

This file contains the accepted manuscript of the following article:

James Merrick, On representation of temporal variability in electricity capacity planning models (2016). *Energy Economics* 59, 261-274.

Published version available at: <https://doi.org/10.1016/j.eneco.2016.08.001>

© 2016. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

On Representation of Temporal Variability in Electricity Capacity Planning Models

James H. Merrick

Stanford University, Department of Management Science and Engineering, 475 Via Ortega, Stanford, CA 94305, United States

Abstract

This paper systematically investigates how to represent intra-annual temporal variability in models of optimum electricity capacity investment. Inappropriate aggregation of temporal resolution can introduce substantial error into model outputs and associated economic insight. The mechanisms underlying the introduction of this error are shown. How many representative periods are needed to fully capture the variability is then investigated. For a sample dataset, a scenario-robust aggregation of hourly (8760) resolution is possible in the order of 10 representative hours when electricity demand is the only source of variability. The inclusion of wind and solar supply variability increases the resolution of the robust aggregation to the order of 1000. A similar scale of expansion is shown for representative days and weeks. These concepts can be applied to any such temporal dataset, providing, at the least, a benchmark that any other aggregation method can aim to emulate. How prior information about peak pricing hours can potentially reduce resolution further is also discussed.

Keywords: Electricity, Investment, Optimisation, Variability, Renewables

1. Introduction

The International Energy Agency (2014) reports that \$415 billion was invested globally in electricity generation capacity in 2013 and estimates, for a central scenario, that approximately \$11 trillion cumulative investment in generation capacity will be required in the period 2014-2040. Models of optimum investment, i.e. capacity planning models, can inform such investment decisions, and given both current emissions from the electricity sector and the existence of emission-free electricity technology options, can also inform the policy discussions relating to the reduction of societal greenhouse gas emissions. In so doing, the models can exist as standalone models or as components of broader Integrated Assessment Models (IAMs).

To reduce computational burden, intra-annual temporal variability in these numeric models is often represented in an aggregate form. This paper systematically considers the representation of this variability, and in so doing, four questions in particular are explored:

(a) How can aggregate representations of temporal variability introduce errors that affect model insights?

- (b) What conditions guarantee such errors are not introduced when aggregating temporal resolution?
- (c) How may an aggregated resolution be found that meets these conditions?
- (d) If the resulting resolution is too great for computational feasibility, how may the model's temporal resolution be reduced further?

An analytic and numeric exposition addresses (a), sets up our exploration of the other questions, and highlights how numeric aggregation schemes used in some applied models today can distort model output.

In addressing (b), we will see that an analytic guarantee of no aggregation error rests on the intuitive notion of finding and removing duplicate time periods. Applying this notion to an example dataset, we will see that, while the 8760 datapoints that comprise hourly electricity demand in a year can be represented in the order of 10 datapoints, the joint profiles of demand, wind, and solar power availability require in the order of a 1000 datapoints for representation to the same degree. To put it another way, the number of electricity goods (peak electricity, off-peak electricity, etc.) can increase from the order of 10 to the order of 1000 when renewable energy potential is fully included. A similar exercise is undertaken for daily and weekly temporal profiles, sug-

Email address: jmerrick@stanford.edu (James H. Merrick)

gesting that the concept of representative days or representative weeks may be inappropriate for certain applications.

For (c), clustering algorithms are a natural way to find the number of unique time periods. Numerical experimentation reveals, however, that unless the members of each cluster are very similar to each other, model outputs will be subject to aggregation error, with the level of error varying by scenario and by the output of interest. A metric for evaluating the magnitude of aggregation error is proposed as part of this discussion. Finally, a number of potential approaches are outlined for (d), but questions remain for future research.

While points related to each contribution have been made in isolated contexts in the literature, an additional contribution of this paper is to combine the ideas with a focus on the primary concern, the non-distortion of the economic insight that electricity capacity planning models are designed to provide.

The robust representation of temporal variability in a tractable model is but one of the challenges facing models for long-term electricity planning. Challenges include conceptual issues such as the appropriate treatment of uncertainty, the reconciliation between private and public discount rates, the representation of various market structures, the changing nature of electricity demand, and the treatment of power system stability in a system with high amounts of renewables. Unlike these conceptual challenges, the issue addressed in this paper can be resolved analytically, conditional on a certain classical modelling paradigm. That is, given the paradigm of representing the shapes of future electricity demand and technological availability on the basis of hourly data from a historical year, there is no ambiguity in writing down on paper a model of optimum investment. Sometimes, the policy or strategy question at hand demands numeric implementation and solution of the model. If the solution is sensitive to choices made in the numeric implementation, for example, in how to aggregate temporal resolution, the premise of this paper is that these choices have become relevant to the economic advice and are worth investigating.

Section 2 provides further context, Section 3 illustrates why the issue is relevant for economic insight, Section 4 assesses how many representative hours/days/weeks are required to provide our simple analytic guarantee, Section 5 discusses options for further reduction, while Section 6 concludes.

2. Background

2.1. Literature

With supply needing to match demand at each instant, temporal variability has always played an important role in planning electricity capacity investments. For example, [Jacoby \(1967\)](#) provides an early discussion of tractable representation of electricity demand variability in investment planning models. With dispatchable thermal generating units on the supply side, variability was only considered on the demand side. The familiar concepts of ‘baseload’, ‘intermediate’, and ‘peaking’ generators came to represent the different types of technologies required to meet the shape of electricity demand in, for example, ‘screening curve’ analysis (see [Masters \(2004\)](#)).

The importance of considering variability on the supply side in valuing variable generation sources has been emphasised before. For example, [Joskow \(2011\)](#) provides a succinct overview of the key issue - that like the availability of wind and sunshine, the value of electricity supplied at any given hour varies significantly over the course of a day, of a week, of a year. Appropriate valuation requires a consideration of how these variable profiles align. [Lamont \(2008\)](#) derives a term for the long term system value of a renewable generator that matches this concept, it being a function of both the average electricity price received by the generator and the covariance of electricity price and the generator’s availability. We will draw upon this work as a guide to the relationships an aggregate temporal representation should maintain.

[Nicolosi \(2011\)](#) compares model output across various temporal representations, finding significant sensitivity of results to degree of aggregation. Wind power capacity is exogenous in Nicolosi’s study, an important distinction we will return to. Similarly, [Ludig et al. \(2011\)](#), [Pina et al. \(2011\)](#), and [Kannan and Turtton \(2013\)](#) also assess the sensitivity of model outputs to temporal resolution, this time assessing various aggregations of hours within seasonal representative days, numerically showing the relevance of higher temporal resolution for policy analysis. While displaying the relevance of temporal representation in a relative sense, the concept of a representative seasonal day is challenging in an absolute sense in the presence of renewables, as will be shown in Section 4.3. With the advance of renewables, the meaning of a seasonal representative day is not obvious. While previously, for example, a spring weekday implied a certain demand profile, now it is not clear whether a sunny or windy spring weekday is being referred to. Section 3.2 contains a numerical exper-

iment similar to the numerical experiments conducted in these papers. It is not conducted, however, with the sole purpose of similarly illustrating the quantitative relevance of temporal representation, but to also illustrate the mechanisms underlying this relevance. The development of smart aggregation strategies will benefit from a clear understanding of these mechanisms.

With the relevance of temporal variability to valuing renewables established in the literature, models in applied settings have been representing this temporal dimension in a variety of ways. Table 1 presents a sample of various representations that have been employed. Note that neither model size nor appropriateness declines monotonically as one looks down the table. In the case of model size, the number of regions in a model, or whether a static or intertemporal optimisation framework is employed, vary across the models present. In the case of appropriateness, the appropriate level of detail is commonly espoused to be a function of the question asked, with more not necessarily better. And of course, for any given number of representative hours, we can choose wisely or poorly. Being a sample, Table 1 is not exhaustive. Its main goal is to illustrate the breadth of different ways model designers have aggregated temporal resolution. A broader categorisation is as follows:

- No representation of temporal variability
- Aggregate temporal resolution designed on basis of demand alone
- Aggregate temporal resolution designed on basis of demand, but with stochastic renewable availability to represent variability
- Aggregate temporal resolution designed on basis of demand, wind, and solar profiles
- Hourly temporal resolution with demand, wind, and solar profiles (no aggregation)

There are two important distinctions that this categorisation does not capture. The first is that, within categories, rather than attempting to represent temporal variability directly, some approaches employ ‘side constraints’ to simulate the representation. For example, even prior to the advent of renewables, the MERGE model (Manne et al., 1995) had ‘share constraints’ on power generation from coal to simulate the economic advantage of gas generation to provide power in the peaking hours not included in the model. Examples of side constraints employed within the models that participated in the Energy Modeling Forum 27 study are available in Luderer et al. (2014). Further examples include

the concept of ‘System LCOE’¹ (Ueckerdt et al., 2013), the inclusion of a calibrated flexibility constraint (Sullivan et al., 2013), and statistical calculations to assess curtailment of renewable power sources (Short et al., 2011). Edenhofer et al. (2013) make the point that there is not always systematic consideration of what these approaches gain or lose in terms of the economic questions the models are attempting to address, and along with Luderer et al. (2014), point out that share constraints on renewable power generation, for example, are challenging to credit in strong decarbonisation scenarios.² To generalise this concern, side constraints that simulate faithfully temporal variability in certain scenarios may lose their validity in other scenarios. This paper does not evaluate various side constraint approaches that exist, or propose new ones, but attempts to make clear what these side constraints are attempting to emulate with respect to temporal variability.

Secondly, of those aggregate temporal resolutions that take into account wind and solar variability in their design, there is an important distinction between those where the renewable capacity is fixed and those where the renewable capacity is endogenous. For example, de Sisternes et al. (2015) model investment in a power system, but with renewable capacity held fixed at a mandated level. In such cases, the temporal representation can be designed around the ‘net load duration curve’. When, however, the level of renewable capacity is endogenous, i.e. investment in renewable capacity can occur, the ‘net load duration curve’ is now impossible to define a priori as it is dependent on the amount of renewable capacity the model chooses to build. An aggregate resolution that is valid for any level of installed renewable capacity will thus typically be of greater magnitude than a resolution designed under the assumption that renewable capacity is fixed.

