Wind data introduce error in time-series reduction for capacity expansion modelling

Lucas Elias Kuepper[a,c], Holger Teichgraeber[a], Nils Baumgärtner[c], André Bardow[b,c,d], Adam R. Brandt[a,*]

[a] *Department of Energy Resources Engineering, Stanford University, Green Earth Sciences Building 065, 367 Panama St., Stanford, California, USA*
[b] *Department of Mechanical and Process Engineering, ETH Zürich, Leonhardstrasse 21, LEE K, 8092 Zurich, Switzerland*
[c] *Institute of Technical Thermodynamics, RWTH Aachen University, 52056 Aachen, Germany*
[d] *Institute of Energy and Climate Research - Energy Systems Engineering (IEK-10), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany*

Abstract

Shares of renewable energy are rapidly increasing in many countries due to emissions policies and declining prices. Investment planning for future renewable deployment often relies on optimization models. Memory usage and solving time restrict these models, leading to tradeoffs in the treatment of temporal complexity, spatial complexity, and physical representation. A common approach is to reduce the temporal complexity of models. Reducing temporal complexity is often achieved by using time-series aggregating and modelling representative periods instead of a complete time series. But the impacts of such approaches are still poorly understood, especially for very low emissions systems with high shares of variable renewable energies. In this paper, the impacts of using time-series aggregation methods on optimal system design are investigated. It is found that the negative impact of time-series aggregation increases for lower emissions. It is also identified that modelling wind time-series data with representative days introduces this negative impact primarily and that representing wind time-series data with representative days decreases the reliability of supply defined as unserved load (0.05% to 18.0%), introduces a bias in installed capacity (-31.15% to +12.2%), and underestimates total system cost (2.5% to 44.9%). These effects are largest in cases with the strongest emission constraints. When designing low emissions systems with a high share of variable renewable energies, it is recommended not to use time-series

aggregation to create representative days for wind power output. This paper contributes an Open Source analysis framework containing time-series aggregation and capacity expansion that should be applied when testing future time-series aggregation methods to reduce the identified negative impacts.
Keywords: energy system, optimization, linear programming, time-series aggregation, emission reduction

1. Introduction

To meet the low-carbon emissions goals of the Paris Agreement [1], today's energy systems – with significant greenhouse gas emissions – need to be transformed into future energy systems with no net or even negative carbon dioxide emissions. Capacity expansion planning (CEP) is an important tool for policy and investment decisions for future energy systems [2]. CEP models represent the investment and operation of a future energy system to determine the optimal installation plan, including decisions on technology mix, location, and investment timing for generation, storage, and transmission capacities [3].

The representation of relevant physical, spatial, and temporal complexities in CEPs is often highly simplified [4] to make a CEP solvable with state-of-the-art computational hardware. For example, the temporal complexity is often reduced by using time-series aggregation to create "typical" periods such as representative days [5].

Time-series aggregation in CEP models has been applied to create representative time series for demand: Dominguez-Munoz et al. use representative demand days [6]. Fazlollahi et al. use representative demand periods with varying lengths in the first step and segment the representative periods into a few consecutive time-steps in a second step [7]. Schütz et al. use different aggregation methods to generate representative demand days [8]. Gabrielli et al. [9] aggregate electricity prices to representative days [10]. Nahmacher et al. [11]

*Corresponding author. Tel: +1 (650) 724-8251
Email addresses: elias.kuepper@rwth-aachen.de (Lucas Elias Kuepper), hteich@stanford.edu (Holger Teichgraeber), nils.baumgaertner@ltt.rwth-aachen.de (Nils Baumgärtner), abardow@ethz.ch (André Bardow), abrandt@stanford.edu (Adam R. Brandt)

extent the approach of representative periods from demand to also include energy availability. Pfenninger et al. [12] use representative periods for the energy availability from wind and solar and highlight the inter-annual variability of multiple year weather data. Multiple methods have been devised to perform time-series aggregation [5] and have been compared with each other: Schütz et al. [8] compare different methods to select representative demand days. Pfenninger [12] compares methods to create representative periods, including combinations of downsampling, k-means, hierarchical, and extreme period selection approaches. Teichgraeber et al. [13] compare hierarchical, k-means, k-medoids, DBA, and k-shape clustering algorithms [14]. All of these methods group periods of the original time-series – like days – into a smaller number of representative periods.

Time-series aggregation methods have been applied widely to different energy system models. First, Time-series aggregation methods have been used for national energy systems: Nahmacher et al. [11] applied it to a European scale energy system. Pfenninger [12], Zeyringer et al. [15], Green et al. [16], and Heuberger et al. [17] model Great Britain. Munoz et al. [18] model the Western Electricity Coordinating Council in the US. Blanford et al. [19] model the US. Baumgärtner et al. [20] model the German energy system. Second, time-series aggregation is used to model regional energy systems: Lara et al. [21] and Merrick et al. [22] model in the Electric Reliability Council of Texas region. Pina et al. [23] model Sao Miguel, a Portuguese island. Third, time-series aggregation methods are used for district energy systems: Fazlollahi et al. [7] model two separate small district models to supply heating and electricity demand. Gabrielli et al. [9] and Kotzur et al. [14] model district energy systems that supply heating and electricity demand in one model. Fourth, Schütz et al. [24] and Kotzur et al. [14] use time-series aggregation to model residential energy systems. Fifth, time-series aggregation is used for industrial energy supply systems: Dominguez-Munoz et al. [6] and Kotzur et al. [14] model combined heat and power industrial energy systems. Bahl et al. [10] model an energy system with heating, cooling, and electricity demand. Brodrick et al. [25] model an energy system and combe a solar thermal, gas-fired, and CO_2 capture system. Broderick et al. [26] model an integrated solar combined cycle. Schilling et al. [27] design an organic rankine cycle. Teichgraeber et al. [28] model a oxyfuel natural gas

plant. Baumgärtner et al. [29] model a industrial synthesis problem. Yokoyama et al. [30] model an energy system with electricity, cold water, and steam demand. Some models are formulated as dispatch problems, which optimize the detailed operation of a given energy system as applied by Green et al. [16] and Brodrick et al. [26]. Other models are formulated as capacity expansion planning problems, which optimize the energy system design while modelling the operation like Nahmacher et al. [11], Pfenninger et al. [12], and Kotzur et al. [14], or Baumgärtner et al. [20] do.

Accurate time-series aggregation would reduce model complexity while generating similar optimization outcomes to those that would have been obtained if aggregation was not applied. One can measure the accuracy of the optimization outcome in a reduced complexity model in objective function value, energy system design, and system constraint violations [31]. For complex optimization problems that are used for policy and decision making, it is often not possible to solve a reference scenario with the full time-series, and only lower and upper bounds can be calculated for some of those models [32]. In the literature, there has been an effort to examine the effects of aggregation on simplified models: Teichgraeber and Brandt [13] use an energy storage and gas turbine generation problem and find that centroid-based clustering methods represent the operational part of the optimization problem more accurately. Kotzur et al. [14] use an industrial, residential, and island energy system and recommend a hierarchical clustering algorithm and that the impact of applying time-series aggregation must be evaluated for each energy system. Pfenninger et al. [12] use a zonal model of Great Britain, find a significant interannual variability of temporal input data depending on the weather year, and recommend the usage of more than one year in the capacity expansion planning. Göke et al. [33] use a zonal model of Germany, find that time-series aggregation should be applied with caution, and recommend including extreme periods, varying temporal resolutions or portioning the problems in smaller parts. Simplified models bring the advantage that the optimization output of a scenario with time-series aggregation can be compared to the optimization output of a reference scenario where the full time-series is used to evaluate the error introduced by time-series aggregation [13]. Studies show that time-series aggregation impact on design decision is stronger for systems with

lower emissions standards [12]. Göke et al. [33] conclude that time-series aggregation is not adequate for designing energy systems with high shares of renewable energy systems. A possible overestimation of electricity produced from wind is shown by Pina et al. [23] when down-sampling is used to reduce the number of time periods per modelled day. However, Poncelet et al. [34] show that increasing the number of time periods per modelled day has a smaller effect on the error of the residual load duration curve than using other methods to aggregate the time series. In the literature, distance measures are used to determine the quality of the clustering comparing the clustering result to the original input data.

Despite the rise in importance of time-series aggregation in recent years, the impacts of time series aggregation on emission constraints and reliability of supply for low-carbon emission systems have not been thoroughly analyzed.

In this work, the impacts of time series aggregation on costs, design decisions, emission constraints, and reliability of supply are analyzed. Compared to existing literature, it is identified why time-series aggregation performs so poorly on low emissions systems with a high share of variable renewable energies. The aggregation of only some attributes is introduced to analyze the source of the error. It is investigated whether 1) adding single day extreme, 2) increasing the number of representative periods, 3) using a medoid instead of a centroid clustering representation, or 4) giving certain attributes more importance in the aggregation process improves optimization outcome. A time-series aggregation analysis framework is provided. The framework uses some case data for two different countries and models the same countries with different numbers of regions.

2. National energy models and time-series aggregation

The electric energy systems of California (CA) and Germany (GER) are modelled, two areas of leading renewable penetration. The geographies of the two electric energy systems are both represented as single- and multiple-node systems, with 10 nodes in CA and 18 nodes in GER. Hourly data from multiple years are used: 2014 to 2017 for CA and 2010 to 2016 for GER, all-inclusive. The time-series data is only available in an hourly resolution across the studied regions and years, limiting the studied temporal granularity.

The temporal input data is split into single years, which are called "full time-series". Three scenarios are distinguished: a) A reference scenario (R), which uses the full time-series for the energy system design and operation. b) A design scenario (D_k), which uses k aggregated time series for the energy system design. c) A design & operation scenario ($D_k\&O_{all}$), which uses a two-stage approach to design and test the energy system. In the two-stage approach, k aggregated representative time series are used first to generate optimal system designs for various CEP models. To measure the resulting system infeasibility, the unmet demand ("lost load") and emissions exceeding the emission constraint ("excess emissions") are then calculated in a second stage operational optimization that uses the full temporal data set and has perfect foresight. This second stage operates and tests the energy system design resulting from the first stage (with time-series aggregation) against the original full data set (without time-series aggregation), somewhat akin to "stress testing" the reduced-form model with real-world, non-aggregated data.

