

The Pareto-Optimal Temporal Aggregation of Energy System Models

Maximilian Hoffmann,^{a,b,1} Leander Kotzur,^a and Detlef Stolten^{a,b}

^a Institute of Energy and Climate Research, Techno-economic Systems Analysis (IEK-3), Forschungszentrum Jülich, 52428 Jülich, Germany

^b Chair for Fuel Cells, RWTH Aachen University, c/o Techno-economic Systems Analysis (IEK-3), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Str., 52428 Jülich, Germany

Abstract

The growing share of intermittent renewable energy sources, storage technologies, and the increasing degree of so-called sector coupling necessitates optimization-based energy system models with high temporal and spatial resolutions, which significantly increases their runtimes and limits their maximum sizes. In order to maintain the computational viability of these models for large-scale application cases, temporal aggregation has emerged as a technique for reducing the number of considered time steps by reducing the original time horizon down to fewer, more representative ones.

This study presents advanced but generally applicable clustering techniques that allow for ad-hoc improvements of current approaches without requiring profound knowledge of the individual energy system model. These improvements comprise a method to find the optimal tradeoff between the number of typical days and inner-daily temporal resolutions, and a representation method that can reproduce the value distribution of the original time series. We prove the superiority of these approaches by applying them to two fundamentally different model types and benchmarking them against state-of-the-art approaches. This is performed for a variety of temporal resolutions, which leads to many hundreds of model runs.

The results show that the proposed improvements on current methods strictly dominate the status quo with respect to Pareto-optimality in terms of runtime and accuracy. Although a speeding up factor of one magnitude could be achieved using traditional aggregation methods within a cost deviation range of two percent, the algorithms proposed herein achieve this accuracy with a runtime speedup by a factor of two orders of magnitude.

Highlights

- The ratio of typical days and time steps therein is crucial for a good optimization result.
- An algorithm for an improved ratio of typical days and segments is introduced.
- The duration curves of time series are crucial for the correct sizing of the components.
- An algorithm is introduced that precisely replicates the original value distribution.
- The methods are open source and available at: <https://github.com/FZJ-IEK3-VSA/tsam>.

Keywords: Typical days, segmentation, time slices, energy system models, time series aggregation, temporal aggregation, temporal resolution, renewable energy systems, clustering, computation time, Pareto optimality

¹ Corresponding author: M. Hoffmann (max.hoffmann@fz-juelich.de)

1. Introduction

Current and future energy systems will be subject to a steadily increasing share of intermittent renewable energy sources due to the necessity of reducing carbon dioxide emissions, which also leads to an intensification of synergies, or ‘coupling’, across different energy sectors [1, 2]. These trends increase system complexity and drive the need for spatially- and temporally-highly-resolved energy system models. In contrast, the growth rate of transistor density, which has been constant over the decades and is known as Moore’s Law [3, 4], has been decreasing in recent years [5].

The conflict between fast-growing energy system models and a decreasing growth rate of computing power has driven the development of algorithms that focus on reducing the complexity of energy system models [6, 7]. One of these approaches has been the reduction of the number of time steps considered in these time-discrete models, for which a large variety of different approaches exist [8]. Among these are aggregation procedures based on clustering, which have emerged as the most accurate methods due to their capability of accounting for the similarity of different time steps or periods. However, most methods described in the literature either focus on reducing the number of time steps by increasing their duration or by representing the time series with a few typical periods, whereas the use of both methods together has been widely neglected. Furthermore, the vast majority of studies have employed well-established clustering and representation methods, such as k-means, k-medoids, and Ward’s hierarchical algorithm, all of which have led to a distortion of the original time series’ extreme values and value distributions (duration curve). Accordingly, this work addresses two major research questions:

Research question 1: Is there an optimal tradeoff between the duration of time steps and the number of typical periods, and are certain technologies affected systematically by either of these temporal aggregation approaches?

Research question 2: Can the aggregation procedure be improved upon by explicitly taking statistical momentums, i.e., the original time series’ value distribution, into account?

Both research questions are investigated using two fundamentally different energy system models, and the methods developed in this study are shown to outperform state-of-the-art approaches using the criterion of Pareto-optimality.

The remainder of this work is structured as follows: In Section 2, we review current developments in temporal aggregation for energy system models. Section 3 introduces the modifications of standard aggregation procedures as proposed herein and provides a short overview of the case studies used to validate the effectiveness of the proposed modifications. In particular, Section 3.2 proposes an approach how to arrive at a balanced number of typical days and inner-daily number of time steps, while Section 3.3 introduces an algorithm that is capable of reproducing the original time series’ value distribution on the aggregated dataset. Section 4 analyzes the impact of the proposed aggregation procedures on the given case study. Finally, the findings are summarized and concluded in Section 5.

2. State-of-the-art

The following section is divided into two parts. Section 2.1 reviews approaches to reducing the number of time steps, whereas Section 0 outlines subsequent methods for improving the quality of the aggregated time series by manually or automatically inserting the statistical features of the fully resolved time series that could be relevant for energy system optimization.

2.1. Typical periods and segments

Time series can be aggregated in two different ways, either by directly reducing the temporal resolution, i.e., merging adjacent time steps, or by aggregating multiple periods of consecutive time

steps into fewer periods. As many time series in the realm of energy system modeling, such as of electricity or photovoltaic feed-in profiles, have a daily pattern, the most common period length is 24 hours. In that case, this method is also referred to as *typical or representative days*. As this approach affects the number of considered periods but not the actual temporal resolution of discrete time steps within these, the temporal resolution can be further reduced in a subsequent step. The impact of aggregation on a typical single time series is depicted in Figure 1. Here, it can be seen that the aggregation to typical periods leads to fewer periods, whereas segmentation leads directly to fewer time steps using a coarser resolution. In the example, the original time series with 8760 hourly time steps was reduced to eight typical days comprising eight segments with irregular lengths each, ending up in 64 time steps, which is equal to a time step reduction of well above 99%. As the runtime of energy system models scales nonlinearly with the number of simultaneously considered time steps, this reduction may lead to an even higher speed-up factor.

Besides, the Root Mean Squared Error (RMSE) is listed in each subplot, which is an error indicator between the original time series and the aggregated one and which is described in detail in Section 3.2. Generally, the error increases if smaller numbers of typical days or segments are chosen to represent the original time series and is not additive. Therefore, the RMSE of the aggregated time series in the lower right figure is smaller than the sum of the RMSEs of the upper right and lower left time series but larger than each of these two RMSEs is, respectively.

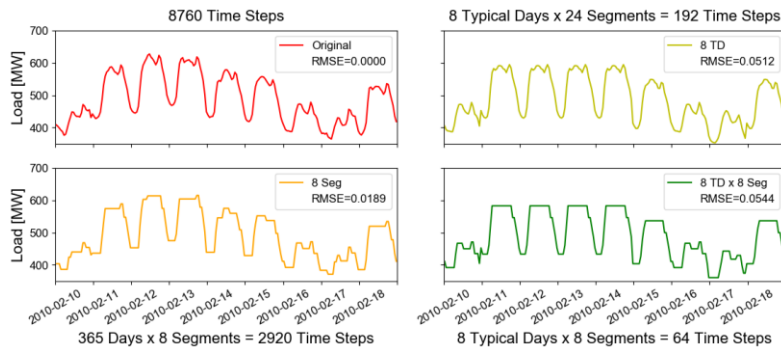


Figure 1. Time step reduction based on a combination of period clustering and segmentation.

Despite the option to freely combine typical periods and a decrease in inner-periodical temporal resolution, very few authors have considered both approaches simultaneously in the literature, which is illustrated in Table 1.

The methods that resemble the approach taken in this study presented in Section 3.2 were mostly developed by Fazlollahi et al. [9], as well as Bahl et al. [10-13]. Fazlollahi et al. employed k-means clustering in order to cluster both periods and the time steps within these periods until a set of input data-based quality criteria were fulfilled. As the segmentation must fulfill the constraint of adjacency, which cannot be directly respected by the k-means algorithm, the segmentation is performed in an iterative manner. In contrast, Bahl et al. perform optimization runs for a certain number of typical days and segments per day in order to obtain upper and lower bounds of the original problem. Then, they either increased the number of segments within each typical day or the number of typical days based on a heuristic to select the more promising direction of refinement in order to obtain tighter bounds on the original model. Finally, Raventos et al. [14] and Wogrin et al. [15] compared the approaches of typical days and segments with one another in order to evaluate their advantages and disadvantages, but did not combine both approaches.

The fact that these approaches did not explicitly compare the performance of various cross-combined aggregation configurations drove the motivation for a large-scale sensitivity analysis, as performed in the following.

Table 1. Studies using either typical periods or segments based on clustering.

Publication	Year	Periods	Segments	Combined	Publication	Year	Periods	Segments	Combined
Domínguez-Muñoz et al. [16]	2011	✗			Schütz et al. [17-19]	2016-18	✗		
Adhau et al. [20]	2014	✗			Gabrielli et al. [21, 22]	2018/19	✗		
Green et al. [23]	2014	✗			Teichgräber et al. [24-27]	2017-20	✗		
Wogrin et al. [28, 29]	2014/16	✗			Kannengießer et al. [30]	2019	✗		
Agapoff et al. [31]	2015	✗			Sun et al. [32]	2019	✗		
Pfenninger et al. [33]	2017	✗			Zatti et al. [34]	2019	✗		
Brodrick et al. [35, 36]	2015/17	✗			Zhang et al. [37]	2019	✗		
Fitiwi et al. [38]	2015	✗			Minh et al. [39]	2020	✗		
Marquant et al. [40, 41]	2015/17	✗			Pecenak et al. [42]	2020	✗		
Munoz et al. [43]	2015	✗			Pinel et al. [44]	2020	✗		
Schiefelbein et al. [45]	2015	✗			Tso et al. [46]	2020	✗		
Lin et al. [47]	2016	✗			Yokoyama et al. 2020	2020	✗		
Ploussard et al. [48]	2016	✗			Hoffmann et al. [49]	2021	✗		
Nahmmacher et al. [50]	2016	✗			Teichgräber et al. [51]	2021	✗		
Bahl et al. [52]	2017	✗			Yokoyama et al. [53]	2021	✗		
Härtel et al. [54]	2017	✗			Göke et al. [55]	2022	✗		
Heuberger et al. [56]	2017	✗			Raventos et al. [14]	2020		✗	
Zhu et al. [57]	2017	✗			Deml et al. [58]	2015		✗	
Stadler et al. [59]	2018	✗			Vom Stein et al. [60]	2017		✗	
Kotzur et al. [61, 62]	2018	✗			Pineda et al. [63]	2018		✗	
Lara et al. [64]	2018	✗			Mavrotas et al. [65]	2008	✗	✗	✗
Liu et al. [66]	2018	✗			Fazlollahi et al. [9]	2014	✗	✗	✗
Reichenberg et al. [67]	2018	✗			Bahl et al. [10, 11]	2018	✗	✗	✗
Tejada-Arango et al. [68, 69]	2018	✗			Baumgärtner et al. [12, 13]	2019	✗	✗	✗
Tupper et al. [70]	2018	✗			Wogrin et al. [15]	2019	✗	✗	
Welder et al. [71]	2018	✗			Raventos et al. [14]	2020	✗	✗	

2.2. Modified clustering techniques

Apart from clustering, time series aggregation also employs a representation step as highlighted in Figure 2. In general, arbitrary value tuples can be used to represent a cluster of time steps from the original time series. However, medoids or centroids are the most frequently used in the literature.

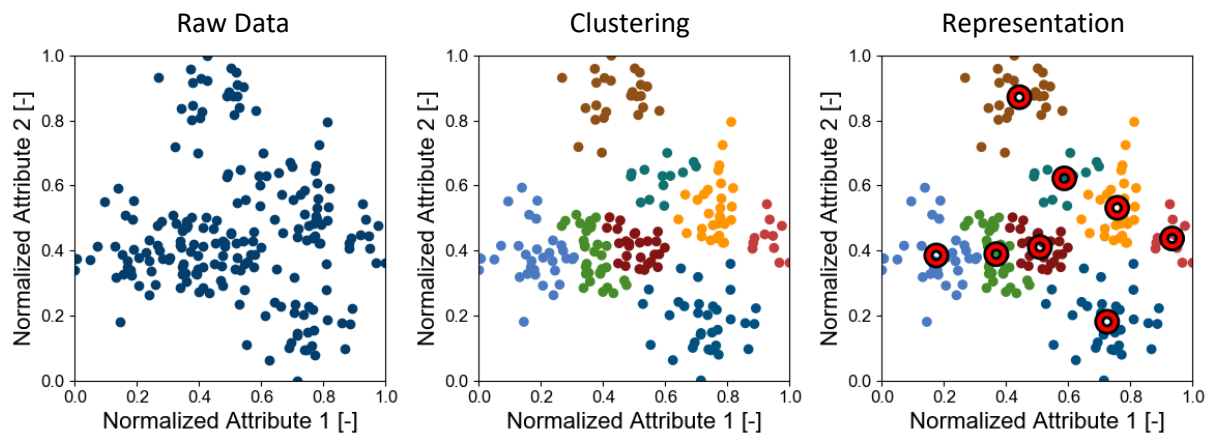


Figure 2. Clustering and representation as separate steps of clustering.

The problem of using centroids or medoids is illustrated in Figure 3 for the case of centroids. It can be seen that centroids lead to an underestimation of the time series' extreme values, which is apparent for both its duration curve (left) and an extraction of the original time series during a week in February.

This underestimation often leads to an underestimation of the system costs as demand maxima are underestimated and supply minima overestimated.

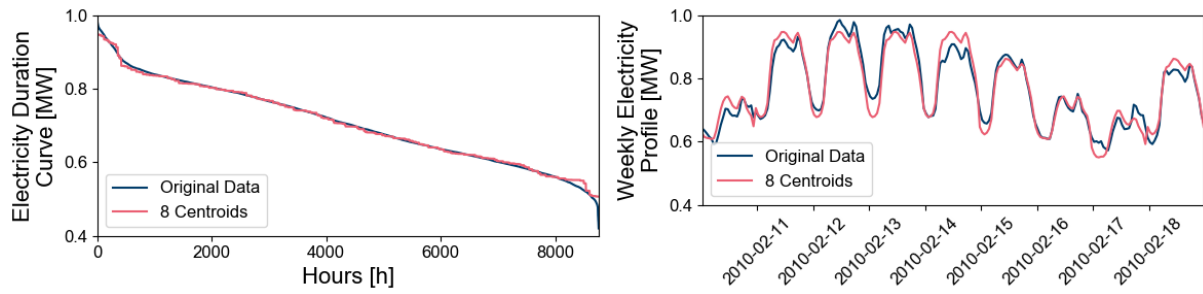


Figure 3. The yearly duration curve of a typical electricity profile, its profile during a week in February, and the corresponding aggregated time series as predicted by eight typical days using hierarchical clustering and centroids as representative values.

To address this issue, many approaches have been developed in the literature that either are based on the manual integration of extreme values, or are algorithms that automatically preserve the desirable features of the original time series. Table 2 provides an overview of the different approaches.

Table 2. Methods employed to subsequently improve the quality of the clustering process.

Publication	Year	Manual Peaks	Algorithmic Peaks	Publication	Year	Manual Peaks	Algorithmic Peaks
Mavrotas et al. [65]	2008	✗		Patteeuw et al. [72]	2016	✗	
Domínguez-Muñoz et al. [16]	2011	✗		Poncelet et al. [73]	2016		✗
Ortiga et al. [74]	2011	✗		Bahl et al. [10, 11, 52]	2017/18		✗
Devogelaer et al. [75]	2012	✗		Härtel et al. [54]	2017		✗
De Sisternes et al. [76, 77]	2013/16	✗		Heuberger et al. [56]	2017	✗	
Simões et al. [78]	2013	✗		Pfenninger et al. [33]	2017	✗	
Voll et al. [79]	2013	✗		Brodrick et al. [36, 80]	2017/18	✗	
Fazlollahi et al. [9, 81]	2014	✗		Gabrielli et al. [21, 22]	2018/19	✗	
Green et al. [23]	2014		✗	Kotzur et al. [62]	2018	✗	
Poncelet et al. [82, 83]	2014/16	✗		Baumgärtner et al. [12, 13]	2019	✗	
Stadler et al. [84]	2014	✗		Hilbers et al. [85]	2019		✗
Wakui et al. [86-88]	2014/16	✗		Zatti et al. [34]	2019		✗
Agapoff et al. [31]	2015		✗	Minh et al. [39]	2020	✗	
Marquant et al. [40, 41]	2015	✗		Pinel [44]	2020	✗	
Munoz et al. [43]	2015	✗		Pecenak et al. [42]	2020		✗
Frew et al. [89]	2016	✗		Teichgräber et al. [51]	2021		✗
Merrick et al. [90]	2016	✗		Göke et al. [55]	2022		✗

Among the algorithm-based approaches, the findings of Agapoff et al. [31] and Poncelet et al. [73] have influenced the method developed in Section 3.3 the most due to their focus on the statistical features of the original time series, rather than single extreme values. Both authors refrained from directly clustering the original time series. Instead, Agapoff et al. selected local price differences, as well as non-controllable demand and generation time series for transmission expansion planning, as well as their statistical features of average, minimum, maximum and standard deviations to directly perform the clustering of statistical attributes instead of the original time series. In contrast, Poncelet et al. introduced a MILP approach for clustering time series in such a way that their respective duration curves are as closely approximated as possible. Based on these approaches, the method presented in Section 3.3 was developed to explicitly preserve the original value distribution and thereby the statistical key features of the original time series.

