is ubiquitous outside the ones shown here. We encourage the interested reader to closely examine the methodology sections of their favorite studies for more examples. While papers provide valuable insights on decarbonization regardless of resolution employed, planners should be conscious of the structural uncertainty involved when relying on investment recommendations from any individual study. Multi-model analyses to mitigate uncertainties should be used where possible.

Decision makers should be informed on their models' structure and associated implications vis-à-vis uncertainty: how many spatial regions are in their system relative to their real-world transmission topology? How many weather years were sampled, with how many hours per year? How were timesteps selected? Are physical systems represented faithfully with high operational resolution? Users should determine a tolerance for error and valuable accuracy metrics based on model use. A planner interested in system-wide investment costs, for example, does not need fine-grained detail on operations or site locations. This user may be satisfied with our high level aggregate results (section 3.1) to explore resolution needed for their work. To a local planner who needs confidence that spending in their county is faithfully modeled in a nationwide study, our maps (figures 7, 9 and 10) will prove more informative in determining minimum appropriate modeling resolution.

It would be reductive to assume that models can be summarized by their spatial, temporal, and operational resolution. No two models are alike: Some do not incorporate consecutive hours [12], some run in a few soft-linked sectoral modules [1, 12] while others are single-sector [12, 27], some are highly user-configurable [12, 20, 27, 45]. Due to endless examples of different design decisions made by separate models, we cannot extrapolate numerical bounds on error from this study to others. Still, the trend that models are impacted by their lowest dimension of resolution should give pause to anyone relying on studies that have 'put all of their eggs in one basket' in allocation of computational resources to a single dimension of resolution. Models that allocate resolution in a balanced manner reduce overall error. Careful model selection when supporting policy will ensure that structural uncertainty does not provide costly and suboptimal recommendations to investors or other decision makers.

Our new model formulation [26, 27, 35] allows higher resolution simulations than previously possible. Despite mathematical advances, it is impossible to increase resolution *ad infinitum*. The value of this write-up is in its explorations of multiple dimensions of resolution and the relationships between them, and metrics that demonstrate resolution sufficient to support policy. Future studies should leverage

Table 4. Some example studies using ESMs with the parameter settings that they were run under. Most models represent CONUS, though two (*) represent California only, and one represents WECC (†). These systems may require less spatial resolution for accuracy, accordingly. ‡ Metrics are shown for the NEMS Electricity Market Module in particular.

Study	Year	Model	Number of Regions	Temporal Resolution
Actions for Reducing US Emissions at Least 50% by 2030 [5]	2022	EDF-NEMS	25	9 time slices
” ”		GCAM-USA-AP	51	Annual. Four non-sequential representative time blocks (electricity module)
” ”		LBNL	134	17 (investments) full annual (operations)
” ”		PATHWAYS	16	960 h
” ”		REGEN	16	Full annual hourly
” ”		USREP-ReEDS	12	17 time slices
Annual Decarbonization Perspective: Carbon-Neutral Pathways for the United States [24]	2023	RIO	27	40 days
Carbon Pricing and Energy Efficiency: Pathways to Deep Decarbonization of the US Electric Sector [11]	2019	NEMS	22‡	9 slices‡
Evaluating Impacts of the Inflation Reduction Act and Bipartisan Infrastructure Law on the US Power System [43]	2023	ReEDS	134	17 slices of 4 - 40 h
Impact of Carbon Dioxide Removal Technologies on Deep Decarbonization of the Electric Power Sector [8]	2021	US-REGEN	16	Full hourly annual
Net Zero America: Potential Pathways, Infrastructure, and Impacts, Final Report Summary [30]	2021	RIO	16	41 days, hourly resolution
Quantifying the Challenge of Reaching a 100% Renewable Energy Power System for the United States [15]	2021	ReEDS	134	17 slices of 4 - 40 h
Robust Decarbonization of the US Power Sector: Policy Options [44]	2021	ReEDS	134	17 slices of 4 - 40 h
What is Different about Different Net-zero Carbon Electricity Systems? [3]	2021	RESOLVE	3*	37 days
” ”		urbs	12*	Full annual hourly
” ”		GenX	9†	16 weeks

mathematical advances on the horizon to more deeply probe the space of resolution with even larger systems once they are tractable.

5. Conclusion

This paper uses new, high resolution modeling for a more in-depth exploration of ESM resolution than was previously possible. We find locational accuracy is more vulnerable to uncertainty in resolution than aggregate installed capacity or cost. Spatial resolution is shown to be more impactful than temporal on the scales and systems tested here. For many metrics, there is no asymptotic behavior to accuracy improvement and non-monotonic trends at mid-level resolutions, implying that more resolution is always preferable: methods for accelerating model performance are confirmed to be critical tools in ensuring accurate recommendations. Because models' accuracy is limited by their lowest-resolution dimension, users are remiss to undervalue any individual dimension (spatial, temporal, operational) of their model. In the face of unavoidable computational limitations, users should carefully balance their allocation of model resolution.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

The simulations and writing comprising this write-up were designed and completed by Anna Jacobson. Jesse Jenkins and Denise Mauzerall provided guidance for the experimental setup and format of the write-up, as well as edits to drafts once written.

The experiments run here would not have been possible without the modification of GenX coded by Filippo Pecci according to the publication jointly written by him and Anna Jacobson.

Additional thanks are due to the ZERO Lab and to Anna Jacobson's thesis committee (professors Simon Levin, Stephen Pacala, Jonathan Levine, Jesse Jenkins, and Denise Mauzerall) who provided feedback throughout the process of this work. The authors have no other pertinent conflicts of interest to share relating to the content of this report.

Funding

This work was supported by the High Meadows Environmental Institute (HMEI) through its fellowship in Science, Technology, and Environmental Policy (HMEI-STEP) and by the Princeton Carbon Mitigation Initiative through a gift from BP.

The authors have no other pertinent conflicts of interest to share relating to the content of this report.

Appendix

A.1. Supplemental methodology

Systems are broken into discrete zones, each representing region in CONUS with existing generation, storage, and transmission capacity; VRE sites and availability; and demand. Regions are based on input for the IPM [25], which subdivides CONUS based on structures established for system operation within the United States.

System aggregations (figure 2) were plugged PowerGenome. PowerGenome uses the Annual Technology Baseline from the National Renewable Energy Lab (NREL), along with data from the Energy Information Agency and the Public Utility Data Liberation Project [41]. Output from PowerGenome was plugged into the GenX [27] for capacity expansion and subsequently operations phase models.

Within each zone for the systems (figure 2) tested, investment sites are divided into separate resource clusters based on levelized cost of electricity (LCOE) and CF. Clustering sites into resource clusters involves:

1. Determining and including spurline costs per-site given the predicted cost of connecting to local urban areas.
2. Selecting candidate sites for the resource (e.g. solar photovoltaic cells) within a given region using a set of filters, see table A2.
3. Grouping sites based on LCOE and CF.

Because zones are included as monoliths within our model scheme, (demand and VRE within a zone has no geographic location,) it is not possible to include intraregional transmission explicitly. Spurline costs attempt to overcome this omission.

Urban areas with population greater than 1 million, as well as the largest one per region, are assumed to have infinite demand such that every site will has somewhere to send its generation. Decreasing systems' level of resolution decreases the number of infinite sinks across the system, thereby altering where sites can connect. This is a prime source of differences across different spatial resolutions.