Let us now consider a number of ways employed to aggregate resolution that take into account wind and solar variability in addition to demand variability alone. van der Weijde and Hobbs (2012) sample 500 of the 8760 hours 10,000 times, then choose the set that most closely matches the statistical properties of the full dataset. While there is no guarantee that the sampling method will not miss important hour combinations, or that 500 is not too many or too few hours to sample, the method does seem to behave well empirically (Munoz and Mills, 2015). Swider and Weber (2007) use repre-

¹LCOE = Levelised Cost of Energy

²As share constraints on a technology imply an infinite price for any new investment in that technology above the arbitrary cap, no matter what the circumstance.

Table 1: Examples of temporal representations in electricity capacity planning models

Method	Model examples
Model every hour	EMMA (Hirth, 2013), SMART (Powell et al., 2012)
13 selected weeks	As presented in Aboumahboub et al. (2012)
500 hours chosen from 10,000 sample combinations	As presented in van der Weijde and Hobbs (2012)
Peak and median load day for each month	Berkeley SWITCH (Johnston et al., 2013; Fripp, 2012)
86 representative hours chosen through combination of clustering and finding extreme hours	EPRI REGEN (Blanford et al., 2014)
6 representative days with 8 time slices each	PIK LIMES-EU (Nahmmacher et al., 2014b)
Representative day for each season, 4 time slices each	NREL ReEDS (Short et al., 2011)
3 time slices from each of 3 seasons	EIA NEMS (EIA (Energy Information Administration), 2013)
4 representative hours (in USA)	JGCRI GCAM (Luderer et al., 2014)
1 average segment	Implicit in an LCOE comparison

representative demand days, but with stochastic availability of wind on those days instead of average availability, increasing the size of the model with probabilities informed by cluster analysis. Munoz et al. (2016) represent varying temporal conditions as different scenarios in a stochastic optimisation framework. Ueckerdt et al. (2015) introduce the concept of a Residual Load Duration Curve, a non-linear formulation where the duration curve is represented by a rectangle and triangle that endogenously change shape with deployment of renewables. Pappas and Webster (2014) formulate the capacity expansion problem as a stochastic optimal control problem, with variability represented by finite state Markov processes. To develop a computationally attractive aggregate problem, singular perturbation theory is drawn upon. The numerical example in the paper takes wind capacity as fixed however, using 4 net load states. We will consider how many temporal states are required when wind capacity is endogenous in Section 4. Continuing with net load duration curve representations, de Sisternes et al. (2015) conduct an exhaustive search of all combinations of 4 selected weeks to find the combination that minimises an error metric and Green et al. (2014) use a k-means clustering algorithm to aggregate hours for an operation model of the Great Britain power system. Returning to models where renewable capacity is endogenous, Munoz et al. (2016), Munoz and Mills (2015), the REGEN model (Blanford et al., 2014), and the LIMES model (Nahmmacher et al., 2014b) adopt clustering methods to choose their representative hours, with LIMES employing a hierarchical clustering method as outlined by Nahmmacher et al. (2014a). Hierarchical clustering will also be used in this

paper, not as a silver-bullet solution to finding an appropriate resolution, but as part of a broader approach to finding robust aggregate temporal structure.

The use of clustering methods has a strong foundation. A section of the operations research literature abstractly considers aggregation problems of the type of concern in this paper. Rogers et al. (1991) and Litvinchev and Vladimir (2003) provide much guidance on our problem, pointing us to results that bound the error introduced in various aggregation strategies, and highlighting the value of clustering techniques for problem reduction. We will draw upon the bounds derived by Zipkin (1980a,b) in Section 4. We can also observe that the original methods to aggregate temporal data prior to the advent of renewables implicitly followed this idea of clustering similar hours, but with hours defined on basis of demand characteristics alone.

For certain questions, maintaining the chronological relationship between hours in an aggregate formulation is an important consideration, allowing, for example, modeling of electricity storage technologies or ramping constraints on generators. A number of modeling approaches have been adopted to maintain this chronological information. The most common approach is to choose representative sequences of hours allowing chronological constraints within each sequence. Many of the example temporal representations in Table 1 refer to such sequences in the form of representative days or representative weeks. We will investigate this approach in Section 4.3. Wogrin et al. (2014) present an alternate novel approach, where chronological information about the aggregated representative hours is contained in a transition matrix. How many temporal states to

represent in the transition matrix can be informed by Section 4.2 below.

Finally, a number of the side constraints mentioned above attempt to simulate more than temporal variability alone, but also features such as generator ramping limitations, unit commitment decisions, and operating reserve requirements. These model features can be broadly categorised as the modelling of ‘system flexibility’, a concept often discussed in the context of modelling systems with large amounts of renewables (Lannoye et al., 2012; Welsch, 2013). If the appropriate representation of temporal variability happened to address system flexibility concerns then we would not need to explicitly model those system flexibility features. Motivating this thought is the observation that a number of the technologies that have the technical characteristics associated with flexibility also have economic characteristics that favour their deployment alongside variable renewable sources in models where temporal variability is captured. For example, gas generation is considered favourable from a flexibility standpoint, but with its traditional low capital costs relative to operating costs, is also economically advantageous for providing power at peak pricing hours, the number of which can increase with greater deployment of renewables. Such a finding has been reported by Bertsch et al. (2016), whereas Palmintier and Webster (2011) motivate further research by presenting a scenario with significant amounts of nuclear investment where the inclusion of unit commitment constraints are shown as relevant to planning. A recent example of such research is that undertaken by Poncelet et al. (2016).

2.2. Model structure

Mathematically, the aggregation process for a simple capacity planning model is the transformation from the detailed form³ to the aggregated form below. To focus on the issue of interest, the model has been stripped to a minimum. For example, the model is deterministic with no treatment of uncertainty, existing power sector capacity and transmission are not included, demand cannot adjust, no storage or ramping constraints are included⁴ and it is a static model in the sense that

³The ‘detailed form’ here could represent any resolution, for example second-by-second resolution. From this point onwards in the paper, hourly resolution will be referred to as the detailed benchmark to which aggregate representations will be compared. Whether more granular resolutions are more relevant for the economic questions at hand is left to further research.

⁴The lack of such constraints imply hours are independent of each other, allowing chronology to be ignored in the aggregation process.

the optimisation is conducted for some future year and not intertemporally across multiple years.

Detailed Form

Objective Function:

$$\text{minimise } Z = \sum_{g \in \mathcal{G}} (I_g cc_g + \sum_{h \in \mathcal{H}} G_{g,h} vc_g) \quad (1)$$

Subject to:

Supply must equal demand:

$$\sum_{g \in \mathcal{G}} G_{g,h} = m.d_h \quad \forall h \in \mathcal{H} \quad (2)$$

Generation must be less than available capacity:

$$G_{g,h} \leq I_g a_{g,h} \quad \forall g \in \mathcal{G}, h \in \mathcal{H} \quad (3)$$

Non-negativity constraints:

$$I_g, G_{g,h} \geq 0 \quad \forall g \in \mathcal{G}, h \in \mathcal{H} \quad (4)$$

Aggregated form

$$\text{minimise } Z = \sum_{g \in \mathcal{G}} (I_g cc_g + \sum_{p \in \mathcal{P}} G_{g,p} vc_g w_p) \quad (5)$$

$$\sum_{g \in \mathcal{G}} G_{g,p} = m.d_p \quad \forall p \in \mathcal{P} \quad (6)$$

$$G_{g,p} \leq I_g a_{g,p} \quad \forall g \in \mathcal{G}, p \in \mathcal{P} \quad (7)$$

$$I_g, G_{g,p} \geq 0 \quad \forall g \in \mathcal{G}, p \in \mathcal{P} \quad (8)$$

Where the sets are as follows:

\mathcal{G} : generators

\mathcal{H} : hours

\mathcal{P} : ‘representative’ periods ($|\mathcal{P}| \ll |\mathcal{H}|$)

And the variables:

$G_{g,h/p}$: Generation by g in hour h / period p

I_g : Investment in generator g

And the parameters:

cc_g : Annualised capital cost of g

vc_g : Variable operating cost of g

$a_{g,h/p}$: Availability factor⁵ of g in h or p

m : Maximum electricity demand

$d_{h/p}$: Fraction of m in h or p

w_p : Weight of representative period p ,

with $\sum_{p \in \mathcal{P}} w_p = |\mathcal{H}|$

Given parameter inputs, the model finds the level of investment and associated operation strategy that minimises total cost. The aggregate formulation reduces the

⁵We will assume a_g is constant for thermal generators. For variable renewable generators, a_g is the hourly availability of capacity as allowed by the weather.

number of constraints from $|\mathcal{H}|.(1+2|\mathcal{G}|)$ to $|\mathcal{P}|.(1+2|\mathcal{G}|)$, and the number of variables from $|\mathcal{G}|.(1+|\mathcal{H}|)$ to $|\mathcal{G}|.(1+|\mathcal{P}|)$. When the structure of the model is more elaborate, this reduction in model size is magnified as a function of $|\mathcal{H}|$ and $|\mathcal{P}|$.

While the relevance of capturing the temporal variability when conducting this aggregation has been recognised in the literature, generally the approaches attempting to do so do not come with analytic guarantees or systematic numerical testing of the effect of the aggregation on model output. We will provide an example of both in the following sections. As a first step however, we will conduct the following analytic and numeric exposition to build understanding of how temporal representation can affect model outputs.