In summary, current methods offer very little information on a good ratio between the number of typical days and their inner-daily resolutions, or more generally approaches to account for extreme situations, without manually manipulating certain time series on the basis of the modeler's experience.

3. Methodology

In the following, the approaches proposed to improve the aggregation process are introduced, as well as the case studies used for their validation.

3.1. Clustering typical days and segments

The clustering process used to aggregate the time series to typical days utilized in this work is based on a procedure that was first introduced by Domínguez-Muñoz et al. [16], and which is frequently described in the literature, e.g., by Nahmmacher et al. [50], as well as the authors of this work [8, 49, 62], who also made this functionality available open-source as part of their tsam modeling tool.² For the sake of completeness, however, the process is also depicted in Figure 4. In a first step, the time series are normalized in order to prevent an overweighing of those with large scales (e.g., total energy demands) compared to those with small scales (e.g., capacity factors). This can either happen using a min-max normalization to values of between 0 and 1, or by using a z-normalization, which leads to time series with a standard deviation of 1 [49]. Then, the time series are rearranged such that all steps of all time series within a certain period, e.g., one day, become a row vector each of which can then be interpreted as a sample point in a hyper-dimensional space and therefore clustered with other sample points in order to achieve groups of days that share a certain degree of similarity. In our study, we employed hierarchical clustering [91] due to its deterministic behavior. Ultimately, one representative is determined for each cluster using either the cluster's centroid, medoid, or a user-specified approach and scaled back. In the energy system model, these typical periods are then weighted according to their respective cluster sizes in order to account for the relative frequency of certain day types throughout the considered time horizon.

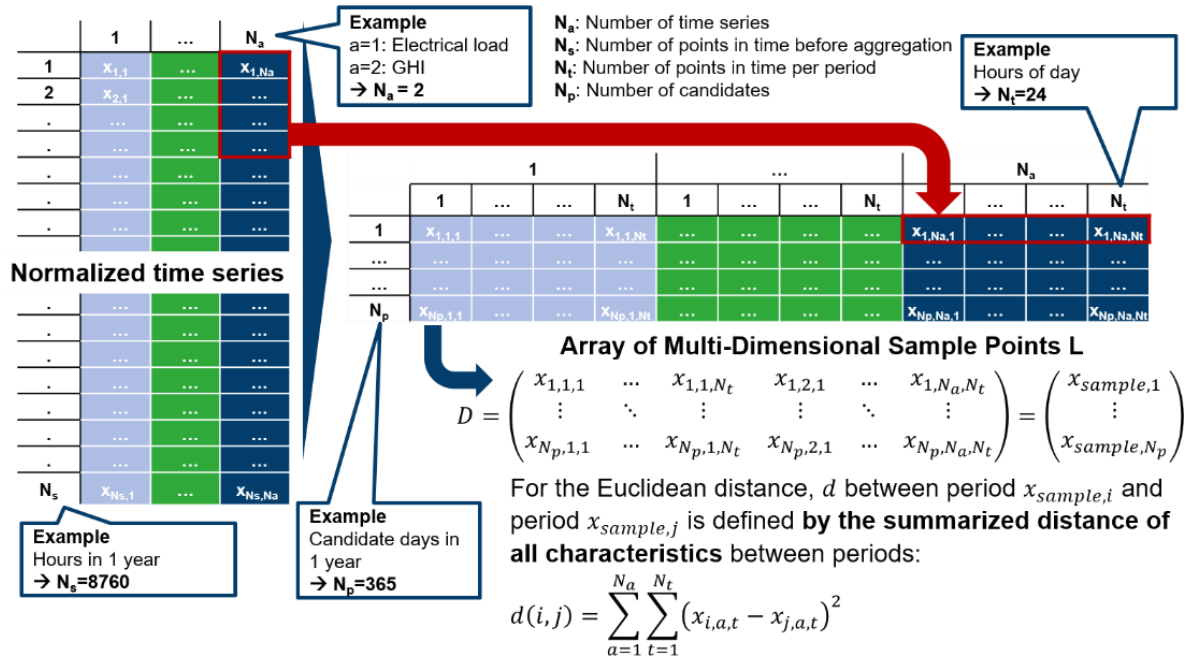


Figure 4. The procedure of period clustering using an arbitrary clustering algorithm. The time series are normalized (left), rearranged to rows containing all data of all time series during a certain time period (middle right) and each row (period) is subsequently clustered (lower right). Afterwards, each cluster is represented by a single period and weighted according to the cluster size.

² tsam package available at: <https://github.com/FZJ-IEK3-VSA/tsam> (date accessed: 11/23/2021)

The method of segmentation applied in this work is also based on clustering and the approach presented by Pineda et al. [63]. The procedure is depicted in Figure 5 and also implemented in the open-source tool, tsam. In contrast to the clustering of typical periods, the normalized time series are rearranged in such a way that each time step represents a single sample point, which is, in the case of multiple time series, multi-dimensional as well. Furthermore, the time steps can only be clustered with adjacent time steps in order to form a larger one, i.e., the clustering procedure must respect the side constraint of adjacency. This constraint can be easily respected using hierarchical clustering because of its sequential algorithm, which only merges those (adjacent) clusters that are closest to each other if a so-called connectivity matrix is defined.

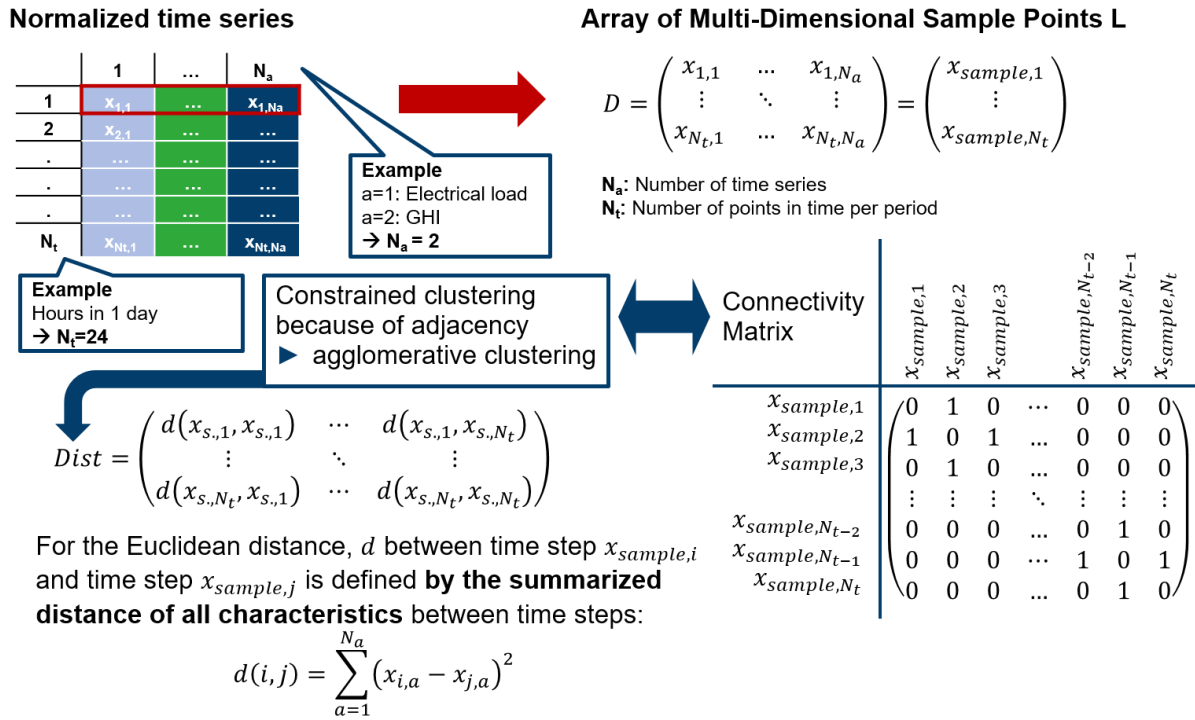


Figure 5. The procedure of segmentation using constrained hierarchical clustering. First, the time series are normalized (upper left) and each time step is interpreted as an own sample point (upper right). Each sample point is only allowed to be clustered together with either its preceding or succeeding sample point, which is defined by the connectivity matrix (lower right). Using hierarchical clustering, those points are clustered which are closest to each other and adjacent using the Euclidean Distance Matrix (lower left). Afterwards, the merged time step clusters are represented by a single time step with a weight according to the new duration of the merged time step.

3.2. The optimal tradeoff between segments and typical days

As was shown in the beginning and could be observed in Figure 1, the number of typical periods and segments can be freely combined. Moreover, the number of constraints and variables of an energy system model scales monotonously with the total number of considered time steps, which is defined by the number of periods times the number of segments per period. Therefore, multiple two-tuples of typical period and segment numbers exist with a comparable number of total time steps, and therefore a comparable, expected optimization runtime. Yet, it is obvious that not every one of these configurations is likely to lead to a good approximation of the original time series or a good optimization result. For example, considering 64 typical days with one segment per day instead of eight typical days with eight segments per day will significantly underestimate the intra-day variability, which is crucial for an appropriate representation of technologies, such as photovoltaic panels or daily storage systems.

In order to estimate the quality of aggregation, different accuracy indicators are used in the literature to quantify the difference between the original time series and the modified one. Among these, the

Root Mean Squared Error (RMSE) is the most popular, which was either used to quantify the difference between the original and aggregated time series [50, 62], an example of which is depicted in the left half of Figure 6, or between the respective duration curves [92-95], as depicted in the right half of Figure 6.

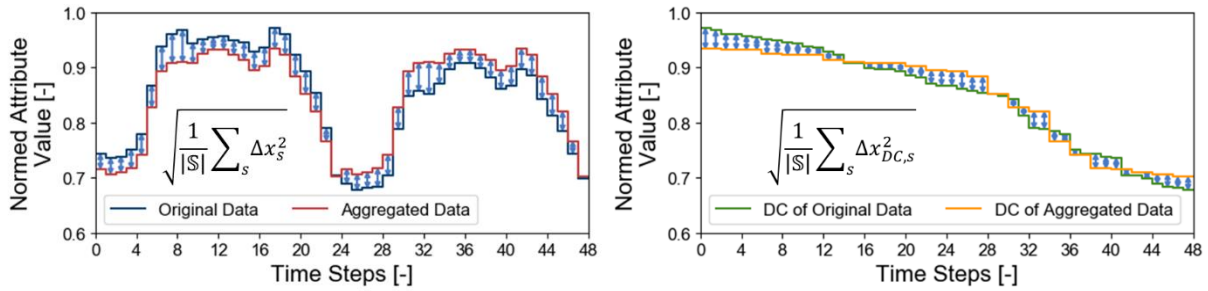


Figure 6. Illustration of squared summands used for calculating the RMSE (left) and between the respective duration curves (right) as defined in [49].

In this study, the RMSE between the normed time series and the corresponding aggregated one was used because it takes into account the difference in each time step, rather than the differences in the value distribution of the time series only, which is agnostic to the value chronology in the time series. For multiple time series that are simultaneously aggregated, the RMSE for all of the time series is defined as follows:

$$\text{RMSE}_{\text{tot}} = \sqrt{\frac{1}{|A| \times |T|} \sum_{|A|} \sum_{|T|} (x_{a,t}^{\text{aggregated}} - x_{a,t}^{\text{original}})^2}$$

Here, $|A|$ is the number of attributes and $|T|$ the number of time steps.

Indicators such as the RMSE are based on the assumption that a small deviation between the original and the aggregated time series leads to a small deviation between the energy system model's optimization result based on the original time series and that based on the aggregated time series. Moreover, the arbitrary combination of typical periods and segments offer the option of searching for a good tradeoff between them in order to reduce the RMSE as much as possible for a given number of time steps. Therefore, we propose an algorithm based on the method of steepest descent, as suggested by Haskell et al. [96]. In this context, the steepest descent is defined as the ratio of an RMSE's decrease over an increase in the total number of time steps. As the latter can be either increased by increasing the number of typical periods or the number of typical segments, the algorithm checks both options and chooses the more efficient one, an example of which is illustrated for three consecutive steps in Figure 7. The decision rule can be described by the following equation:

$$\hat{t}_{n+1} = \hat{t}_n + \delta \cdot \min \left(\frac{\text{RMSE}_{\text{tot}}(\hat{p}_{n+1}, \hat{s}_n) - \text{RMSE}_{\text{tot}}(\hat{p}_n, \hat{s}_n)}{\hat{s}_n \cdot (\hat{p}_{n+1} - \hat{p}_n)}, \frac{\text{RMSE}_{\text{tot}}(\hat{p}_n, \hat{s}_{n+1}) - \text{RMSE}_{\text{tot}}(\hat{p}_n, \hat{s}_n)}{\hat{p}_n \cdot (\hat{s}_{n+1} - \hat{s}_n)} \right)$$

where $\hat{t}_n$ describes the number of total time steps of the aggregated time series at iteration n , δ the finite resolution increase and $\hat{p}_n$ and $\hat{s}_n$ the number of typical periods and number of segments at that iteration step, respectively.

The complete algorithm is described in Figure 8. Starting from a single typical period with one segment, i.e., an aggregated time series consisting of a single time step, the algorithm checks in each step what direction of increase in temporal resolution is the most beneficial with respect to a decrease in the RMSE. Once the number of typical periods equals the resolution of the original time series, only increases in segments are considered, or vice versa. Moreover, it is possible to define a maximum number of total time steps for the aggregated time series, as it defines the remaining size of the

aggregated energy system model. Accordingly, the algorithm may stop once this total number of time steps has been surpassed. In our calculations, the step width between the number of typical days and segments followed a ratio of approximately $\sqrt{2}$ in each direction, as this proved to be suitable for the degressive RMSE decrease for higher temporal resolutions.

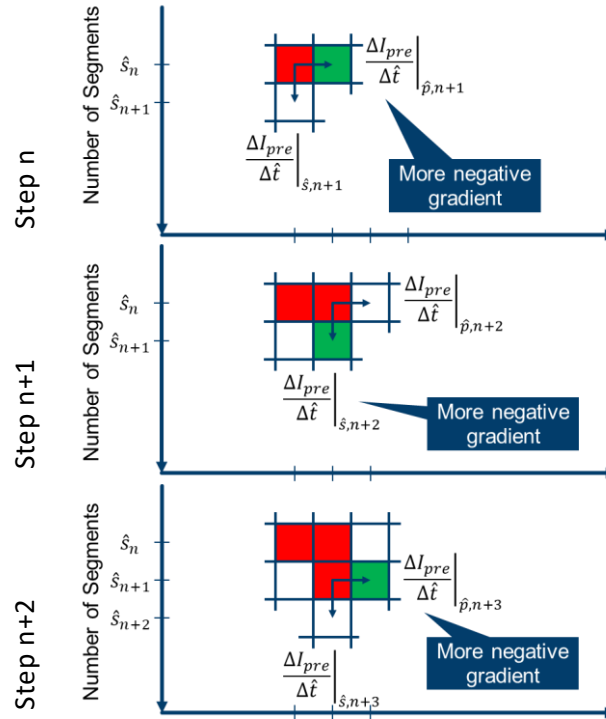


Figure 7. Three subsequent steps of the proposed algorithm to find the optimal number of segment and typical days.

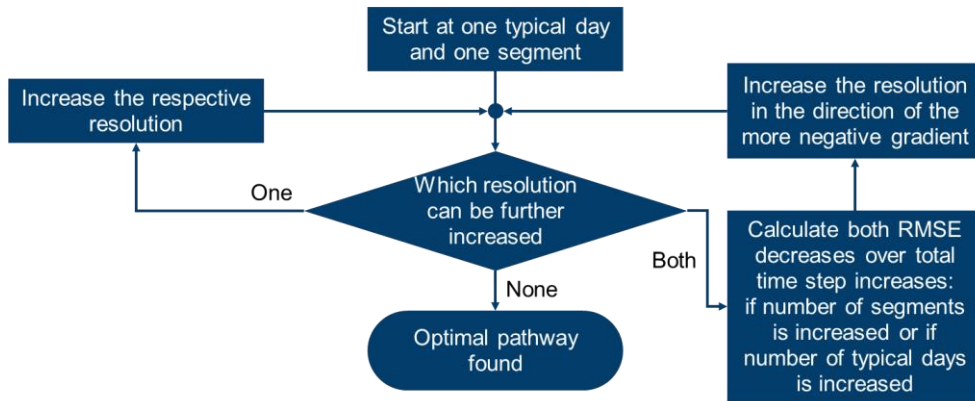


Figure 8. The optimal pathway algorithm from minimum to maximum resolution.

As is displayed in Figure 9, the algorithm is capable of differentiating between different types of time series. In the case of a time series for global a horizontal time series, the algorithm prioritizes an increase in the number of segments up to a level of twelve segments per typical days before choosing to increase the number of typical days. In contrast, the wind time series can be aggregated more efficiently by first increasing the number of typical days before the inner-daily resolution is increased. This emphasizes the fact that the algorithm takes the specific features of the time series into account. While solar irradiance profiles exhibit a clear daily pattern, are close to zero during the nighttime hours, and have steep gradients during the morning and evening hours, wind time series feature an aperiodic pattern with fairly smooth hourly transitions. Therefore, solar irradiance profiles require a high intra-daily temporal resolution to appropriately cover about 12–16 hours of the day. During the nighttime

hours, a significantly smaller temporal resolution is sufficient. In contrast, wind speed time series require many typical days due to their aperiodic behavior, leading to possible wind peaks at any point in the daytime, whereas a relatively small temporal resolution within each day is sufficient due to the relatively slow transition in the wind speed. In an energy system optimization, multiple time series are considered together. Therefore, the observed examples constitute extreme cases. In general, the algorithm adapts to the specific types of considered time series due to the definition of the $RMSE_{tot}$, which enables the simultaneous evaluation of all time series.