For our work, sites with LCOE over \$200 – \$300 per MW (technology-dependent) are excluded. Remaining sites are put into bins by LCOE (weighted by capacity) and CF. For offshore wind turbines, only preferential sites given sociopolitical constraints are considered (table A2).

Our weather year from PowerGenome is calibrated to 2012 with VRE profiles from Vibrant Clean Energy and demand from NREL [41].

A.2. Site selection

Within GenX, a solar cluster may stand for hundreds of individual investment sites; a coal powerplant may be comprised of 10 discrete generating units; a transmission line may represent many real-world connections. We derive information on individual sites from capacity expansion phase clusters using the following methods:

Table A1. Glossary of region names. ∗ TRE represents a smaller spatial extent in the 26-zone case than in the 3 - 22-zone cases (figure 2).

Abbrev	Region	3z	7z	12z	16z	22z	26z
BASN	WECC-Basin					✓	✓
CA	California			✓	✓		
CANO	California-North					✓	✓
CASO	California-South					✓	✓
EASC	Eastern Central			✓			
EIC	Eastern Interconnect	✓					
EICW	Eastern Interconnect-West		✓				
FRCC	Florida Reliability Coordinating Council						✓
ISNE	Independent System Operator New England			✓	✓	✓	✓
MISC	Midcontinent Independent System Operator-Central				✓	✓	✓
MISE	Midcontinent Independent System Operator-East					✓	✓
MISW	Midcontinent Independent System Operator-West			✓	✓	✓	✓
MISS	Midcontinent Independent System Operator-South			✓	✓	✓	✓
NE	Northeast		✓				
NWPP	Northwest Power Pool			✓	✓	✓	✓
NY	New York			✓			
NYUP	New York Upstate				✓	✓	✓
NYCW	New York City-West				✓	✓	✓
PJMC	PJM-Commonwealth Edison					✓	✓
PJMD	PJM-Dominion					✓	✓
PJME	PJM-East					✓	✓
PJMW	PJM-West					✓	✓
PJM	Pennsylvania-New Jersey-Maryland Interconnection		✓	✓	✓		
RMRG	WECC-Rockies					✓	✓
SOU	South		✓	✓	✓	✓	
SPP	Southwest Power Pool		✓	✓	✓	✓	
SPPN	Southwest Power Pool-North						✓
SPPC	Southwest Power Pool-Central						✓
SPPS	Southwest Power Pool-South						✓
SRCA	SERC Reliability Corporation-East				✓	✓	✓
SRCE	SERC Reliability Corporation-Central				✓	✓	✓
SRSE	SERC Reliability Corporation-Southeast						✓
SRSG	Southwest Reserve Sharing Group				✓	✓	✓
TRE	Texas Reliability Entity	✓	✓	✓	✓	✓	✓*
TREW	Texas Reliability Entity-West						✓
WECC	Western Electricity Coordinating Council	✓	✓				
WECCC	Western Electricity Coordinating Council-Central				✓		
WECCE	Western Electricity Coordinating Council-East			✓			

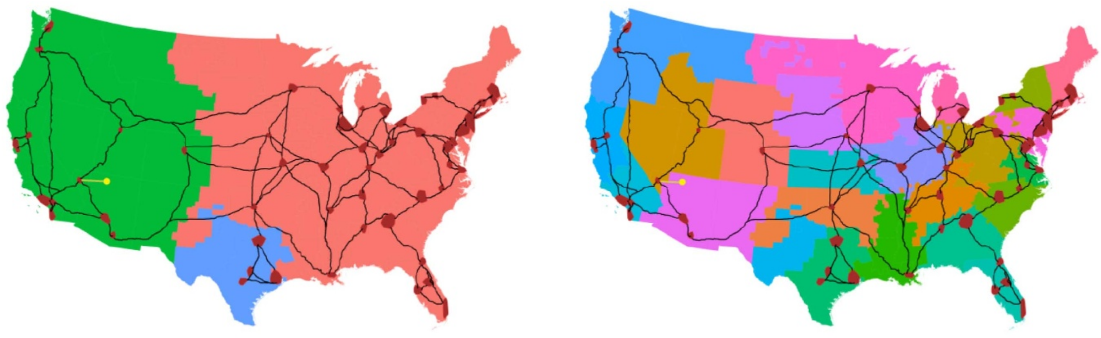
Table A2. Clustering and filter parameters used to create the resource clusters used for optimizations.

Technology	LCOE Filter	LCOE Bins (#)	CF Bins (#)	Total Bins (#)
Utility PV	⩽200 $/MW	5	3	15
Onshore wind	⩽200 $/MW	5	3	15
Offshore wind (Floating)	⩽300 $/MW	3	2	6
Offshore wind (Fixed)	⩽300 $/MW	3	2	6

VRE: Select sites belonging to VRE resource clusters according to ascending LCOE. Sites providing cheaper power receive investment first. For offshore wind, fill fixed turbines' capacity first and allocate remaining sites amongst floating clusters. We allow partial investment in the final VRE site if its full capacity would, after a cumulative sum, exceed the amount of investment of the cluster.

Thermal investments: Allocate thermal investments proportionally to subregions' demand, where a region's 'subregions' are any that, when going to a more spatially diaggregated case, lie geographically within it. (e.g. FRCC and SRSE in figure 2(f) for SOU in figure 2(e)). If, for example, FRCC has twice the demand of SRSE, 66% of thermal investment per resource in SOU is allocated to FRCC in this step, and SRSE receives the remainder.

Battery: Battery storage largely offsets the variability of VRE. We assign storage to subregions proportionally to VRE capacity. Storage is reallocated identically to thermal investments, but weighting by VRE capacity instead of demand.



(a) 3-zone case (b) 26-zone case

Figure A1. Transmission topology under two spatial resolutions. When increasing spatial resolution, some intraregional capacity becomes interregional. Black lines are backbone transmission capacity. Yellow circle represents an example renewable resource investment site connecting via a spurline to an urban area (red polygon). Real systems include tens of thousands of such sites for wind and hundreds of thousands for solar.

Thermal retirements: Because fuel used by each thermal sub unit within a resource cluster is identical, heat rates are inversely proportional to revenue; more efficient units are mobilized first. We retire the highest heat-rate (lowest-revenue / most inefficient) units per thermal plant according to a cumulative sum for retired capacity.

Transmission investments: Interregional transmission is included in GenX as pairwise connections between regions. Intraregional transmission, subdivided into 'spurlines' and 'backbone lines,' is considered when creating costs for VRE and interregional transmission but hidden from GenX itself. Spurlines, incorporated as investment costs for VRE clusters, connect individual VRE investment sites to urban areas with line capacity equal to that of the site. Backbone lines, a network of larger intraregional lines, ensure power can flow within zones to sites of demand; their capacity is assumed proportional to the interregional lines connecting to their region (figure A1).

We say a line is 'redistricted to interregional' if it lies fully within one region in lower spatial resolution but would be in multiple were the boundaries of the 26-zone system enforced. The spurline connecting to the VRE site in figure A1, for example, is redistricted interregionally when comparing figures A1(a) and (b). For a line which is redistricted to interregional, its 'relevant interregional line' is that which connects the 26-zone case regions at its endpoints.

