Long term storage in generation expansion planning models with a reduced temporal scope

Sebastian Gonzato *, Kenneth Bruninx, Erik Delarue

Department of Mechanical Engineering, KU Leuven, Leuven, Belgium
Energy Ville, Genk, Belgium

ARTICLE INFO

ABSTRACT

To reduce the computation time of Energy System Optimization Models and Generation Expansion Planning Models operational detail is typically limited to several hours, days, or weeks in a year selected using Time Series Aggregation methods. We compare time series aggregation methods and generation expansion planning models which aim to capture the value of long-term storage for the first time in the literature. Time Series Aggregations methods were compared by varying the number of representative periods and then running a full year generation expansion planning model on novel synthetic time series. Generation Expansion Planning Models were run on selections and ordering of representative periods in order to compare them. Our results suggest that approximating the full-year time series does not necessarily translate to approximating the full-year generation expansion planning solution and that selecting hours or days is a greater determinant of performance than the time series aggregation method itself. Two of the generation expansion planning models considered, Enhanced Representative Days and Chronological Time Period Clustering, could capture the value of long-term storage, though over or underinvestment in long-term storage by more than a factor of 2 was also possible and the latter formulation exhibited a clear bias towards long-term storage. Based on these results we formulate recommendations for modelers seeking to include long-term storage in generation expansion planning models.

1. Introduction

Bottom-up Energy System Optimization Models (ESOMs), such as TIMES [1] and OSeMOSYS [2], and Generation Expansion Planning Models (GEPMs), such as ReEDS [3] and LIMES [4], are frequently used to aid decision-makers in shaping the transition towards a low carbon energy or power system [5]. This transition will likely entail an increased penetration of Variable Renewable Energy Sources (VRES) and energy storage, such as pumped hydro or power to gas, into the power system [6].

To reduce the computational burden of an ESOM,[1] intra-annual temporal variability is often represented in an aggregate form. Given that speed-ups of several orders of magnitude are possible, it is clear why there is an abundance of literature on the subject of choosing the hours, days, or weeks of a year to be fed as input to such models, as illustrated by a recent review on the subject which includes 130 different publications [7]. A difficulty that occurs when selecting such representative periods is modeling the inter-period arbitrage typically performed by mid to long-term energy storage technologies such as pumped hydro and power to gas. Ignoring such arbitrage can lead to a difference of more than an order of magnitude compared to when it is considered [8]. This motivates research on combinations of Time Series Aggregation (TSA) methods and GEPM formulations that can capture the value of such arbitrage.

Some of the literature on TSAs presents novel aggregation methods, such as optimization based procedures as was done in [9] in which representative days were selected based on approximating duration curves. More often clustering algorithms are used, as in [10] who compared the performance of different clustering algorithms. Some papers present heuristic additions to said algorithms, such as including "extreme" periods in the case of [11], or scaling the aggregated time series to preserve the correct annual average in the case of [12]. Attempts have also been made to rigorously compare these TSA methods see, e.g., [10,13] and [14]. It is nonetheless difficult to formulate consistent

---

* Corresponding author at: Department of Mechanical Engineering, KU Leuven, Leuven, Belgium.
 E-mail address: sebastian.gonzato@kuleuven.be (S. Gonzato).
 [1] For brevity, the term Energy System Optimization Models (ESOMs) is used here to denote both ESOMs and Generation Expansion Planning Models (GEPMs) unless noted otherwise.

Nomenclature

Variables and parameters are denoted by lower and uppercase letters respectively, with the variables of Time Series Aggregation methods becoming parameters in Generation Expansion Planning Models

Sets

$a \in A$	Set of Time Series Aggregation methods
$b \in \mathcal{B}$	Set of bins used to discretize duration curves
$c \in \mathcal{C}$	Set of clusters
$g \in \mathcal{G}$	Set of generation technologies
$h \in \mathcal{H}$	Set of storage technologies
$m \in M$	Set of Generation Expansion Planning Models
$i \in \mathcal{P}$	Set of all periods in a year
$j \in \mathcal{R}$	Set of representative periods in a year
$s \in \mathcal{S}$	Set of time series
$t \in \mathcal{T}$	Set of time steps in a period
$y \in Y$	Set of test systems