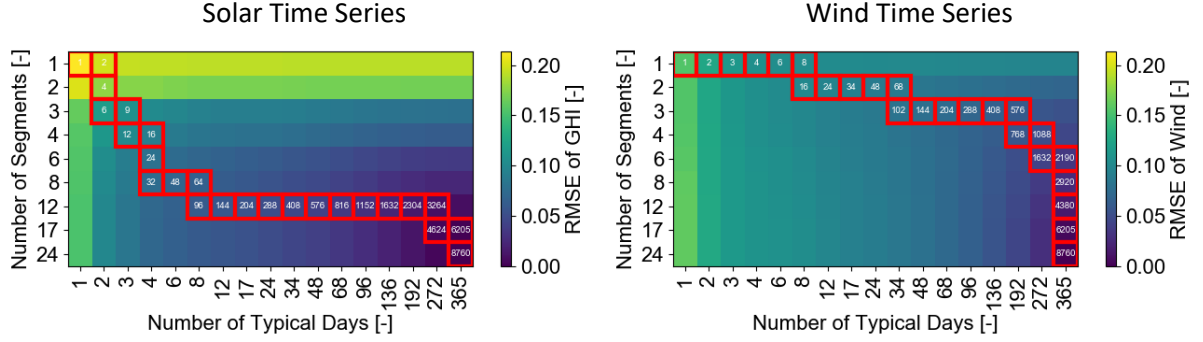


Figure 9. Proposed algorithm for the RMSE of a typical global horizontal irradiance and a wind speed profile.

3.3. Preserving the original time series' value distribution

The second algorithm proposed in this work in order to improve the quality of temporal aggregation for energy system models addresses the underestimation of extreme values as observed for k-means and k-medoids. Although numerous studies have manually included extreme values from the original time series into the aggregated ones, this procedure requires profound knowledge regarding the potentially critical system situations and suffers from the combinatorial problem that the number of extreme values grows in conjunction with the number of time series [8]. Therefore, the algorithm proposed below strives for the precise reproduction of the original time series' value distributions (value curves).

The algorithm consists of the following steps:

1. For a specific attribute a , sort all values of those periods that are assigned to a specific cluster k and yield the cluster's duration curve $\tilde{X}_{a,k}$. This is illustrated by the green duration curve shown in Figure 10.
2. Average every $|C_k|$ values of the cluster's duration curve $\tilde{X}_{a,k}$ in order to obtain the duration curve of the cluster's representative $\tilde{Y}_{a,k}$, which is represented by the brown line in Figure 10.
3. At the same time, calculate the mean (i.e., centroid) profile $\mathbb{M}_{a,k}$ for the specific attribute a using all periods assigned to cluster k for each time step represented by the dark blue line in Figure 10.
4. Determine the mean profile's duration curve $\tilde{\mathbb{M}}_{a,k}$ by sorting its values and extract the order of time steps t in which they appear in the sorted curve $\tilde{\mathbb{M}}_{a,k}$. The sorted mean profile is represented by the yellow line in Figure 10.
5. Assign the index order from $\tilde{\mathbb{M}}_{a,k}$ to the sorted values of $\tilde{Y}_{a,k}$ and sort them such that the attached indices of $\tilde{Y}_{a,k}$ are in ascending order. The result is the representative profile $\mathbb{Y}_{a,k}$ and is depicted as a red line in Figure 10.

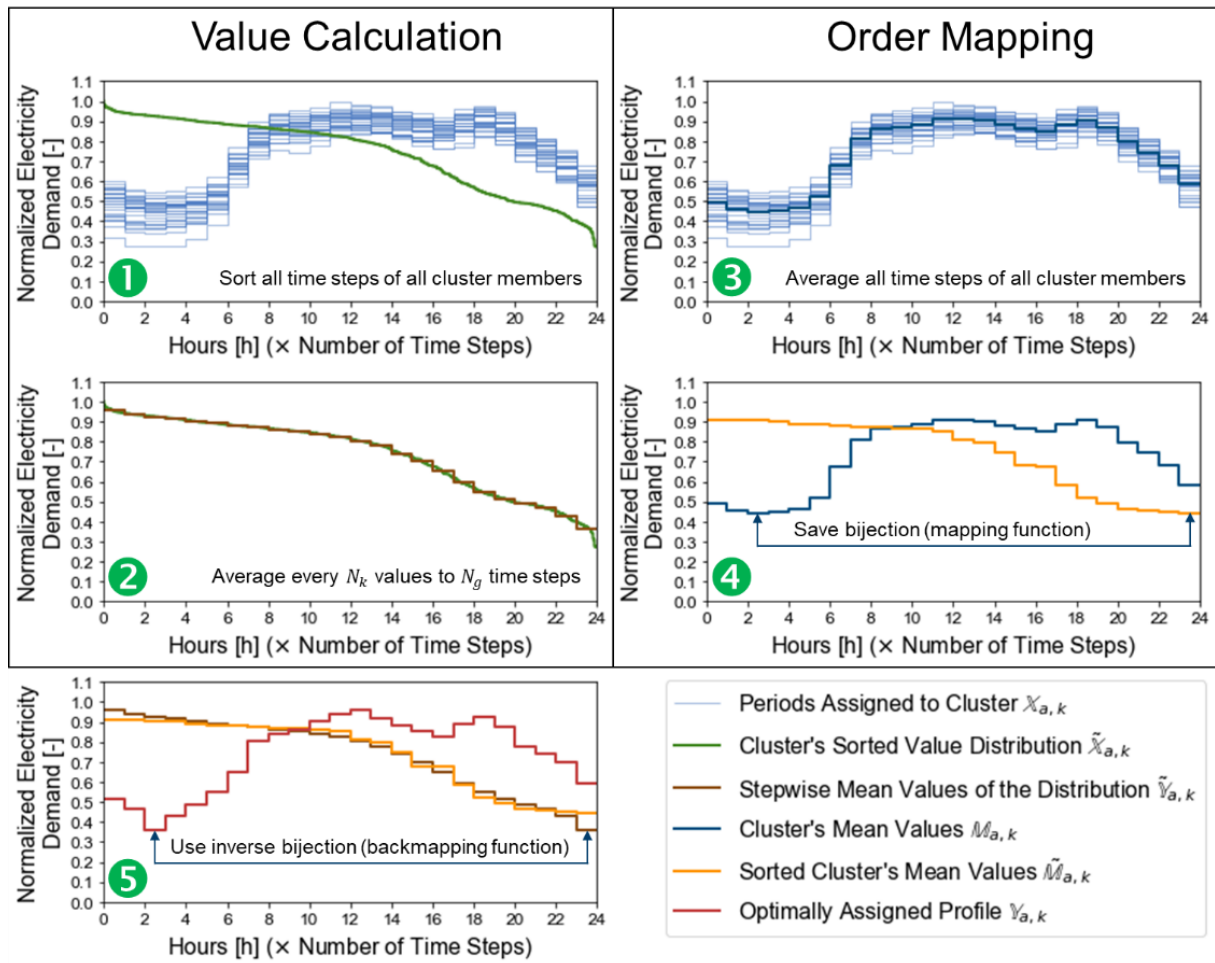


Figure 10. Graphical interpretation of the algorithm.

In contrast to a cluster representation by centroids or medoids, this approach is based on a direct approximation of the original time series' value distribution. The impact of this approach on a typical electricity time series with hourly resolution can be observed in Figure 11. On the left-hand side, the duration curve of the original and aggregated time series, consisting of eight typical days at hourly resolutions, are depicted. It can be observed that the original time series' duration curve is almost perfectly replicated. The right-hand side of Figure 11 depicts the original and aggregated electricity profiles during nine days in February. As the proposed method preserves the original time series' value distributions and therefore approximates the extreme values as well, the aggregated profile covers the cluster's extreme periods within its respective representative, which leads to a more peaky profile.

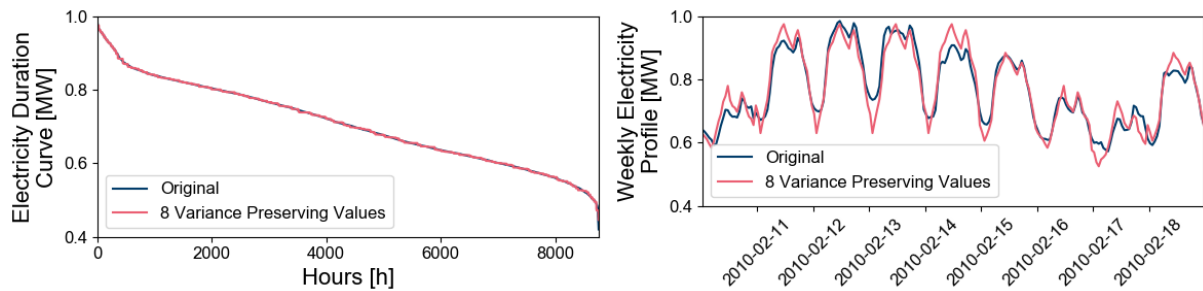


Figure 11. Yearly duration curve of a typical electricity profile, its profile during a week in February, and the corresponding aggregated time series as predicted by eight typical days using hierarchical clustering and the approximate distribution-preserving representation method

3.4. Case studies

In order to validate the effectiveness of our proposed methods, we introduce two different capacity expansion models that differ significantly with respect to their number of regions, their mathematical model structure, and their considered technologies. Both models have been implemented in the FINE³ optimization framework [71].

3.4.1. The self-sufficient building model

The first case study used for validation was introduced by Kotzur et al. [97] and is a model for a self-sufficient building that relies exclusively on photovoltaic electricity feed-in, which is also used to cover household heat demand using a heat pump and an electric boiler. In order to account for daily and seasonal variations in the electricity supply, the system comprises a battery, a heat and two hydrogen storages, as depicted in Figure 12. The hydrogen subsystem is linked to the heat and electricity system by a reversible solid-oxide fuel cell (rSOC). Furthermore, the system has been used as an application case in earlier studies [49, 98] and its techno-economic data is provided in Appendix A.1, and was made publicly available in [99].

The system is a single-node model considering two time series for heat and electricity demand, as well as two rooftop photovoltaic and one open field photovoltaic time series generated using the open-source software tsib [100, 101]. For this, the weather year of 2013 in the location of Berlin was taken from the COSMO dataset [102-104]. Additionally, the energy system model considers discrete investment decisions for certain components using binary variables, turning the model into an MILP.

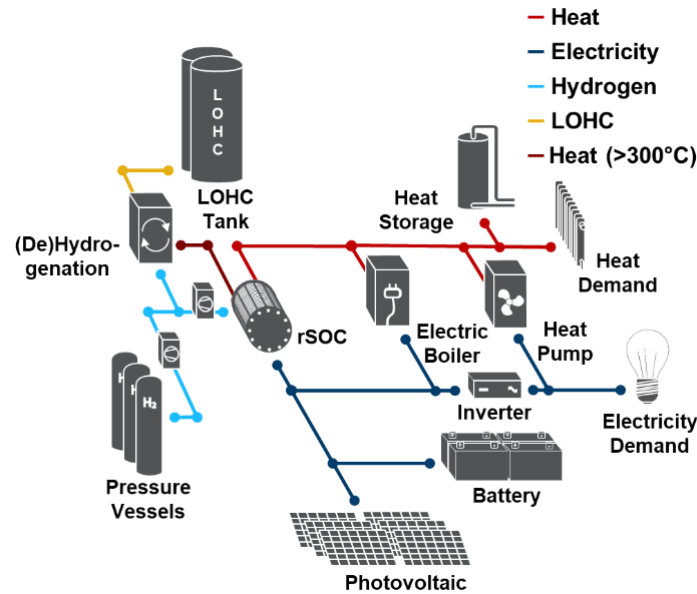


Figure 12. Simplified scheme of the self-sufficient building model.

3.4.2. The European model

In contrast to the self-sufficient building model, the European model developed by Çağlayan et al. [105, 106] is a carbon-neutral capacity expansion model with the target year 2050, and considers 96 European regions. Moreover, it is not a greenfield study and accounts for preexisting infrastructure.

As is shown in Figure 13, the system primarily focuses on a hydrogen and electricity network based on renewable energy sources, such as PV, wind energy, biomass, and hydroelectricity. The hydrogen and

³ FINE package available at: <https://github.com/FZJ-IEK3-VSA/FINE> (date accessed: 11/23/2021)

electricity networks are connected by electrolyzers, gas engines, and different types of fuel cells and gas turbines. Furthermore, hydrogen and battery storage is considered in order to account for fluctuating supply and demand. The hydrogen demand is assumed to originate exclusively from fuel cell-electric vehicles with a market penetration of 75% [107], whereas the electricity demand was obtained from the E-Highway study [108] for a 100% renewable scenario. This demand is considered to account for electrified heating in the residential and industrial sectors, as well as the operation of battery-electric and plug-in hybrid vehicles.

For the fully resolved case, the European model is on the edge of computational feasibility, with a calculation time of approximately four days, and therefore does not consider binary decision variables, i.e., it is a large LP. Furthermore, the multitude of regions and technologies leads to more than 900 time series, which are then simultaneously clustered. These constitute fundamental differences from the self-sufficient building model with respect to both the application case and mathematical structure of the model.

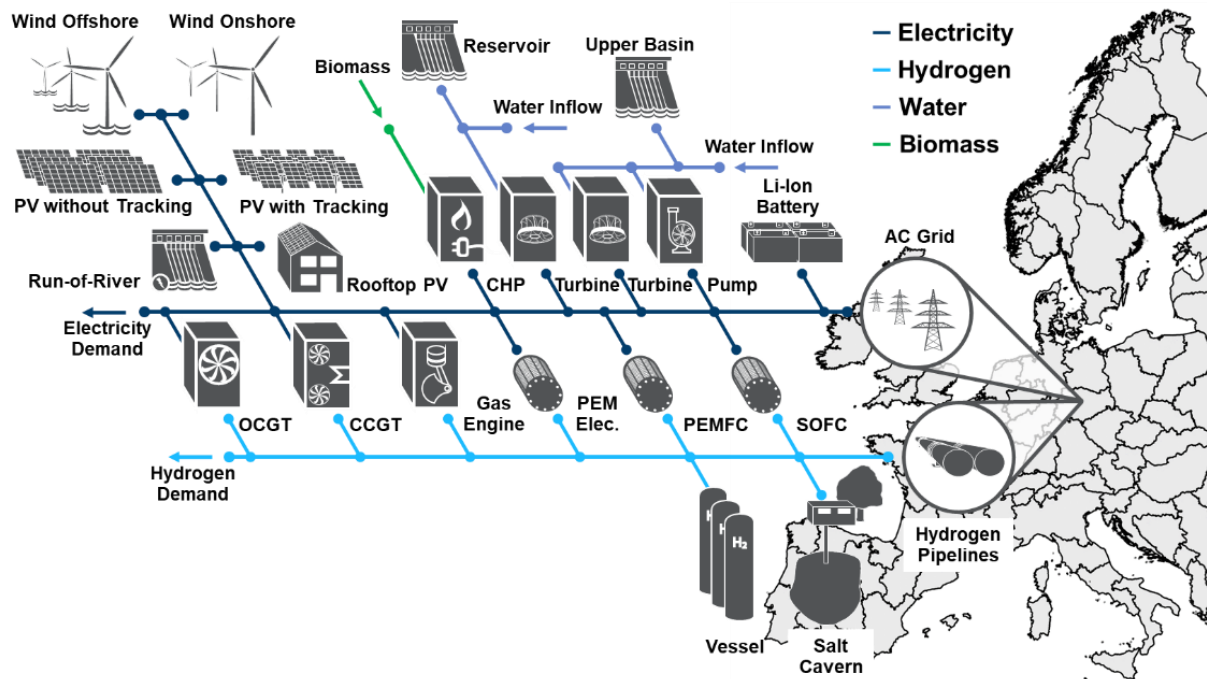


Figure 13. Layout of the European model.

3.5. Computational resources and comparison framework

As previously noted, the computational runtime not only depends on the size of the model but also on the hardware used. For that reason, the computational resources used to conduct the analyses in Section 4 are also listed in Table 3. All calculations were performed on a computer cluster in parallel, but with an individual core assigned to each calculation. Gurobi 9.0.0 was used to carry out the optimization, with the thread number set to one in order to prevent unpredictable variations in the calculation times arising due to data traffic between the threads.

Table 3. Computational resources used for Section 4.

CPU Model	Intel(R) Xeon(R) Gold 6144 CPU
Number of Cores per Computing Node	16
Threads per Core	2
CPU Max Frequency [MHz]	3.5
Shared Memory [GB]	1024

4. Results

In this section, the models introduced in Section 3.4 are optimized for the large set of different aggregation configurations depicted in Figure 14. Each field describes an aggregation setup, with the white number representing the total number of remaining time steps for that configuration, .e.g., in the case of eight typical days and eight segments, the time series were aggregated from 8760 to 64 time steps. The ratio between two neighboring typical period or segment numbers was selected to be approximately $\sqrt{2}$, as this accounts for a degressive trend of accuracy over an increase in the total number of time steps. Moreover, only typical periods with a length of one day were considered due to the daily charging pattern of short-term storage in the respective models. A more detailed discussion on the optimal period length for temporal aggregation is provided in an earlier work by the authors [49].