A hierarchical clustering algorithm is used because it is deterministic, in contrast to k-means. Within the study, different cases are investigated, and the base case uses a centroid representation for the cluster centres, the same attribute weights for all time-series, and no extreme value selection. First, the time-series aggregation methods used are explained, second, the capacity expansion model is introduced, and third, the scenarios that combine time-series aggregation and capacity expansion model are introduced.

2.1. Time-series aggregation

A mean-preserving time-series aggregation method to represent all original periods of a time-series with fewer representative periods is applied (Figure 1). It is referred to Teichgraeber and Brandt [13] for a framework describing clustering methods for time-series aggregation methods applied to CEP. The julia implementation "TimeSeriesClustering.jl" [35] is used in this paper. The time-series aggregation method consists of the following steps:

1. Splitting and reshaping of the original time-series, which is a data preparation step that brings the data in the necessary format.

2. Normalization, which ensures the same scale for all attributes, e.g. demand has the unit MW and availability is a factor between 0 and 1.
3. Optional: Extreme period selection, which selects periods with particular low wind availability, low solar availability, or high demand and adds them to the representative periods.
4. Hierarchical clustering with Euclidean distance measure and representation of each cluster, which reduces the number of periods to a few representative periods.
5. Denormalization of the selected extreme periods and clustered periods.

A hierarchical clustering algorithm is chosen over a partitional clustering algorithm to ensure reproducibility: Partitional clustering algorithms use a random initial distribution of starting points. In initial experiments, it is found that even selecting the best result from 10,000 partitional clustering runs is not sufficient to ensure the exact reproducibility of the results. Hierarchical clustering algorithms yield reproducible results. The initial analysis has shown that the qualitative statements of this paper are the same for partitional and hierarchical clustering algorithms. That partitional and hierarchical clustering algorithms lead to structurally similar results is also in accordance with the findings of Teichgraeber and Brandt [13], Kotzur et al. [14], Schütz et al. [8], and Liu et al. [36]. Teichgraeber and Brandt [13] observe the similar performance of hierarchical and k-means clustering algorithms. Only k-medioid appears less predictable than the others. Kotzur et al. [14] also recommend hierarchical algorithm as the choice of aggregation method k-medoid or hierarchical has a small influence, and a hierarchical aggregation can be reproduced and has a lower computation effort.

Li et al. [37] use the Davies-Bouldin index to measure the clustering quality and determine the number of representative periods. The Euclidean distance, which is a specific form of the Davies-Bouldin index, is common in literature [13] and can be used to determine the numerical quality of the clustered input data generated via the data reduction algorithm. However, the numerical quality of the input data preparation is not equivalent to the quality of the optimization result. Prior work has shown that reduced clustering error does not

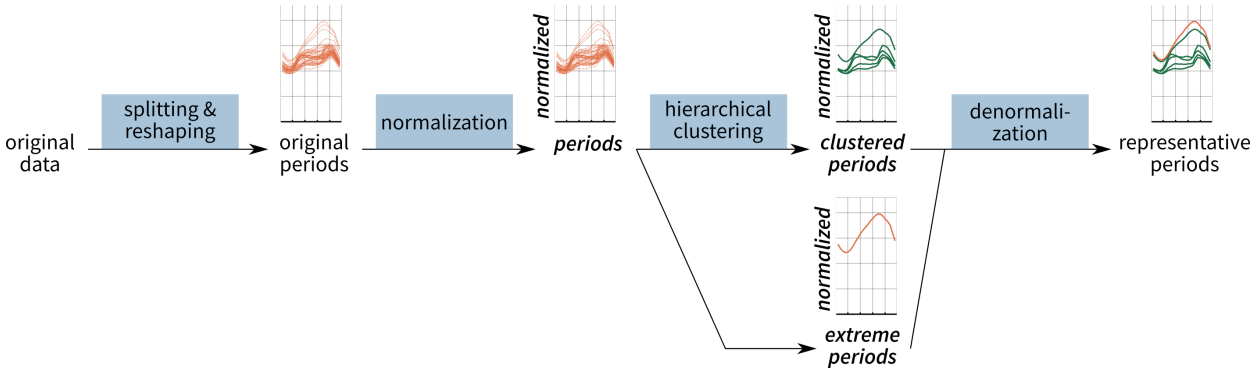

Figure 1: First, the original time-series data is split and reshaped to the necessary format. Second, the original periods are normalized, ensuring the same sale for all attributes. Third, extreme periods are optionally selected. Fourth, the remaining periods are clustered. Fifth, the extreme and clustered periods are denormalized to the representative periods used in the optimization.

always result in reduced optimization error relative to the optimization solution with no clustering. We, therefore, do not determine the required number of representative periods based on the numerical quality of the input data preparation using the Davies-Bouldin index but instead investigate how sensitive the quality of the result is to the number of representative periods.

First, the regular aggregation (used by Teichgraeber and Brandt [13]) is described, second, attribute weighting is introduced, and third extreme period selection is explained.

2.1.1. Regular aggregation

The full time-series is shaped into a matrix (rows are features and columns are periods, Figure 2) to find similarities between the different original days, which are used as the original periods. A single column contains all time-step values of one period, called time period, for each attribute and each region. A z-normalization is applied on the values of each attribute and region to achieve comparable scales between different attributes, which may have different units. The values of each attribute and region are transformed to have a mean of zero and standard deviation of one to allow comparison across different attributes and regions:

$$\hat{P}_{a,n} = \frac{1}{\sigma_{a,n}}(P_{a,n} - \mu_{a,n}) \tag{1}$$

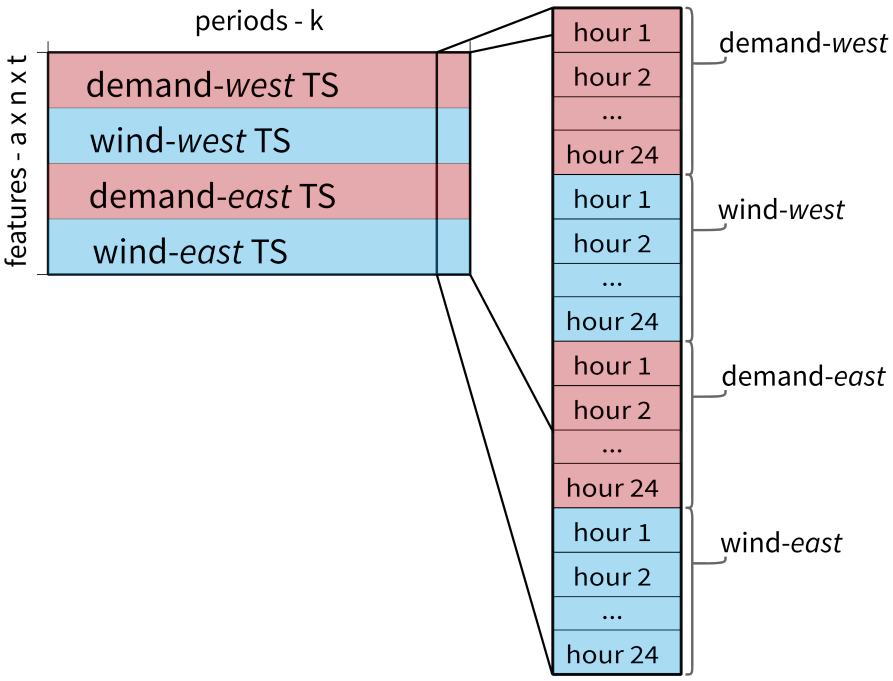

Figure 2: The original time-series is reorganized in a matrix to apply the clustering. Every single row is one feature, which is compared across the different periods k in the following clustering process. Each column is another period like e.g. another day. All time-series input data of one day is represented in a single column. An exemplary column is highlighted. The exemplary column contains multiple features, which are the time-step t values for each attribute a and region n of this day. The attribute "demand" is coloured in red, and the attribute "wind" is coloured in blue. Both attributes exist in the region "west" and "east".

where $\hat{P}_{a,n}$ is the normalized value for an attribute a and region n, $P_{a,n}$ is the original value, $\mu_{a,n}$ is the mean of the attribute and region, and $\sigma_{a,n}$ is the standard deviation of the attribute and region.

A hierarchical clustering method is used to assign the normalized, original periods to fewer cluster groups. The hierarchical clustering method minimizes a distance measure between the aggregated and the full time-series. A Euclidean distance measure $\mathrm{ED}(x, y)$ is used. The Euclidean distance quantifies the dissimilarity of two periods and is common in literature [13]. Euclidean distance $\mathrm{ED}(x, y)$ compares each hourly time-step value $x_{a,n,t}$ of one day with the hourly time-step value $y_{a,n,t}$ of another day:

$$\mathrm{dist}(\vec{x}, \vec{y}) = \mathrm{ED}(\vec{x}, \vec{y}) = \sqrt{\sum_{a \in \mathbf{a}} \sum_{n \in \mathbf{n}} \sum_{t \in \mathbf{t}} (x_{a,n,t} - y_{a,n,t})^2} \tag{2}$$

where $\mathbf{t}$ is the set of all time steps t within a period, $\mathbf{a}$ is the set of all attributes a, and $\mathbf{n}$ is the set of all regions n. The differences between each pair of hourly time-step values are hereby weighted the same. The Euclidean distance measure is used in the iterative hierarchical clustering using Ward's algorithm [38]. The hierarchical clustering algorithm is initialized with a single cluster for each original period, which leads to N cluster sets $\mathbf{C}_k$ $k \in 1, 2, ..., N$. In an iterative loop, the two clusters are merged so that the merging minimizes the Euclidean distance ED measure between the centres of each cluster $\vec{z}_k$ and the points of each cluster $\vec{p}_i$ $\forall i \in \mathbf{C}_k$ [13] [14]:

$$\vec{c}_k = \underset{z}{\mathrm{argmin}} \sum_{i \in \mathbf{C}_k} \mathrm{ED}(\vec{p}_i, \vec{z})^2. \tag{3}$$

Each iteration reduces the total number of cluster sets by 1, and the iterative loop is repeated until the desired number of cluster sets is achieved $\mathbf{C}_k$ $k \in 1, 2, ..., K$.