3. Why temporal representation is relevant

3.1. Mechanism

The mechanism by which temporal representation affects model outputs can be understood through the lens of how temporal variability itself affects the economics of electricity sources. The marginal value of an additional unit of investment in an electricity generator can be derived from the first order conditions of the optimisation problem. As per [Lamont \(2008\)](#), with the only change being in notation, the following is an expression of the marginal value of a new unit of wind or solar power capacity in our hourly model structure:

$$\text{Marginal Value}_g = 8760[E(\lambda).CF_g + Cov(\lambda, \mathbf{a}_g)] \quad (9)$$

Where $\lambda \in \mathfrak{R}^{|\mathcal{H}|}$ is the vector of dual variables associated with the supply-demand constraint (Statement 2), or more particularly, the price of electricity. The $E(\cdot)$ operator is the time-weighted mean, $Cov(\cdot)$ is the time-weighted covariance, and CF_g (the capacity factor) = $E(\mathbf{a}_g)$ where $\mathbf{a}_g \in \mathfrak{R}^{|\mathcal{H}|}$ is the availability vector of wind or solar technology g . Applying a similar logic, we can develop the following expression for the aggregated model. The derivation is included in the supplementary material.

$$\text{Marginal Value}_g = 8760[E(\mathbf{\Lambda}).\overline{CF}_g + Cov(\mathbf{\Lambda}, \overline{\mathbf{a}}_g)] \quad (10)$$

Here $\mathbf{\Lambda} \in \mathfrak{R}^{|\mathcal{P}|}$ is the vector of the dual variables of the aggregated supply demand constraint (Statement 6), each divided by its associated weight w_p , while $\overline{\mathbf{a}}_g \in \mathfrak{R}^{|\mathcal{P}|}$ is the aggregated vector of technology availability, with $\overline{CF}_g = E(\overline{\mathbf{a}}_g)$. Time-weighted expectation and covariance in the aggregated case implies weighting the

associated vectors in the calculations with the $\mathbf{w} \in \mathfrak{R}^{|\mathcal{P}|}$ vector.

By considering when these two statements diverge, we can walk through a number of cases where the aggregate structure misvalues investment in a wind or solar power generator. For this exercise, let us assume we have an aggregate structure that captures the variability of demand. For example, assume weighted baseload, intermediate, and peaking hours can fully express the variability of electricity demand. In addition, assume we are considering the marginal value of a renewable generator before any renewable generation is installed, and that non-renewable generation has constant availability. In such a world, electricity price varies solely with electricity demand, and the solution to the aggregate model then reproduces the temporal profile of the hourly electricity price. We can then expect $E(\mathbf{\Lambda}) = E(\lambda)$. Additionally, as a weighted average of averages preserves the overall average, \overline{CF}_g equals CF_g . Returning to Statements 9 and 10, the constituent factors of the first term are then equal. We can then consider two paths to misvaluation of the wind and solar power generator through the second term:

- Under-valuation ($Cov(\mathbf{\Lambda}, \overline{\mathbf{a}}_g) < Cov(\lambda, \mathbf{a}_g)$): This can occur if hours that contribute toward a positive covariance between price (demand) and availability are missing in the aggregate representation. For example, a high demand hour with low solar availability is chosen but the high demand hour with high solar availability is not.
- Over-valuation ($Cov(\mathbf{\Lambda}, \overline{\mathbf{a}}_g) > Cov(\lambda, \mathbf{a}_g)$): This can occur if hours that contribute toward a negative covariance between price (demand) and availability are missing. For example, a high demand hour when the sun is not shining is not included.

For our next case, let us assume that Statements 9 and 10 exactly match when no renewable generation is installed, but let us relax the assumption that it is at this point alone that we are considering the marginal value. With the level of renewables deployment now endogenous, the electricity price distribution changes, and becomes no longer a function of demand alone. Both the first and second terms of the statements can then diverge leading to either under- or over- valuation in the aggregate model. Hours that are not included in the aggregate model can now become relevant to the marginal value as renewable capacity changes. The change of marginal capacity value of wind and solar power with increasing deployment has been noted of course by the aforementioned [Lamont \(2008\)](#) and, for example, by [Hirth](#)

(2013). The idea has been extended here to illustrate how the value under the aggregate numeric implementation can diverge from the true value as expressed conceptually by these papers.

We will see examples of our ‘under-valuation’, ‘over-valuation’, and ‘endogenous’ cases in the next section.

3.2. Numeric illustration

We will now illustrate the mechanisms for misvaluation numerically. To undertake this assessment, the model structures seen in statements {1 - 4} and {5 - 8} are implemented, with the associated GAMS code available in the supplementary material. To parameterise, the model of the Texas power system in Merrick (2010) is drawn upon to provide electricity demand, wind and solar data, investment costs and fuel prices.⁶ To focus on the relevance of temporal representation to investment in new generation capacity, existing capacity and transmission constraints are not included.

It is important to note that the main goal of this exercise is not to study the absolute numbers, but to provide intuition by studying the relative changes across different temporal representations, holding everything else constant. The representations chosen for this illustrative experiment are as follows:

- *8760*: Hourly resolution of demand, wind, and solar data - no aggregation.
- *S*: 144 periods, aggregation performed as inferred from the SWITCH model documentation (Johnston et al., 2013). Here, 2 days were chosen from each month: one the peak electricity demand day, the other the median. These 24 days were then sampled at 4 hour periods to arrive at 144 periods.
- *M*: 1 period, with demand and wind/solar availability represented by the mean of the hourly data. This representation is implicit when technologies are selected by comparison of Levelised Cost Of Energy (LCOE), for example as in the MERGE model (Manne et al., 1995).

The code that undertook these aggregations is included in the supplementary material. Figure 1(a)

⁶ For this exercise there is one wind profile and one solar profile that the model can choose from. Electricity demand and wind production data comes from ERCOT (Electricity Reliability Council Of Texas), while solar data comes from the National Solar Radiation Database. All temporal data is for the year 2008. Technology cost and performance data comes from EIA and NREL.

Table 2: Influence of temporal resolution on model runtime

Temporal resolution (number of periods)	Runtime for single run (s)
8760	15
144	0.09
1	<0.01

Table 3: Correlation between demand and solar availability under each sample temporal representation

8760	S	M
0.39	0.51	0

shows the load duration curve⁷ associated with each representation, while Figure 1(b) shows the ‘solar insolation duration curve’ for the potential solar investment.⁸ As expected given its design, the S representation captures the demand variability well. Additionally, it appears from the solar insolation duration curve that it also captures the solar availability reasonably well. A proposed aggregation methodology could present such charts as proof of a good fit. As we will see in our numerical experiment, what seems like a reasonable fit in terms of duration curves may in fact introduce aggregation error in model outputs. For example, a challenge with any given duration curve is that it does not provide information about correlations with other temporal profiles. Table 3 compares the correlation between electricity demand and solar availability across the representations. We can see that choosing hours based on demand alone happens to overestimate the correlation under the S representation for this particular dataset (the same aggregation methodology could underestimate this correlation for a different dataset). Meanwhile the simple M representation underestimates the correlation in this case. As we have seen however, it is the correlation with price that counts, for which correlation with demand is only a proxy that holds at low levels of installed wind and solar power capacity.

Our first comparison, the results of which are displayed in Table 2, illustrates part, if not all, of the motivation for aggregated temporal representation - the

⁷The 2008 data is multiplied by a factor of 1.5 to represent load growth. This may or may not strike the reader as high, but it is important to emphasise that it is the shape that counts for our discussion.

⁸The duration curve is developed for each representation by simply sorting the elements of each associated profile.

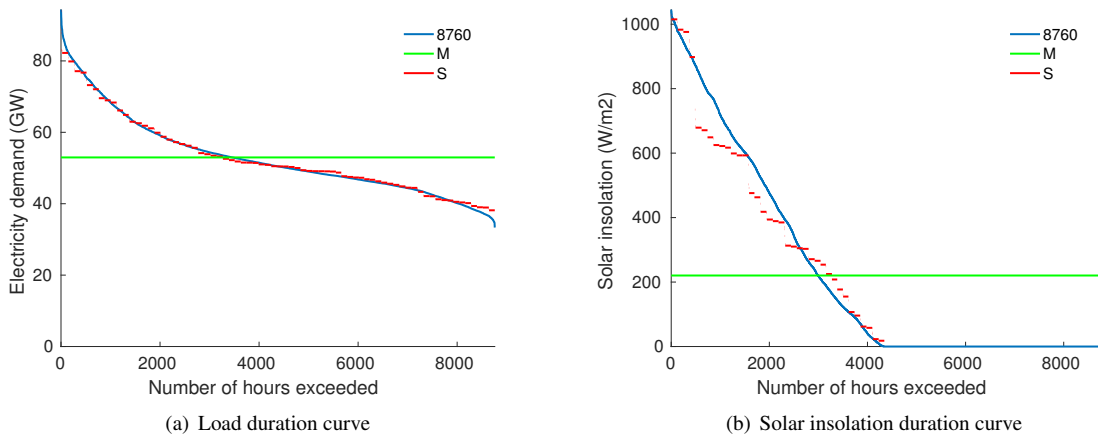


Figure 1: Sample duration curves in numerical model under each temporal representation

possibility of significant decreases in model runtime. The relative, and not the absolute, numbers are important here, as the computational penalty associated with greater resolution in Table 2 will increase for models of greater detail and geographic scope.⁹

In this illustrative experiment, everything in the model is held constant across runs with the exception of the temporal representation employed, and outputs are subsequently compared. This comparison is undertaken for two scenarios. In our first scenario, which we now consider, we have no carbon price and no price on air pollutants, but a low capacity cost for photovoltaics (PV) of approximately \$1/W. Figure 2(a) provides an example of how temporal representation can matter in appropriate valuation of PV technology. An order of magnitude more PV capacity is installed under the S representation than under the 8760 representation, while no PV is installed under the M representation. The difference between the model outputs under the 8760 and S representations could be critically important for some questions, or within the range of acceptable tolerance for others. As seen in Table 3, the S representation happens to overestimate the correlation between PV production and electricity demand, missing hours when demand is high but PV production is low, and thus mis-values as in our over-valuation case in Section 3.1. The M representation on the other hand, only considering the average PV availability across the whole year, underestimates the correlation between photovoltaic production and electricity demand, under-valuing PV as in our under-valuation case in Section 3.1.

⁹These runs were conducted on an 8 core 2.7GHz machine with 32GB RAM.