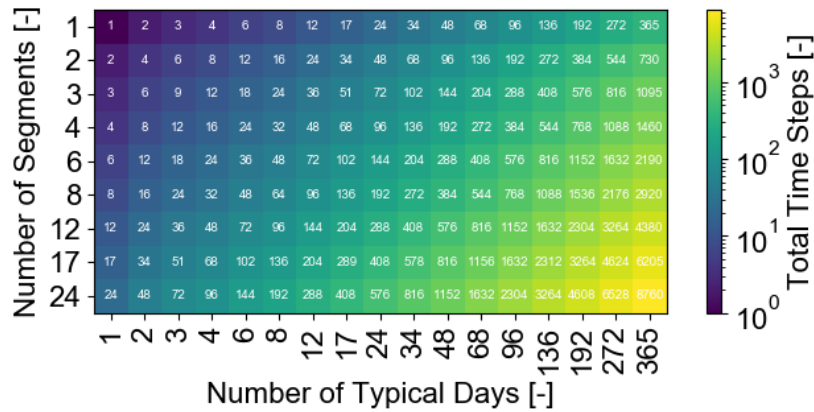


Figure 14. Number of total time steps depending on typical day and segment configurations that were considered in the sensitivity analysis for four different representation methods

Furthermore, each typical day and segment configuration was represented by centroids, medoids, and the proposed distribution-preserving algorithm, leading to $9 \times 17 \times 3 = 459$ different optimization runs per model. Throughout the entire chapter, Ward's hierarchical clustering algorithm was used due to its good runtime scaling behavior with respect to both the sample number and sample dimensionality. For the considered multi-regional model, it was the only computationally-tractable clustering algorithm and, for the single-regional model, it was found that the results were less sensitive to the clustering technique than the representation method. In addition, the reproducibility of the results was given due to its deterministic algorithm.

After all optimizations were run, the most commonly used aggregation configurations from the literature, i.e., typical days at hourly resolution, represented by either centroids or medoids, were benchmarked against the optimization results obtained from the aggregation configurations, along with the optimal ratio of typical days and segments using the distribution-preserving algorithm as the representation method. In particular, these analyses will illustrate the effectiveness of the novel algorithms proposed in Section 3.2 and 3.3 with respect to their capability to reduce aggregation-induced errors in both models for an arbitrary number of aggregated time steps.

4.1. Results for the self-sufficient building

First, the large-scale parameter variation was applied to the self-sufficient building model. The right-hand side of Figure 15 depicts the individual model runtimes and relative cost deviations of the respective aggregated models from the fully resolved reference case for all aggregation configurations. Here, the chosen representation methods are highlighted in different colors.

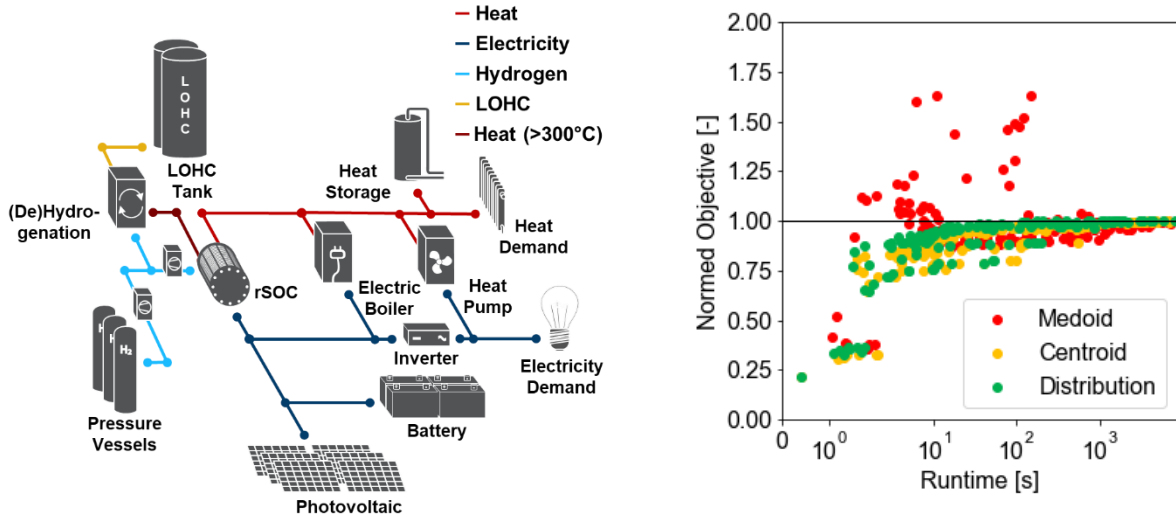


Figure 15. The self-sufficient building model and corresponding normed optimal objectives over the computation times for all considered configurations.

As Figure 15 indicates, there is a clear tradeoff between the runtime of the models and their respective accuracies, which is represented by the aggregation-induced deviation from the optimal objective function values of the reference case. The more time steps that are considered, the more the system costs approach those of the reference case. Moreover, it can be seen that a representation by centroids and the distribution-preserving algorithm lead to a consistent underestimation of the total system costs, but converge somewhat monotonously on the system costs of the reference case for higher resolved temporal aggregation configurations with longer runtimes. In contrast, this trend cannot be observed for a representation by medoids, which may lead to either an over- or underestimation of the total system costs.

The left-hand column of Figure 16 depicts the pathways of the aggregation configurations as proposed by the algorithm introduced in Section 3.2. It is worth noting that the clusters as defined by Ward's hierarchical algorithm for segments and typical days are identical for all three representation methods. However, the different representation methods lead to slightly differing aggregated time series, and therefore different RMSE values and proposed ratios between the typical days and segments. For all representation methods, it can be clearly seen that the algorithm initially favors a relatively high number of segments for relatively strongly aggregated time series. For twelve typical days, the proposed number of segments is between six and eight. However, for larger numbers of total time steps, the number of necessary segments per typical day is at most twelve, except for the configurations with 365 typical days, which is equal to the original number of days. Both observations can be explained by the heavy reliance of the self-sufficient building model on PV, for which this effect was already explained in Section 3.2. Moreover, the remaining time series pertain to heat and electricity, which also have a strong daily pattern and fairly smooth profile intervals between midnight and 6:00 AM. Accordingly, the segments can be comparably large during the night, which leads to a sufficient model aggregation with no more than twelve time steps per day.

The graphs shown in the right half of Figure 16 map the aggregation-induced deviations of the optimal objective onto the respective aggregation configurations. Here, the convergent behavior for a representation by centroids or the distribution-preserving algorithm can again be observed, indicating that a higher number of time steps indeed leads to smaller deviations. Likewise, it can also be seen that an increase in either the number of typical days or segments per typical day always drives the need to increase the complement resolution, e.g., in the case of four typical days represented by centroids, it does not make a big difference whether four segments per day are chosen or 24. If the total "runtime budget" would not allow for more than 96 total time steps, a better choice would

therefore be to choose twelve typical days with eight segments each instead of four typical days with 24 each. For a representation by medoids, however, an unpredictable behavioral pattern can be observed that impedes an intuitive choice for a sufficient number of typical days or segments. Assuming that a system is not solvable in its fully resolved configuration, a sensitivity analysis in the neighborhood of a specific temporal aggregation configuration would therefore not lead to easily interpretable results, and it is therefore not reliable.

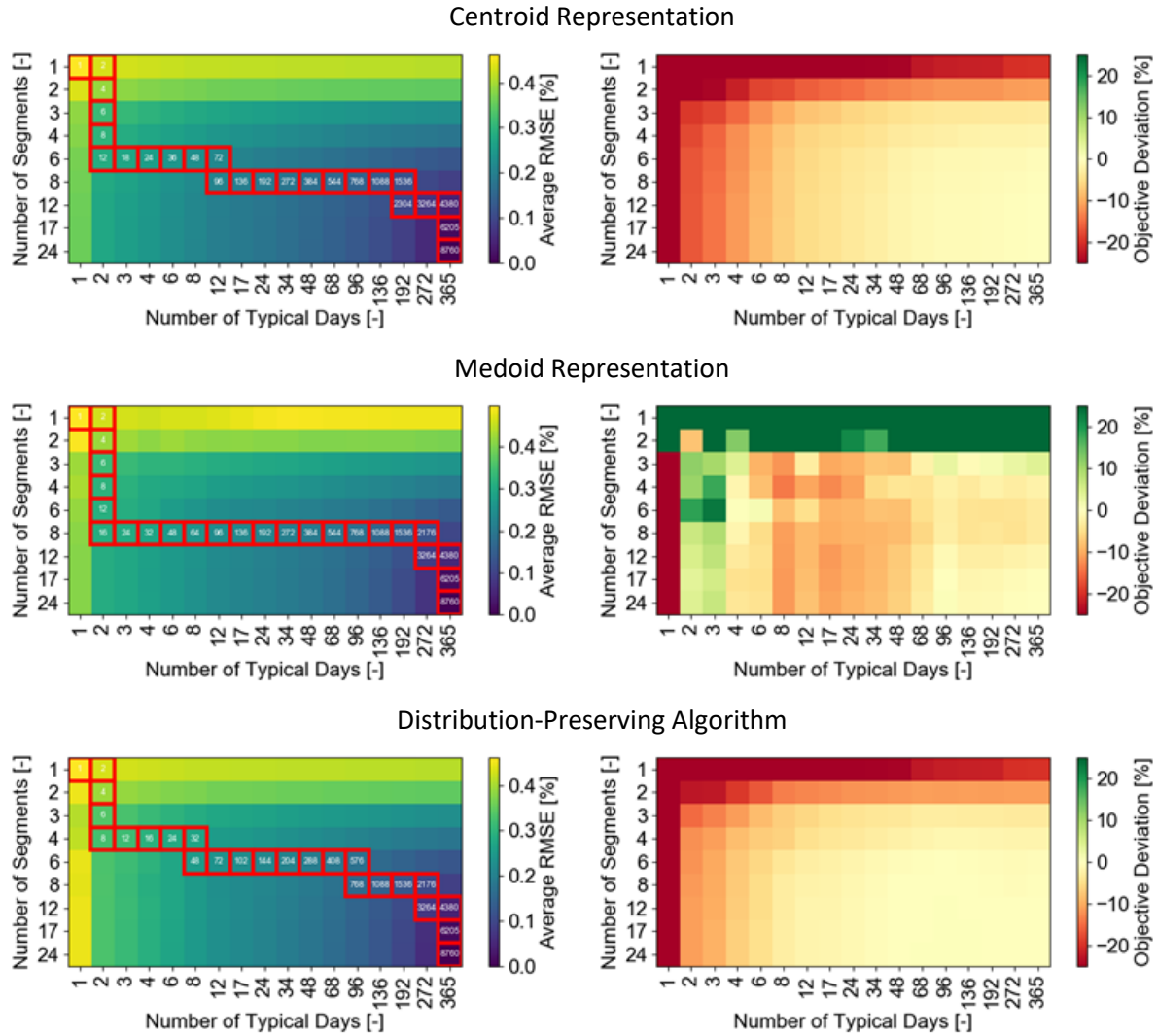


Figure 16. The pathway found by the proposed algorithm for an optimal ratio between the number of typical days and segments depending on the representation method (left) and deviation from the optimal objective of the fully resolved case depending on the configuration of the number of typical days and segments, as well as the representation method (right).

In order to evaluate the effectiveness of choosing a balanced number of typical days and segments for each typical day, Figure 17 separately depicts the results shown in the right half of Figure 15 for each representation method. Moreover, the aggregation configurations that are most frequently used in the literature, i.e., typical days with an hourly resolution, are highlighted as violet dots, whereas the configurations as obtained by an improved ratio between the number of typical days and the number of segments are highlighted as green dots.

As medoids do not lead to consistent results that monotonously converge upon the reference case, it is self-explanatory that choosing a balanced number of typical days and segments within each typical day does not lead to predictable results either. However, those configurations requiring more than 10 seconds of optimization runtimes are indeed closer to the objective functional value of the reference case than those obtained by typical days with an hourly resolution. In contrast, the algorithm proposed

in Section 3.2 not only outperforms the configurations of typical days with an hourly resolution but also leads to Pareto-optimal results in the case of a representation by centroids or when using the distribution-preserving algorithm.

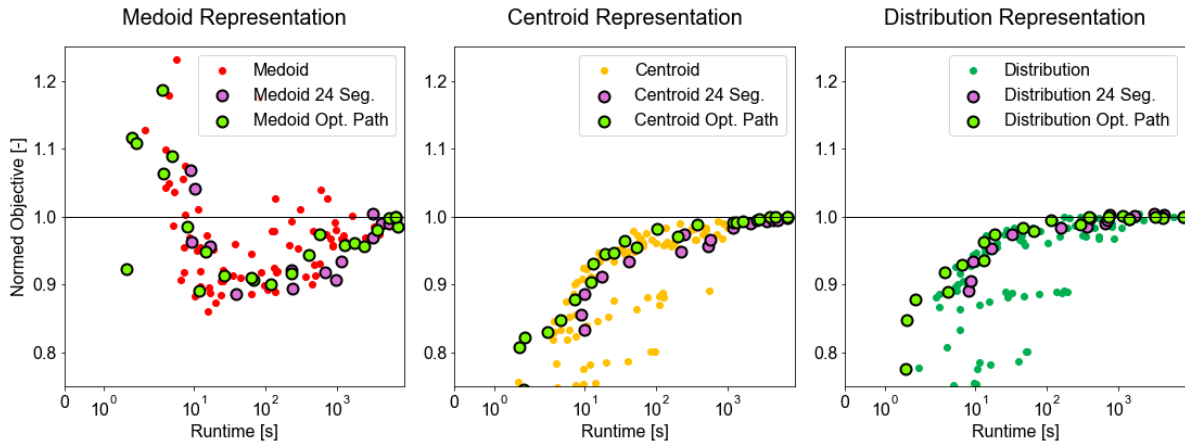


Figure 17. Convergence behavior of the different representation methods if an optimal ratio between the number of typical days and the number of segments is chosen based on the RMSE compared to the convergence behavior of different typical day numbers at hourly resolutions.

Lastly, the combination of both of the proposed improvements on current aggregation approaches is compared to all other model runs in the left graph of Figure 18. Here, it can be seen that those aggregation configurations that were obtained using the proposed number of typical days and segments, as well as the distribution-preserving algorithm, leads to the Pareto-optimal model aggregations. The left graph of Figure 18 compares the same aggregation configurations to the status quo of most aggregation approaches, i.e., a representation by centroids or medoids of typical days with hourly resolutions, which are clearly dominated. Compared to a representation by centroids, it can be seen that the proposed aggregation configurations lead to less than half of the deviations from the optimal objective of the reference case at comparable runtimes. Because of the decreasing gain in accuracy for longer runtimes, this means that the proposed algorithm can obtain results of comparable accuracy in about one tenth the runtime required for representation by centroids.

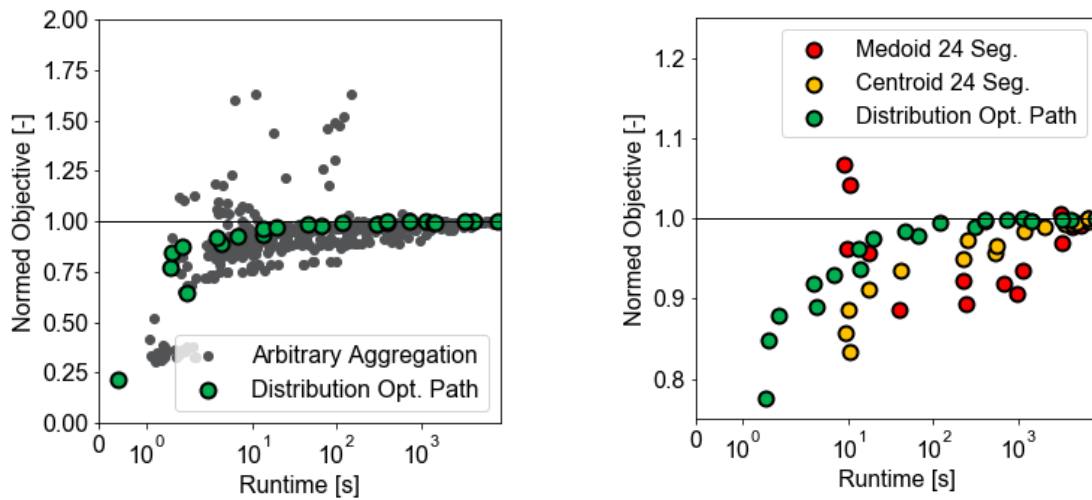


Figure 18. The identified optimal aggregation method amongst all tested aggregation approaches (left) and a direct comparison to state-of-the-art aggregation methods as used in *FINE* (right).

4.2. Results for the European model

The European system model differs considerably from the self-sufficient building one with respect to the number of time series, the number of regions, and the fact that numerous aperiodic wind time series are considered. Therefore, it can be expected that the results of the different aggregation methods significantly differ. Analogously to the previous section, the right half of Figure 19 depicts the normed objective of all model runs over their respective runtimes separated by the representation method. Again, the representation by centroids leads to a significant underestimation of the system costs, whereas the distribution-preserving algorithm quickly converges on the objective function value of the reference case for increasing numbers of time steps and higher runtimes. In contrast to the self-sufficient building, however, the representation by medoids leads to far less unpredictable behavior. Although medoids still lead to the highest approximations of total system costs, the range of medoid-based optimization runs resembles those of the representation by centroids or the distribution-preserving algorithm. One reason for this is the fact that the representation by medoids does not preserve the original time series' mean values, and can therefore distort the cumulative energy consumption of small island systems, such as the self-sufficient building. For large-scale problems such as the European model, however, the selection of real days from the original dataset accounts for real correlations between time series, and therefore has advantages for large and complex datasets.