The cluster centers $\vec{z}_k$ are calculated as the arithmetic mean of all points within the cluster set $\mathbf{C}_k$:

$$\vec{z}_k = \overline{\vec{p}_i} \; \forall i \in \mathbf{C}_k \tag{4}$$

where $\vec{p}_i$ are the different points and $\mathbf{C}_k$ the set of points within cluster k.

A cluster is either respresented by its centroid or by its medoid [13]. The centroid representation uses the arithmetic centre of the cluster $\vec{z}_k$ for the representation. The medoid

representation uses a real period, which is the one closest to the arithmetic centre. A rescaling of the medoid is applied to preserve the original mean [11].

To reverse the normalization after obtaining the normalized representation $\hat{E}_{a,n}$ by clustering, a denormalization step is applied:

$$E_{a,n} = \sigma_{a,n} \cdot \hat{E}_{a,n} + \mu_{a,n} \tag{5}$$

where $E_{a,n}$ is the representative period in the original unit.

2.1.2. Attribute weighting

Within this paper, it is investigated whether giving certain attributes more importance in the aggregation process improves optimization outcomes. Therefore a new weighted Euclidean distance measure is introduced, and the application, which varies the importance of certain attributes in the aggregation process, is called attribute weighting: The Euclidean distance measure (2) is expanded by attribute weights w_a to investigate their influence. A higher attribute weight for one attribute such as e.g. wind (w_{wind}) in comparison to the others means a higher influence of this attributes distance ($\text{dist}(\vec{x}_{\text{wind}}, \vec{y}_{\text{wind}})$) on the total distance ($\text{dist}_{\text{WED}}(\vec{x}, \vec{y})$). Accordingly, a lower attribute weight for one attribute means a lower influence of this attributes distance on the total distance. Each value $x_{a,n,t}$ and $y_{a,n,t}$ is multiplied by an attribute weight w_a in the weighted Euclidean distance:

$$\text{dist}_{\text{WED}}(\vec{x}, \vec{y}) = \text{WED}(\vec{x}, \vec{y}) = \sqrt{\sum_{a \in \mathbf{a}} \sum_{n \in \mathbf{n}} \sum_{t \in \mathbf{t}} \left(w_a \cdot x_{a,n,t} - w_a \cdot y_{a,n,t}\right)^2} \tag{6}$$

The weighted Euclidean distance 6 is used in the hierarchical clustering algorithm:

$$\begin{aligned} \vec{c}_k &= \underset{z}{\text{argmin}} \sum_{i \in \mathbf{C}_k} \text{WED}(\vec{p}_i, \vec{z})^2 \\ &= \underset{z}{\text{argmin}} \sum_{i \in \mathbf{C}_k} \left(\sum_{a \in \mathbf{a}} w_a^2 \cdot \sum_{n \in \mathbf{n}} \sum_{t \in \mathbf{t}} (p_{i,a,n,t} - z_{a,n,t})^2 \right) \end{aligned} \tag{7}$$

2.1.3. Extreme period inclusion

Within this paper, it is also investigated whether the addition of single day extreme periods reduces the error in optimization outcome [39]. It is expected that the extreme periods

maintain some key characteristics of the original time-series and increase the reliability of the supply of the designed energy system. Three extreme periods are selected based on their statistical character. 1) The period with the highest demand, 2) the period with lowest solar availability, and 3) the period with lowest wind availability are chosen:

$$\begin{aligned} &\underset{i}{\text{argmax}} \| \sum_{a,n,t} p_{i,a,n,t} \|_1 \qquad \forall a \in (demand) \\ &\underset{i}{\text{argmin}} \| \sum_{a,n,t} p_{i,a,n,t} \|_1 \qquad \forall a \in (wind, solar) \end{aligned} \tag{8}$$

where $p_{i,a,n,t}$ is a time step value for a time period i, an attribute a, node n, and time t. The selected periods (index i) are excluded from the periods clustered with the hierarchical algorithm and added to the set of representative periods.

2.2. Capacity expansion planning

The CEP is designed to represent electric energy systems. The generation, demand, and availability factors are spatially aggregated to regions and represented by single energy system nodes. Different dispatchable, non-dispatchable, and storage technologies can be set up at each node, and transmission can be set up on the lines between nodes to meet the demand.

The CEP is formulated as a linear optimization model, and the CEP is implemented in the open-source modelling language "julia" [40] using the "JuMP"' package for the optimization formulation [41]. The model and all data used in this paper are provided in the publicly available git repository "CapacityExpansion.jl"[1] [42].

The following sets are used. **n** represents the nodes, **l** represents the lines between nodes, **tech** represents all technologies, **imp** represents impact categories, which can be monetary or life-cycle assessment impact categories [43], **acc** represents the type of costs, which can be annually fixed or variable per produced electric energy, **k** includes the different representative periods, **t** represents the time steps within a period, **i** represents the time periods within the full input data. The subsets of the full sets are indicated as $\mathbf{set}_{subset}$.

[1] `https://github.com/YoungFaithful/CapacityExpansion.jl`

Variables are capitalized for consistency with the published CEP source code. Sets are indicated in bold. It is assumed that an equation is applied for the entire set with the same name, except if explicitly written out another set: The usage of, e.g. VAR_{set1} implicates $VAR_{set1} \forall\ set1 \in \mathbf{set1}$. The objective function minimizes total system costs, where $COST$ is cost of different technologies, LL is lost load, c_{ll} the variable costs for lost load, LE is excess emissions, and c_{le} is the variable costs for excess emissions:

$$\min \sum_{acc,tech} COST_{acc,imp_{money},tech} + \sum_{n} (LL_n \cdot c_{ll}) + \sum_{imp} (LE_{imp} \cdot c_{le,imp}) \tag{9}$$

This minimization is subject to the following constraints:

$$COST_{acc,imp,tech} = \sum_{t,k,n} GEN_{tech,t,k,n} \cdot w_k \cdot \Delta t_{t,k} \cdot c_{acc,tech,imp} \quad \forall\ tech \in \mathbf{tech}_{pw},\ acc \in \{var\} \tag{10}$$

$$COST_{acc,imp,tech} = \sum_{n} CAP_{tech,n} \cdot (c_{acc,tech,imp}) \quad \forall\ tech \in \mathbf{tech}_{pw},\ acc \in \{fix\} \tag{11}$$

$$0 \leq GEN_{tech,t,k,n} \leq CAP_{tech,n} \quad \forall\ tech \in \mathbf{tech}_{disp} \tag{12}$$

$$0 \leq GEN_{tech,t,k,n} \leq CAP_{tech,n} \cdot z_{tech,n,t,k} \quad \forall\ tech \in \mathbf{tech}_{nondisp} \tag{13}$$

$$0 \leq INTRAS_{tech_e,t,k,n} = INTRAS_{tech_e,t-1,k,n} \cdot \eta_{tech_e}^{\Delta t_{t,k}/732h} - \\ - \Delta t_{t,k} \cdot (GEN_{tech_{in},t,k,n} \cdot \eta_{tech_{in}} + \frac{GEN_{tech_{out},t,k,n}}{\eta_{tech_{out}}}) \\ \forall\ tech_{out} \in \mathbf{tech}_{stor,out}, tech_{in} \in \mathbf{tech}_{stor,in}, tech_e \in \mathbf{tech}_{stor,e} \tag{14}$$

$$INTRAS_{sc,tech,t,k,n} \leq CAP_{tech,n} \quad \forall\ tech \in \mathbf{tech}_{stor} \tag{15}$$

$$INTRAS_{sc,tech,te_0,k,n} = INTRAS_{sc,tech,te_{end},k,n} \quad \forall\ tech \in \mathbf{tech}_{stor,e} \tag{16}$$

$$0 \leq FLOW_{dir,tech_l,t,k,l} \tag{17}$$

$$\sum_{dir} |FLOW_{dir,tech,t,k,l}| \leq TRANS_{tech,l} \quad \forall\ tech \in \mathbf{tech}_{trans} \tag{18}$$

$$0 = COST_{acc,imp,tech} \quad \forall\ acc \in \{var\}, tech \in \mathbf{tech}_{trans} \tag{19}$$

$$COST_{acc,imp,tech} = \sum_{n}(TRANS_{tech,l} \cdot len_l) \cdot (c_{fix,tech,imp}) \quad \forall\ tech \in \mathbf{tech}_{trans},\ acc \in \{fix\} \tag{20}$$

$$GEN_{tech,t,k,n} = \sum_{\mathbf{l}_{end}(n)} \left(FLOW_{uniform,tech,t,k,l} - \frac{\sum_l FLOW_{opposite,tech,t,k,l}}{\eta_{tech,l}} \right) - \sum_{\mathbf{l}_{start}(n)} \left(\frac{\sum_l FLOW_{uniform,tech,t,k,l}}{\eta_{tech,l}} - FLOW_{opposite,tech,t,k,l} \right) \quad \forall\ tech \in \mathbf{tech}_{trans} \tag{21}$$

$$\sum_{acc,tech} COST_{acc,imp,tech} \leq LE_{imp} + lim_{imp} \cdot \sum_{n,t,k} (w_k \cdot \Delta t_{t,k} \cdot z_{demand,n,t,k}) \, \forall imp \in \mathbf{imp}_{lca} \tag{22}$$

$$LL_n = \sum_{t,k} (SLACK_{t,k,n} \cdot w_k \cdot \Delta t_{t,k}) \qquad \sum_{tech} GEN_{tech,t,k,n} = z_{demand,n,t,k} - SLACK_{t,k,n} \tag{23}$$

The variable costs are calculated in Equation 10, where GEN is the generation, Δt is the time step length and $c_{acc,tech,imp}$ is the variable cost per electric energy. The fixed costs are calculated in Equation 11, where CAP is the installed capacity, calculating how many years are represented by the original time series. The generation is limited for dispatchable technologies by their installed capacities in Equation 12 and for non-dispatchable technologies by their installed capacities and temporal availability factor z in Equation 13. The storage level throughout each period $INTRAS$ is calculated in Equation 14, where η_{tech} is the efficiency. The total storage is limited to the installed storage capacity in Equation 15 and the storage level is fixed to be the same at the beginning and end of each period in Equation 16. $FLOW$ is the flow over each transmission line and defined positive in Equation 17. The absolute flow is limited to the transmission capacity in Equation 18, where $TRANS$ is the installed transmission capacity. The variable costs of transmission are fixed to zero in Equation 19 and the costs for the installed transmission capacity are calculated in Equation 20. The total flow per node is calculated in Equation 21. The emissions are limited to the emission constraints, which can be exceeded by the excess emissions, in Equation 22. The demand is fixed to the generation and can be exceeded by the lost load in Equation 23.