Variables

u_i	Binary variable denoting selection of period i as representative
w_i	Weight ascribed to a period i
v_{ij}	Ordering variable denoting selection of period j to represent period i
c_{hjt}	Charging of storage
d_{hjt}	Discharging of storage
e_{hit}	State of charge of storage
e_{hi}^0	State of charge of storage at the start of a representative period
ec_h	Energy capacity of storage
k_g	Generator capacity
q_{gjt}	Generator power output
ls_{jt}	Load shedding

Parameters

X_{sit}	Normalized value of time series [–]
$\bar{X}_{sit}^a$	Normalized value of synthetic time series [–]
N_{total}	Total number of periods in the year
N_{total}	Number of representative periods to be selected
PE_h	Power to energy ratio of storage [1/h]
AF_{gjt}	Availability factor of generator [–]
η_h	Round trip efficiency of storage [–]
$\tau_{j,t}$	Length of time step [h]
D_{jt}	Electric power demand [MW]
ϕ	Renewable penetration target [–]
C^{shed}	Value of lost load [€/MWh]

Acronyms

ESOM	Energy System Optimization Model
GEPM	Generation Expansion Planning Model
VRES	Variable Renewable Energy Sources
ERD	Enhanced Representative Days
ERH	Enhanced Representative Hours
URD	Unlinked Representative Days
LRD	Linked Representative Days

FY	Full Year
CTPC	Chronological Time Period Clustering
HC	Hierarchical Clustering
ORDF	Optimal Representative Days Finder
ORDO	Optimal Representative Days Orderer
RR	Optimal Representative Days Re Orderer
TSA	Time Series Aggregation

takeaways regarding TSA methods, other than to avoid averaging (or "temporal smoothing") and that it is difficult (if not impossible) to determine a TSA technique's performance without testing it on an ESOM.[2] This last fact makes a priori selection of representative periods particularly difficult.

Other publications have focused on the effect of temporal aggregation in ESOMs, where it has been noted that the temporal scope is more important than the operational detail of a model [17,18]. [19] highlighted the sensitivity of a GEPM's solution to the number of clustered hours since the solution is heavily influenced by a handful of scarcity (or peak pricing) hours. The author concludes by noting that for models which include VRES, care should be taken when not using a full year of data. [20] reached a similar conclusion but for inter-annual variability, remarking that the performance of a TSA depended on the year it was applied to.

Recent work by [21] attempts to overcome these issue issues by determining which hours most affect a GEPM solution and selecting these as input to a GEPM.[3] This method led to more accurate results while including fewer time steps, capturing the variability of 36 years of weather data in 8760 h.

A drawback common to the methods cited above is the previously mentioned difficulty of including storage technologies that arbitrage over time frames longer than the length of a representative period. They could therefore gloss over the value of longer-term storage technologies such as pumped hydro and power to gas, of which the latter is particularly valuable at high ($\geq$ 80%) VRES penetrations [24].

To overcome this, novel methods that allow for long-term storage in ESOMs with a reduced temporal scope have recently been proposed by [25–29]. These methods make use of representative periods being ordered throughout the year in order to model inter-period arbitrage. While all of the GEPM formulations cited above have been compared to a full-year model and or a GEPM in which inter-period arbitrage is not allowed, only [30] has compared two of these formulations, Enhanced Representative Days and Chronological Time Period Clustering, and then only in an operational setting. It may also be difficult to generalize any conclusions as these may be test system specific as noted by [20] and [13].

Given the above, the contributions of this paper are thus threefold:

1. We compare for the first time existing TSA methods and GEPM formulations which allow for long-term storage in an investment setting. We do so on 16 different test systems in order to draw robust conclusions since the results could otherwise be heavily influenced by input data [20].
2. We introduce the novel concept of synthetic time series to distinguish between errors that arise from a TSA method and those from the GEPM formulation since the constraints introduced to allow inter-period arbitrage lead to a distinction between the two. This distinction, a first in the literature, produced new insights, for example, that it is the GEPM formulation and not the TSA in the Enhanced Representative Days model which drives investment in long-term storage.

---

[2] Theoretical bounds on the value of the objective function can however be derived [15,16].