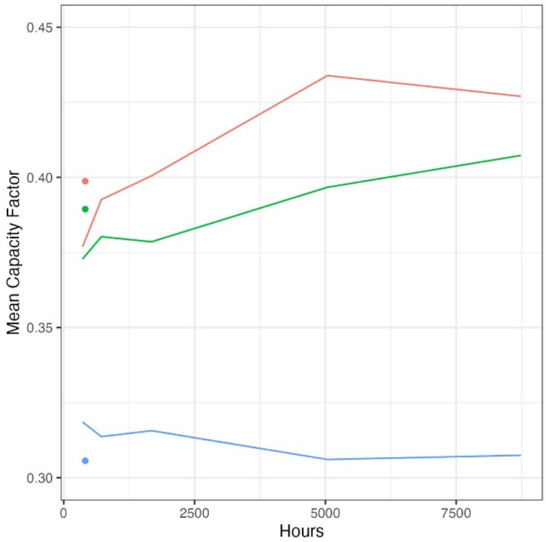
When rescaling transmission, an interregional line's capacity is the sum of:

- *Spurline transmission:* For spurlines which are redistricted to interregional, add invested capacity to the relevant interregional line.
- *Interregional transmission:* If several interregional lines in the 26-zone case correspond to one interregional line in the low resolution case, interregional investment is allocated to 26-zone lines proportionally to the total population of urban areas at their ends.
- *Backbone transmission:* For backbone lines that are redistricted to interregional, add invested capacity to the relevant interregional line.

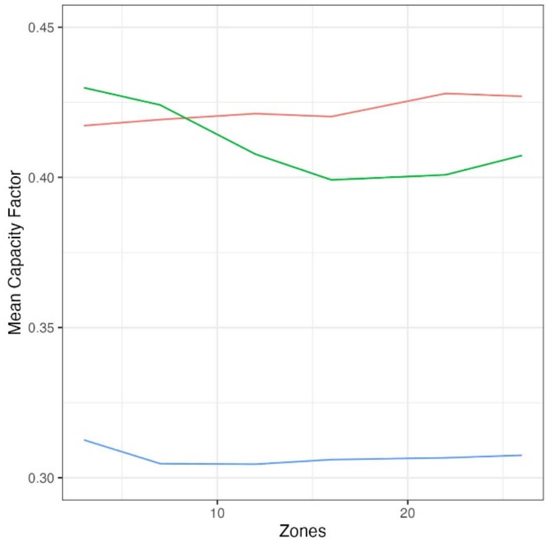
A.3. Creation of the high resolution operational model

Once sites, units, and lines have been selected, we translate to 26 zones. For VRE, we determine the cluster invested sites belong to in the 26-zone case, and add capacity accordingly. For thermal retirements, we find the 26-zone thermal clusters that units belong to, and subtract their capacity accordingly. Investments for transmission, thermal, and storage were reweighted and according to the metrics listed in section A.2. To derive new capacities, we incorporate these reweightings and allocate capacity accordingly.

A.4. Supplemental results



(a) Mean capacity factor vs temporal resolution



(b) Mean capacity factor vs spatial resolution

(c)

Figure A2. Mean capacity factor (CF) per technology by resolution. CF shows the average amount of generation (per-MW of capacity) across the entire year that a given resource can expect to generate. CF of 0.5, for example, denotes that a resource is generating an average of 0.5 MW of generation per MW of capacity across the entire timeseries. In (a), the dots represent the 15-day case with no extreme timesteps included.

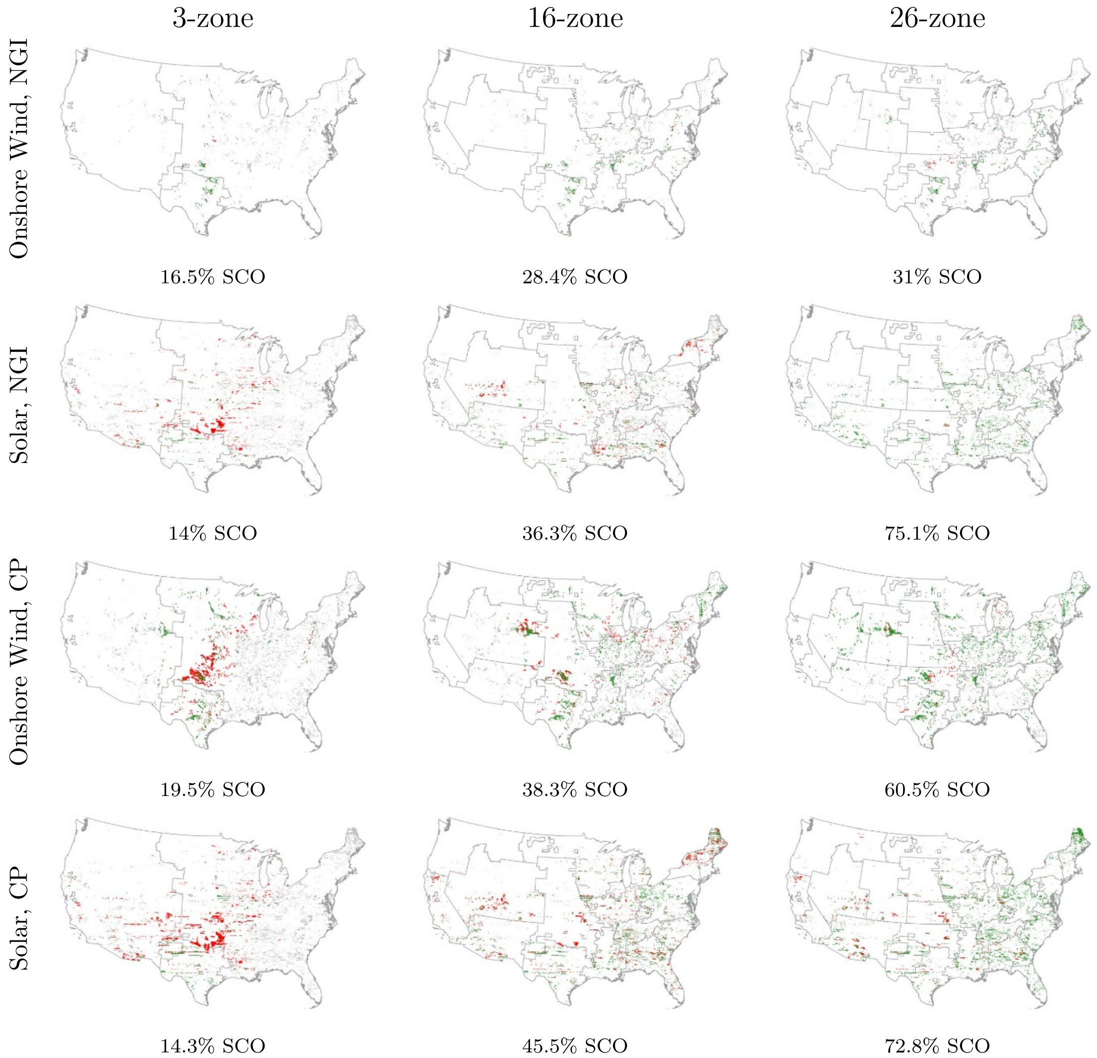


Figure A3. Difference plot for VRE capacity for the cases without unit commitment (UC) constraints for both the no government incentives (NGI) and carbon penalty (CP) cases. We consider sites selected by the highest resolution system (HRS) to be optimal; site capacity overlap (SCO) is the percentage of onshore wind sites, weighted by capacity, that is invested in by both the HRS and a given low resolution case. SCO $= 100\% \cdot \frac{\text{sites}_{\text{HRS}} \cap \text{sites}}{\text{sites}_{\text{HRS}} \cup \text{sites}}$. Sites selected by low resolution cases are shown in color: those also selected by the HRS (optimal) are in green, while those not chosen by the HRS (suboptimal) are in red. (HRS-only investments are in gray). Cases with higher resolution are better able to recapitulate siting results, selecting fewer sites that were suboptimal in (i.e. not selected by) the HRS. Cases withincreased spatial and operational resolution are better able to recapitulate siting results.