2.3. Framework

The CEP model is used to formulate a time-series aggregation testing framework (Figure 3). Three different scenarios are calculated for each case of different input data:

- Reference scenario - $R \equiv D_{all} \equiv D_{all}\&O_{all}$
- Design scenario - D_k
- Design & operational scenario - $D_k\&O_{all}$

For the reference scenario (R), the energy system design is optimized using all original time periods of one year. While this is the best temporal representation possible for the given data, modelling all original time periods is computationally complex and only feasible for a "simple" CEP.

The original time periods are typically aggregated to representative periods to reduce the computational complexity. The different time-series aggregation methods introduced in Section 2.1 are applied. The output of the time-series aggregation method applied to the original time periods is a set of k representative periods. The reduced temporal information of k representative periods are used as input data for the design scenario (D_k), which is also formulated as a CEP. The optimization is computationally less expensive than the reference scenario, and the energy system design can vary from the reference scenario because fewer periods are modelled.

The capacities determined in the design scenario D_k are used as an input for the design & operational scenario ($D_k\&O_{all}$), which is formulated as a dispatch problem. In a dispatch problem formulation, the energy system design is fixed, and only the operation is optimized. The design & operational scenario is computationally less complex than the reference scenario because fixing the design variables reduces the computational complexity significantly. A lost load variable 23 is introduced for the case that the design has unreliability of supply and the demand exceeds the installed generation. A lost emission variable 22 is introduced for the case that the policy target is not met and the total emissions exceed the emission constraint. All original time periods are modelled in the design & operational scenario.

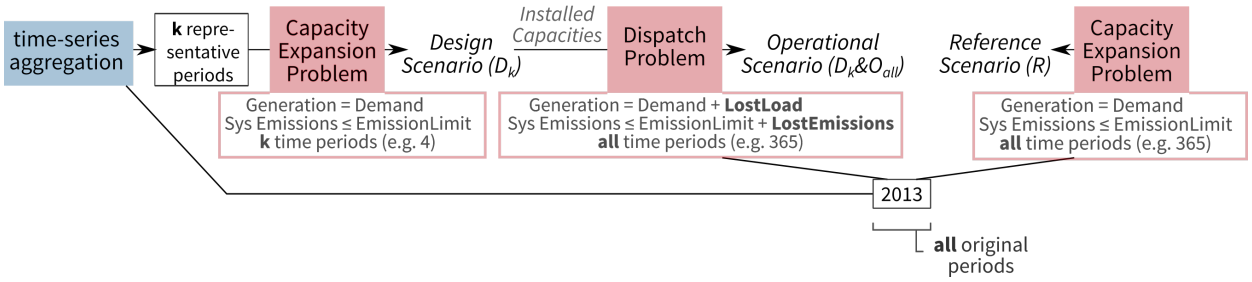

Figure 3: For each case, three scenarios are defined and the results from the reference scenario (R) are compared with the results from the design & operational scenario (D_k&O_{all}). The results for the reference scenario ($R \equiv D_{all} \equiv D_{all}$&$O_{all}$) are found by optimizing design and operation with all original periods in a CEP. The design scenario (D_k) only uses k representative periods to determine installed capacities in a CEP. The results for the design & operational scenario (D_k&O_{all}) by running a dispatch problem (operation only with fixed installed capacities, which were determined by the design scenario (D_k)) and allowing both lost load and lost emissions.

The same single year original time-series is used as an input for the time-series aggregation, design & operational scenario, and reference scenario.

3. Results: Impact investigation

The impact of time-series clustering on capacity expansion is analyzed.

3.1. Hypothesis: Errors mainly result from wind aggregation

The CEP model has hourly time-series input data for solar availability, wind availability, and electricity demand for each node. The wind is suspected to be especially difficult to aggregate because the daily patterns do not share the same regularity of peaks or valleys as found in solar availability and electricity demand. The hypothesis is formulated that aggregating the wind time-series leads to the largest errors in optimization outcome, whereas aggregating the two other attributes, solar and electricity demand, will lead to smaller errors in optimization outcome. The basis of the hypothesis is that solar availability and electricity demand have much less inter daily variability compared to wind and more similar shapes at each day: The solar availability always peaks around noon and electricity demand is governed by working hours.

As a first demonstration, the results of time-series aggregation into $k = 5$ representative periods for the California 1-node 2016 data show that the aggregated availability factors exhibit different patterns for solar (a), electricity demand (b), and wind (c) (Figure 4). The daily curves of solar availability and especially electricity demand are much more similar to each other than the daily curves of wind availability within each cluster because the peaks and bases of solar availability and electricity demand occur at similar hours each day. For example, daily curves of solar availability have their daily maximum near the middle of the light hours. The wind availability has a much more random appearance, and no clear pattern within each cluster emerges. The aggregation of wind availability to 5 representative periods significantly smoothes the original wind periods, and the sum of squared errors is more than four times worse for the clustering of wind availability compared to the clustering of solar availability. The smoothing leads to more constant wind availability in the aggregated representation than it exists in reality. Further, the aggregated representations of wind time series lack extremes.

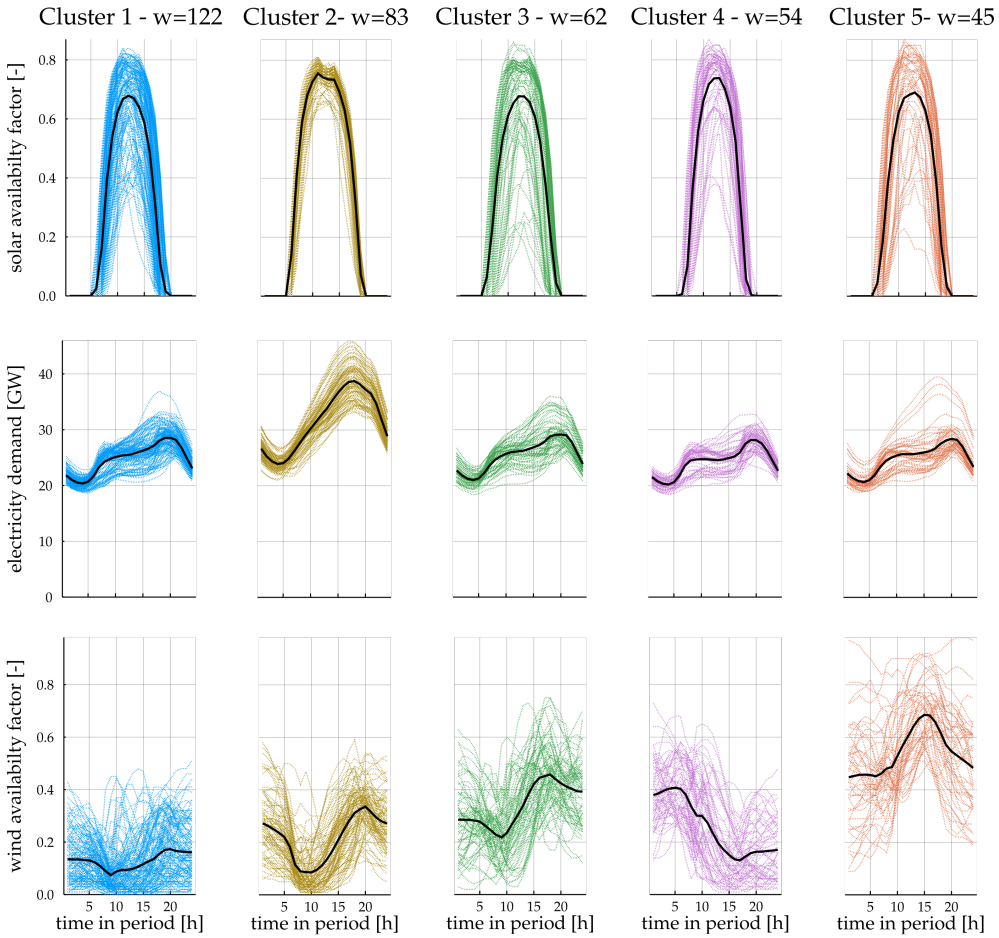

Figure 4: Aggregation of 365 days of solar (a), demand (b), and wind (c) data to 5 representative periods using a centroid hierarchical-clustering. 5 representative periods are used to show all clusters next to each other in one figure. The black lines indicate the representative days (centroid). w indicates the weight of the representative day, which indicates how many original time periods (indicated as dotted lines) are aggregated in this representative period. The closer the original, colored periods are to the representative, black period, the lower the clustering error is. The cluster measures in z-normalized sum of squared distances are a) $SSD \simeq 974.0$, b) $SSD \simeq 1353.2$, and c) $SSD \simeq 3999.0$. It is observed that the daily solar availability factors and electricity demand profiles within each cluster are much closer to each other and have a lower sum of squared distances than the daily wind availability profiles within each cluster.