Figure 2(a) also allows us to see capacity investment in other technologies, while Figure 2(b) presents generation by technology across our temporal representations. These figures indicate the importance of temporal representation for valuing ‘conventional’ technologies also. Under the most aggregate temporal representation, M, we see that the technology with the lowest LCOE is chosen, which for this set of assumptions is coal. Under our disaggregated S and 8760 representations, we see gas capacity is installed in addition to coal, to contribute power for the peak demand hours that the aggregate M representation does not include.

We can see in Figure 3(a) installed capacity in our second scenario where PV costs are even lower than our first scenario (to \$0.5/W). We can notice that under the M representation in this scenario, the model now builds out more PV than under the 8760 representation. This is the *opposite* result to our first case. Here we have an example similar to our endogenous case in Section 3.1. The M representation is missing hours that become important to the valuation of solar power as solar deployment increases, resulting in the model both under-valuing and over-valuing solar power, depending on the scenario, under the same M temporal representation. This somewhat extreme example shows that it is not the case that a particular aggregation always under-values or over-values, but that it can do either, depending on the particular parameterisation. In contrast to the M representation in this case, the model ‘knows’ under the 8760 and S representations that hours exist where the sun is not shining, and that other electricity sources must be installed to provide electricity in those hours. Considering this case alone, the S representation seems

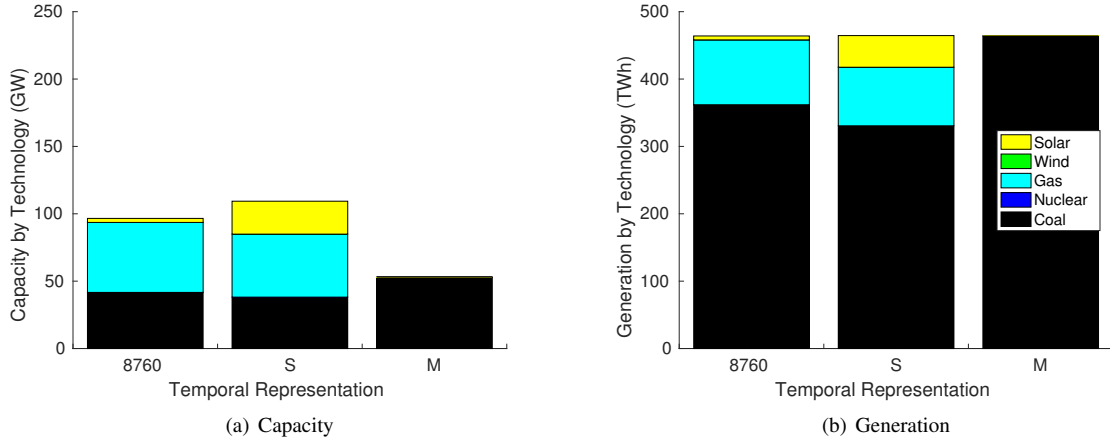


Figure 2: Optimal capacity and generation mix by temporal representation, \$1/W PV case

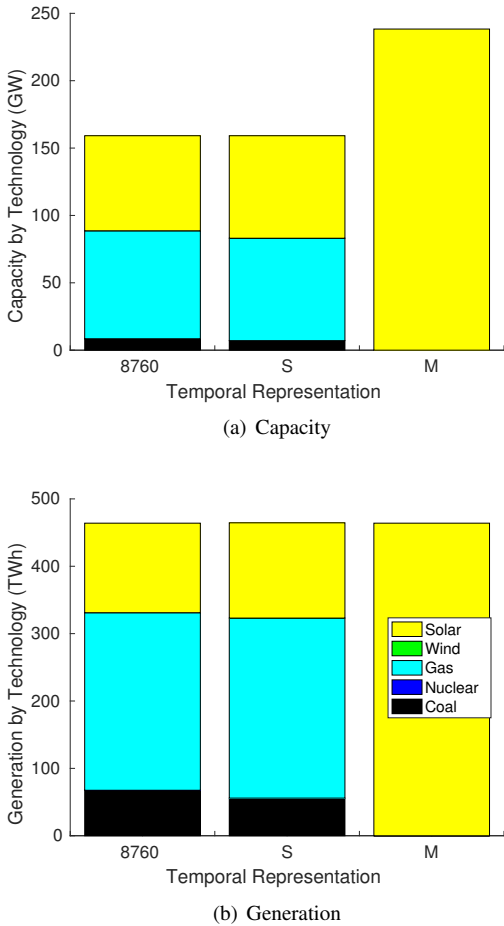


Figure 3: Optimal capacity and generation mix by temporal resolution, \$0.5/W PV case

appropriate. Importantly, we can only discuss the accuracy of the S representation for this case *ex post* and with the solution of the non-aggregated model on hand. As we have seen in our first scenario, the accuracy of the S representation is not guaranteed across all possible input sets.

In this section, we have seen how temporal representations used in existing policy models can introduce error, how error can be introduced for some input sets and not others, how the error can be relevant for some questions and not others, and, importantly, the mechanism underlying these outcomes. With this knowledge in hand, we will now consider how to systematically find temporal representations with a guarantee of not introducing error under any input set. Additionally, for those cases where exact reproduction of the non-aggregated model is required, we will consider a metric that allows us to avoid checking every model output one-by-one.

4. How many representative periods?

4.1. When does an aggregate problem reproduce the original problem?

The structure of the electricity capacity planning problem can be represented in a general abstract form as follows:

$$\text{minimise } \mathbf{c}^T \mathbf{x} \quad \text{Subject to: } \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} > 0 \quad (11)$$

Where \mathbf{x} is the vector of investment and generation variables, \mathbf{c} is the vector of costs of investment and generation, and $\mathbf{A}\mathbf{x} = \mathbf{b}$ represents the system of constraints. The aggregation of temporal resolution discussed in this paper, the transformation from the set of

statements {1-4} in Section 2.2 to the statements {5-8}, can be discussed in the context of this more abstract structure. In particular, our aggregation of hourly resolution can be considered as column aggregation followed by row aggregation. Column aggregation (aggregation of variables) results in a reduction in the size of the \mathbf{c} and \mathbf{x} vectors and the number of columns in the A matrix, while the subsequent row aggregation (aggregation of constraints) reduces the size of the \mathbf{b} vector and the number of rows in the column-reduced \bar{A} matrix.

Zipkin (1980a,b) shows that in conducting this aggregation, the more similar the aggregated columns and rows are to each other, the lower the bound on how much the objective function value will change from the original problem. In the context of our problem, this translates to the intuitive idea of finding hours that are duplicates of each other in terms of demand and technology availability, removing them, and replacing them with a weighted representative. When duplicates are aggregated, not only will the objective value be maintained, but the solution to the original problem in terms of investment and total generation by technology will also be reproduced. For example, if every hour had the same profile, $|\mathcal{P}|$ could be 1 with an associated weight, w , equal to 8760, and the exact same model outcome would occur with the aggregate model as with the full hourly resolution model. With real data, exact duplicates may be challenging to find, so we can look for hours similar to each other. Just how similar, and how many such hours exist, is numerically considered next for the dataset underlying our earlier illustrative numerical model. Additionally, we will extend the analysis to when the representative periods in question are days and weeks.

4.2. How many representative hours?

4.2.1. How many hours are unique?

To find those hours that are closely similar, we first need a metric for similarity. The nature of the data lends itself to using Euclidean distance¹⁰ as this metric, with the greater the similarity between two hours, the smaller the Euclidean distance between them. To calculate the distance between a pair of hours, each hour is represented as an n -dimensional data point comprising demand¹¹ and availability of each renewable resource¹² in that hour. While electricity demand is represented

¹⁰ The Euclidean distance between the n -dimensional points x and y is: $\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$

¹¹ d_h in earlier notation

¹² $a_{g,h}$ in earlier notation

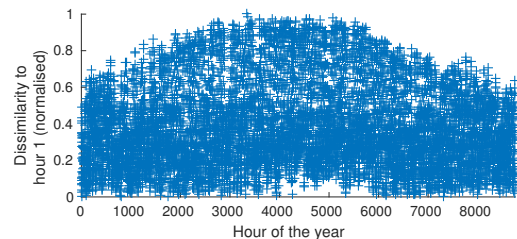


Figure 4: Normalised distance between hour 1 of dataset and other hours, zero means exact duplicate

Table 4: Sample of temporal data used in numerical analysis

	Load	Wind	Solar
H1	0.4480	0.4186	0
⋮	⋮	⋮	⋮
H8	0.5615	0.3810	0
H9	0.5588	0.3867	0.0130
H10	0.5490	0.3208	0.0590
H11	0.5334	0.2310	0.1260
⋮	⋮	⋮	⋮
h5225	1.0000	0.1207	0.2550
⋮	⋮	⋮	⋮
h8759	0.5285	0.5339	0
h8760	0.5025	0.4721	0

in units of power, endogenous renewable capacity implies that the availability of renewables is a dimensionless parameter, representing the availability per unit of installed capacity in each hour. To enable comparison between these different units, all temporal profiles are normalised such that each vector is divided by the maximum value of that vector.¹³ Table 4 presents a sample of the data when prepared in this format. As an example application of this test of similarity between hours, Figure 4 displays the Euclidean distance between hour 1 of the dataset and every other hour, with the results normalised by the maximum distance. As can be seen from the figure, the majority of the hours are different from hour 1, but hours do exist that are quite similar. To undertake this assessment on a systematic basis across all hours, a hierarchical clustering approach is employed.

In particular, a form of hierarchical clustering known

¹³ Numerical experiments revealed sensitivity in results to choice of normalisation methods. However while having a quantitative impact they did not change the qualitative findings here. See Milligan and Cooper (1988) for more on normalisation in cluster analysis.

as agglomerative clustering is employed. The advantage of hierarchical clustering methods over the commonly applied k-means clustering method is that there is no need to provide the clustering algorithm with a set of k initial points. See [Hastie et al. \(2009\)](#) for more details on the method. The basic idea is to start with clusters consisting of one data point each and sequentially merge the clusters that are closest to each other. While the distance between two points can be determined in a straightforward manner once a distance metric has been defined, it is not necessarily obvious how to determine the distance between two clusters of points. A variety of methods exist - for example, the clusters with the two closest points are merged, or clusters with the shortest longest distance between points are merged. In this application, ‘‘Ward’s’’ method is applied.¹⁴ This method merges the clusters that produce the minimum variance in the merged cluster. In the context of this application, this is desirable as we want to keep the clusters as tight as possible, which is consistent with the goal of finding hours that are duplicates of, or extremely similar to, each other.