[3] Similar methods have also been developed by [22] and [23].
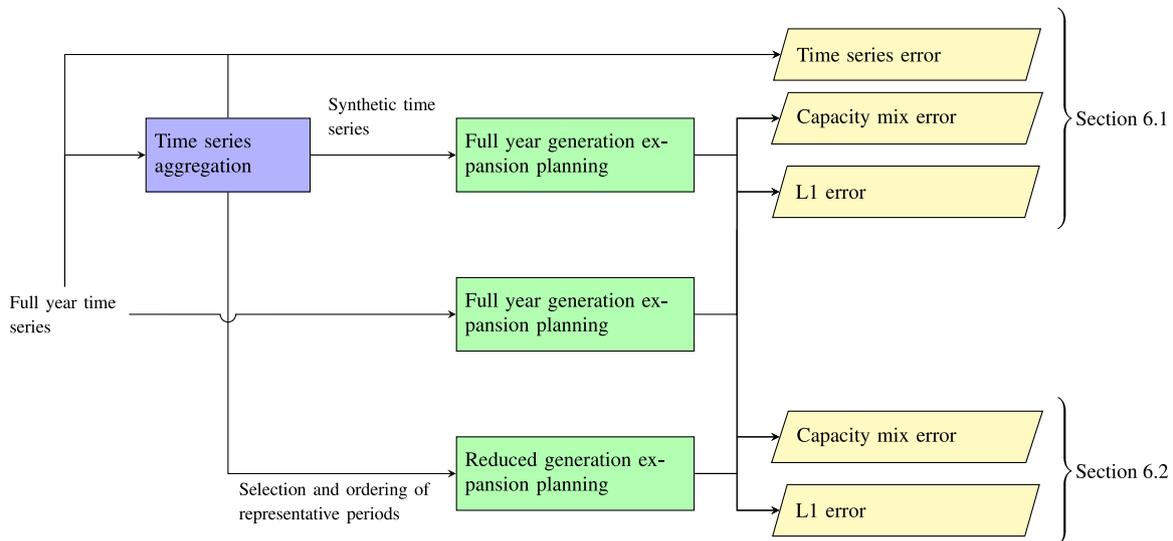
Fig. 1. Flowchart of the investigation procedure carried out in this paper. The rectangles indicate time series aggregation methods (see Section 2 and Table 1) and generation expansion planning models (see Section 3 and Table 2) while trapezia the performance metrics used to compare them (see Section 4.3). Curly braces on the right hand side indicate in which section these results are investigated. For a textual explanation of the investigation procedure please refer to Section 4.1.

3. We use the obtained results to formulate recommendations for modelers wishing to include long-term storage in an ESOM or GEPM. This is valuable information for modelers tasked with envisioning pathways to a highly decarbonized society, in which long-term storage technologies may be crucial [31].

In addition to the above, a novel optimization based TSA, Optimal Representative Days Orderer, is employed for the first time in the literature.

This paper thus serves as an initial benchmark of the aforementioned novel TSAs and GEPMs combinations. We do this in an attempt to rank the performance of several TSA methods that allow for long-term storage and compare different GEPM formulations that also allow for long-term storage. In both cases, we additionally investigate the presence of technology biases due to the TSA or GEPM formulation.

Regarding the comparison of TSA methods, we found that better approximating the full-year time series does not necessarily lead to a better approximation of the full-year GEPM solution. This is not yet common knowledge in the literature, as evidenced by the detailed statistical analysis of TSA results in [9] for example. We found minimal differences between the TSA methods compared using synthetic time series, save that selecting and ordering hours instead of days led to more accurate GEPM solutions. This is in contrast to the results of [14], highlighting the utility of differentiating between TSA and GEPM errors. Since the performance of a TSA method varies depending on the test system, we conclude that ranking particular methods should be treated with caution.

Regarding the comparison of GEPMs, two formulations, Enhanced Representative Days and Chronological Time Period Clustering, were able to well approximate the full-year solution in our analysis, but only for a high number of representative periods (128 days). This true on average, though for specific test systems the installed capacity of short or long-term storage could differ by a factor of 2 with respect to the full-year solution. A lower number (32 days) led to significant biases in storage capacity, such as little to no investment in short-term storage in the case of Chronological Time Period Clustering or overinvestment in short and long-term storage for Enhanced Representative Days. Other formulations, such as Unlinked Representative Days or Linked Representative Days, did not capture the value of long-term storage or were actually slower to run than the full-year model in the case of Enhanced Representative Hours. This last GEPM was thus unable to leverage the better performance of its corresponding TSA formulation, which selected and ordered hours instead of days.

Due to the simplicity of the GEPM used, these results may not be directly translatable to quantitative energy transition studies using ESOMs. However, they do indicate that caution should be taken when reducing the temporal detail in such models. This is not immediately apparent in the current literature, where temporal aggregation is typically presented as a simple way of achieving computational speed-ups. Given this we formulate recommendations to modelers and suggest avenues of future research, stressing that incremental improvements to existing TSA methods or GEPM formulations are unlikely to prove fruitful.