3.2. Aggregation impact on costs, lost load, and excess emissions

The costs of the reference scenarios (R), which use the full time-series to design the energy system, are presented in Figure 5, increasing the stringency of the emissions target on the x-axis. The total annual reference case costs, including installation and operation for the different dispatchable and non-dispatchable generation technologies, show two key trends: both (1) total costs and (2) share of non-dispatchable generation increase with tighter emission constraints (right side of x-axis in each figure). These trends are the same for the multi-node (a & b) and the single-node (c & d) modelling approaches.

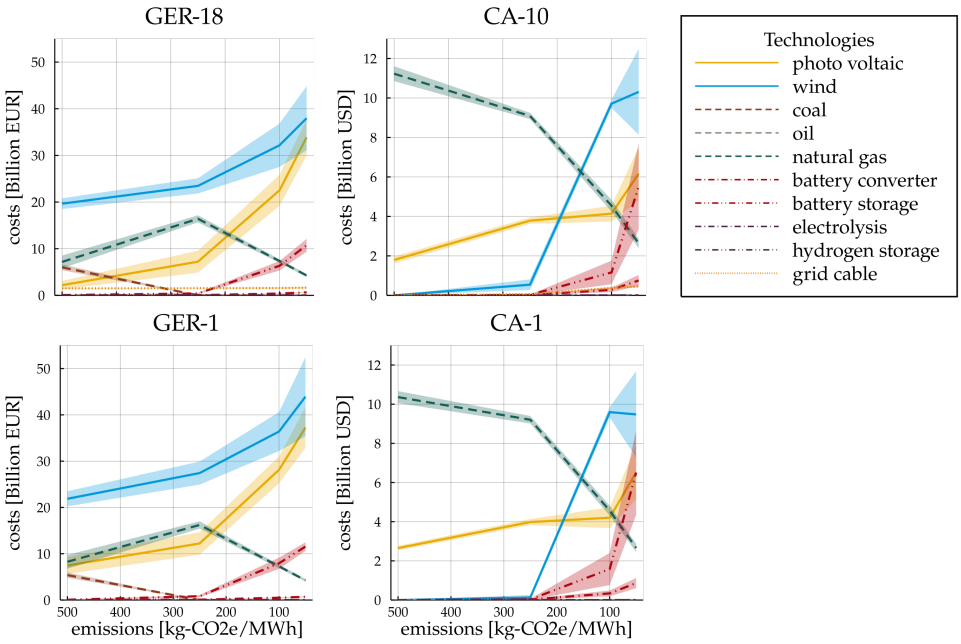

Figure 5: Costs of the reference scenarios (R) with multiple nodes and a single node for Germany (GER) (a & c) and California (CA) (b & d): The annual costs for the reference scenarios for 500 to 50 $\frac{kg-CO2e}{MWh}$. The cost per technology is plotted as a range for the single years of 2010 to 2016 for Germany and the single years of 2014 to 2017 for California. The range is plotted as a shaded area, and the average is plotted as a line. A smaller shaded area indicates less variance of the energy system design based on different weather years. An increase in the shares of photovoltaic costs, storage costs, and especially wind costs as well a decrease in the shares of coal and natural gas costs are observed for tighter emission constraints.

The impact of time-series aggregation is analyzed by examining the results of the two-stage design & operational scenarios (D_{20}&O_{all}). The two-stage design & operational scenarios use the aggregated time series with an exemplary $k = 20$ representative days for the energy system design and operate the resulting fixed energy system design with the full time-series. The results of the two-stage design & operational scenarios (D_{20}&O_{all}) is compared with the results of reference scenarios (R), which use the full time-series for both the energy system design and operation and calculate the relative total system costs, the lost load, and excess emissions (Figure 6).

The reference costs are the total system costs of the optimal energy system that has no lost load or excess emissions and is referenced as 1.0 for each case. The relative costs (a) of cases where time-series aggregation is applied diverge more from the reference total system costs for tighter emission constraints. The relative costs (a) of cases with aggregated time series are smaller if less capacity is installed than necessary to meet demand and emission limits.

The lost load (b) and excess emissions (c) increase for tighter emission constraints and indicate that major policy targets may not be achieved when using energy system designs based on aggregated time series data. The GER-18 cases with an emission constraint of $50\frac{kg-CO2e}{MWh}$ on average estimate less than 75% of total costs, have more than 5% lost load and exceed emissions by more than 100%. The impact on these multi-node cases is on average more than three times the impact on the single-node cases of the same region, GER-1 with an emission constraint of $50\frac{kg-CO2e}{MWh}$. Also, the impacts on the multi-node model CA-10 are higher than the impacts on the single-node model CA-1 for the emission constraints 100 and $50\frac{kg-CO2e}{MWh}$. More nodes within a model represent the spatial complexity of the real energy system more accurately. However, the results indicate that more nodes are also more sensitive to time-series aggregation. The number of features scales linearly with the number of nodes. The features are compared and combined in the clustering process, and the increase in features is one reason for the increased sensitivity of multi-node energy systems. The impact of new time-series aggregation methods should always be tested on multi-node models with tight emission constraints, and higher temporal resolution is necessary when

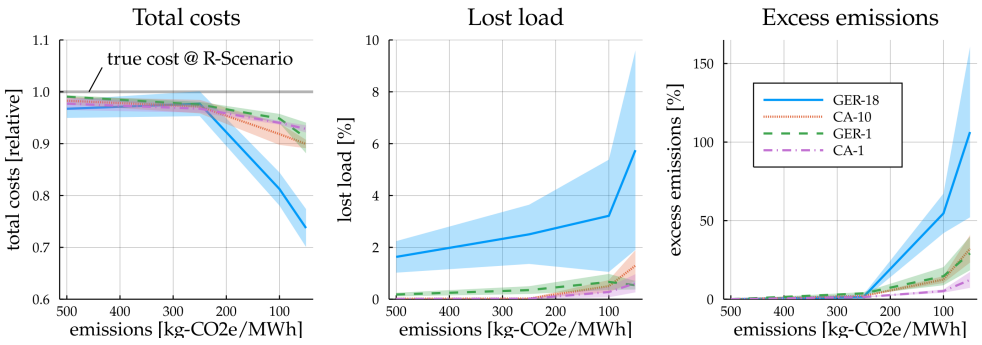

Figure 6: Impact of time-series aggregation on optimization outcome: The total system costs (a) of design & operational scenarios (D_{20}&O_{all}) are shown relative to the total system costs of the reference scenario (R_{all}). The lost load (b) is shown relative to the total demand, and excess emissions (c) are shown relative to the total emissions. The optimization outcome is shown for cases from 500 to 50 $\frac{kg-CO2e}{MWh}$ using 20 representative periods and Germany with 18 nodes, Germany with 1 node, California with 10 nodes, and California with 1 node. An increase in the error in the optimization outcome is observed for tighter emission constraints. Especially the reliability of supply indicated by lost load and the fulfilment of the policy target indicated by lost emission decreases for tighter emission constraints.

modelling more nodes.

3.3. Cause of lost load and excess emissions

The cause of lost load and excess emissions is explained to better understand errors in time-series aggregation.

Excess emissions are caused if the design case (D_k) with time-series aggregation underestimates the actual dispatchable generation required in the operations run (O_{all}) because in the model, only dispatchable generation is carbon-emitting. The underestimation of the required dispatchable generation could arise from an error in foreseeing long duration of low wind and solar availability, and therefore too little storage capacity ("storage error"), but could also arise due to insufficient extreme period representation and too small dispatchable generation capacities ("power error").

The lost load is caused if the difference between the demand and non-dispatchable generation in the full time-series cases – net load – is greater than the installed dispatchable generation and available storage. Lower generation from renewable generation sources and

the same demand leads to a higher "actual" netload. A net load greater than the installed dispatchable generation and available storage leads to a lost load.

The time-series aggregation errors are introduced to the CEP model by the demand time-series and the availability time-series of non-dispatchable generation (Equations 23 and 13). The error in expected demand is independent of the emission constraint, but the error in expected non-dispatchable generation scales with the installed non-dispatchable capacity. The correlation of error in the expected non-dispatchable generation and installed non-dispatchable capacity is a possible explanation why both the share of non-dispatchable generation and the time-series aggregation impacts on lost load and excess emissions increase for tighter emission constraints (Figures 5 and 6).

The effects of lost load and excess emissions are hidden in CEP planning results if the resulting system design is not operated with the full time-series.

3.4. Wind aggregation impact on optimization outcome

Four groups of cases are compared to study the hypothesis that aggregating the wind time-series causes most of the error in optimization outcome: In the first group, none of the three attributes is represented by its full time-series. In the other groups out of the three attributes (demand, solar, and wind), one is represented by the full time-series, and the other two are reduced to $k = 20$ representative periods. It is observed that cases with a full representation of the wind time series have relative costs much closer to 1.0 and significantly reduced lost load and excess emissions compared over multiple years of weather data and regions (GER-1, GER-18, CA-1, CA-10)(Figure 7). Compared to the other cases with a full representation of no time series ("none"), the demand time series, or solar time series, representing the wind data without time-series aggregation reduces the error in relative costs from up to 34% to less than 5.6%, the lost load from up to 7.3% to less than 0.22% and excess emissions from up to 177% to less than 22.5%. It can be concluded that the hypothesis holds: aggregating – and thereby smoothing – the wind time-series mainly leads to errors in optimization outcome, whereas aggregating solar and demand lead to significantly smaller errors in optimization outcome.