A number of methods can be employed to provide a rule for when this agglomeration process stops. For example, the algorithm can simply be stopped when a desired number of clusters is reached, or alternately when some metric of disunity within clusters is reached. For our investigation, the hierarchical clustering algorithm is run until a desired number of clusters is reached, where the desired number changes in increments of 50 from 8760 to 60, and increments of 10 from 60 to 10. This approach allows a visual representation of the performance of the clustering algorithm across resolutions, with a goal of revealing some properties of the temporal variability. The clustering portion of this investigation was completed using the MATLAB ([The MathWorks, Inc., 2013](#)) `clusterdata` function.

The following metric will allow us to assess the performance of a given clustered resolution, where performance can be defined as how similar the members of each cluster are to each other. The same metric, in a rearranged format, is employed by [Ludig et al. \(2011\)](#) to assess how much variability is captured in a variety of time slice configurations. In the context of this paper, the metric allows us to determine how similar the members of our clusters are, which by the reasoning of [Section 4.1](#) can guide us to find an aggregate temporal representation that is robust across all policy and cost

¹⁴Further information, including the exact distance equation, is available in the supplementary material.

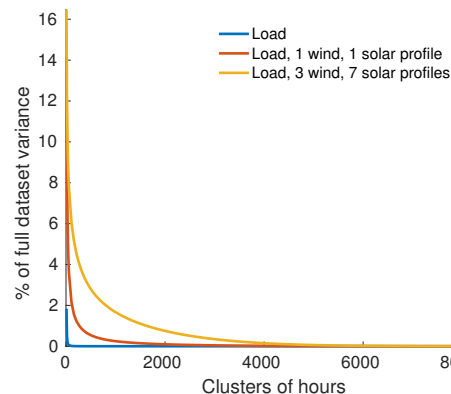


Figure 5: Relative variance by clustered resolution. This metric allows illustration of the number of unique hours in dataset. When zero, each cluster consists only of duplicate hours.

scenarios. The ‘% of full dataset variance’ can be defined as follows:

$$100 \cdot \frac{\text{Sum across clusters of within-cluster variance}}{\text{Variance across all hours}} \quad (12)$$

When this metric reaches zero, the members of each cluster are identical to each other. [Figure 5](#) presents the results of this investigation. When only demand is considered, less than 40 clusters capture the vast majority of the variance, implying that 8720 hours of the year can be discarded. Many models designed before the advent of large scale renewables have temporal resolutions within this order of magnitude. The inclusion of wind and solar profiles into our hourly data changes the curve, with approximately 1000 hours required to capture the variance when one wind and one solar profile are included, and around 3000 hours required to capture the variance when 3 regional wind and 7 regional solar profiles are included. To express it another way, while in the absence of variable renewables we had about 40 electricity goods, we now have on the order of 1000 when variable renewables are included in the set of investment options. While it will not be a qualitative surprise that increasing the number of temporal profiles from 1, to 3, to 11, leads to an increase in the number of unique hours, here we have quantified the increase for a sample dataset, and have seen a set of ideas that can be applied to any temporal dataset for any region.

4.2.2. Sensitivity of model outputs

We will now consider the sensitivity of what we really care about, the economic insights generated by model

outputs, to the variance metric displayed in Figure 5. For a sample of parameter configurations, we will compare the output generated by our model of Section 3.2 under an hourly resolution with the output generated under each of the resolutions associated with the clustering process that underlie Figure 5.¹⁵ The parameter configurations chosen for this numerical experiment are shown in Table 5.

The bounds of Zipkin (1980a,b) indicate that the more similar the rows and columns being aggregated, the smaller the errors introduced into the total value of the objective function, which in this case is total system costs. Figure 6(a) shows the absolute error in total system costs associated with each clustered resolution relative to the total system costs associated with the non-aggregated hourly resolution. We can observe a broad agreement with the bounds of Zipkin, and with Figure 5, in that the difference in total system costs in all scenarios approaches zero at the same resolution where the variance metric approached zero, i.e. where only very similar hours are aggregated. In particular, lines indicating 11 temporal profiles have reached zero at a clustered resolution of 3000, lines indicating 3 temporal profiles have reached zero at a clustered resolution close to 1000, and the solid line indicating the no renewables case reaches zero at a clustered resolution of tens of hours. This is as expected from our previous discussion. What may not necessarily be expected are the changes across lines within each colour group. The extent of the error in total system costs introduced by temporal aggregation is sensitive to how the model is parameterised. We will expand on this effect later in this section.

Frequently, we are interested in the components of the objective function as much as the objective value itself. In the application at hand, we may be as interested in the technologies the model chooses to meet electricity demand as we are in the total system costs. For example, we could be interested in how a technology-specific subsidy affects deployment of that technology. However, we can imagine cases where components of the objective function could change, but the total value of the objective function remains the same. In our application, this could particularly be the case if technologies have similar total lifetime costs, but a different distribution between whether the costs arise upfront or in operation. In choosing one such technology over the other,

¹⁵Each clustered resolution was represented in the model as follows. For each cluster of hours, the hour closest to the centroid was chosen as the representative hour, and assigned a weight equal to the number of members of the cluster.

the total system costs may not change significantly, but we could be considering a very different type of power system.¹⁶ One option to address this possible limitation of considering the objective value alone is to consider other metrics. To assess every model output individually, for example, would be comprehensive, but it would also be burdensome on the analyst and it could be a process where insight could get buried in the mass of data. The ‘L1-distance’ is thus proposed here as a means of comparison between model solution values. This metric treats the model’s objective function at solution as an n -dimensional point, where each of the n points represents an element of the decomposed objective function. The distance measurement between two model solutions is then the sum of absolute differences between each of the n values that comprise each objective function at solution. Equivalently, the metric can be considered the L1-norm of the vector of n differences in objective function elements. In our case here, the objective function is decomposed into investment costs and generation costs by technology, with the ‘L1-distance’ measuring the difference between two model solutions by summing the absolute differences in our decomposed terms across model solutions. Using our earlier notation, the metric is defined as follows:

$$\begin{aligned} \text{L1-distance} = & \sum_{g \in \mathcal{G}} \left(cc_g (|I_g^{*h} - I_g^{*c}|) \right. \\ & \left. + vc_g (|\sum_{h \in \mathcal{H}} G_{g,h}^{*h} - \sum_{p \in \mathcal{P}} G_{g,p}^{*c} w_p|) \right) \end{aligned} \quad (13)$$

Where the superscript h denotes a variable from the model with hourly resolution, and the superscript c denotes a variable from the model with the clustered resolution. Like the comparison of objective value (total system costs), this metric will be zero when generation and investment in the aggregated model match the hourly model exactly, but unlike the total system cost comparison, will be sensitive to substitution of one technology type over another with similar lifetime costs. As in Figure 6(a) with total system costs, for each resolution found by agglomerative clustering, the model was solved and the outputs were compared to the outputs from the hourly model using the L1-distance metric. The outcome is plotted in Figure 6(b).¹⁷

¹⁶Another way of expressing this is that the model solution is approaching non-uniqueness. If there was one very clear unique optimum in terms of technology choice, then system costs would be a sufficient measure to compare model outputs. DeCarolis (2011) discusses the exploration of near-optimal solution spaces.

¹⁷For the interested reader, the breakdown in contribution between investment and generation for each configuration is available in the supplementary material.

Table 5: Parameter configurations in numerical comparison experiment

Number of variable profiles in model	Technology cases
Load profile only (no renewables)	Default assumptions
Load, 1 wind, 1 solar profile	\$1/W Photovoltaics, \$0.5/W Photovoltaics
Load, 3 wind, 7 solar profiles	\$1/W Photovoltaics, \$0.5/W Photovoltaics

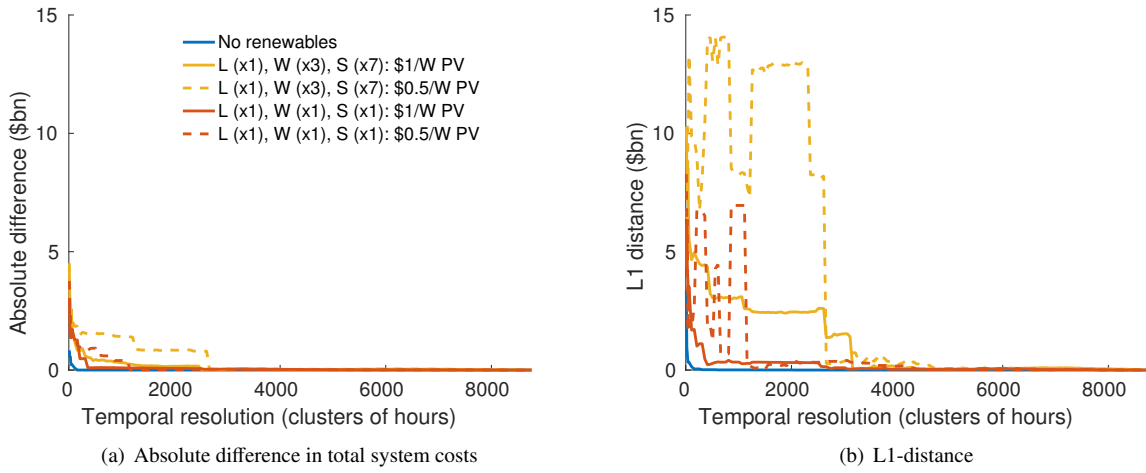


Figure 6: Distance from solution under hourly resolution to solution under clustered resolution. When the L1-distance is zero, each investment and generation component in the model solution under the clustered resolution exactly matches the hourly model solution

Consistent with our earlier discussion, the L1-distance metric in Figure 6(b) approaches zero at a clustered temporal resolution in the order of tens of hours when no renewables are in the model, a clustered resolution of approximately 1000 when two renewable profiles are included, and a clustered resolution of approximately 3000 when ten renewable profiles are included. Comparing lines within each colour group, we again see sensitivity across the parameterisation of the model, this time in greater amounts. We see that the \$0.5/W solar PV scenario is more sensitive to temporal resolution aggregation than the \$1/W scenario, in particular displaying strong non-monotonicity. Finally, as our metric approaches zero as expected at the clustered resolutions of 1000 and 3000, some residual error remains beyond these resolutions in certain configurations.