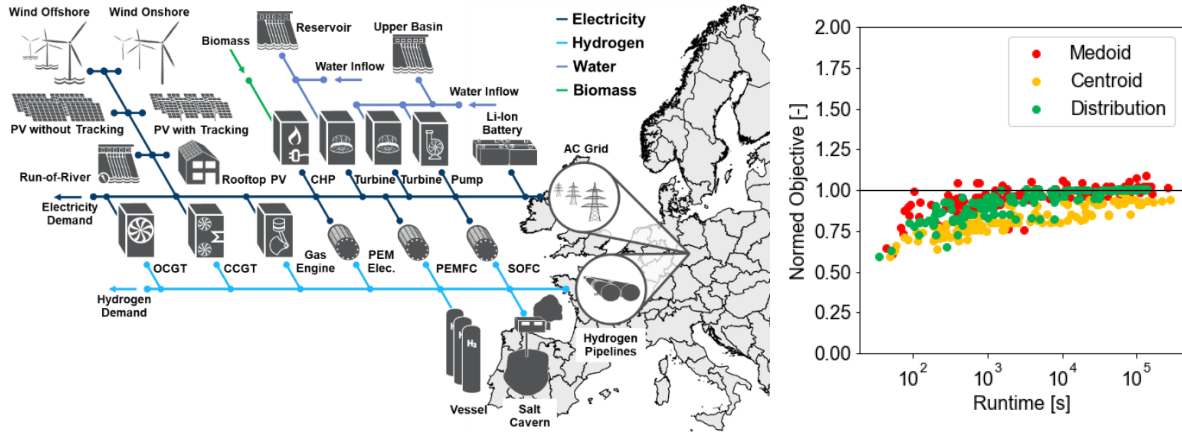


Figure 19. The European model and corresponding normed optimal objectives over computation times for all considered configurations.

The left column of Figure 20 depicts the configurations of typical days and segments as proposed by the algorithm presented in Section 3.2. The white fields are optimization runs that were terminated due to numerical instabilities, infeasibility, and non-convergence within the runtime limit set to five days. Here, it can be seen that the algorithm opts to increase the number of typical days first and after that, the number of typical days and segments in a more balanced way than for the self-sufficient building, due to the simultaneous consideration of time series with and without a daily pattern.

The right column of Figure 20 depicts the deviations of the objective function values depending on the aggregation configuration. Again, it can be seen that the total annual costs are relatively indifferent to the number of segments if at least three of them per typical day are chosen. Additionally, medoids do not exhibit a monotonous convergence behavior for higher resolved configurations, and the distribution-preserving algorithm provides a close approximation of the reference case if at least 34 typical days with three segments per day are chosen, which equals a time step reduction of almost 99% compared to the 8760 original time steps, and outperforms all other representation methods.

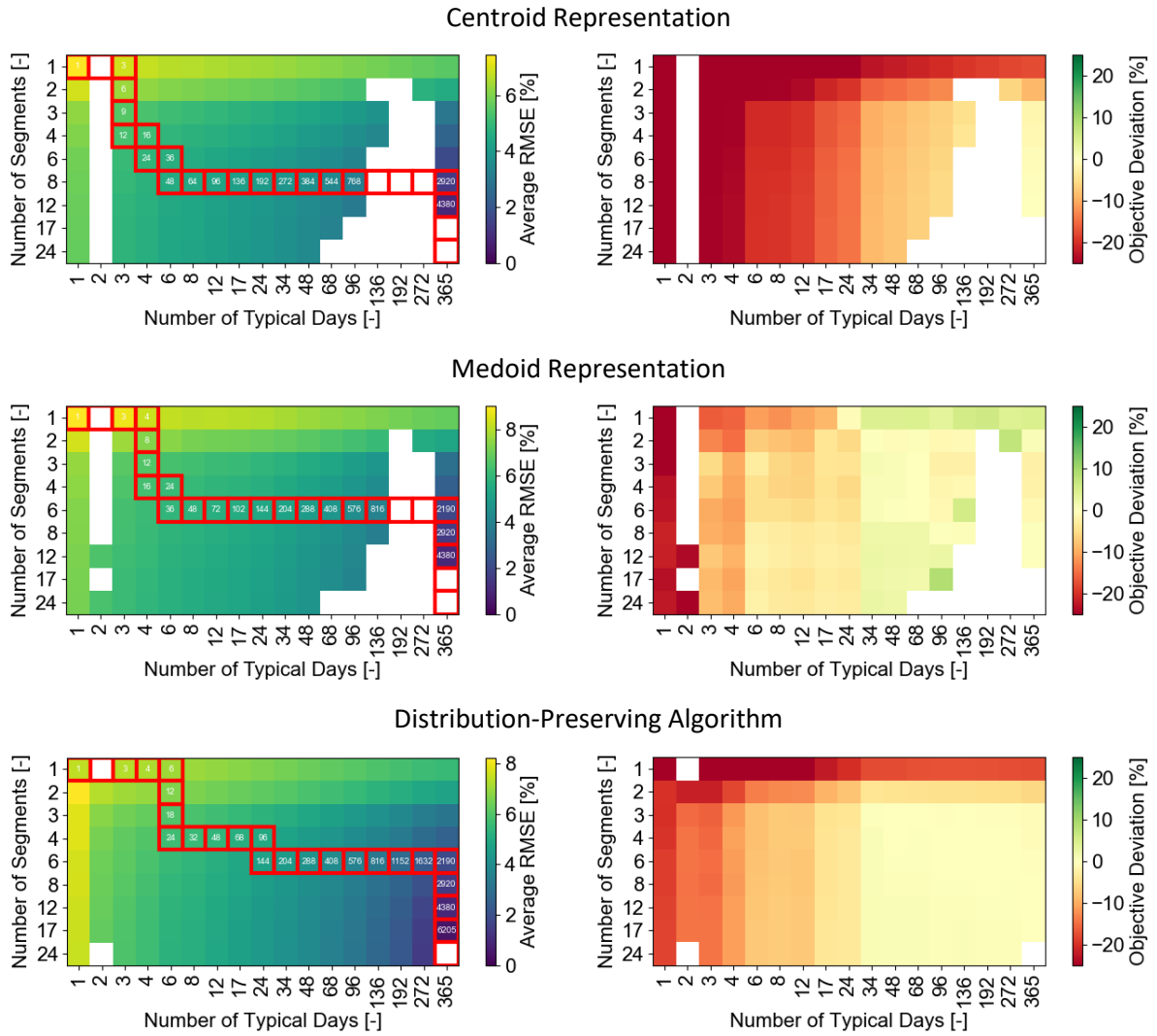


Figure 20. Pathway found by the proposed algorithm for an optimal ratio between the number of typical days and segments depending on the representation method (left) and deviation from the optimal objective of the fully-resolved case depending on the configuration of the number of typical days and segments, as well as the representation method (right).

Figure 21 compares the results of the aggregation configurations as proposed by the algorithm in Section 3.2 to those obtained by using typical days with an hourly resolution only, as well as all remaining aggregation configurations. Although the approach does not provide the Pareto-optimal configurations in the case of medoids and centroids, it clearly outperforms the configurations considering typical days at an hourly resolution only. It is worth noting that the green dot lying significantly over the normed objective function value in the case of medoids was a suboptimally-terminated optimization run, more precisely the run with 136 typical days and six segments per typical day, and can therefore be neglected. These findings emphasize that the algorithm proposed in Section 3.2 is agnostic to the financial impact of the individual time series types, as they are an outcome of the optimization itself and therefore do not necessarily choose the optimal configurations from an output data perspective, i.e., with respect to the deviation of the objective function value. However, this should not be the only quality criterion, as a small overall cost deviation does not necessarily indicate meaningful technology sizing, as will be demonstrated in Section 4.3.

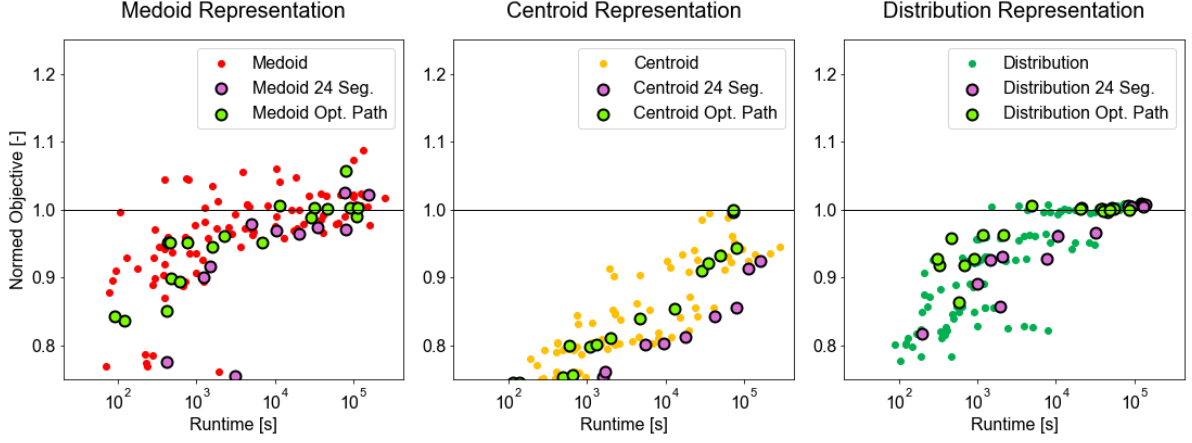


Figure 21. Convergence behavior of the different representation methods if an optimal ratio between the number of typical days and the number of segments is chosen based on the RMSE compared to the convergence behavior of different typical day numbers at an hourly resolution.

Finally, Figure 22 visualizes the impact of the proposed improvements compared to all tested aggregation configurations and the status quo of centroid- or medoid-based typical days at hourly resolutions. Here, it can be seen that the proposed methods lead again to the Pareto-optimal solutions for those configurations, with a runtime of more than 1000 seconds. For shorter runtimes and therefore stronger aggregated configurations, the result is not as clear because the distribution-preserving algorithm is based on synthesized time step values, and in the case of too strong an aggregation, the creation of a few typical days with the value distribution of the original time series may distort the correlations between them. In summary, however, the proposed improvements significantly outperform the status quo of the aggregation approaches.

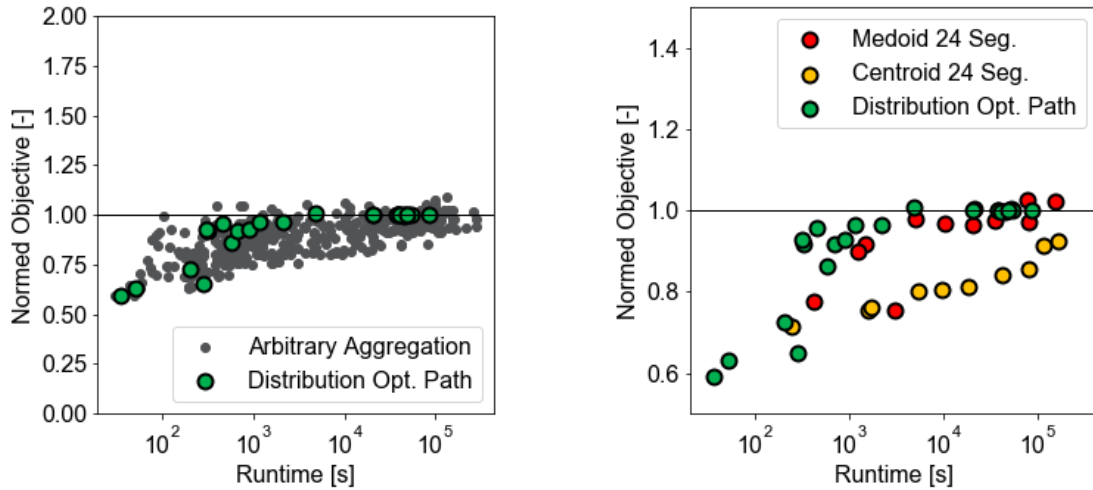


Figure 22. The identified optimal aggregation method among all tested aggregation approaches (left) and a direct comparison to state-of-the-art aggregation methods as used in FINE (right).

4.3. Additional observations and discussion

In addition to the optimal objective function value, which can always be used in order to estimate the quality of an aggregation, there are numerous other model-specific features that indicate the quality of an aggregation and may relate to both, system design and operation. For example, Poncelet et al. [83] used the deviation of energy supply by each technology as indicator as an indicator to quantify the quality of an aggregation, which they defines as generation mix error. In case of the presented capacity expansion models, especially the exact component capacities by technology may be of

interest. Analogously to Hoffmann et al.[49], one may define the RMSE of aggregation-induced cost deviations across all technologies as follows:

$$SSE_{tech} = \sqrt{\sum_{i \in \mathbb{M}_{tech}} \left(\frac{cost_i - cost_{ref,i}}{cost_{ref,total}} \right)^2}$$

In case of capacity expansion models, this metric mainly refers to annualized investment costs. However, as shown in prior publications, it can also be applied to unit commitment or dispatch models, in which the cost components of a system are mainly determined by operational costs as well as costs and revenues due to import, export of consumption of energy carriers.

Figure 23 depicts the sum of squared relative cost deviations by technology for both, the self-sufficient building and the European model depending on the chosen representation method for both, typical periods only (24 segments each) and along the optimal pathway as proposed in Section 3.2. Here, it can be seen that, for a given number of total time steps, those aggregation configurations proposed by the method presented in Section 3.2 and using the value-distribution preserving algorithm presented in Section 3.3 also lead to the smallest relative cost deviations by technology in both models. Further, it can be observed that each proposed method alone already improves the status quo of aggregation in the literature, i.e. medoid- or centroid-based typical days at an hourly resolution.

First, the configurations consisting typical days with 24 segments each, which are represented by markers with black edges in Figure 23, lead almost consistently to a higher SSE_{tech} than the configurations with a balanced reduction of inner-daily resolution and period number. Second, the distribution-preserving algorithm presented in Section 3.3 dominates a representation by medoids or centroids in almost any case. In summary, the combination of both methods developed in this work lead to a significantly improved aggregation with respect to both, objective function approximation and technology-wise investments.

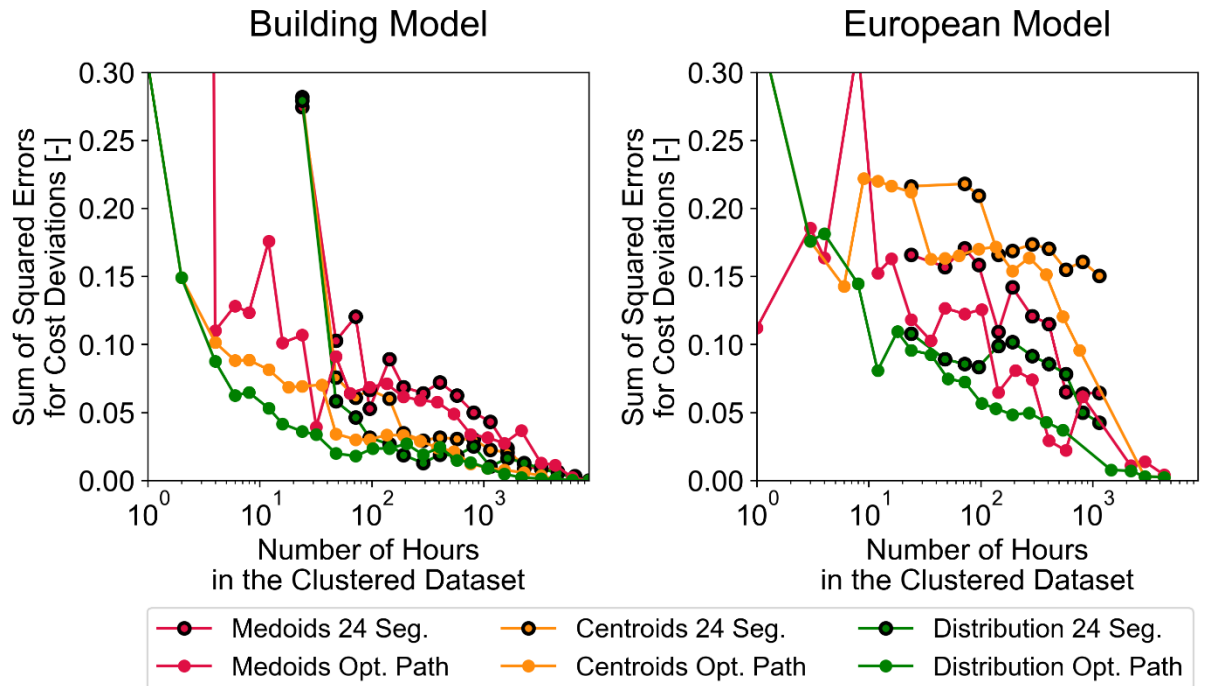
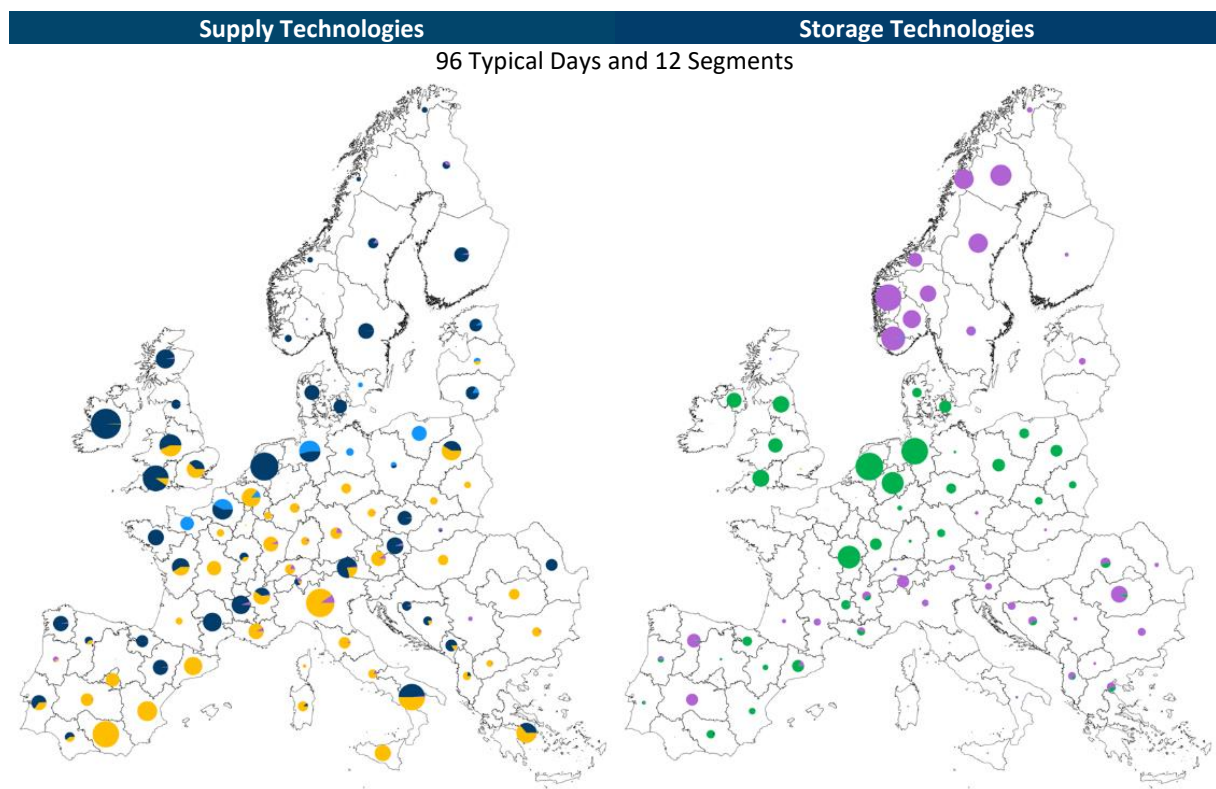