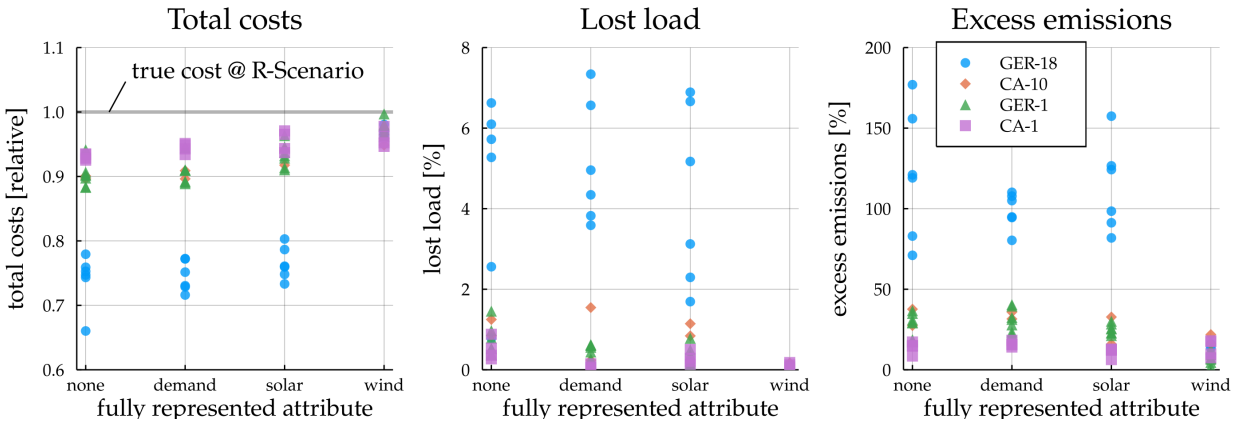

Figure 7: Single fully represented time-series: Out of the three attributes, one is represented by the full time-series, and the other two are aggregated to 20 representative periods. Each fully represented attribute forms one group of cases. An additional group of base cases (*none*) is modelled, where no attribute is represented by a full time-series and all three attributes are aggregated to 20 representative periods. The error in optimization outcome is the difference of relative costs (a), lost load (b), and excess emissions (c) compared to the reference scenario (relative COST = 1, LL = 0, LE = 0). The figure shows $50\frac{kg-CO2e}{MWh}$ cases with multiple years of time-series data and Germany and California being represented as single node and multiple nodes models. It is observed that the error in optimization outcome is lowest in cases of a fully represented wind time series. It is concluded that aggregating and thereby smoothing of the wind time-series mainly leads to errors in optimization outcome, whereas aggregating the two other attributes, solar and demand leads to significantly smaller errors in optimization outcome.

4. Addressing the errors introduced by aggregating the wind time-series

Given the above errors introduced by clustering time periods to generate representative days, 4 options are explored for reducing the impacts of the aggregation: (1) adding extreme periods to cluster representatives; (2) increasing the number of representative periods k; (3) using the medoid instead of the centroid as cluster representation; and (4) weighting of different attributes like wind and solar during clustering.

4.1. Adding single day extreme periods

Because unmet load usually occurs at times of extreme conditions, it is intuitively expected that the addition of extreme periods improves the energy system reliability of supply. Three extreme periods are selected: 1) the period with the highest demand, 2) the period with the lowest solar availability, and 3) the period with the lowest wind availability. These three days are added to the set of $k = 20$ days to represent worst-case outcomes requiring more backup dispatchable capacity investment. Figure 8 shows optimization outcomes in terms of total system costs, lost load, and excess emissions for cases with and without simple extreme values. It is found some improvement with extreme days: the error in total system costs (the difference to the reference scenario (R)) is slightly smaller, while the lost load is significantly reduced by adding extreme days. However, excess emissions increase for cases with extreme values compared to cases without extreme values.

The change in installed capacities can explain the increase in excess emissions and decrease in lost load without a significant change in total cost. More dispatchable capacity, in this case, natural gas, is installed to meet the single extreme days. The installation of more dispatchable generation is necessary to meet the demand and decrease lost load. However, the necessary total dispatchable generation and its emissions are underestimated in the design optimization because the extreme days only represent 3 of total 365 days, and each has the weight according to $\frac{1}{365}$.

Extreme period selection is used in the remainder of this paper due to its significant reduction of lost load.

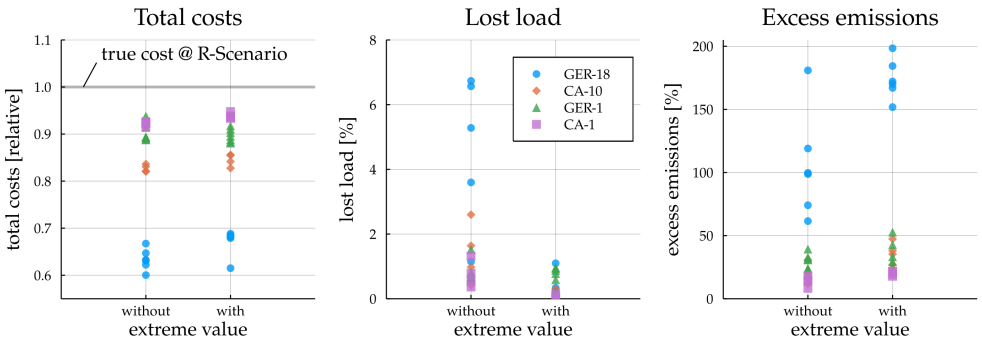

Figure 8: Extreme period selection: The error in optimization outcome is shown as relative costs (a), lost load (b), and excess emissions (c) for $50\frac{kg-CO2e}{MWh}$ cases with and without extreme period selection and a total of 20 representative periods. Cases with multiple years of time series are shown for each representation of Germany and California as single node and multiple nodes models. It is observed that the errors in total system costs and lost load decrease for cases with extreme period selection compared to cases without extreme period selection. However, the significant decrease in lost load comes with an increase in excess emissions.

4.2. Increasing the number of representative periods

The errors in relative total system costs and lost load on the average decrease with more representative periods (Figure 9). However, more representative periods only reduce the excess emissions on average for all modelled regions when more than 30 representative periods are used. The lost load is still up to 0.3% of the total demand, and the excess emissions exceed 100% even in some cases with 80 representative periods. It can be concluded that using more representative periods generally decreases the error in optimization outcome, as would be expected. But improvement is slow with the increasing number of representative periods (k), and relatively high numbers of representative periods are required. Thus the reduction in outcome error comes at high computational costs.

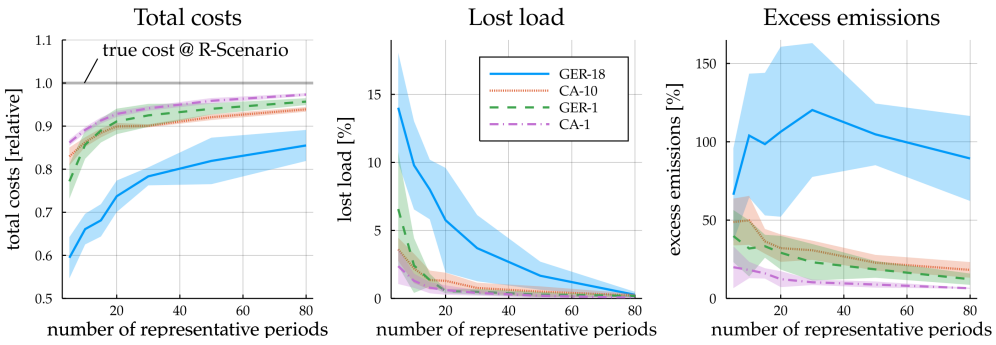

Figure 9: Numbers of representative periods: The error in optimization outcome is shown as relative costs (a), lost load (b), and excess emissions (c) for $50\frac{kg-CO2e}{MWh}$ cases with varying numbers of representative periods. Cases with multiple years of time series are shown as areas for each representation of Germany and California as single node and multiple nodes models. It is observed that the errors in total system costs and lost load decrease with more representative periods at high computational expenses.

4.3. Using the medoid instead of a centroid cluster representation

In the centroid cluster representation, each group of aggregated time periods is represented by the average of all aggregated time periods. However, in the medoid cluster representation, each group of aggregated time periods is represented by the original time period closest to the average of the group, which constraints the cluster representation to one of the original time periods. Figure 10 shows the relative total system costs, the lost loads, and excess emissions for cases with a centroid and medoid cluster representation. The errors in total system costs and lost load slightly decrease using a medoid instead of a centroid cluster representation for the cases modelling California. However, the errors in total system costs, lost load, and excess emissions significantly increase for all cases modelling Germany with multiple nodes and using a medoid instead of a centroid cluster representation.

The single, original time period, which is chosen to represent the group of time periods, has to sufficiently represent the entire group in each attribute and at each node. Representing the group sufficiently for each attribute and at each node is more difficult for more nodes and greater inter-spatial variation of the time periods. It is found that the error in relative total costs almost doubles from up to 38.6% to up to 67.1%. The error in lost load increases more than 50 times from up to 1.29% to up to 70.7%. The error in excess emissions more

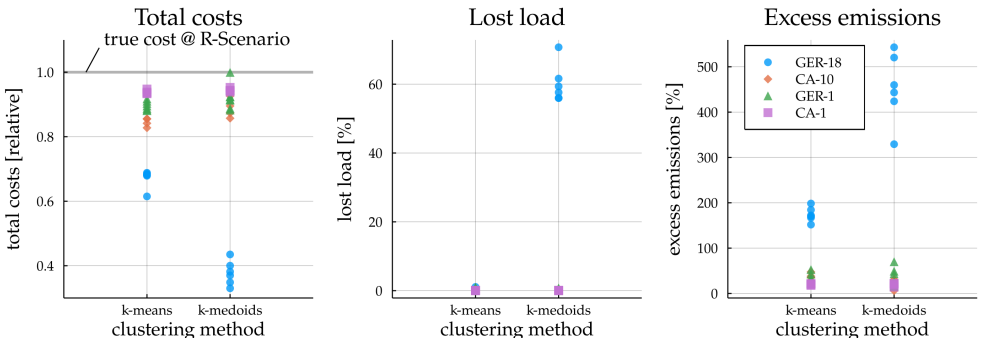

Figure 10: Centroid compared to medoid cluster representation: The error in optimization outcome is shown as relative total cost (a), lost load (b), and excess emissions (c) for cases using either a centroid or a medoid cluster representation. $50\frac{kg-CO2e}{MWh}$ cases with 20 representative periods and multiple years of data for the representations of Germany and California with multiple nodes are investigated. It is observed that the error in optimization outcome is greater for the multiple nodes representation of Germany if the medoid instead of the centroid cluster representation is used. Representing the group sufficiently for each attribute and at each node is more difficult for more nodes and greater inter spacial variation of the time periods.

than doubles from less than 200% to up to 543.5% using a medoid instead of a centroid cluster representation when modelling Germany with multiple nodes. Constraining the cluster representation to one of the original time periods can introduce significant errors to optimization outcome, and it is time-series input data dependent if a model is significantly affected by the increase in error. A priori measures of the time-series can not unveil the entire impact on the energy system. So to avoid the possibility of significant error introduction, the cluster centroid is a better but not totally accurate representation.