To ensure replicability and transparency of the results, all the Julia code used to produce these is freely available under a permissive MIT license at https://gitlab.kuleuven.be/u0128861/rep-days-4-storage. The TSA methods and GEPM formulations available as individual Julia packages at https://gitlab.kuleuven.be/UCM/representativeperiodsfinder.jl and https://gitlab.kuleuven.be/u0128861/GEPPR.jl respectively. The TSA package SpinePeriods.jl was also developed as part of this paper, in addition to modifications to SpineOpt.jl which implement the Enhanced Representative Hours method. Both can be found at https://github.com/Spine-project.

The rest of this paper is organized as follows. Sections 2 and 3 present the TSAs and GEPMs compared in this paper. Section 4 elaborates on how these are compared, while Section 5 describes the test systems and data used. Section 6 presents the results of the comparison of the TSA methods and GEPM formulations. Section 7 discusses limitations of the study and recommendations to modelers, along with suggestions for future research. Section 8 concludes.

A flowchart illustrating the investigation procedure can be found in Fig. 1.

2. Time Series Aggregation Methods

The purpose of Time Series Aggregation is to select a subset of representative periods $\mathcal{R} \subset \mathcal{P}$ in a time series with corresponding weights W_j, in the hope that running an ESOM on this subset will approximate the solution had the full set of periods $\mathcal{P}$ been used. In this paper, the full set $\mathcal{P}$ is comprised of the $|\mathcal{P}| = 365$ days or $|\mathcal{P}| = 8760$ h in a given year.[4] Each period i contains one or more chronologically

---

[4] Recent literature has highlighted the importance of including multiple years worth of data in the set $\mathcal{P}$ for energy transition studies, see, e.g., [20,21, 32]. Though an important observation, this is beyond the scope of this paper.

Table 1

Summary of TSA methods compared in this paper. **Criteria** indicates what is approximated in order to select and order days, the original time series or its duration curve. **References** indicates papers where (versions) of this method have been presented, though the list is not exhaustive.

Name		Algorithm	Selects periods?	Orders periods?	Can select hours?	Criteria		Reference
						Time series?	Duration curve?	
Optimal Representative Days Finder	*ORDF*	Optimization	✓	✗	✓	✗	✓	[9]
Optimal Representative Days Orderer	*ORDO*	Optimization	✓	✓	✗	✓	✗	This paper
Optimal Representative Days Re Orderer	*RR*	Optimization	✗	✓	✗	✓	✗	[29]
Hierarchical Clustering	*HC*	Clustering	✓	✓	✓	✓	✗	[26]
Chronological Time Period Clustering	*CTPC*	Clustering	✓	✓	✓	✓	✗	[28]
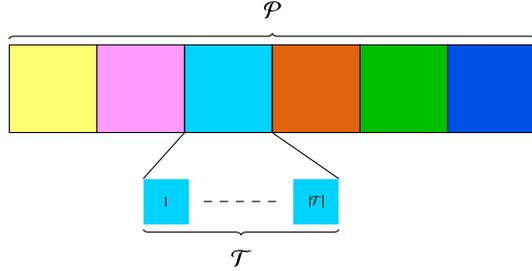
Fig. 2. Temporal representation employed in this paper. Top row is a fictitious year containing $|\mathcal{P}| = 6$ periods, each containing $|\mathcal{T}|$ time steps.
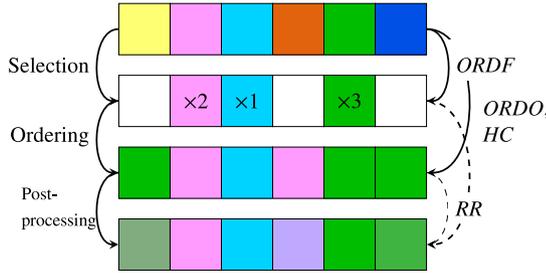
Fig. 3. Illustration of selection and ordering of representative days for a fictitious year of 6 days, of which 3 are chosen to be representative. The optional post processing step *RR* is also shown. The last two rows depict synthetic time series which can be obtained from ordering representative periods (see Section 4.2).

continuous time steps t, here of hourly length, as illustrated in Fig. 2. When reporting results we convert the total number of hours into a number of days i.e. 768 h is reported as 32 days regardless of the number of time steps per period.