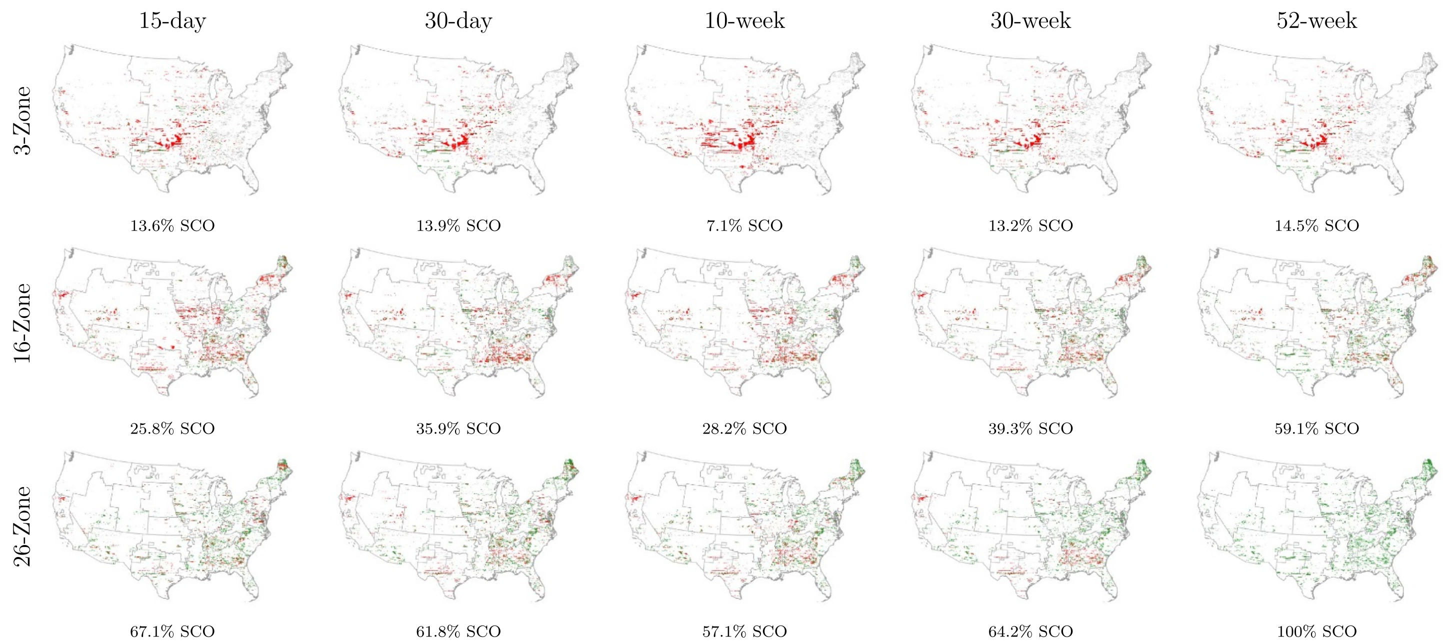
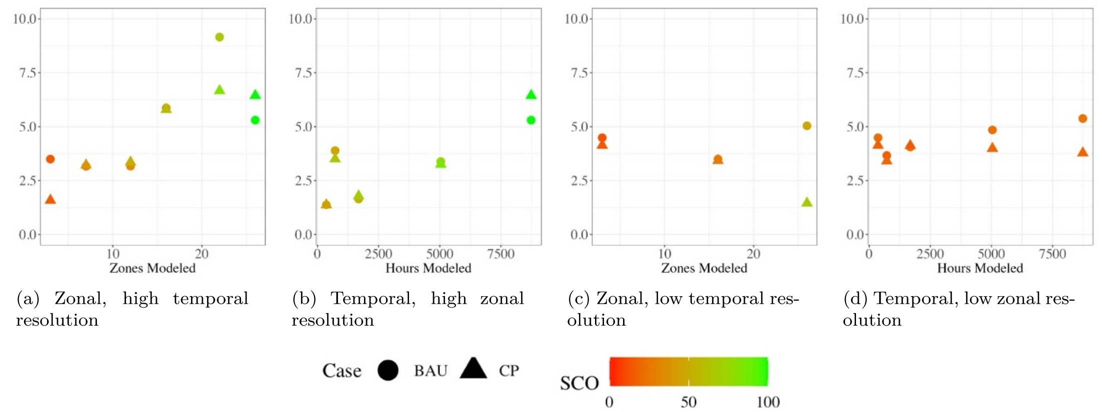
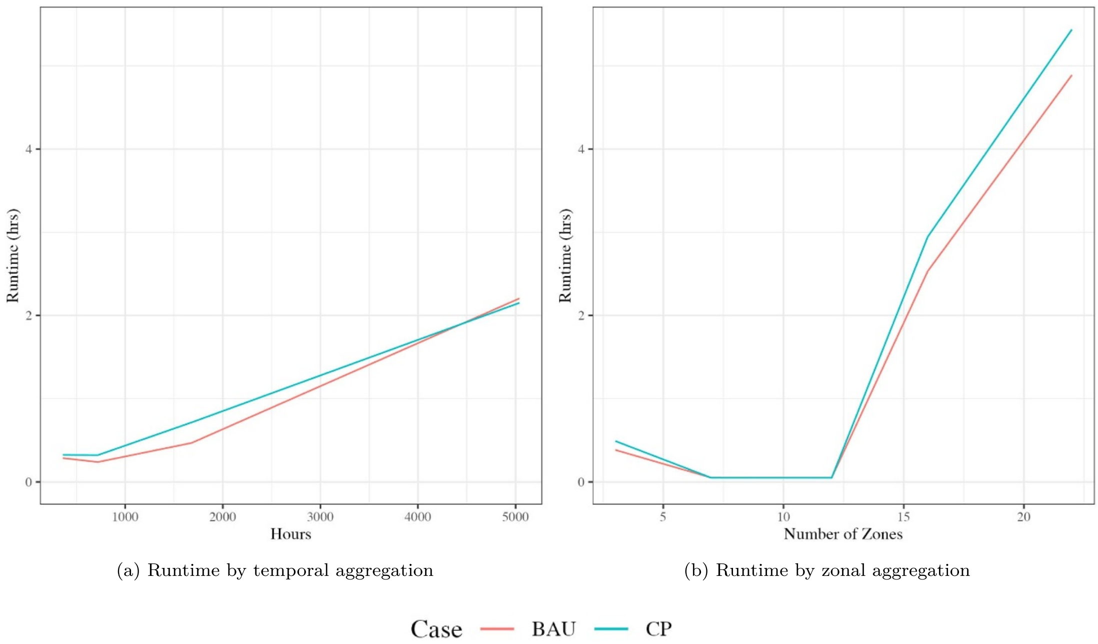


Figure A4. Difference plot for solar capacity by case. Carbon penalty case. We consider sites selected by the highest resolution system (HRS) to be optimal; site capacity overlap (SCO) is the percentage of onshore wind sites, weighted by capacity, that is invested in by both the HRS and a given low resolution case. SCO $= 100\% \cdot \frac{\text{sites}_{\text{HRS}} \cap \text{sites}}{\text{sites}_{\text{HRS}} \cup \text{sites}}$. Sites selected by low resolution cases are shown in color: those also selected by the HRS (optimal) are in green, while those not chosen by the HRS (suboptimal) are in red. (HRS-only investments are in gray). Cases with higher resolution are better able to recapitulate siting results, selecting fewer sites that were suboptimal in (i.e. not selected by) the HRS. Cases with increased spatial resolution and temporal resolution are better able to recapitulate siting results.