4.4. Can re-weighting the attributes during clustering improve optimization outcomes?

Lastly, giving each attribute a variable weight in the clustering method is explored. The higher the weight, the more important the algorithm will give to matching that attribute during clustering. One might imagine, for instance, that weighting errors in wind clustering more heavily will result in clusters that better represent wind variability and thus solve the above problem. The error in optimization outcome is analyzed as the relative total system costs, lost load, excess emissions and differences in installed capacities compared to the reference scenario. Detailed results are provided in the supplementary information (SI).

It is concluded that the neutral weighting of attributes is the best choice to avoid that the error in either total system costs, lost load, or excess emissions more than double. Using weighting other than neutral can reduce the error of either total system costs, lost load, excess emissions, or a certain installed capacity can. Still, the reduction always correlates with the increase of another error.

5. Conclusion

Time-series aggregation methods are often applied to reduce the computational complexity of CEP models. These methods have been investigated for cases of current or near-future electricity system studies. In this work, it is shown that the error in optimization outcome increases for energy systems that have ambitious policy targets with tighter emission constraints and have higher shares of non-dispatchable (i.e. renewable) generation. The designed energy systems have higher unmet demand and emit more emissions than initially planned for. Thus, the energy systems are less reliable in regards to served load and do not meet emission policy targets. It is identified that aggregating and thereby smoothing the wind time-series mainly leads to errors in optimization outcome, whereas aggregating the two other attributes, solar and demand, leads to significantly smaller errors in optimization outcome. It is, therefore, recommended not using time-series aggregation to representative days but using the original, non-aggregated time series in CEP models with ambitious emissions targets if at all possible. The drawback of this method is that complex CEP models will not be solvable on given computational hardware.

To better understand the error introduced by time-series aggregation, furthermore (1) adding extreme days, (2) increasing the number of representative periods, (3) using a medoid instead of a centroid cluster representation, and (4) an attribute specific clustering method was analyzed. (1) Adding extreme days addresses lost load at low computational costs but does not solve or slightly worsens excess emissions. (2) Increasing the number of representative periods can decrease error in total system costs and lost load but comes at high computational costs. (3) It was identified that using a medoid instead of a centroid cluster representation can lead to significant errors in optimization outcome if the single, original

time period is not able to represent all attributes and nodes sufficiently. (4) It is found that using the same attribute weights for wind, solar, and demand is best to avoid greater errors in optimization outcomes. None of the analyzed methods was able to sufficiently address the errors in optimization outcome, which are primarily introduced by the aggregation. It is thus emphasized that one should always validate the optimization output in the form of the energy system design in a second dispatch step. The second dispatch step identifies infeasibility (constraint violation) in the form of lost load or excess emissions. It determines the quality of the energy system design resulting from time-series aggregation.

This paper provides insights into the problems arising during the design of reliable energy systems with a high share of renewable generation sources in combination with current methods of time-series aggregation to reduce the computational complexities. An Open Source analysis framework combining time-series aggregation and capacity expansion is provided. This framework is useful for future research to test newly developed time-series aggregation methods and avoid overseeing significant errors. Future research includes the development and investigation of more complex methods to reduce the error in optimization outcome mainly introduced by the aggregation of the wind time series. Other methods like the seasonal storage coupling approach presented by Kotzur et al. [44] should be tested to reduce errors in sizing long-term storage more accurately and thereby reduce lost load and excess emissions. It should be investigated whether aggregation to representative hours instead of representative days reduces the smoothing effect and thereby decreases lost load and excess emissions.

Acknowledgment

The Stanford Center for Computational Earth and Environmental Science (CEES) provided the computational resources used in this work.

References

[1] UNFCCC. Conference of the Parties (COP), ADOPTION OF THE PARIS AGREEMENT - Conference of the Parties COP 21, Adoption of the Paris Agreement. Proposal by the President. 21932 (2015) 32.

arXiv:arXiv:1011.1669v3, doi:FCCC/CP/2015/L.9/Rev.1.
URL http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf

[2] V. Krey, Global energy-climate scenarios and models: A review, Wiley Interdisciplinary Reviews: Energy and Environment 3 (4) (2014) 363–383. doi:10.1002/wene.98.

[3] L. Gacitua, P. Gallegos, R. Henriquez-Auba, Lorca, M. Negrete-Pincetic, D. Olivares, A. Valenzuela, G. Wenzel, A comprehensive review on expansion planning: Models and tools for energy policy analysis, Renewable and Sustainable Energy Reviews 98 (October 2017) (2018) 346–360. doi:10.1016/j.rser.2018.08.043.
URL https://doi.org/10.1016/j.rser.2018.08.043

[4] N. E. Koltsaklis, A. S. Dagoumas, State-of-the-art generation expansion planning: A review, Applied Energy 230 (April) (2018) 563–589. doi:10.1016/j.apenergy.2018.08.087.

[5] M. Hoffmann, L. Kotzur, D. Stolten, M. Robinius, A review on time series aggregation methods for energy system models, Energies 13. doi:10.3390/en13030641.

[6] F. Domínguez-Muñoz, J. M. Cejudo-López, A. Carrillo-Andrés, M. Gallardo-Salazar, Selection of typical demand days for CHP optimization, Energy and Buildingsdoi:10.1016/j.enbuild.2011.07.024.

[7] S. Fazlollahi, S. L. Bungener, P. Mandel, G. Becker, F. Maréchal, Multi-objectives, multi-period optimization of district energy systems: I. Selection of typical operating periods, Computers and Chemical Engineering 65 (2014) 54–66. doi:10.1016/j.compchemeng.2014.03.005.

[8] T. Schütz, M. H. Schraven, M. Fuchs, P. Remmen, D. Müller, Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis, Renewable Energy 129 (2018) 570–582. doi:10.1016/j.renene.2018.06.028.

[9] P. Gabrielli, M. Gazzani, E. Martelli, M. Mazzotti, Optimal design of multi-energy systems with seasonal storage, Applied Energy 219 (2018) 408–424. doi:10.1016/j.apenergy.2017.07.142.
URL https://doi.org/10.1016/j.apenergy.2017.07.142

[10] B. Bahl, A. Kümpel, H. Seele, M. Lampe, A. Bardow, Time-series aggregation for synthesis problems by bounding error in the objective function, Energy 135 (2017) 900–912. doi:10.1016/j.energy.2017.06.082.

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

[12] S. Pfenninger, Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability, Applied Energy 197 (2017) 1–13. doi:10.1016/j.apenergy.2017.03.051.

URL http://dx.doi.org/10.1016/j.apenergy.2017.03.051

[13] H. Teichgraeber, A. R. Brandt, Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison, Applied Energy (2019) 1283–1293doi:10.1016/j.apenergy.2019.02.012.

[14] L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, Impact of different time series aggregation methods on optimal energy system design, Renewable Energy 117 (2018) 474–487. arXiv:1708.00420, doi:10.1016/j.renene.2017.10.017.

[15] M. Zeyringer, J. Price, B. Fais, P. H. Li, E. Sharp, Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather, Nature Energy 3 (5) (2018) 395–403. doi:10.1038/s41560-018-0128-x.
URL http://dx.doi.org/10.1038/s41560-018-0128-x

[16] R. Green, I. Staffell, N. Vasilakos, 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) (2014) 251–260. doi:10.1109/TEM.2013.2284386.

[17] C. F. Heuberger, E. S. Rubin, I. Staffell, N. Shah, N. Mac Dowell, Power capacity expansion planning considering endogenous technology cost learning, Applied Energydoi:10.1016/j.apenergy.2017.07.075.

[18] F. D. Munoz, B. F. Hobbs, J. P. Watson, New bounding and decomposition approaches for MILP investment problems: Multi-area transmission and generation planning under policy constraints, European Journal of Operational Researchdoi:10.1016/j.ejor.2015.07.057.

[19] G. J. Blanford, J. H. Merrick, J. E. Bistline, D. T. Young, Simulating Annual Variation in Load, Wind, and Solar by Representative Hour Selection, The Energy Journal 39. doi:10.5547/01956574.39.3.gbla.

[20] N. Baumgärtner, S. Deutz, C. Reinert, N. Nolzen, L. E. Kuepper, M. Hennen, D. E. Hollermann, A. Bardow, Life-cycle assessment of sector-coupled national energy systems: Environmental impacts of electricity, heat, and transportation in germany till 2050, Frontiers in Energy Research 9. doi:10.3929/ethz-b-000481968.

[21] C. L. Lara, D. S. Mallapragada, D. J. Papageorgiou, A. Venkatesh, I. E. Grossmann, Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm, European Journal of Operational Researchdoi:10.1016/j.ejor.2018.05.039.

[22] J. H. Merrick, On representation of temporal variability in electricity capacity planning models, Energy Economics 59 (2016) 261–274. doi:10.1016/j.eneco.2016.08.001.
URL http://dx.doi.org/10.1016/j.eneco.2016.08.001

[23] A. Pina, C. Silva, P. Ferrão, Modeling hourly electricity dynamics for policy making in long-term

scenarios, Energy Policy 39 (9) (2011) 4692–4702. doi:10.1016/j.enpol.2011.06.062.

[24] T. Schütz, M. H. Schraven, H. Harb, M. Fuchs, D. Müller, Clustering algorithms for the selection of typical demand days for the optimal design of building energy systems, Proceedings of the 29th International Conference on Efficiency, Cost, Optimisation, Simulation and Environmental Impact of Energy Systems (ECOS 2016).