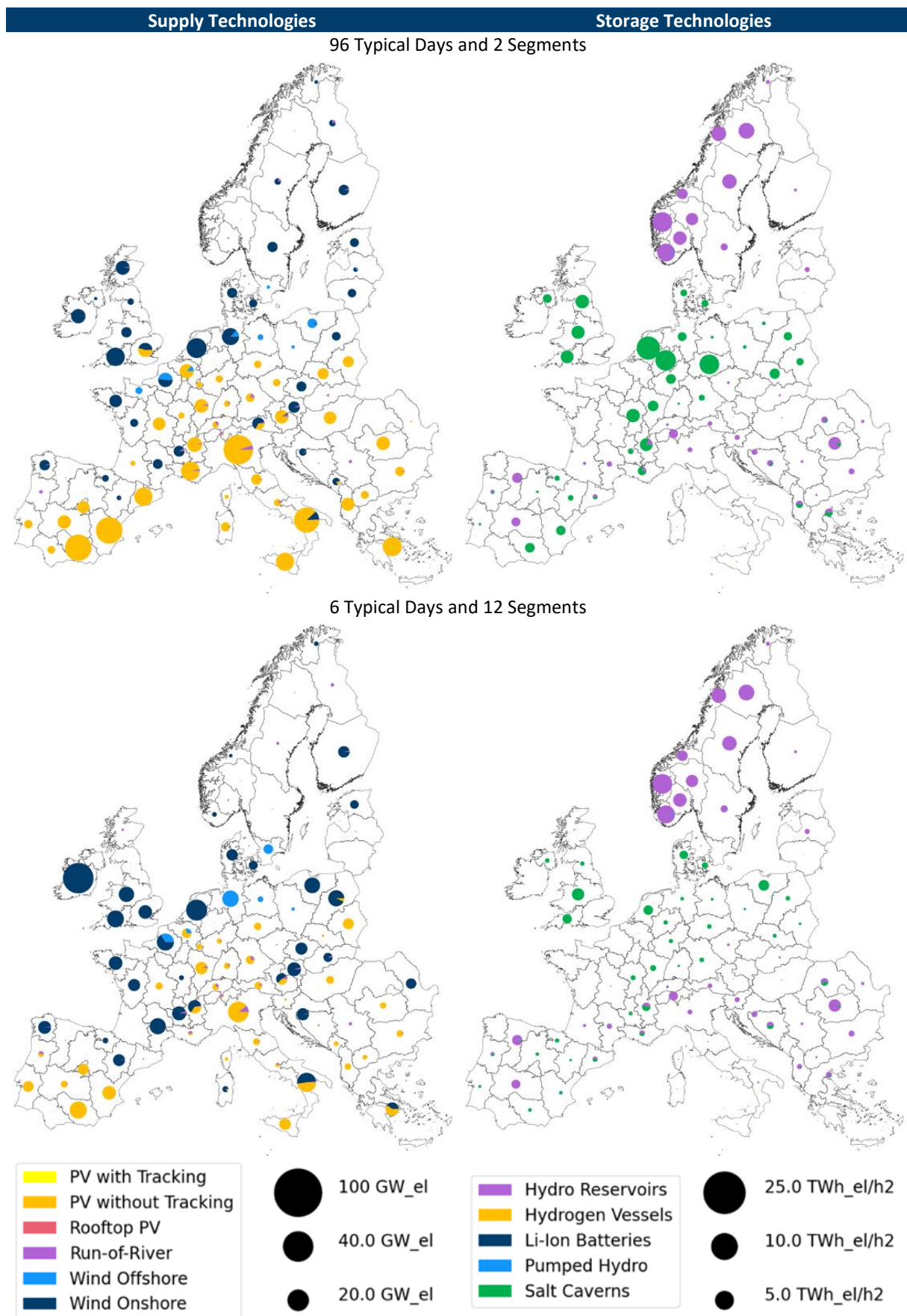
Figure 23: The sum of squared relative cost deviations by technology for both, the self-sufficient building and the European model depending on the chosen representation method for both, typical periods only (24 segments each) and along the optimal pathway as proposed in Section 3.2

Furthermore, it can also be observed in our case studies that the proposed method is directly applicable to both, linear programs (the second model) and mixed-integer linear programs (the first model) because the reduction of considered time steps directly decreases the number of variables and constraints in the optimization problem. As MILPs are solved by a sequential solving of linear programs (e.g. by a branch-and-bound approach), a method that accelerates an approximated solution of LPs also accelerates MILPs. Moreover, if binary (operational) variables are also aggregated to fewer ones, temporal aggregation may accelerate MILPs even stronger due to a reduced branch-and-bound tree.

Apart from the analyses that could be individually conducted for each model, some observations arise from the comparison of both models and the time series employed. First, the European system model differs considerably from the self-sufficient building one with respect to the number of time series, the number of regions, and the fact that numerous aperiodic wind time series are considered. Therefore, it can be used to study the effect of different temporal aggregation procedures on the technology choice. As a major part of the time series is given by the capacity factor time series for wind turbines and photovoltaic panels, these technologies are of particular interest.

Therefore, Figure 24 depicts the energy supply and storage capacities depending on a varying number of typical days and segments using a representation by centroids. Considering the configuration in the first row with 96 typical days and 12 segments per day a highly resolved model, a decrease to two segments per typical day, as shown in the second row, leads to a significant oversizing of the photovoltaic capacities. The reason for this is the disproportionate underestimation of solar intermittency in this case, which makes photovoltaic panels more economically attractive. In contrast, twelve typical days with twelve segments per day leads to a predominant underestimation of the wind time series' variance. Accordingly, the capacities for wind turbines are overestimated, as can be seen in the third row of Figure 24. Apart from that, a small number of typical days leads to an underestimation of seasonality effects. For that reason, the capacities for seasonal hydrogen storage are underestimated in the corresponding right column. Accordingly, temporal aggregation and resolution in general have a significant and heterogeneous impact on different technologies, which must be taken into account in the modeling process.





Hydro Reservoirs

 Hydrogen Vessels

 Li-Ion Batteries
 Pumped Hydro
 25.0 TWh_el/h2

Figure 24. Energy supply and storage capacities depending on a varying number of typical days and segments using a representation by centroids.

Apart from that, both models vary with respect to the overall number of considered time series. Accordingly, it can be expected that the number of “redundant” time steps, which can be aggregated without a major loss of accuracy, may vary. For that, Figure 25 depicts the convergence behavior of the two models for an increasing number of typical days, with hourly time steps represented by centroids. The x-axis represents the total number of time steps of the respective aggregation configuration, with 8760 (365 typical days) serving as the reference case.

Due to the logarithmic scale of the x-axis, the marginal gain in accuracy decreases with an increasing number of typical days. However, the normed objective function value of the reference case forms an asymptote in the case of the self-sufficient building, which is not the case for the European model. For more than 1000 time steps, i.e., about 40 typical days, there is little change in the objective function value, which means that about 90% of the original time series values are indeed redundant. In contrast to that, the results for 1000 time steps are still far from those of the reference case for the European model, and it can be expected that the objective function value would further increase if more than 365 days were considered as original inputs.

The reason for this observation can be found in the multidimensionality of the respective datasets: In the case of the European model, much more low-correlated time series are simultaneously clustered and aggregated. Therefore, the resemblance of different time steps to each other is much lower, as a close similarity of two separate time steps becomes less likely as more low-correlated time series are taken into account. From a practical point of view, this also indicates that the number of potential extreme situations increases with the overall system size.

Therefore, the quality of temporal aggregation also depends on the redundancy, and ultimately the number and cross-correlation of the considered time series. This must be taken into account when choosing an appropriate method for reducing a model’s mathematical complexity.

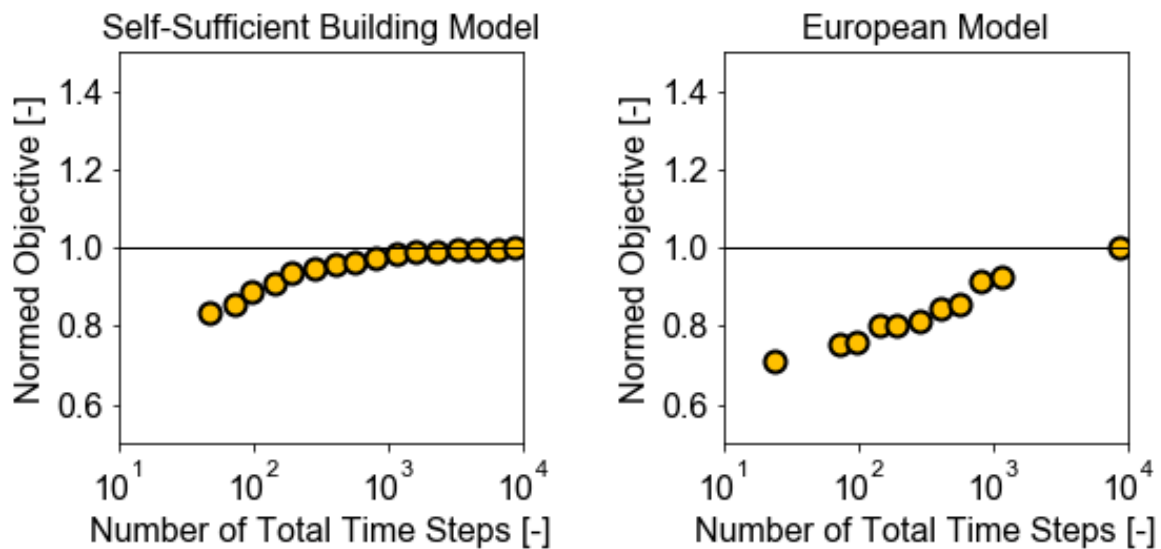


Figure 25. Convergence behavior of the two different aggregated energy systems’ optimal objective values to the reference value if centroid-based typical days with an hourly resolution (no segmentation) are used.

5. Conclusions

Temporal aggregation is an approach to represent time series by smaller sets of time steps while maintaining as many features of the fully resolved ones as possible. It is frequently applied in time-discrete energy system modelling in order to handle computationally intractable optimization runtimes as it limits the number of time-discrete variables and constraints.

In this study, a broad sensitivity analysis for different temporal aggregation techniques was applied to two different models and statistically-analyzed with respect to the desirable properties of an aggregation procedure, which led to small deviations in the respective energy system models. Concisely, two hypotheses were deduced stating that an imbalanced ratio between the number of typical time steps and the inner-daily temporal resolution may systematically bias the appropriate modeling of certain technologies, and that the preservation of the original time series' statistical momentums, and therefore its individual value distribution, is of the utmost importance for a good temporal aggregation. The individual research questions presented in the introduction of this study can be answered as follows:

Research question 1: Is there an optimal tradeoff between the duration of time steps and the number of typical periods, and are certain technologies affected systematically by either of these temporal aggregation approaches?

Our results imply that *there is* a systematic bias introduced by too low a number of inner-daily segments or typical days. We found that too low an inner-daily temporal resolution disproportionately underestimates the daily variance of photovoltaic time series and makes them more economically viable due to this underestimated intermittency. In contrast, too small a number of typical days more strongly affects the aperiodic behavior of wind profiles, with the same effect of overestimating wind capacities if only a few typical days are considered. We proposed an algorithm that automatically differentiates between the different time series and searches for a good tradeoff between the number of typical days and inner-daily segments for a given maximum number of total time steps.

Research question 2: Can the aggregation procedure be improved by explicitly taking statistical momentums, i.e., the original time series' value distributions, into account?

We developed an algorithm that is capable of aggregating time series in such a way that the original time series' value distribution is very closely approximated. The algorithm significantly outperformed a representation by centroids and medoids for both models. Therefore, it can be stated that not only the mean of a time series but also other statistical momentums of higher order are of importance for achieving an appropriate aggregation.

Both of the approaches proposed for improving on current temporal aggregation techniques led to a Pareto-optimal tradeoff between the calculation runtime and the energy system models' aggregation-induced objective function deviation. This means that for arbitrarily strongly aggregated, and therefore accelerated models, the results using the proposed methods were strictly more accurate than any comparably strongly aggregated model using traditional methods. On average, the speedup to traditionally-employed aggregation approaches was about one order of magnitude given the same accuracy level. Future research could focus on an improvement in the algorithms to account for the importance of certain time series for the energy system model to be aggregated when determining a ratio between the number of segments and typical days, as well as the development of a representation method that also preserves the original set of time series' covariance.

The methods are open source and available at: <https://github.com/FZJ-IEK3-VSA/tsam>

Acknowledgements

This work was supported by the Federal Ministry for Economic Affairs and Energy of Germany as part of the METIS project (project number: 03ET4064A).

A. Appendix

A.1. Techno-economic data of the self-sufficient building

Table 4. Cost parameters of the self-sufficient building model.

	Capex				Opex		Lifetime		Source
Components	Fixed		Capacity-Specific		Fixed+Capacity-Specific				
Photovoltaic Ground	—	—	4000.00	€/kW _p	1.00	% Inv./a	20	a	own assumption
Photovoltaic Rooftop	—	—	769.00	€/kW _p	1.00	% Inv./a	20	a	[109]
Inverter	—	—	75.00	€/kW _p	—	—	20	a	[110]
Battery	—	—	301.00	€/kWh _p	—	—	15	a	[109]
Reversible Solid Oxide Cell	5,000.00	€	2,400.00	€/kW _{el}	1.00	% Inv./a	15	a	[111]
Heat Pump	4,230.00	€	504.90	€/kW _{th}	1.50	% Inv./a	20	a	[112]
Thermal Storage	—	—	90.00	€/kWh _{th}	0.01	% Inv./a	25	a	[113]
E-Heater & E-Boiler	—	—	60.00	€/kW _{th}	2.00	% Inv./a	30	a	[113]
Tank	—	—	0.79	€/kWh _{H₂}	—	—	25	a	[114]
Dibenzyltoluene	—	—	1.25	€/kWh _{H₂}	—	—	25	a	[115, 116]
Hydrogen Vessels	—	—	15.00	€/kWh _{H₂}	—	—	25	a	[117]
Hydrogenizer	2,123.30	€	761.10	€/kW _{H₂}	1.00	% Inv./a	20	a	[115]
Dehydrogenizer	1,140.00	€	408.60	€/kW _{H₂}	1.00	% Inv./a	20	a	[115]
Low Pressure Compressor	—	—	1716.71	€/kW _p	1.00	% Inv./a	25	a	[118]
High Pressure Compressor	560.00	€	1329.80	€/kW _p	1.00	% Inv./a	25	a	[118]
Heat-Exchangers 1 and 2	—	—	1.00	€/kW _{th}	1.00	% Inv./a	—	a	own assumption
Expanders 1 and 2	—	—	1.00	€/kW _{th}	1.00	% Inv./a	25	a	own assumption

References

- Robinius, M., A. Otto, P. Heuser, L. Welder, K. Syranidis, D.S. Ryberg, T. Grube, P. Markewitz, R. Peters, and D. Stolten, *Linking the Power and Transport Sectors—Part 1: The Principle of Sector Coupling*. Energies, 2017. **10**(7): p. 956, DOI: <https://doi.org/10.3390/en10070956>.
- Robinius, M., A. Otto, K. Syranidis, D.S. Ryberg, P. Heuser, L. Welder, T. Grube, P. Markewitz, V. Tietze, and D. Stolten, *Linking the Power and Transport Sectors—Part 2: Modelling a Sector Coupling Scenario for Germany*. Energies, 2017. **10**(7): p. 957, DOI: <https://doi.org/10.3390/en10070957>.
- Schaller, R.R., *Moore's law: past, present and future*. IEEE Spectrum, 1997. **34**(6): pp. 52-59, DOI: <https://doi.org/10.1109/6.591665>.
- Robison, R.A., *Moore's Law: Predictor and Driver of the Silicon Era*. World Neurosurgery, 2012. **78**(5): pp. 399-403, DOI: <https://doi.org/10.1016/j.wneu.2012.08.019>.
- Koch, T., A. Martin, and M.E. Pfetsch, *Progress in Academic Computational Integer Programming*, in *Facets of Combinatorial Optimization: Festschrift for Martin Grötschel*, M. Jünger and G. Reinelt, Editors. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 483-506, DOI: https://doi.org/10.1007/978-3-642-38189-8_19.
- Priesmann, J., L. Nolting, and A. Praktijnjo, *Are complex energy system models more accurate? An intra-model comparison of power system optimization models*. Applied Energy, 2019. **255**: p. 113783, DOI: <https://doi.org/10.1016/j.apenergy.2019.113783>.