(a) Zonal, high temporal resolution (b) Temporal, high zonal resolution (c) Zonal, low temporal resolution (d) Temporal, low zonal resolution

Figure A5. Resolution vs runtime in hours. Points are colored reflecting the percentage of site capacity overlap for VRE technologies as a metric for accuracy, (see figures A4 and 7. SCO $= 100\% \cdot \frac{\text{sites}_{\text{HRS}} \cap \text{sites}}{\text{sites}_{\text{HRS}} \cup \text{sites}}$) figures include data for both the carbon penalty (CP) and no government incentives (NGI) cases. (a) and (b) have highest resolution (26 zones or 8736 h) for the dimension not included in the x-axis. (c) and (d) have lowest resolution (3 zones or 15 days) for the dimension not included in the x-axis. Increasing resolution improves accuracy, though this trend is severely masked in cases where the non-considered dimension of resolution is too low. An equivalent figure for onshore wind is included in the main text as figure 8.



(a) Runtime by temporal aggregation (b) Runtime by zonal aggregation

Figure A6. Runtime by zonal (b) and temporal aggregation (a) for the capacity expansion optimizations of varying resolution. Figure shows increasing runtime alongside resolution with much stronger impact of zonal resolution, in accordance with expectations laid out in [26].

Table A3. Computational resources needed for select cases, capacity expansion phase optimization.

			Temporal resolution		
	Zones	Resources	15-day	15-week	52-week
Spatial Resolution	3	244	5 min 20 GB	7 min 23 GB	3 min 20 GB
	16	976	7 min 20 GB	30 min 45 GB	3 h 100 GB
	26	1555	19 min 45 GB	40 min 50 GB	6.5 h 150 GB

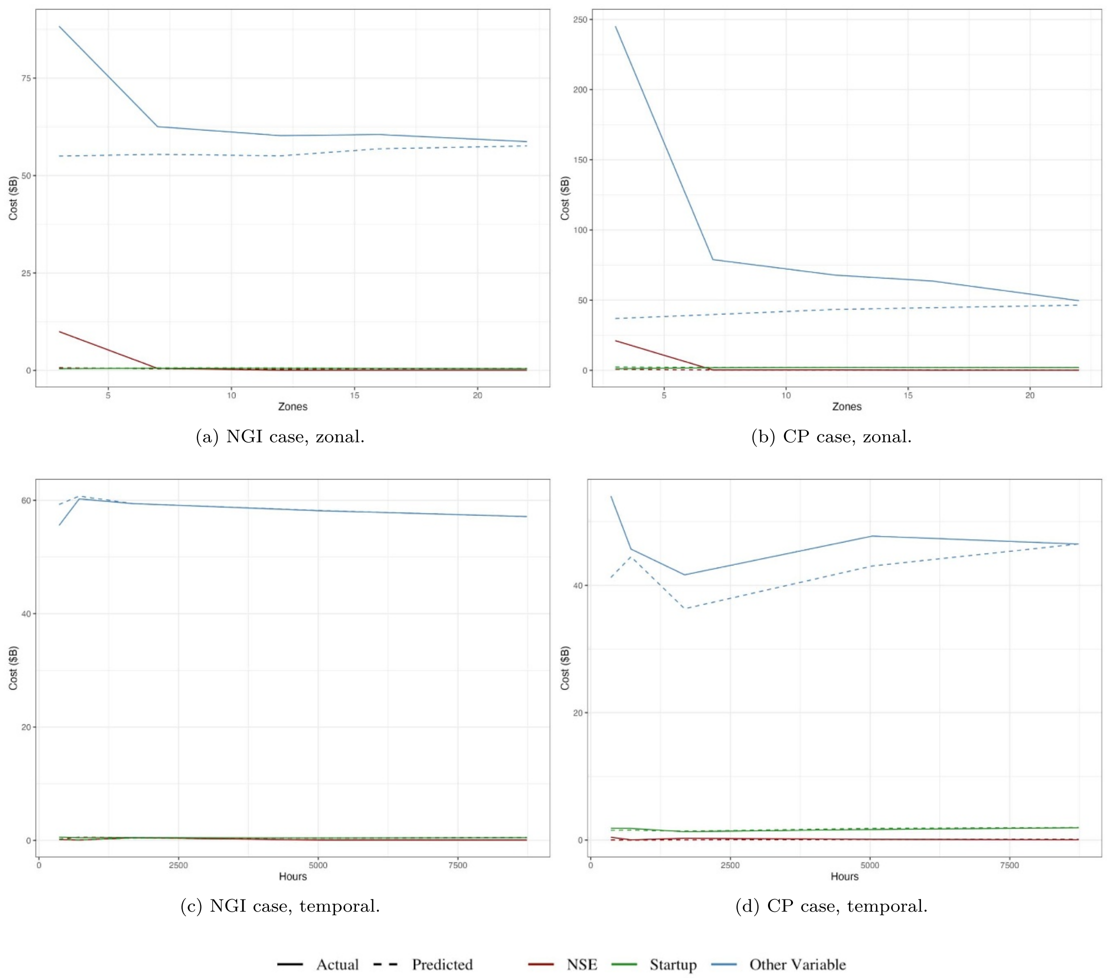


Figure A7. Comparison of operational costs between the abstracted models of capacity expansion's prediction of variable costs, and actual costs as output from operations phase. Data show that operational models consistently underestimated cost for operations when compared to actual system costs. Trends are stronger for zonal resolutions.