What are the reasons for the residual error and the non-monotonicity? Investigation reveals that the answer rests in the role of ‘extreme hours’, i.e. those hours where the capacity charge can rest, or equivalently, the hours where one unit of extra demand can require investment in a new unit of capacity. With residual error, though overall variance can be tiny, model outputs can be very sensitive to any variance associated with the extreme hours. This residual error disappears when one reaches the true duplicates on the right hand side of the figure. We will illustrate this role of extreme hours by a more detailed investigation of non-monotonicity. The non-monotonic jump in L1-distance occurs when, as a ‘parent’ cluster splits into two as the resolution becomes more disaggregate, one of the ‘child’ clusters can end up with the peak electricity price. However, if this representative hour is not the true peak electricity price hour, and the true hour remains hidden within a cluster, then the disaggregation can lead to increased error. As disaggregation continues, the true peak electricity price hour eventually becomes a representative hour, at which point the error relative to the hourly solution can rapidly drop. To illustrate this mechanism, we will consider the sharp non-monotonic increase in the dashed line representing the \$0.5/W PV case with one load, wind, solar profile, that happens to occur across the clustered resolutions of 857 and 858.¹⁸ The jump in L1-distance to the hourly solution in this case is due to substitution of coal technology for solar and gas technology.¹⁹ Table 6 illustrates the difference in temporal profile between the

¹⁸While the figures are associated with discrete jumps in resolution of 50, these particular cases were run to illustrate the drivers of the observed non-monotonicity.

¹⁹For the interested reader, the exact underlying capacities across these resolutions are available in the supplementary material.

Table 6: Difference in temporal profile between the 857 and 858 clustered resolutions

Hour	Weight	Demand	Solar	Price (\$/MWh)
4649 (parent)	5	0.915	0.149	41.5
4648	3	0.906	0.121	41.5
5201 (children)	2	0.985	0.101	54421.5

Table 7: Difference in temporal profile between the 1110 and 1111 clustered resolutions

Hour	Weight	Demand	Solar	Price (\$/MWh)
5179 (parent)	5	0.924	0.056	41.5
5179	3	0.924	0.056	41.5
5228 (children)	2	0.942	0.004	54421.5

857 resolution and the 858 resolution. With the splitting of the non-peak price parent cluster, child cluster 5201 has a higher demand and it turns out that it takes on the peak electricity price. Even though cluster 5201 has a low solar availability of 10%, this level of solar availability under such a high electricity price can increase the valuation of solar, leading to the misvaluation, and thus to the non-monotonic increase in the L1-distance.²⁰

Along the same line a sharp drop in L1-distance occurs across the 1110 and 1111 clustered resolutions. Table 7 illustrates the difference in temporal profile between these resolutions. What causes this sharp drop in error as measured by the L1-distance? Here the peak pricing hour emerges when the parent cluster splits. 4% availability instead of 10% at the peak pricing hour reduces the value of solar back to that of the 8760 non-aggregated case.

This is but an example where small changes in the temporal data change the nature of optimum capacity investment. Our dashed line cases, with PV so affordable, are particularly sensitive to small changes in temporal

²⁰In the terms of our expression of marginal value of a renewable generator in Section 3.1 above, there is an increase in the covariance between solar availability and electricity price.

resolution. While they may be extreme cases in a model where electricity demand cannot adjust, they are helpful in illustrating how aggregation strategies that may perform under one set of conditions perform less well under others. This discussion illustrates that the value of clustering in aggregation of temporal resolution rests as a tool to find duplicates or periods extremely similar to each other. More aggregate resolutions found by clustering algorithms, with greater variance within clusters, are not generally robust to introducing aggregation error into model results.

4.2.3. How many hours required?

Referring to the title of this section, how many representative hours are required to represent intra-annual temporal variability appropriately? For the example dataset here, the implication is an increase from tens to thousands in the number of representative hours required to not introduce aggregation error along the temporal dimension. Agglomerative clustering, when used appropriately, is a way of finding a set of representative hours that guarantee no aggregation error is introduced. Importantly however, the search for the appropriate clustering resolution should be conducted with caution - settling on an appropriate resolution on the basis on one or two model outputs for one particular scenario could provide one with a false sense of security with regard to aggregation error in other scenarios, as our use of the L1-distance metric showed. In the exercising of this caution, the variance metric introduced above can be used to check the level of similarity within clusters. To illustrate the applicability of this approach to other datasets, the supplementary material contains a similar cluster analysis for Irish power system data, showing a similar number of unique demand hours and approximately 500 unique hours when one wind profile is included.

The approach has been discussed in the context of one geographic region thus far, however, it extends naturally to multiple regions. Simply every additional temporal profile is an additional column in the data matrix to which clustering is applied. In this manner, multiple wind and solar profiles were included in the numerical analysis above.

4.3. How many representative days/weeks?

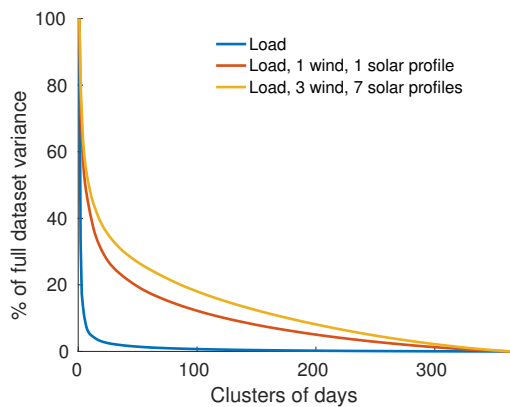
The numerical analysis thus far has assumed that no constraints span across hours, and, therefore, that hours can be treated independently without concern for chronology. Constraints that require chronology to be maintained are used to model such features as smart grid

technologies, ramping ability of generators, and short term energy storage. One approach to include such constraints in aggregated models is to employ representative days or weeks, with the constraints spanning across hours within each day or week, but not across days or weeks. Examples of models with such a temporal structure can be seen in Table 1. The two key assumptions of this approach are a) that the constraints are relevant for the question being asked of the model and b) that constraints across representative days or weeks are not relevant, such as those that model interseasonal storage options. We will not consider here whether the constraints that the representative day/week approach enables are necessary for economic insight. We will investigate, however, how many representative days/weeks are required to guarantee that the approach, when employed, does not introduce aggregation error.

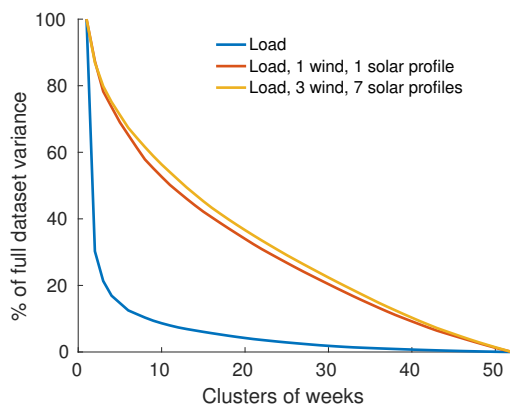
Using the same logic as Section 4.2, we will now consider how many similar days and weeks are in our example temporal dataset. While previously the dataset consisted of 8760 points with associated load/wind/solar values, for this analysis, the dataset is restructured to consist of 365 (daily assessment) or 52 (weekly assessment) points with associated load/wind/solar vectors. For example, with one wind and one solar availability profile in the daily assessment, each of the 365 points has a dimension of 72 (the number of hours, 24, multiplied by the number of load, wind and solar profiles, 3).

Figures 7(a) and 7(b) show how the number of clustered days and weeks required to capture the temporal variability increases when wind and solar availability profiles are included in addition to demand alone. For example, in the case of clustered days, the level of variance associated with 20 days, when demand alone was considered, is now associated with approximately 300 days as renewables are included. In the case of weeks, we see that there are essentially no weeks that are duplicates of each other in terms of electricity demand and renewables availability. While it is not qualitatively surprising that there will be more unique sequences of hours than unique hours themselves, and that there will be more unique periods as more temporal profiles are included, the approach here has quantitatively shown how many periods are required to represent our sample dataset, and can be applied to any such dataset. For example, the same approach is applied to Irish power system data in the supplementary material, showing similar increases in the number of unique periods required to represent the increased temporal variability introduced by wind power.

As in our investigation of representative hours, we



(a) Assessment of the similarity of days



(b) Assessment of the similarity of weeks

Figure 7: Relative variance by clustered daily and weekly resolutions. When zero, each cluster consists of only duplicate days/weeks.

will compare model outputs associated with each clustered resolution to the non-aggregated model solution in terms of both difference in system costs and the L1-distance. Figures 8(a) and 8(b) show the results for the clustered daily resolutions, while Figures 9(a) and 9(b) show the results for the clustered weekly resolutions. As expected from our earlier discussions, we see that error in both total system costs and L1-distance reaches zero in each scenario as the variance metric in Figures 7(a) and 7(b) approaches zero, with the error in system costs reaching zero at more aggregate resolutions.

Figures 6, 8 and 9 all indicate the existence of solutions that are very close to each other in terms of what the model is optimising, total cost, but are some distance away from each other in terms of our L1-distance measure, highlighting the sensitivity of technology choice, and potentially associated emissions calculations, to temporal representation.

If a temporal representation robust to aggregation error for any scenario is required, the representative day and week aggregation approach offers little potential for reducing model size. Representative days and weeks are simply not that representative when temporal profiles of renewable resource availability are considered in conjunction with the electricity demand profile. Equivalently, models that do employ this approach in cases of endogenous renewable capacity are at risk of introducing aggregation error, depending on the scenarios run and the questions asked of the model.