7. Kotzur, L., L. Nolting, M. Hoffmann, T. Groß, A. Smolenko, J. Priesmann, H. Büsing, R. Beer, F. Kullmann, B. Singh, A. Praktiknjo, D. Stolten, and M. Robinius, *A modeler's guide to handle complexity in energy systems optimization*. Advances in Applied Energy, 2021. **4**: p. 100063, DOI: <https://doi.org/10.1016/j.adapen.2021.100063>.
8. Hoffmann, M., L. Kotzur, D. Stolten, and M. Robinius, *A Review on Time Series Aggregation Methods for Energy System Models*. Energies, 2020. **13**(3), DOI: <https://doi.org/10.3390/en13030641>.
9. Fazlollahi, S., S.L. Bungener, P. Mandel, G. Becker, and F. Maréchal, *Multi-objectives, multi-period optimization of district energy systems: I. Selection of typical operating periods*. Computers & Chemical Engineering, 2014. **65**: pp. 54-66, DOI: <https://doi.org/10.1016/j.compchemeng.2014.03.005>.
10. Bahl, B., J. Lützwow, D. Shu, D.E. Hollermann, M. Lampe, M. Hennen, and A. Bardow, *Rigorous synthesis of energy systems by decomposition via time-series aggregation*. Computers & Chemical Engineering, 2018. **112**: pp. 70-81, DOI: <https://doi.org/10.1016/j.compchemeng.2018.01.023>.
11. Bahl, B., T. Söhler, M. Hennen, and A. Bardow, *Typical Periods for Two-Stage Synthesis by Time-Series Aggregation with Bounded Error in Objective Function*. Frontiers in Energy Research, 2018. **5**(35), DOI: <https://doi.org/10.3389/fenrg.2017.00035>.
12. Baumgärtner, N., F. Temme, B. Bahl, M. Hennen, D. Hollermann, and A. Bardow. *RiSES4 Rigorous Synthesis of Energy Supply Systems with Seasonal Storage by relaxation and time-series aggregation to typical periods*. in *Ecos 2019 - The 32nd International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems*. 2019. Wroclaw, Poland.
13. Baumgärtner, N., B. Bahl, M. Hennen, and A. Bardow, *RiSES3: Rigorous Synthesis of Energy Supply and Storage Systems via time-series relaxation and aggregation*. Computers & Chemical Engineering, 2019. **127**: pp. 127-139, DOI: <https://doi.org/10.1016/j.compchemeng.2019.02.006>.
14. Raventós, O. and J. Bartels, *Evaluation of temporal complexity reduction techniques applied to storage expansion planning in power system models*. Energies, 2020. **13**(4): p. 988, DOI: <https://doi.org/10.3390/en13040988>.
15. Wogrin, S., D.A. Tejada-Arango, S. Pineda, and J.M. Morales. *What time-period aggregation method works best for power system operation models with renewables and storage?* in *2019 International Conference on Smart Energy Systems and Technologies (SEST)*. 2019. DOI: 10.1109/SEST.2019.8849027.
16. Domínguez-Muñoz, F., J.M. Cejudo-López, A. Carrillo-Andrés, and M. Gallardo-Salazar, *Selection of typical demand days for CHP optimization*. Energy and Buildings, 2011. **43**(11): pp. 3036-3043, DOI: <https://doi.org/10.1016/j.enbuild.2011.07.024>.
17. Schütz, T., L. Schiffer, H. Harb, M. Fuchs, and D. Müller, *Optimal design of energy conversion units and envelopes for residential building retrofits using a comprehensive MILP model*. Applied Energy, 2017. **185**: pp. 1-15, DOI: <https://doi.org/10.1016/j.apenergy.2016.10.049>.
18. Schütz, T., M. Schraven, H. Harb, M. Fuchs, and D. Mueller. *Clustering algorithms for the selection of typical demand days for the optimal design of building energy systems*. in *29th International Conference on Efficiency, Cost, Optimisation, Simulation and Environmental Impact of Energy Systems*. 2016. Portorož, Slovenia: University of Ljubljana / Faculty of Mechanical Engineering.
19. Schütz, T., M.H. Schraven, M. Fuchs, P. Remmen, and D. Müller, *Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis*. Renewable Energy, 2018. **129**: pp. 570-582, DOI: <https://doi.org/10.1016/j.renene.2018.06.028>.
20. Adhau, S.P., R.M. Moharil, and P.G. Adhau, *K-Means clustering technique applied to availability of micro hydro power*. Sustainable Energy Technologies and Assessments, 2014. **8**: pp. 191-201, DOI: <https://doi.org/10.1016/j.seta.2014.09.001>.

21. Gabrielli, P., F. Furer, G. Mavromatidis, and M. Mazzotti, *Robust and optimal design of multi-energy systems with seasonal storage through uncertainty analysis*. Applied Energy, 2019. **238**: pp. 1192-1210, DOI: <https://doi.org/10.1016/j.apenergy.2019.01.064>.
22. Gabrielli, P., M. Gazzani, E. Martelli, and M. Mazzotti, *Optimal design of multi-energy systems with seasonal storage*. Applied Energy, 2018. **219**: pp. 408-424, DOI: <https://doi.org/10.1016/j.apenergy.2017.07.142>.
23. Green, R., I. Staffell, and 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, 2014. **61**(2): pp. 251-260, DOI: <https://doi.org/10.1109/TEM.2013.2284386>.
24. Teichgraeber, H. and A.R. Brandt, *Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison*. Applied Energy, 2019. **239**: pp. 1283-1293, DOI: <https://doi.org/10.1016/j.apenergy.2019.02.012>.
25. Teichgraeber, H., C. Lindenmeyer, N. Baumgärtner, L. Kotzur, D. Stolten, M. Robinius, A. Bardow, and A. Brandt, *Extreme events in time series aggregation: A case study for optimal residential energy supply systems*. 2020.
26. Teichgraeber, H., C.P. Lindenmeyer, N. Baumgärtner, L. Kotzur, D. Stolten, M. Robinius, A. Bardow, and A.R. Brandt, *Extreme events in time series aggregation: A case study for optimal residential energy supply systems*. Applied Energy, 2020. **275**: p. 115223, DOI: <https://doi.org/10.1016/j.apenergy.2020.115223>.
27. Teichgraeber, H., P.G. Brodrick, and A.R. Brandt, *Optimal design and operations of a flexible oxyfuel natural gas plant*. Energy, 2017. **141**: pp. 506-518, DOI: <https://doi.org/10.1016/j.energy.2017.09.087>.
28. Wogrin, S., P. Dueñas, A. Delgadillo, and J. Reneses, *A New Approach to Model Load Levels in Electric Power Systems With High Renewable Penetration*. IEEE Transactions on Power Systems, 2014. **29**(5): pp. 2210-2218, DOI: <https://doi.org/10.1109/TPWRS.2014.2300697>.
29. Wogrin, S., D. Galbally, and J. Reneses, *Optimizing Storage Operations in Medium- and Long-Term Power System Models*. IEEE Transactions on Power Systems, 2016. **31**(4): pp. 3129-3138, DOI: <https://doi.org/10.1109/TPWRS.2015.2471099>.
30. Kannengießer, T., M. Hoffmann, L. Kotzur, P. Stenzel, F. Schuetz, K. Peters, S. Nykamp, D. Stolten, and M. Robinius, *Reducing Computational Load for Mixed Integer Linear Programming: An Example for a District and an Island Energy System*. Energies, 2019. **12**(14): p. 2825, DOI: <https://doi.org/10.3390/en12142825>.
31. Agapoff, S., C. Pache, P. Panciatici, L. Warland, and S. Lumbrellas, *Snapshot selection based on statistical clustering for Transmission Expansion Planning*. in 2015 IEEE Eindhoven PowerTech. 2015. DOI: <https://doi.org/10.1109/PTC.2015.7232393>.
32. Sun, M., F. Teng, X. Zhang, G. Strbac, and D. Pudjianto, *Data-Driven Representative Day Selection for Investment Decisions: A Cost-Oriented Approach*. IEEE Transactions on Power Systems, 2019: pp. 1-1, DOI: <https://doi.org/10.1109/TPWRS.2019.2892619>.
33. Pfenninger, S., *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, 2017. **197**: pp. 1-13, DOI: <https://doi.org/10.1016/j.apenergy.2017.03.051>.
34. Zatti, M., M. Gabba, M. Freschini, M. Rossi, A. Gambarotta, M. Morini, and E. Martelli, *k-MILP: A novel clustering approach to select typical and extreme days for multi-energy systems design optimization*. Energy, 2019. **181**: pp. 1051-1063, DOI: <https://doi.org/10.1016/j.energy.2019.05.044>.
35. Brodrick, P.G., C.A. Kang, A.R. Brandt, and L.J. Durlofsky, *Optimization of carbon-capture-enabled coal-gas-solar power generation*. Energy, 2015. **79**: pp. 149-162, DOI: <https://doi.org/10.1016/j.energy.2014.11.003>.
36. Brodrick, P.G., A.R. Brandt, and L.J. Durlofsky, *Operational optimization of an integrated solar combined cycle under practical time-dependent constraints*. Energy, 2017. **141**: pp. 1569-1584, DOI: <https://doi.org/10.1016/j.energy.2017.11.059>.

37. Zhang, H., Z. Lu, W. Hu, Y. Wang, L. Dong, and J. Zhang, *Coordinated optimal operation of hydro–wind–solar integrated systems*. Applied Energy, 2019. **242**: pp. 883-896, DOI: <https://doi.org/10.1016/j.apenergy.2019.03.064>.
38. Fitiwi, D.Z., F. de Cuadra, L. Olmos, and M. Rivier, *A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty*. Energy, 2015. **90**: pp. 1360-1376, DOI: <https://doi.org/10.1016/j.energy.2015.06.078>.
39. Minh Khuong, P., H. Ü. Yilmaz, R. McKenna, and D. Keles, *A statistical approach to identify asynchronous extreme events for multi-regional energy system models*. International Journal of Energy Sector Management, 2020. **ahead-of-print**(ahead-of-print), DOI: 10.1108/IJESM-11-2019-0010.
40. Marquant, J.F.O., Akomeno; Orehounig, Kristina; Evins, Ralph; Carmeliet, Jan, *Application of Spatial-Temporal Clustering to Facilitate Energy System Modelling*, in *14th International Conference of IBPSA Building Simulation 2015*, V.R. Khare, Chaudhary, Gaurav, Editor. 2015, IIIT Hyderabad: Hyderabad, India. p. 551 - 558.
41. Marquant, J.F., G. Mavromatidis, R. Evins, and J. Carmeliet, *Comparing different temporal dimension representations in distributed energy system design models*. Energy Procedia, 2017. **122**: pp. 907-912, DOI: <https://doi.org/10.1016/j.egypro.2017.07.403>.
42. Pecenak, Z.K., M. Stadler, P. Mathiesen, K. Fahy, and J. Kleissl, *Robust design of microgrids using a hybrid minimum investment optimization*. Applied Energy, 2020. **276**: p. 115400, DOI: <https://doi.org/10.1016/j.apenergy.2020.115400>.
43. Munoz, F.D. and A.D. Mills, *Endogenous Assessment of the Capacity Value of Solar PV in Generation Investment Planning Studies*. IEEE Transactions on Sustainable Energy, 2015. **6**(4): pp. 1574-1585, DOI: <https://doi.org/10.1109/TSTE.2015.2456019>.
44. Pinel, D., *Clustering methods assessment for investment in zero emission neighborhoods' energy system*. International Journal of Electrical Power & Energy Systems, 2020. **121**: p. 106088, DOI: <https://doi.org/10.1016/j.ijepes.2020.106088>.
45. Schiefelbein, J., J. Tesfaegzi, R. Streblow, and D. Müller, *Design of an optimization algorithm for the distribution of thermal energy systems and local heating networks within a city district*. Proceedings of ECOS, 2015, DOI: <https://doi.org/10.13140/RG.2.1.1546.3521>.
46. Tso, W.W., C.D. Demirhan, C.F. Heuberger, J.B. Powell, and E.N. Pistikopoulos, *A hierarchical clustering decomposition algorithm for optimizing renewable power systems with storage*. Applied Energy, 2020. **270**: p. 115190, DOI: <https://doi.org/10.1016/j.apenergy.2020.115190>.
47. Lin, F., S. Leyffer, and T. Munson, *A two-level approach to large mixed-integer programs with application to cogeneration in energy-efficient buildings*. Computational Optimization and Applications, 2016. **65**(1): pp. 1-46, DOI: <https://doi.org/10.1007/s10589-016-9842-0>.
48. Ploussard, Q., L. Olmos, and A. Ramos, *An operational state aggregation technique for transmission expansion planning based on line benefits*. IEEE Transactions on Power Systems, 2016. **32**(4): pp. 2744-2755, DOI: <https://doi.org/10.1109/TPWRS.2016.2614368>.
49. Hoffmann, M., J. Priesmann, L. Nolting, A. Praktiknjo, L. Kotzur, and D. Stolten, *Typical periods or typical time steps? A multi-model analysis to determine the optimal temporal aggregation for energy system models*. Applied Energy, 2021. **304**: p. 117825, DOI: <https://doi.org/10.1016/j.apenergy.2021.117825>.
50. Nahmmacher, P., E. Schmid, L. Hirth, and B. Knopf, *Carpe diem: A novel approach to select representative days for long-term power system modeling*. Energy, 2016. **112**: pp. 430-442, DOI: <https://doi.org/10.1016/j.energy.2016.06.081>.
51. Teichgraber, H., L.E. Küpper, and A.R. Brandt, *Designing reliable future energy systems by iteratively including extreme periods in time-series aggregation*. Applied Energy, 2021. **304**: p. 117696, DOI: <https://doi.org/10.1016/j.apenergy.2021.117696>.
52. Bahl, B., A. Kümpel, H. Seele, M. Lampe, and A. Bardow, *Time-series aggregation for synthesis problems by bounding error in the objective function*. Energy, 2017. **135**: pp. 900-912, DOI: <https://doi.org/10.1016/j.energy.2017.06.082>.