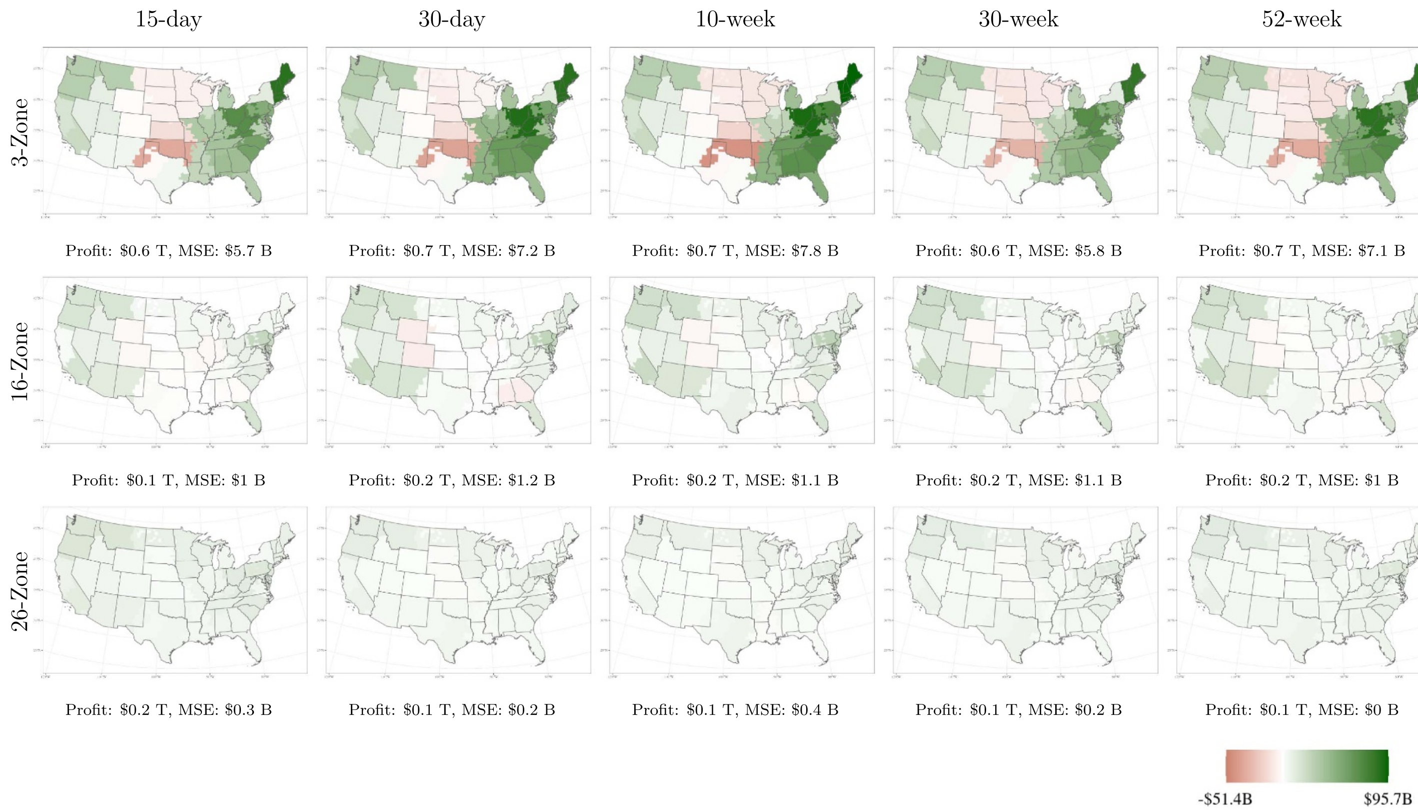
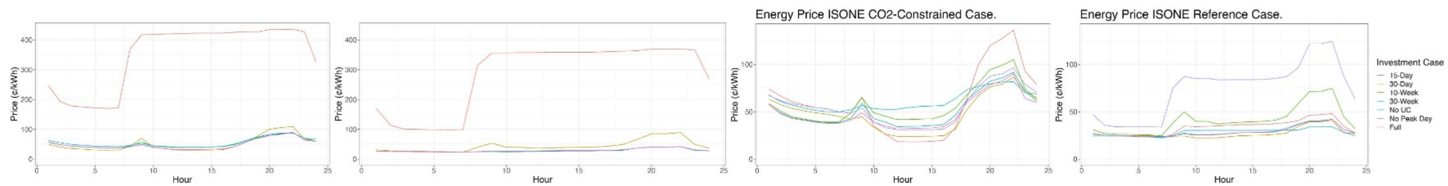
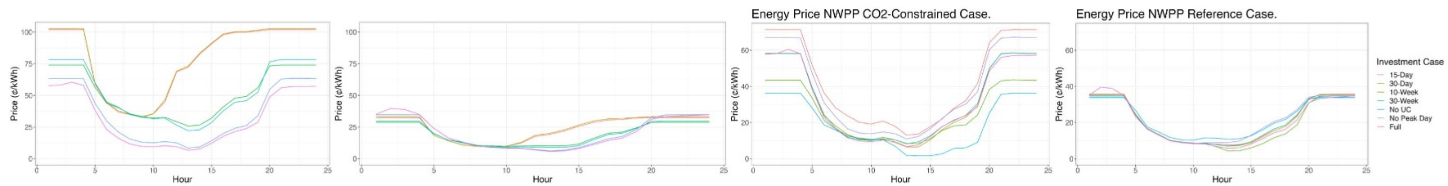


Figure A8. Total profit for all resources by region. Carbon penalty case. Figure includes the total profit in USD per system, as well as the average per-region mean squared error (MSE) of profit when compared to the highest resolution case. $\text{MSE}_{\text{profit}} = \sqrt{\sum (\text{profit} - \text{profit}_{\text{HRS}})^2} \div 26$. Trends here are due to scarcity pricing.



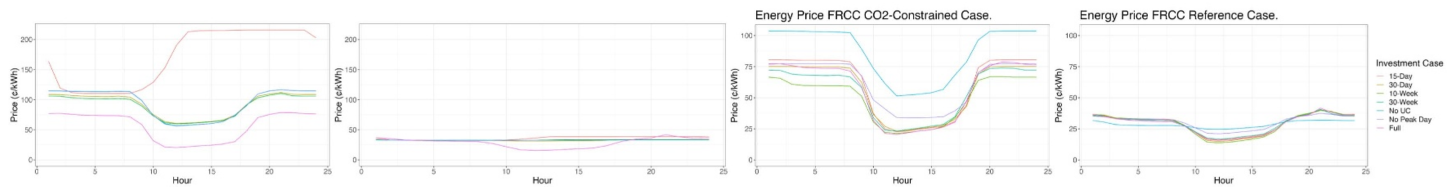
(a) Average energy price by hour, ISONE. Zonal aggregations.

(b) Average energy price by hour, ISONE. Temporal / operational aggregations.



(c) Average energy price by hour, NWPP. Zonal abstractions.

(d) Average energy price by hour, NWPP. Temporal / operational abstractions.



(e) Average energy price by hour, FRCC. Zonal abstractions.

(f) Average energy price by hour, FRCC. Temporal / operational abstractions.