The algorithms and criteria for selecting the set $\mathcal{R}$ are what differentiate between TSA methods, and the many possibilities have led to the abundance of literature on the matter [7]. In this paper we restrict ourselves to two types of algorithms, optimization and clustering, and two criteria, approximating the full-year time series' values, or the distribution of these values, i.e. the time series' duration curves. Common to all methods are the input time series of electric load and VRES availability factors, which are normalized to lie between -1 and 1 as is common practice in the literature [33].

What sets these TSA methods apart is that all (save *ORDF*) produce an ordering of representative periods as well as selecting them, as illustrated by Fig. 3. This ordering can be thought of as a mapping of the representative periods to all periods (representative or not) in the year, and it is used in the novel GEPM formulations presented in Section 3.

The TSA methods used in this paper are described in the following Sections 2.1 and 2.2, and are summarized in Table 1. They were selected on the basis that they represent a variety of selection algorithms and criteria and they provide an ordering of periods in the year. Figs. 3 and 4 illustrate the selection and ordering of representative days and hours by said TSA methods.

2.1. Optimization based methods

Three different optimization based methods are used in this paper, namely the Optimal Representative Days Finder (*ORDF*), the Optimal Representative Days Orderer (*ORDO*) and the Optimal Representative Days Re Orderer (*RR*). By their nature, optimization methods should perform at least as well at approximating the original time series when compared to clustering algorithms, though in this paper their high computational cost restricted them to ordering days or weeks but not hours. Computational limitations also led to a two-step procedure (selection followed by ordering) being used in the case of *ORDF-RR*, as illustrated in Fig. 3.

The investigated optimization models all share the same underlying Mixed Integer Linear Programming structure shown below.

$$\min_{u_i, w_i, v_{ij}} \sum_{s \in S} W_s \cdot (W^{err} \cdot ts_s^{err} + (1 - W^{err}) \cdot dc_s^{err}) \tag{1}$$

$$\sum_{i \in \mathcal{P}} u_i \leq N_{rep} \tag{2}$$

$$\sum_{j \in \mathcal{P}} w_i = N_{total} \tag{3}$$

$$\sum_{j \in \mathcal{P}} v_{ij} = 1 \quad i \in \mathcal{P} \tag{4}$$

$$\sum_{i \in \mathcal{P}} v_{ij} = w_i \quad j \in \mathcal{P} \tag{5}$$

$$v_{ij} \leq u_i \quad i \in \mathcal{P}, \; j \in \mathcal{P} \tag{6}$$

$$w_i \leq u_i \cdot N_{total} \quad i \in \mathcal{P} \tag{7}$$

$$u_i \in \{0, 1\}, \quad w_i \in \mathbb{R}_0^+, \quad v_{ij} \in [0, 1] \tag{8}$$

This optimization problem has the two objectives of minimizing the squared difference between a) the full-year and aggregated time series' values, ts_s^{err}, and b) the full-year and aggregated duration curves, dc_s^{err}. These will be defined later when discussing the differences between *ORDF*, *ORDO*, and *RR*. The time series can be weighted by setting different values of W_s.

The variables in this problem are the binary selection of representative periods u_i, their real valued weights w_i and their mapping to non-representative periods in the year v_{ij}, a number between 0 and 1. The first Constraint (2) ensures that at least N_{rep} periods are selected as representative, while Constraint (3) ensures that the sum of the weights adds up to the total number of periods in the year N_{total}. Constraint (4) ensures that each period in the year is represented by another period, while Constraint (5) states that the number of *non-representative* periods mapped to a *representative* period is equal to that representative period's weight w. Constraints (6) and (7) ensure that a period can only represent other periods or have a non zero weight if said period has been selected as representative.

2.1.1. Optimal Representative Days Finder (ORDF)

In the case of *ORDF*, only the duration curve error is minimized ($W^{err} = 0$). This is done by discretizing the original load duration curve into a set of bins $\mathcal{B}$ and taking the squared difference between it and the aggregate duration curve for each bin:

$$dc_s^{err} = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} (dc_{sb}^{agg} - dc_{sb})^2 \tag{9}$$

The aggregate duration curve dc_{sb}^{agg} can be constructed from the representative period weights w_i as described in the original paper by [9].