Our finding that approximately 300 representative days are required to capture the temporal variability in our example dataset stands in contrast to the 20 days that Nahmmacher et al. (2014a) find is adequate for capturing the temporal variability in the Limes-EU model, a number also derived through the application of a hierarchical clustering algorithm. In addition to the difference in datasets, each day in Limes-EU has 8 time slices relative to the 24 in this analysis, implying aggregation that smooths the within-day variability, making Limes-EU days more similar pre-clustering. Additionally Nahmmacher et al. (2014a) numerically show the quality of the 20 day resolution at one particular parameterisation scenario and with two model output metrics (total costs and variable renewable share). As we have seen, there could be discrepancies in components that underlie these aggregate outputs, discrepancies that could matter for certain questions. Similarly if the 20 days do not represent all the unique days, other parameterisations could theoretically be found where the missing days produce aggregation error for these two output metrics. While one can say that these issues may only arise for extreme scenarios, sometimes it is those sce-

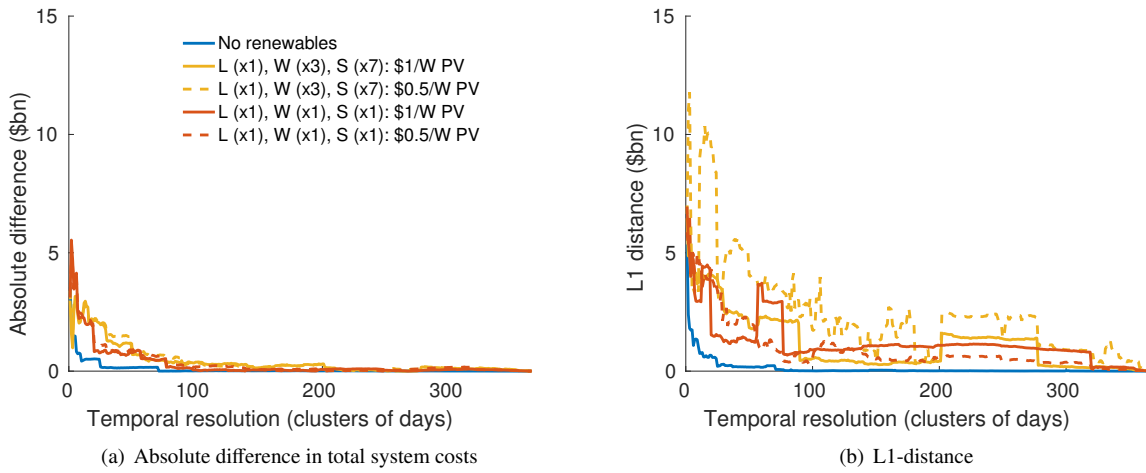


Figure 8: Comparing clustered daily resolutions to the non-aggregated resolution

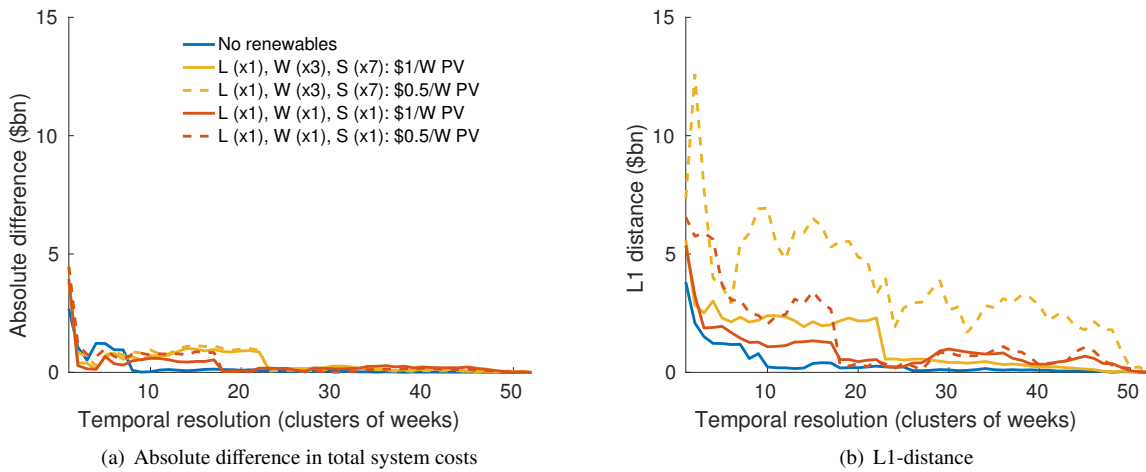


Figure 9: Comparing clustered weekly resolutions to the non-aggregated resolution

narios that are of interest.

The relatively small model used in our numeric calculations has provided us with a platform to conduct an extensive number of numerical model runs to explore this issue of temporal aggregation. The experiments conducted here would not be possible in any of the large scale models used for policy analysis without extraordinary runtime. The findings of these numerical experiments, supported by the earlier analytic reasoning, can be implemented by those maintaining large models or designing new models by systematically considering the number of duplicates/highly similar periods in the temporal input data they are aggregating. If the resulting resolution is too great for computational implementation, the next section will consider how we can reduce it further.

5. On reducing resolution further

5.1. Operating conditions

We have seen how aggregation can introduce error for some model outputs and not for others. If our question of economic interest relates to the unaffected output only, then we only need to disaggregate to the level of temporal resolution that is relevant for that output, and not for all outputs. Similarly, we have seen that aggregation error only occurs for certain model parameterisations and not for others. If the question relates to the unaffected range of parameterisation, then the coarser level aggregation will be appropriate. Additionally, tolerance for what level of error is acceptable can vary by model output and the question asked. These approaches imply the necessity for ‘model operating conditions’, i.e. conditions for what model runs are appropriate and which are not given the chosen temporal aggregation scheme.²¹ For example, computationally tractable ‘side constraint’ approaches, as discussed earlier, would benefit from explicit accompanying conditions outlining when they are valid and when they are not. Such operating conditions can be derived analytically or through numerical experiment.

5.2. Problem structure and prior information

For a given problem structure, the observation that finding duplicates provides a guarantee is not the same as stating that finding duplicates produces the minimum possible guaranteed resolution. Rather than represent all the hourly input data flawlessly in our aggregation

scheme, could we instead focus on representing the subset of hourly input data that drives model results? An analogy is the encoding of MP3 audio files, where the goal is to capture a waveform ‘as it sounds’ and not ‘as it is’ (Sayood, 2012). What sort of prior information could allow us to find subsets of hours that are more important than others in driving model behaviour? We have already seen in Section 4.2 how hours with peak prices (prices that are often orders of magnitude greater than the average electricity price) can drive model decisions through their large role in the covariance terms earlier derived. Aggregation accuracy was particularly sensitive to these hours. An aggregation system that treats potential peak pricing hours differently thus may be a promising direction. The challenge, however, is knowing a-priori which hours those peak pricing hours might be. While before the advent of variable renewables the hour with the peak electricity price was that with maximum demand, the peak electricity hour now endogenously depends on the level of renewables deployment. If we had such prior information however, we could partition our hours into hours where this peak price could rest, and where it could not. In the latter partition, we could have a coarser threshold for what counts as a duplicate, allowing a reduction in the number of periods. For example, a clustering algorithm could be modified to penalise variance within clusters that contain those hours identified as potential peak pricing hours. The REGEN (Blanford et al., 2014) temporal representation design implicitly follows such an approach. In this case, ‘bubble’ partitions are defined around extreme load/wind/solar periods, with at least 1 period required to be chosen from each bubble, before a clustering algorithm is implemented to aggregate the remaining hours outside the bubble partitions. The REGEN approach, however, is a heuristic as synthetic examples can be constructed such that peak pricing hours will not be located at one of the extreme vertices.

5.3. Use of bounds enabled by a stochastic framework

Framing temporal variability as a space of stochastic incidences allows the machinery of stochastic programming to be applied to the problem. Munoz et al. (2016) harness this machinery to potentially reduce model resolution. Starting with a coarse clustered resolution and progressively trying more disaggregate resolutions, the model allows calculation of upper and lower bounds on total cost at each resolution, until such point as the bounds converge. If bounds converge at a coarse resolution, the more detailed model run can be avoided. As we have discussed with regard to the L1-distance metric, capacity planning models can have numerous tech-

²¹This and related points are discussed in greater detail in Merrick (2016).

nology choice options close to the optimum, an issue this approach would not directly address. In addition, this procedure would have to be applied each time the model runs a new scenario.

6. To conclude

The importance of temporal variability in valuing different electricity supply technologies has been recognised in the literature. We have considered in this study how this variability is represented in numeric models of electricity capacity planning. We have analysed how representations can misvalue technology options, seen how to guarantee a robust aggregate resolution, and discussed the intricacies of finding such a resolution numerically. Finally, potential approaches to reduce the resolution further have been outlined. In addition to the individual contributions, a contribution of this study as a whole is to bring these ideas together in one paper.

This study does not change the qualitative understanding of how temporal variability affects the economics of power system investments, but is highly relevant for any quantitative understanding, such as that required for applied policy analysis. There is simply greater information content in the temporal variability data when variable renewable profiles are considered in addition to demand variability alone. Equivalently, from an economic perspective of viewing unique hours as unique electricity goods in the absence of large scale electricity storage, the number of electricity goods increases significantly when variable renewable potential is included. For our sample dataset, the increase is by two orders of magnitude.

This work has indicated a number of directions for further research additional to those outlined in Section 5. Following an oft-adopted paradigm, this paper assumed a fixed demand shape. Electricity demand, however, is potentially on the cusp of change not seen since the foundation of power systems, straining this assumption. Also following the same paradigm is the assumption that future years' supply and demand variability can be represented by historical data. With a changing climate, wind and solar temporal profiles may also be subject to change. As these issues are researched, this study aims to aid the advancement of the models that are used today to address important societal challenges relating to the electricity system and the broader climate system.

Acknowledgements

The author wishes to thank Professor James L. Sweeney, Professor John P. Weyant, participants at

the Stanford University Policy and Economic Research Roundtable, and two anonymous reviewers for feedback that has greatly improved this paper. Primary funding for this project came from the Department of Energy, Office of Science PIAMDDI Grant (DE-SC005171) to the Energy Modeling Forum at Stanford University.