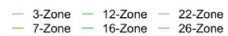
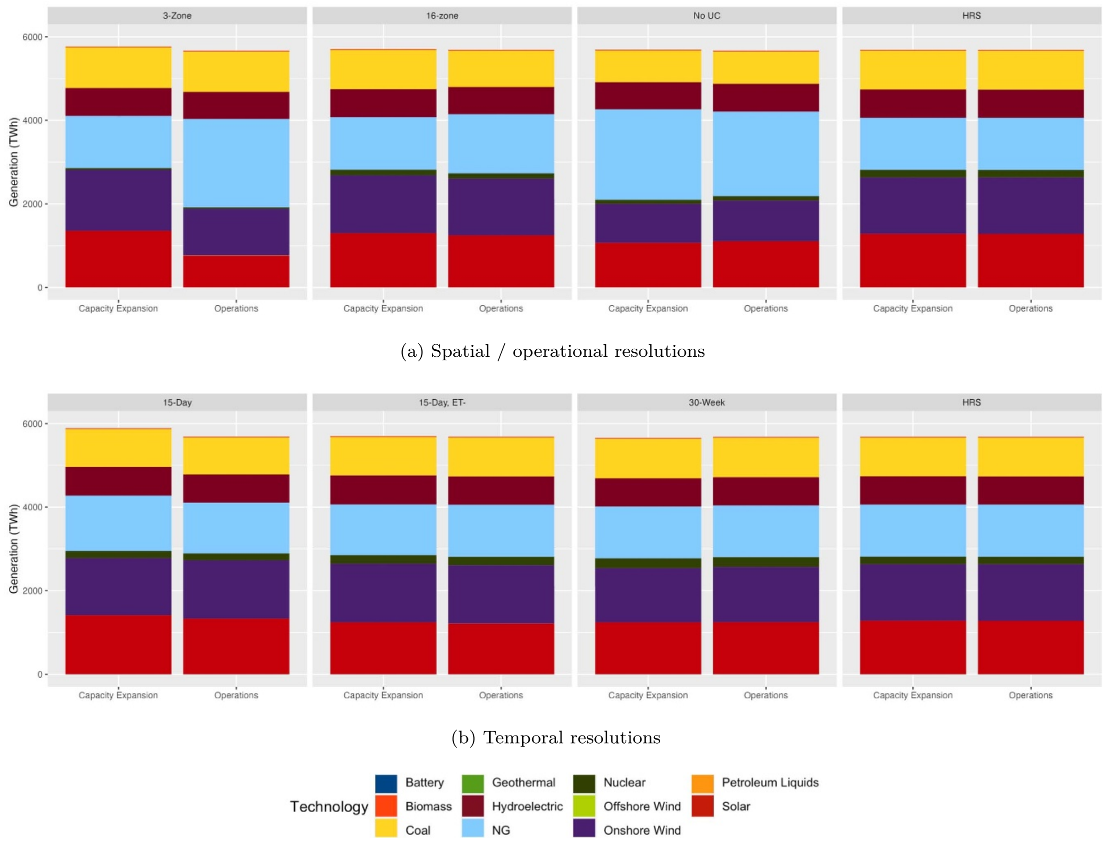


Figure A9. Average energy price by hour across the entire timeseries (time of day) for four selected regions (see figure 2). Data shown for Northeast (a) and (b) northwest (c) and (d) and Florida (e) and (f) figure shows that higher resolution models tend to have cheaper energy for consumers. Impact is region-specific and abstraction-specific in strength.



(a) Spatial / operational resolutions

(b) Temporal resolutions

Figure A10. Generation by technology. No government incentives case. Figure compares operations from the capacity expansion phase from operations with those from the operations phase. For low resolution cases, first phase optimizations (missing information about systems and their physical constraints) are misinformed about how much they can mobilize resources when operating them. Results here are for the no government incentives case. Results for the carbon penalty case are in the main text. HRS: High resolution system. UC: Unit commitment.

Table A4. A summary of some of the error metrics included in section 3. When combined with model information in table 4, this table illustrates the magnitude of some of the errors potentially being used in policy relevant modeling. Data includes site capacity overlap (SCO) for VRE (SCO $= 100\% \cdot \frac{\text{cap}_{\text{full}} \cap \text{cap}_{\text{abstracted}}}{\text{cap}_{\text{full}} \cup \text{cap}_{\text{abstracted}}}$) and per-region mean squared error (MSE) CO_2 emissions and generator profit.

	15-day	30-day	15-week	30-week	52-week
3-Zone	Solar SCO: 13.6% Onshore SCO: 17.1% Profit MSE: \$5.7 B CO_2 MSE: 7.3×10^{-6} T	Solar SCO: 13.9% Onshore SCO: 19.1% Profit MSE: \$7.2 B CO_2 MSE: 8.2×10^{-6} T	Solar SCO: 7.1% Onshore SCO: 14.6% Profit MSE: \$7.8 B CO_2 MSE: 7.7×10^{-6} T	Solar SCO: 13.2% Onshore SCO: 18.5% Profit MSE: \$5.8 B CO_2 MSE: 7.4×10^{-6} T	Solar SCO: 14.5% Onshore SCO: 17.8% Profit MSE: \$7.1 B CO_2 MSE: 8.0×10^{-6} T
16-Zone	Solar SCO: 25.8% Onshore SCO: 37.1% Profit MSE: \$1 B CO_2 MSE: 1.9×10^{-6} T	Solar SCO: 35.9% Onshore SCO: 37% Profit MSE: \$1.2 B CO_2 MSE: 2.8×10^{-6} T	Solar SCO: 28.2% Onshore SCO: 42.9% Profit MSE: \$1.1 B CO_2 MSE: 2.4×10^{-6} T	Solar SCO: 39.3% Onshore SCO: 46.2% Profit MSE: \$1.1 B CO_2 MSE: 1.8×10^{-6} T	Solar SCO: 59.1% Onshore SCO: 48.2% Profit MSE: \$1 B CO_2 MSE: 1.6×10^{-6} T
26-Zone	Solar SCO: 67.1% Onshore SCO: 43.8% Profit MSE: \$0.3 B CO_2 MSE: 4.4×10^{-5} T	Solar SCO: 61.8% Onshore SCO: 63.7% Profit MSE: \$0.2 B CO_2 MSE: 2.0×10^{-5} T	Solar SCO: 57.1% Onshore SCO: 70% Profit MSE: \$0.4 B CO_2 MSE: 3.9×10^{-5} T	Solar SCO: 64.2% Onshore SCO: 85.7% Profit MSE: \$0.2 B CO_2 MSE: 1.5×10^{-5} T	Solar SCO: 100% Onshore SCO: 100% Profit MSE: \$0 B CO_2 MSE: 0.0×10^{-0} T

ORCID iDs

Anna F Jacobson https://orcid.org/0000-0002-3929-4742
Denise L Mauzerall https://orcid.org/0000-0003-3479-1798
Jesse D Jenkins https://orcid.org/0000-0002-9670-7793

References

[1] US Energy Information Administration 2022 Electricity market module of the national energy modeling system: Model documentation 2022 (Accessed 19 January 2023)

[2] Aryanpur V, O'Gallachoir B, Dai H, Chen W and Glynn J 2021 A review of spatial resolution and regionalisation in national-scale energy systems optimisation models *Energy Strategy Rev.* **37** 100702

[3] Baik E, Chawla K P, Jenkins J D, Kolster C, Patankar N S, Olson A, Benson S M and Long J C S 2021 What is different about different net-zero carbon electricity systems? *Energy Clim. Change* **2** 100046

[4] Balta-Ozkan N, Watson T and Mocca E 2015 Spatially uneven development and low carbon transitions: insights from urban and regional planning *Energy Policy* **85** 500–10

[5] Bistline J *et al* 2022 Actions for reducing us emissions at least 50 percent by 2030 *Science* **376** 922–4

[6] Bistline J *et al* 2023 Emissions and energy impacts of the inflation reduction act *Science* **380** 1324–7

[7] Bistline J E T 2021 The importance of temporal resolution in modeling deep decarbonization of the electric power sector *Environ. Res. Lett.* **16** 084005

[8] Bistline J E T and Blanford G J 2021 Impact of carbon dioxide removal technologies on deep decarbonization of the electric power sector *Nat. Commun.* **12** 3732

[9] Bojke L, Claxton K, Sculpher M and Palmer S 2009 Characterizing structural uncertainty in decision analytic models: a review and application of methods *Value Health* **12** 739–49

[10] Brinkerink M, Mayfield E and Deane P 2024 The role of spatial resolution in global electricity systems modelling *Energy Strategy Rev.* **53** 101370

[11] Brown M A and Yufei Li 2019 Carbon pricing and energy efficiency: pathways to deep decarbonization of the US electric sector *Energy Efficiency* **12** 463–81