[25] P. G. Brodrick, C. A. Kang, A. R. Brandt, L. J. Durlofsky, Optimization of carbon-capture-enabled coal-gas-solar power generation, Energydoi:10.1016/j.energy.2014.11.003.

[26] P. G. Brodrick, A. R. Brandt, L. J. Durlofsky, Operational optimization of an integrated solar combined cycle under practical time-dependent constraints, Energydoi:10.1016/j.energy.2017.11.059.

[27] J. Schilling, K. Eichler, S. Pischinger, A. Bardow, Integrated design of ORC process and working fluid for transient waste-heat recovery from heavy-duty vehicles, AppliedEnergy 255 (2019) 113207. doi:10.1016/j.apenergy.2019.05.010.

[28] H. Teichgraeber, P. G. Brodrick, A. R. Brandt, Optimal design and operations of a flexible oxyfuel natural gas plant, Energy 141 (2017) 506–518. doi:10.1016/j.energy.2017.09.087.
URL https://doi.org/10.1016/j.energy.2017.09.087

[29] N. Baumgärtner, B. Bahl, M. Hennen, A. Bardow, RiSES3: Rigorous Synthesis of Energy Supply and Storage Systems via time-series relaxation and aggregation, Computers and Chemical Engineering 127 (2019) 127–139. doi:10.1016/j.compchemeng.2019.02.006.

[30] R. Yokoyama, K. Takeuchi, Y. Shinano, T. Wakui, Effect of model reduction by time aggregation in multiobjective optimal design of energy supply systems by a hierarchical milp method, Energy 228 (2021) 120505. doi:https://doi.org/10.1016/j.energy.2021.120505.
URL https://www.sciencedirect.com/science/article/pii/S0360544221007544

[31] B. Bahl, J. Lützow, D. Shu, D. Elena, M. Lampe, M. Hennen, A. Bardow, Rigorous synthesis of energy systems by decomposition via time-series aggregation, Computers and Chemical Engineering 112 (2018) 70–81. doi:10.1016/j.compchemeng.2018.01.023.
URL https://doi.org/10.1016/j.compchemeng.2018.01.023

[32] N. Baumgärtner, M. Leisin, B. Bahl, M. Hennen, A. Bardow, Rigorous synthesis of energy systems by relaxation and time-series aggregation to typical periods, Computer Aided Chemical Engineering 44 (2018) 793–798. doi:10.1016/B978-0-444-64241-7.50127-0.

[33] L. Göke, M. Kendziorski, The adequacy of time-series reduction for renewable energy systems (2021). arXiv:2101.06221.

[34] K. Poncelet, E. Delarue, D. Six, J. Duerinck, W. D'haeseleer, Impact of the level of temporal and operational detail in energy-system planning models, Applied Energy 162 (2016) 631–643. doi:10.1016/j.apenergy.2015.10.100.

[35] H. Teichgraeber, L. Kuepper, A. Brandt, TimeSeriesClustering: An extensible framework in Julia, Journal of Open Source Software 4 (41) (2019) 1573. doi:10.21105/joss.01573.

[36] Y. Liu, R. Sioshansi, A. J. Conejo, Hierarchical Clustering to Find Representative Operating Periods for Capacity-Expansion Modeling, IEEE Transactions on Power Systems 33 (3) (2018) 3029–3039. doi:10.1109/TPWRS.2017.2746379.

[37] Y. Li, B. Wang, Z. Yang, J. Li, C. Chen, Hierarchical stochastic scheduling of multi-community integrated energy systems in uncertain environments via stackelberg game, Applied Energy 308 (2022) 118392. doi:https://doi.org/10.1016/j.apenergy.2021.118392.
URL https://www.sciencedirect.com/science/article/pii/S0306261921016299

[38] J. Ward, JH, Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association.

[39] H. Teichgraeber, C. P. Lindenmeyer, N. Baumgärtner, L. Kotzur, D. Stolten, M. Robinius, A. Bardow, A. R. Brandt, Extreme events in time series aggregation: A case study for optimal residential energy supply systems, Applied Energy 275 (2020) 115223. doi:https://doi.org/10.1016/j.apenergy.2020.115223.
URL https://www.sciencedirect.com/science/article/pii/S0306261920307352

[40] J. Bezanson, A. Edelman, S. Karpinski, V. B. Shah, Julia: A fresh approach to numerical computing, SIAM ReviewarXiv:1411.1607, doi:10.1137/141000671.

[41] I. Dunning, J. Huchette, M. Lubin, JuMP: A Modeling Language for Mathematical Optimization, MIT 59 (2) (2015) 295–320. arXiv:1508.01982, doi:10.1137/15M1020575.

[42] L. E. Kuepper, H. Teichgraeber, A. R. Brandt, Capacityexpansion: A capacity expansion modeling framework in julia, Journal of Open Source Software 5 (52) (2020) 2034. doi:10.21105/joss.02034.
URL https://doi.org/10.21105/joss.02034

[43] M. Pehl, A. Arvesen, F. Humpenöder, A. Popp, E. G. Hertwich, G. Luderer, Understanding future emissions from low-carbon power systems by integration of life-cycle assessment and integrated energy modelling, Nature Energy 2 (12) (2017) 939–945. doi:10.1038/s41560-017-0032-9.
URL http://dx.doi.org/10.1038/s41560-017-0032-9

[44] L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, Time series aggregation for energy system design: Modeling seasonal storage, Applied Energy 213 (October 2017) (2018) 123–135. arXiv:1710.07593, doi:10.1016/j.apenergy.2018.01.023.

Supplementary information

Additional information about the re-weighting of attributes during the clustering is provided in this section.

Can re-weighting the attributes during clustering result in improved optimization outcomes?

Giving each attribute a variable weight in the clustering method is explored. The higher the weight, the more important the algorithm will give to matching that attribute during clustering. The error in optimization outcome is analyzed as the relative total system costs, lost load, excess emissions and differences in installed capacities compared to the reference scenario and divided by the total installed power capacity. Figure 11 shows average error across all years in optimization outcome for $50\frac{kg-CO2e}{MWh}$ cases with varying weights for each attribute in the design & operational scenarios ($D_{20}O_{all}$). The darker the blue colour, the more positive the difference is from the reference scenario, and equivalently the darker red, the more negative the difference is. The best case with no difference has a light grey colour. The scales for the colours change for each plot and are shown in the colour bar next to each plot. The attribute weights vary along the x- and y-axis of the plots. In the direction of the positive x-axis, the attribute weights for wind are increased, and in the direction of the positive y-axis, similarly, the attribute weights of solar are increased. The base case has an attribute weight of 1.0 for solar and wind as well as for the electricity demand. The base case is found in the centre of each figure, and the base case is equivalent to using a non-weighted Euclidean distance measure during clustering. All previously reported results correspond to using attribute weights as in the base case (centre of each plot).

The errors in total system costs and excess emissions are greatest when weighting solar higher than demand and wind, as Figure 11 shows. On the other hand, errors in total system costs, lost load, and excess emissions also increase for a higher weighting of wind than demand and solar.

It is observed that the reduction of one specific error using attribute weighting often leads to an increase of another specific error. This contradictory effect applies to the installed capacities of wind and PV shown in Figure 11. The summation of necessary capacities is

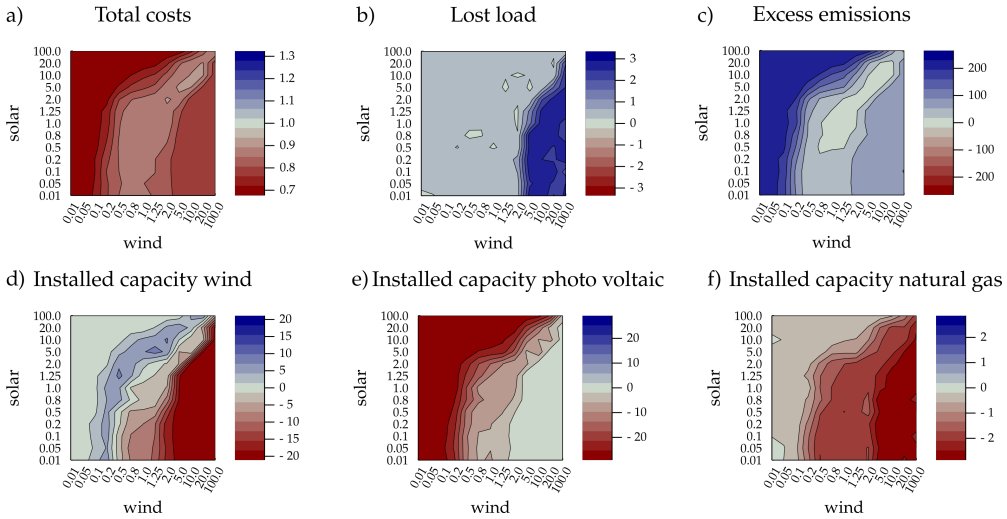

Figure 11: Attribute specific weighting: The error in optimization output is shown as (a) relative total costs, (b) lost load, (c) excess emissions and difference in installed capacities of (d) wind, (e) PV, and (f) natural gas between the design & operational scenario (D_{20}&O_{all}) with an aggregated time series and the reference scenario (R) with the full time-series. The attribute weights vary along the x-axis for wind and along the y-axis for solar, while the attribute weight of demand is 1.0 in all cases. The base case has an attribute weight of 1.0 for all attributes and is found in the centre of each figure. It is observed that the reduction of one specific error using attribute weighting often leads to an increase of another specific error. It is also observed that the centre does not contain any immoderate errors in either direction and is, therefore, the best choice of attribute weights.

too little in any attribute weighting configuration. The overall effect that time-series data are smoothed by aggregation remains because increasing the weighting of the attribute (a) leads to less smoothing of the time-series data of attribute (a), but more smoothing of the time-series data of the attributes (b) and (c). Attribute weighting only shifts the capacities from one generation or storage technology to another.