- Aboumahboub, T., Schaber, K., Wagner, U., Hamacher, T., 2012. On the CO₂ emissions of the global electricity supply sector and the influence of renewable power-modeling and optimization. *Energy Policy* 42, 297–314.
- Bertsch, J., Growitsch, C., Lorenczik, S., Nagl, S., 2016. Flexibility in Europe's power sector - An additional requirement or an automatic complement? *Energy Economics* 53, 118–131.
- Blanford, G. J., Merrick, J. H., Young, D., Sep. 2014. A Clean Energy Standard Analysis with the US-REGEN Model. *The Energy Journal* 35 (01), 137–164.
- de Sisternes, F. J., Webster, M. D., Perez-Arriaga, I. J., 2015. The Impact of Bidding Rules on Electricity Markets with Intermittent Renewables. *IEEE Transactions on Power Systems* 30 (3), 1603–1613.
- DeCarolis, J. F., 2011. Using modeling to generate alternatives (MGA) to expand our thinking on energy futures. *Energy Economics* 33 (2), 145–152.
- Edenhofer, O., Hirth, L., Knopf, B., Pahle, M., Schlömer, S., Schmid, E., Ueckerdt, F., 2013. On the economics of renewable energy sources. *Energy Economics* 40, S12–S23.
- EIA (Energy Information Administration), 2013. The Electricity Market Module of the National Energy Modeling System Model Documentation 2013. Tech. rep., U.S. Energy Information Administration, Washington DC.
URL [http://www.eia.gov/forecasts/aeo/nems/documentation/electricity/pdf/m068\(2013\).pdf](http://www.eia.gov/forecasts/aeo/nems/documentation/electricity/pdf/m068(2013).pdf)
- Fripp, M., Jun. 2012. SWITCH: A Planning Tool for Power Systems with Large Shares of Intermittent Renewable Energy. *Environmental science & technology* 46 (11), 6371–8.
- Green, R., Staffell, I., Vasilakos, N., 2014. Divide and Conquer? k-Means Clustering of Demand Data Allows Rapid and Accurate Simulations of the British Electricity System. *IEEE Transactions on Engineering Management* 61 (2), 251–260.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning*, 2nd Edition. Springer-Verlag, New York.
- Hirth, L., 2013. The market value of variable renewables. The effect of solar wind power variability on their relative price. *Energy Economics* 38, 218–236.
- IEA (International Energy Agency), 2014. *World Energy Outlook 2014*. OECD, Paris.
- Jacoby, H. D., 1967. *Analysis of Investment in Electric Power*. Ph.D. thesis, Harvard University.
- Johnston, J., Mileva, A., Nelson, J. H., Kammen, D. M., 2013. SWITCH-WECC: Data, Assumptions, and Model Formulation. Tech. rep., University of California, Berkeley, Berkeley, CA.
URL http://rael.berkeley.edu/old_drupal/sites/default/files/SWITCH-WECC_Documentation_October_2013.pdf
- Joskow, P. L., 2011. Comparing the Costs of Intermittent and Dispatchable Electricity Generating Technologies. *American Economic Review* 101 (3), 238–241.
- Kannan, R., Turton, H., 2013. A Long-Term Electricity Dispatch Model with the TIMES Framework. *Environmental Modeling & Assessment* 18 (3), 325–343.
- Lamont, A. D., 2008. Assessing the long-term system value of intermittent electric generation technologies. *Energy Economics*, 1208–1231.

- Lannoye, E., Flynn, D., O'Malley, M., 2012. Evaluation of power system flexibility. *IEEE Transactions on Power Systems* 27 (2), 922–931.
- Litvinchev, I. S., Vladimir, T., 2003. *Aggregation in large-scale optimization*. Kluwer Academic Publishers, Dordrecht.
- Luderer, G., Krey, V., Calvin, K., Merrick, J., Mima, S., Pietzcker, R., Van Vliet, J., Wada, K., Oct. 2014. The Role of Renewable Energy in Climate Stabilization: Results from the EMF27 Scenarios. *Climatic Change* 123 (3-4), 427–441.
- Ludig, S., Haller, M., Schmid, E., Bauer, N., 2011. Fluctuating renewables in a long-term climate change mitigation strategy. *Energy* 36 (11), 6674–6685.
- Manne, A., Mendelsohn, R., Richels, R., 1995. MERGE: A model for evaluating regional and global effects of GHG reduction policies. *Energy Policy* 23 (1), 17–34.
- Masters, G. M., 2004. *Renewable and efficient electric power systems*. John Wiley & Sons, Hoboken NJ.
- Merrick, J. H., 2010. An Assessment of the Economic, Regulatory, and Technical Implications of Large-Scale Solar Power Deployment. Master's thesis, Massachusetts Institute of Technology.
- Merrick, J. H., 2016. On Assessing Appropriateness of Normative Model Formulation, available at SSRN: <http://ssrn.com/abstract=2740843>.
- Milligan, G. W., Cooper, M. C., 1988. A study of standardization of variables in cluster analysis. *Journal of Classification* 5 (2), 181–204.
- Munoz, F. D., Hobbs, B. F., Watson, J. P., 2016. New bounding and decomposition approaches for MILP investment problems: Multi-area transmission and generation planning under policy constraints. *European Journal of Operational Research* 248 (3), 888–898.
- Munoz, F. D., Mills, A. D., 2015. Endogenous Assessment of the Capacity Value of Solar PV in Generation Investment Planning Studies. *IEEE Transactions on Sustainable Energy* 6 (4), 1574–1585.
- Nahmmacher, P., Schmid, E., Hirth, L., Knopf, B., 2014a. Carpe diem: A novel approach to select representative days for long-term power system models with high shares of renewable energy sources, USAEE Working Paper No. 14-194. URL <http://ssrn.com/abstract=2537072>
- Nahmmacher, P., Schmid, E., Knopf, B., 2014b. Documentation of LIMES-EU - A long-term electricity system model for Europe. Tech. rep., Potsdam Institute for Climate Impact Research, Potsdam. URL <https://www.pik-potsdam.de/members/paulnah/limes-eu-documentation-2014.pdf>
- Nicolosi, M., 2011. The Importance of High Temporal Resolution in Modeling Renewable Energy Penetration Scenarios. In: 9th Conference on Applied Infrastructure Research. Berlin.
- Palmitier, B., Webster, M., 2011. Impact of unit commitment constraints on generation expansion planning with renewables. In: IEEE Power and Energy Society General Meeting. Detroit.
- Parpas, P., Webster, M., 2014. A stochastic multiscale model for electricity generation capacity expansion. *European Journal of Operational Research* 232 (2), 359–374.
- Pina, A., Silva, C., Ferrão, P., 2011. Modeling hourly electricity dynamics for policy making in long-term scenarios. *Energy Policy* 39, 4692–4702.
- Poncellet, K., Delarue, E., Six, D., Duerinck, J., D'haeseleer, W., 2016. Impact of the level of temporal and operational detail in energy-system planning models. *Applied Energy* 162, 631–643.
- Powell, W., George, A., Simao, H., Scott, W., Lamont, A., Stewart, J., 2012. SMART: a stochastic multiscale model for the analysis of energy resources, technology, and policy. *INFORMS Journal on Computing* 24 (4), 665–682.
- Rogers, D. F., Plante, R. D., Wong, R. T., Evans, J. R., 1991. Aggregation and Disaggregation Techniques and Methodology in Optimization. *Operations Research* 39 (4), 553–582.
- Sayood, K., 2012. Audio Coding. In: Sayood, K. (Ed.), *Introduction to Data Compression*, 4th Edition. Morgan Kaufmann, Boston, pp. 569–590.
- Short, W., Sullivan, P., Mai, T., Mowers, M., Uriarte, C., Blair, N., Heimiller, D., Martinez, A., 2011. Regional Energy Deployment System. Tech. Rep. NREL/TP-6A20-46534, National Renewable Energy Laboratory, Golden, Colorado. URL <http://www.nrel.gov/docs/fy12osti/46534.pdf>
- Sullivan, P., Krey, V., Riahi, K., Mar. 2013. Impacts of considering electric sector variability and reliability in the MESSAGE model. *Energy Strategy Reviews* 1 (3), 157–163.
- Swider, D. J., Weber, C., 2007. The costs of wind's intermittency in Germany: Application of a stochastic electricity market model. *European Transactions on Electrical Power* 17, 151–172.
- The MathWorks, Inc., 2013. *MATLAB and Statistics Toolbox Release 2013a*. Natick, Massachusetts.
- Ueckerdt, F., Brecha, R., Luderer, G., Sullivan, P., Schmid, E., Bauer, N., Böttger, D., Pietzcker, R., 2015. Representing power sector variability and the integration of variable renewables in long-term energy-economy models using residual load duration curves. *Energy* 90, 1799–1814.
- Ueckerdt, F., Hirth, L., Luderer, G., Edenhofer, O., Dec. 2013. System LCOE: What are the costs of variable renewables? *Energy* 63, 61–75.
- van der Weijde, A. H., Hobbs, B. F., 2012. The economics of planning electricity transmission to accommodate renewables: Using two-stage optimisation to evaluate flexibility and the cost of disregarding uncertainty. *Energy Economics* 34 (6), 2089–2101.
- Welsch, M., 2013. *Enhancing the Treatment of Systems Integration in Long-term Energy Models*. Ph.D. thesis, KTH Royal Institute of Technology.
- Wogrin, S., Duenas, P., Delgadillo, A., Reneses, J., 2014. A New Approach to Model Load Levels in Electric Power Systems With High Renewable Penetration. *IEEE Transactions on Power Systems* 29 (5), 2210–2218.
- Zipkin, P. H., 1980a. Bounds for Row-Aggregation in Linear Programming. *Operations Research* 28 (4), 903–916.
- Zipkin, P. H., 1980b. Bounds on the Effect of Aggregating Variables in Linear Programs. *Operations Research* 28 (2), 403–418.