This approach captures the distribution, i.e. the duration curve, of the time series, of which the correlation between these is only partially maintained due to days being selected.[5] When solving the *ORDF* problem, an optimality gap of less than 1% was never reached within 20 min using the Gurobi solver [34], compared to the several seconds required to solve the *ORDO* and *RR* models on a single core of an Intel Xeon Gold 6140 CPU at 2.30 GHz.

Since the duration curves hold no chronological information, it is not possible to obtain an ordering variable v_{ij} from this problem. To obtain a value for v_{ij} a second optimization problem is solved, *RR*.

2.1.2. Optimal Representative Days Re Orderer (RR)

The *RR* problem is inspired by [29], in which the selection variable u_i has been fixed leaving only the weights w_i and ordering v_{ij} to be found. This means that the set of representative periods, $\mathcal{R}$, is also defined. It has two weighted objective terms, since along with dc_s^{err} given in (9) ts_s^{err} is defined as follows:

$$ts_s^{err} = \sum_{i \in \mathcal{P}} \sum_{j \in \mathcal{R}} v_{ij} \cdot (X_{sjt} - X_{sit})^2 \qquad (10)$$

Where X_{sit} is the normalized time series.

In the paper by [29], the ordering in *RR* assigns a one-to-one mapping of *non*-representative periods to their representative counterparts. We relax this such that a non-representative day can be represented by a linear combination of representative days,[6] allowing for a better approximation of the original time series.

RR would ideally be solved without fixing u_i since any two-step optimization is inherently sub-optimal. However, computational tractability issues meant that it could only be solved with u_i fixed however, after which it could be solved within several seconds.

2.1.3. Optimal Representative Days Orderer (ORDO)

For the *ORDO* formulation, only the time series error is minimized ($W^{err} = 1$) which is defined as follows:

$$ts_s^{err} = \frac{1}{2 \cdot |\mathcal{P}| \cdot |\mathcal{T}|} \sum_{i \in \mathcal{P}, j \in \mathcal{P}} v_{ij} \cdot \sum_{t \in \mathcal{T}} (X_{sit} - X_{sjt})^2 \qquad (11)$$

For the above error to correct, v_{ij} must be binary i.e. the following constraint is added:

$$v_{ij} \in \{0, 1\} \qquad (12)$$

Inspired by the facility location problem [35], *ORDO* is not an entirely novel TSA method. As noted by [33], it is hierarchical clustering formulated as an optimization problem. However, to the best of the authors' knowledge, this problem has not been solved in the literature despite it taking less than 10 s to solve using the Gurobi solver [34].

Unlike *ORDF*, *ORDO* has the advantage of providing an ordering variable v_{ij}. By minimizing the time series error ts_s^{err} it implicitly captures the correlation between time series. The disadvantages compared to the two-step method of *ORDF-RR* is that it may be worse at reproducing the original duration curves,[7] and, since the ordering variable v_{ij} must be binary, it may also be worse at reproducing the original time series. In the results, the combination of *ORDO* with *RR* was compared with just *ORDO* to investigate this last point.

2.2. Clustering based methods

As noted by [33], clustering-based TSA methods can be seen as heuristics for solving optimization based problems. They should therefore perform worse at the task of approximating the original time series, though their low computational cost allows them to select and order hours as well as days or weeks.

Both clustering methods considered here, Hierarchical Clustering (*HC*) and Chronological Time Period Clustering (*CTPC*), are based on Ward's method for agglomerative hierarchical clustering [36], whose algorithm is presented below:

1. Set the number of clusters n to the total number of periods in the year N_{total}.
2. Determine the centroid $\bar{x}_{cst}$ of each cluster c:

$$\bar{x}_{cst} = \frac{1}{|\mathcal{P}_c| \cdot |\mathcal{T}|} \sum_{i \in \mathcal{P}_c} X_{sit} \qquad (13)$$

   Where $\mathcal{P}_c$ is the set of periods in cluster c.
3. Compute the dissimilarity $M_{c_1 c_2}$ between each pair of clusters c_1, c_2 following Ward's method:

$$M_{c_1 c_2} = \frac{2 \cdot |\mathcal{P}_{c_1}| \cdot |\mathcal{P}_{c_2}| \cdot |\mathcal{T}|}{|\mathcal{P}_{c_1}| + |\mathcal{P}_{c_1}|} \sum_{s \in S, t \in \mathcal{T}} W_s \cdot (\bar{x}_{c_1 st} - \bar{x}_{c_2 st})^2 \qquad (14)$$