[12] Brown M et al 2020 Regional energy deployment system (reeds) model documentation: version 2019 *Technical Report* (National Renewable Energy Lab. (NREL), Golden, Co (United States))

[13] Brown T W, Bischof-Niemz T, Blok K, Breyer C, Lund H and Vad Mathiesen B V 2018 Response to 'burden of proof: a comprehensive review of the feasibility of 100 percent renewable-electricity systems' *Renew. Sustain. Energy Rev.* **92** 834–47

[14] Cho S, Li C and Grossmann I E 2022 Recent advances and challenges in optimization models for expansion planning of power systems and reliability optimization *Comput. Chem. Eng.* **165** 107924

[15] Cole W J, Greer D, Denholm P, Will Frazier A, Machen S, Mai T, Vincent N and Baldwin S F 2021 Quantifying the challenge of reaching a 100 percent renewable energy power system for the United States *Joule* **5** 1732–48

[16] Nature Ecology and Evolution 2021 A decisive decade *Nat. Ecol. Evol.* **5** 1465–1465

[17] Erans M, Sanz-Perez E S, Hanak D P, Clulow Z, Reiner D M and Mutch G A 2022 Direct air capture: process technology, techno-economic and socio-political challenges *Energy Environ. Sci.* **15** 1360–405

[18] Frew B A and Jacobson M Z 2016 Temporal and spatial tradeoffs in power system modeling with assumptions about storage: an application of the power model *Energy* **117** 198–213

[19] Friedmann J, Fan Z, Byrum Z, Ochu E, Bhardwaj A and Sheerazi H 2020 Levelized cost of carbon abatement: an improved cost-assessment methodology for a net-zero emissions world *Columbia University SIPA Center on Global Energy Policy: New York, NY, USA*

[20] Johnston J *et al* 2019 Switch 2.0: A modern platform for planning high-renewable power systems *SoftwareX 10* 100251 (available at: www.sciencedirect.com/science/article/pii/S2352711018301547)

[21] Frysztacki M M, Hagenmeyer V and Brown T 2023 Inverse methods: how feasible are spatially low-resolved capacity expansion modelling results when disaggregated at high spatial resolution? *Energy* **281** 128133

[22] Frysztacki M M, Horsch J, Hagenmeyer V and Brown T 2021 The strong effect of network resolution on electricity system models with high shares of wind and solar *Appl. Energy* **291** 116726

[23] Grubert E and Hastings-Simon S 2022 Designing the mid-transition: a review of medium-term challenges for coordinated decarbonization in the united states *Wiley Interdiscip. Rev.-Clim. Change* **13** e768

[24] Haley B, Jones R, Williams J, Kwok G, Farbes J, Bentz D, Schivley G and Jenkins J 2023 Annual decarbonization perspective: carbon-neutral pathways for the United States *Evol. Energy Res.* **2** e2020AV000284

[25] US EPA 2018 Documentation for epa's power sector modeling platform v6 using the integrated planning model (United States Environmental Protection Agency) (available at: www.epa.gov/sites/default/files/2018-08/documents/epa_platform_v6_documentation_-_all_chapters_august_23_2018_updated_table_6-2.pdf) (Accessed 7 January 2024)

[26] Jacobson A, Pecci F, Sepulveda N, Qingyu X and Jenkins J 2023 A computationally efficient benders decomposition for energy systems planning problems with detailed operations and time-coupling constraints *INFORMS J. Optim.* **6** 1–62

[27] Jenkins J D and Sepulveda N A 2017 Enhanced decision support for a changing electricity landscape: the GenX configurable electricity resource capacity expansion model

[28] Kota S, Bayne S B and Nimmagadda S 2015 Offshore wind energy: a comparative analysis of UK, USA and India *Renew. Sustain. Energy Rev.* **41** 685–94

[29] Lara C L, Mallapragada D S, Papageorgiou D J, Venkatesh A and Grossmann I E 2018 Deterministic electric power infrastructure planning: mixed-integer programming model and nested decomposition algorithm *Eur. J. Oper. Res.* **271** 1037–54

[30] Larson E *et al* 2021 Net-Zero America: potential pathways, infrastructure, and impacts *Final Report Summary* (Princeton University) (available at: https://netzeroamerica.princeton.edu/img/Princeton%20NZA%20FINAL%20REPORT%20SUMMARY%20(29Oct2021).pdf)

[31] Li C, Conejo A J, Liu P, Omell B P, Siirola J D and Grossmann I E 2022 Mixed-integer linear programming models and algorithms for generation and transmission expansion planning of power systems *Eur. J. Oper. Res.* **297** 1071–82

[32] Lohmann T and Rebennack S 2017 Tailored benders decomposition for a long-term power expansion model with short-term demand response *Manage. Sci.* **63** 2027–48

[33] Marcy C, Goforth T, Nock D and Brown M 2022 Comparison of temporal resolution selection approaches in energy systems models *Energy* **251** 123969

[34] Meus J, Poncelet K and Delarue E 2017 Applicability of a clustered unit commitment model in power system modeling *IEEE Trans. Power Syst.* **33** 2195–204

[35] Pecci F and Jenkins J D 2024 Regularized benders decomposition for high performance capacity expansion models (arXiv:2403.02559)

[36] Pfenninger S, Hawkes A and Keirstead J 2014 Energy systems modeling for twenty-first century energy challenges *Renew. Sustain. Energy Rev.* **33** 74–86
[37] Poncelet K, Delarue E and D'haeseleer W 2020 Unit commitment constraints in long-term planning models: relevance, pitfalls and the role of assumptions on flexibility *Appl. Energy* **258** 113843
[38] Poncelet K, Delarue E, Six D, Duerinck J and D'haeseleer W 2016 Impact of the level of temporal and operational detail in energy-system planning models *Appl. Energy* **162** 631–43
[39] Prina M G, Manzolini G, Moser D, Nastasi B and Sparber W 2020 Classification and challenges of bottom-up energy system models-a review *Renew. Sustain. Energy Rev.* **129** 109917
[40] Ringkjob H-K, Haugan P M and Marie Solbrekke I M 2018 A review of modelling tools for energy and electricity systems with large shares of variable renewables *Renew. Sustain. Energy Rev.* **96** 440–59
[41] Schivley G *et al* 2022 PowerGenome/PowerGenome: v0. 5.4 Zenodo (available at: https://ui.adsabs.harvard.edu/abs/2022zndo...6092712S/abstract)
[42] Shirizadeh B and Quirion P 2022 Do multi-sector energy system optimization models need hourly temporal resolution? a case study with an investment and dispatch model applied to france *Appl. Energy* **305** 117951
[43] Steinberg D C, Brown M, Wiser R, Donohoo-Vallett P, Gagnon P, Hamilton A, Mowers M, Murphy C and Prasanna A 2023 Evaluating impacts of the inflation reduction act and bipartisan infrastructure law on the us power system *Technical Report* (National Renewable Energy Lab.(NREL), Golden, CO (United States))
[44] Stock J H and Stuart D N 2021 Robust decarbonization of the us power sector: policy options *Technical Report* National Bureau of Economic Research
[45] Hunter K, Sreepathi S and DeCarolis J F 2013 Modeling for insight using tools for energy model optimization and analysis (Temoa) *Energy Econ.* **40** 339–49
[46] Wuijts R H, van den Akker M and van den Broek M 2024 Effect of modelling choices in the unit commitment problem *Energy Syst.* **15** 1–63