53. Yokoyama, R., K. Takeuchi, Y. Shinano, and T. Wakui, *Effect of model reduction by time aggregation in multiobjective optimal design of energy supply systems by a hierarchical MILP method*. Energy, 2021. **228**: p. 120505, DOI: <https://doi.org/10.1016/j.energy.2021.120505>.
54. Härtel, P., M. Kristiansen, and M. Korpås, *Assessing the impact of sampling and clustering techniques on offshore grid expansion planning*. Energy Procedia, 2017. **137**: pp. 152-161, DOI: <https://doi.org/10.1016/j.egypro.2017.10.342>.
55. Göke, L. and M. Kendzioriski, *Adequacy of time-series reduction for renewable energy systems*. Energy, 2022. **238**: p. 121701, DOI: <https://doi.org/10.1016/j.energy.2021.121701>.
56. Heuberger, C.F., I. Staffell, N. Shah, and N.M. Dowell, *A systems approach to quantifying the value of power generation and energy storage technologies in future electricity networks*. Computers & Chemical Engineering, 2017. **107**: pp. 247-256, DOI: <https://doi.org/10.1016/j.compchemeng.2017.05.012>.
57. Zhu, Q., X. Luo, B. Zhang, and Y. Chen, *Mathematical modelling and optimization of a large-scale combined cooling, heat, and power system that incorporates unit changeover and time-of-use electricity price*. Energy Conversion and Management, 2017. **133**: pp. 385-398, DOI: <https://doi.org/10.1016/j.enconman.2016.10.056>.
58. Deml, S., A. Ulbig, T. Borsche, and G. Andersson. *The role of aggregation in power system simulation*. in 2015 IEEE Eindhoven PowerTech. 2015. DOI: <https://doi.org/10.1109/PTC.2015.7232755>.
59. Stadler, P., L. Girardin, A. Ashouri, and F. Maréchal, *Contribution of Model Predictive Control in the Integration of Renewable Energy Sources within the Built Environment*. Frontiers in Energy Research, 2018. **6**(22), DOI: <https://doi.org/10.3389/fenrg.2018.00022>.
60. Stein, D.v., N.v. Bracht, A. Maaz, and A. Moser. *Development of adaptive time patterns for multi-dimensional power system simulations*. in 2017 14th International Conference on the European Energy Market (EEM). 2017. DOI: <https://doi.org/10.1109/EEM.2017.7981868>.
61. Kotzur, L., P. Markewitz, M. Robinius, and D. Stolten, *Time series aggregation for energy system design: Modeling seasonal storage*. Applied Energy, 2018. **213**: pp. 123-135, DOI: <https://doi.org/10.1016/j.apenergy.2018.01.023>.
62. Kotzur, L., P. Markewitz, M. Robinius, and D. Stolten, *Impact of different time series aggregation methods on optimal energy system design*. Renewable Energy, 2018. **117**: pp. 474-487, DOI: <https://doi.org/10.1016/j.renene.2017.10.017>.
63. Pineda, S. and J.M. Morales, *Chronological Time-Period Clustering for Optimal Capacity Expansion Planning With Storage*. IEEE Transactions on Power Systems, 2018. **33**(6): pp. 7162-7170, DOI: <https://doi.org/10.1109/TPWRS.2018.2842093>.
64. Lara, C.L., D.S. Mallapragada, D.J. Papageorgiou, A. Venkatesh, and I.E. Grossmann, *Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm*. European Journal of Operational Research, 2018. **271**(3): pp. 1037-1054, DOI: <https://doi.org/10.1016/j.ejor.2018.05.039>.
65. Mavrotas, G., D. Diakoulaki, K. Florios, and P. Georgiou, *A mathematical programming framework for energy planning in services' sector buildings under uncertainty in load demand: The case of a hospital in Athens*. Energy Policy, 2008. **36**(7): pp. 2415-2429, DOI: <https://doi.org/10.1016/j.enpol.2008.01.011>.
66. Liu, Y., R. Sioshansi, and A.J. Conejo, *Hierarchical Clustering to Find Representative Operating Periods for Capacity-Expansion Modeling*. IEEE Transactions on Power Systems, 2018. **33**(3): pp. 3029-3039, DOI: <https://doi.org/10.1109/TPWRS.2017.2746379>.
67. Reichenberg, L., A.S. Siddiqui, and S. Wogrin, *Policy implications of downscaling the time dimension in power system planning models to represent variability in renewable output*. Energy, 2018. **159**: pp. 870-877, DOI: <https://doi.org/10.1016/j.energy.2018.06.160>.
68. Tejada-Arango, D.A., M. Domeshek, S. Wogrin, and E. Centeno, *Enhanced Representative Days and System States Modeling for Energy Storage Investment Analysis*. IEEE Transactions on Power Systems, 2018. **33**(6): pp. 6534-6544, DOI: <https://doi.org/10.1109/TPWRS.2018.2819578>.

69. Tejada-Arango, D.A., S. Wogrin, and E. Centeno, *Representation of Storage Operations in Network-Constrained Optimization Models for Medium- and Long-Term Operation*. IEEE Transactions on Power Systems, 2018. **33**(1): pp. 386-396, DOI: <https://doi.org/10.1109/TPWRS.2017.2691359>.
70. Tupper, L.L., D.S. Matteson, C.L. Anderson, and L. Zephyr, *Band Depth Clustering for Nonstationary Time Series and Wind Speed Behavior*. Technometrics, 2018. **60**(2): pp. 245-254, DOI: <https://doi.org/10.1080/00401706.2017.1345700>.
71. Welder, L., D. Ryberg, L. Kotzur, T. Grube, M. Robinus, and D. Stolten, *Spatio-Temporal Optimization of a Future Energy System for Power-to-Hydrogen Applications in Germany*. 2018, DOI: <https://doi.org/10.1016/j.energy.2018.05.059>.
72. Patteeuw, D. and L. Helsen, *Combined design and control optimization of residential heating systems in a smart-grid context*. Energy and Buildings, 2016. **133**: pp. 640-657, DOI: <https://doi.org/10.1016/j.enbuild.2016.09.030>.
73. Poncelet, K., H. Höschle, E. Delarue, A. Virag, and W. D'haeseleer, *Selecting Representative Days for Capturing the Implications of Integrating Intermittent Renewables in Generation Expansion Planning Problems*. IEEE Transactions on Power Systems, 2016. **32**(3): pp. 1936-1948, DOI: <https://doi.org/10.1109/TPWRS.2016.2596803>.
74. Ortega, J., J.C. Bruno, and A. Coronas, *Selection of typical days for the characterisation of energy demand in cogeneration and trigeneration optimisation models for buildings*. Energy Conversion and Management, 2011. **52**(4): pp. 1934-1942, DOI: <https://doi.org/10.1016/j.enconman.2010.11.022>.
75. Devogelaer, D., *Towards 100% renewable energy in Belgium by 2050*. 2012, Brussels, Belgium: Federal Planning Bureau.
76. De Sisternes Jimenez, F. and M.D. Webster, *Optimal selection of sample weeks for approximating the net load in generation planning problems*, in *ESD Working Papers*. 2013, DOI: <https://doi.org/10.1721.1/102959>.
77. de Sisternes, F.J., J.D. Jenkins, and A. Botterud, *The value of energy storage in decarbonizing the electricity sector*. Applied Energy, 2016. **175**: pp. 368-379, DOI: <https://doi.org/10.1016/j.apenergy.2016.05.014>.
78. Simões, S., W. Nijs, P. Ruiz, A. Sgobbi, d. radu, P. Yilmaz Bolat, c. thiel, and e. peteves, *The JRC-EU-TIMES model - Assessing the long-term role of the SET Plan Energy technologies*. 2013, DOI: <https://doi.org/10.2790/97596>.
79. Voll, P., C. Klaffke, M. Hennen, and A. Bardow, *Automated superstructure-based synthesis and optimization of distributed energy supply systems*. Energy, 2013. **50**: pp. 374-388, DOI: <https://doi.org/10.1016/j.energy.2012.10.045>.
80. Brodrick, P.G., A.R. Brandt, and L.J. Durlofsky, *Optimal design and operation of integrated solar combined cycles under emissions intensity constraints*. Applied Energy, 2018. **226**: pp. 979-990, DOI: <https://doi.org/10.1016/j.apenergy.2018.06.052>.
81. Fazlollahi, S., L. Girardin, and F. Maréchal, *Clustering Urban Areas for Optimizing the Design and the Operation of District Energy Systems*, in *Computer Aided Chemical Engineering*, J.J. Klemeš, P.S. Varbanov, and P.Y. Liew, Editors. 2014, Elsevier. p. 1291-1296, DOI: <https://doi.org/10.1016/B978-0-444-63455-9.50050-7>.
82. Poncelet, K., E. Delarue, J. Duerinck, D. Six, and W. D'haeseleer, *The Importance of Integrating the Variability of Renewables in Long-term Energy Planning Models*, in *TME Working Paper - Energy and Environment*. 2014, KU Leuven: Leuven.
83. Poncelet, K., E. Delarue, D. Six, J. Duerinck, and W. D'haeseleer, *Impact of the level of temporal and operational detail in energy-system planning models*. Applied Energy, 2016. **162**: pp. 631-643, DOI: <https://doi.org/10.1016/j.apenergy.2015.10.100>.
84. Stadler, M., M. Groissböck, G. Cardoso, and C. Marnay, *Optimizing Distributed Energy Resources and building retrofits with the strategic DER-CAModel*. Applied Energy, 2014. **132**: pp. 557-567, DOI: <https://doi.org/10.1016/j.apenergy.2014.07.041>.

85. Hilbers, A.P., D.J. Brayshaw, and A. Gandy, *Importance subsampling: improving power system planning under climate-based uncertainty*. Applied Energy, 2019. **251**: p. 113114, DOI: <https://doi.org/10.1016/j.apenergy.2019.04.110>.
86. Wakui, T. and R. Yokoyama, *Optimal structural design of residential cogeneration systems in consideration of their operating restrictions*. Energy, 2014. **64**: pp. 719-733, DOI: <https://doi.org/10.1016/j.energy.2013.10.002>.
87. Wakui, T. and R. Yokoyama, *Optimal structural design of residential cogeneration systems with battery based on improved solution method for mixed-integer linear programming*. Energy, 2015. **84**: pp. 106-120, DOI: <https://doi.org/10.1016/j.energy.2015.02.056>.
88. Wakui, T., H. Kawayoshi, and R. Yokoyama, *Optimal structural design of residential power and heat supply devices in consideration of operational and capital recovery constraints*. Applied Energy, 2016. **163**: pp. 118-133, DOI: <https://doi.org/10.1016/j.apenergy.2015.10.154>.
89. Frew, B.A. and M.Z. Jacobson, *Temporal and spatial tradeoffs in power system modeling with assumptions about storage: An application of the POWER model*. Energy, 2016. **117**: pp. 198-213, DOI: <https://doi.org/10.1016/j.energy.2016.10.074>.
90. Merrick, J.H., *On representation of temporal variability in electricity capacity planning models*. Energy Economics, 2016. **59**: pp. 261-274, DOI: <https://doi.org/10.1016/j.eneco.2016.08.001>.
91. Ward, J.H., *Hierarchical Grouping to Optimize an Objective Function* AU - Ward, Joe H. Journal of the American Statistical Association, 1963. **58**(301): pp. 236-244, DOI: <https://doi.org/10.1080/01621459.1963.10500845>.
92. van der Heijde, B., L. Scapino, A. Vandermeulen, D. Patteeuw, L. Helsen, and R. Salenbien. *Using representative time slices for optimization of thermal energy storage systems in low-temperature district heating systems*. in *31st International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems*. 2018. Guimarães, Portugal: University of Minhos; Guimarães.
93. van der Heijde, B., A. Vandermeulen, R. Salenbien, and L. Helsen, *Representative days selection for district energy system optimisation: a solar district heating system with seasonal storage*. Vol. 248. 2019. 79-94, DOI: <https://doi.org/10.1016/j.apenergy.2019.04.030>.
94. Poncelet, K., H. Höschle, E. Delarue, and W. D'haeseleer, *Selecting representative days for investment planning models*, in *TME Working Paper - Energy and Environment*. 2015, KU Leuven: Leuven, DOI: <https://doi.org/10.1109/TPWRS.2016.2596803>.
95. Scott, I.J., P.M.S. Carvalho, A. Botterud, and C.A. Silva, *Clustering representative days for power systems generation expansion planning: Capturing the effects of variable renewables and energy storage*. Applied Energy, 2019. **253**: p. 113603, DOI: <https://doi.org/10.1016/j.apenergy.2019.113603>.
96. Curry, H.B., *The method of steepest descent for non-linear minimization problems*. Quarterly of Applied Mathematics, 1944. **2**(3): pp. 258-261.
97. Kotzur, L., P. Markewitz, M. Robinius, and D. Stolten. *Kostenoptimale Versorgungssysteme für ein vollautarkes Einfamilienhaus*. in *Internationale Energiewirtschaftstagung*. 2017. Graz, Austria.
98. Knosala, K., L. Kotzur, F.T.C. Röben, P. Stenzel, L. Blum, M. Robinius, and D. Stolten, *Hybrid Hydrogen Home Storage for Decentralized Energy Autonomy*. International Journal of Hydrogen Energy, 2021. **46**(42): pp. 21748-21763, DOI: <https://doi.org/10.1016/j.ijhydene.2021.04.036>.
99. Knosala, K., L. Kotzur, F.T.C. Röben, P. Stenzel, L. Blum, M. Robinius, and D. Stolten, *Hybrid Hydrogen Home Storage for Decentralized Energy Autonomy*. 2021: Mendeley Data, DOI: <https://doi.org/10.17632/zhwkr6k93.1>.
100. Kotzur, L., K. Knosala, P. Markewitz, P. Stenzel, M. Robinius, and D. Stolten, *tsib - Time Series Initialization for Buildings*. 2019.

101. Kotzur, L., P. Markewitz, M. Robinius, G. Cardoso, P. Stenzel, M. Heleno, and D. Stolten, *Bottom-up energy supply optimization of a national building stock*. Energy and Buildings, 2020. **209**: p. 109667, DOI: <https://doi.org/10.1016/j.enbuild.2019.109667>.
102. Wahl, S., C. Bollmeyer, S. Crewell, C. Figura, P. Friederichs, A. Hense, J.D. Keller, and C. Ohlwein, *A novel convective-scale regional reanalysis COSMO-REA2: Improving the representation of precipitation*. 2017, DOI: <https://doi.org/10.1127/metz/2017/0824>.
103. Bollmeyer, C., J. Keller, C. Ohlwein, S. Wahl, S. Crewell, P. Friederichs, A. Hense, J. Keune, S. Kneifel, and I.J.Q.J.o.t.R.M.S. Pscheidt, *Towards a high-resolution regional reanalysis for the European CORDEX domain*. 2015. **141**(686): pp. 1-15, DOI: <https://doi.org/10.1002/qj.2486>.
104. Frank, C.W., S. Wahl, J.D. Keller, B. Pospichal, A. Hense, and S. Crewell, *Bias correction of a novel European reanalysis data set for solar energy applications*. Solar Energy, 2018. **164**: pp. 12-24, DOI: <https://doi.org/10.1016/j.solener.2018.02.012>.
105. Caglayan, D.G., *Robust Design of a Future 100% Renewable European Energy System with Hydrogen Infrastructure*, in *Faculty of Mechanical Engineering*. 2020, RWTH Aachen University: Aachen, Germany.
106. Caglayan, D.G., *A Robust Design of a Renewable European Energy System Encompassing a Hydrogen Infrastructure*, in *Schriften des Forschungszentrums Jülich. Reihe Energie & Umwelt / Energy & Environment*. 2020, RWTH Aachen University: Jülich. p. xxii, 312 S.
107. Robinius, M., G. Erdmann, and D. Stolten, *Strom-und Gasmaktdesign zur Versorgung des deutschen Straßenverkehrs mit Wasserstoff*. 2015, Lehrstuhl für Brennstoffzellen (FZ Jülich).
108. Bruninx, K., D. Orlic, D. Couckuyt, N. Grisey, B. Betraoui, T. Anderski, Y. Surmann, N. Franck, G. Keane, and B. Hickman, *Modular development plan of the Pan-European transmission system 2050: Data sets of scenarios for 2050*, in *Technical Report*. 2015, The e-HIGHWAY 2050 Project.
109. Krampe, L., M. Wünsch, and M. Koepp, *Eigenversorgung aus Solaranlagen.: Das Potenzial für Photovoltaik-Speicher-Systeme in Ein-und Zweifamilienhäusern Landwirtschaft sowie im Lebensmittelhandel*. 2016.
110. Mayer, J.N., P. Simon, N.S.H. Philipps, T. Schlegl, and C. Senkpiel, *Current and future cost of photovoltaics. Long-term Scenarios for Market Development, System Prices LCOE of Utility-Scale PV Systems*. 2015.
111. Röben, F., *Techno-economical potential of reversible Solid Oxide Cells for autarkic Buildings and Districts*. 2017, RWTH Aachen University.
112. Wolf, *Preisliste Heizsysteme*. 2016 [cited 2021 June 11]; Available from: <https://www.ck-heiztechnik.de/media/kataloge/wolf/Wolf-Preisliste-Heizsysteme-03-2016.pdf>.
113. Lindberg, K.B., D. Fischer, G. Doorman, M. Korpås, and I. Sartori, *Cost-optimal energy system design in Zero Energy Buildings with resulting grid impact: A case study of a German multi-family house*. Energy and Buildings, 2016. **127**: pp. 830-845, DOI: <https://doi.org/10.1016/j.enbuild.2016.05.063>.
114. Energy, S. Schütz *Heizöl-Lagerbehälter und Tank-Ersatzteil Preisliste*. 2017 [cited 2021 June 11]; Available from: <https://www.schuetz-energy.net/downloads/heizoel-lagerbehaeltersysteme/tank-in-tank/heizoel-lagerbehaelter-preisliste-und-tank-ersatzteil-preisliste/schuetz-es-heizoel-lagerbehaelter-preisliste-und-tank-ersatzteil-preisliste-de.pdf?cid=617>.
115. Eypasch, M., M. Schimpe, A. Kanwar, T. Hartmann, S. Herzog, T. Frank, and T. Hamacher, *Model-based techno-economic evaluation of an electricity storage system based on Liquid Organic Hydrogen Carriers*. Applied Energy, 2017. **185**: pp. 320-330, DOI: <https://doi.org/10.1016/j.apenergy.2016.10.068>.
116. Teichmann, D., *Konzeption und Bewertung einer nachhaltigen Energieversorgung auf Basis flüssiger Wasserstoffträger*. 2015, Friedrich Alexander Universität Erlangen-Nürnberg.
117. US Department of Energy, *The Fuel Cell Technologies Office Multi-Year Research, Development, and Demonstration Plan*. 2016, Technical report: US Department of Energy.

118. Yang, C. and J. Ogden, *Determining the lowest-cost hydrogen delivery mode*. International Journal of Hydrogen Energy, 2007. **32**(2): pp. 268-286, DOI: <https://doi.org/10.1016/j.ijhydene.2006.05.009>.