4. Merge the two closest clusters according to the dissimilarity matrix.
5. Update the number of clusters $n = n - 1$.
6. If $n = N_{rep}$ then go to step 7. Else go to step 2.
7. Translate the results into the values for the variables u_i, w_i and v_{ij}:

   (a) Set u_i to be one for the medioid of each cluster, zero otherwise — these are the selected representative periods. The medioid is defined as the period with the minimum dissimilarity to the rest of periods in each cluster.
   (b) Set the weight of a representative period w_i equal to the number of periods in its cluster.
   (c) Populate v_{ij} as so: for each cluster c, identify the period j of v_{ij} corresponding to the cluster's medioid, then set $v_{ij} = 1 \ \forall i \in \mathcal{P}_c$.

2.2.1. Hierarchical Clustering (HC)

The Hierarchical Clustering (HC) algorithm (defined in the previous section) is particularly popular in the literature, see, e.g., [13,26–28]. In the context of long-term storage and ESOMs, it has the advantage over other clustering methods (such as k means) of providing an ordering of representative periods throughout the year. Note that in this paper *HCD* and *HCH* differentiate between HC applied to days of 24 h (Fig. 3) or single hours respectively (Fig. 4a). As with the optimization based methods, the post-processing *RR* optimization can be applied if days are selected and this combination is denoted as *HCD-RR* in the results.

2.2.2. Chronological Time Period Clustering (CTPC)

Chronological Time Period Clustering (CTPC) is the same as *HCH* save that only adjacent periods can be clustered together. CTPC can be seen as a sophisticated temporal resolution reduction method, where instead of merging a fixed number of time steps, the length of the time steps is variable. This is illustrated in Fig. 4b, in which it is clear that this method could smooth out short-term dynamics.

Similarly to temporal resolution reduction, continuity throughout the year is preserved and because of this it is the only method where there is no distinction between TSA and GEPM errors.

---

[5] In the original paper [9] a method for capturing the correlation between time series is also presented, though this was not implemented here.
[6] It is possible to do so because the energy balances are linear equations, and linear superposition means that any convex combination of the energy balances will also hold.
[7] The importance of doing so was the motivation for developing *ORDF* in [9].

(a) Hierarchical clustering for $|\mathcal{T}| = 1$, in which non-representative hours can be mapped to any representative hour in the year.

(b) Chronological time period clustering for $|\mathcal{T}| = 1$, in which non-representative hours can only be mapped to adjacent representative hours.

Fig. 4. Illustration of ordering hours in the year for a fictitious year containing 30 h. Top row: original time series, bottom row: synthetic time series (see Section 4.2). Note that ordering hours (as opposed to days) in the year is possible with clustering but not optimization based methods due to computational limitations.

Table 2
Summary of the reduced GEPM formulations compared in this paper. References indicates articles where a method has appeared before, though the list is not exhaustive. TSA methods shown in bold if they were used in conjunction with the GEPM in Section 6.2.

Name		Allows intra-period arbitrage?	Uses ordering parameter?	Compatibility with TSA	Reference
Unlinked Representative Days	*URD*	×	×	*ORDF* (-RR), *ORDO*(-RR), *HCD* (-RR)	[37]
Linked Representative Days	*LRD*	✓	×	*ORDF* (-RR), *ORDO*(-RR), *HCD* (-RR)	[8]
					[26]
Enhanced Representative Days	*ERD*	✓	✓	*ORDF* (-RR), *ORDO*(-RR), *HCD* (-RR)	[27]
					[29]
Enhanced Representative Hours	*ERH*	✓	✓	*ORDO*, *HCD*, *HCH*	[25]
Chronological Time Period Clustering	*CTPC-GEP*	✓	✓	*CTPC*	[28]

3. Generation Expansion Planning Models

Generation Expansion Planning Models are generally formulated as optimization problems which determine the investments required to satisfy electrical loads at minimum cost, typically subject to a carbon emissions budget or VRES penetration target [38]. They are a subset of Energy System Optimization Models, and due to their reduced scope (in comparison to ESOMs) they are more likely to include greater detail such as, e.g., unit commitment constraints of generators [17] or full-year time series [31]. While here we investigate long-term storage in GEPMs with reduced temporal scope, the findings are therefore more relevant for ESOMs that cover a large geographical and sectoral scope and several decades planning horizon (see, e.g., [39]), since reducing the temporal scope is essential to keep computation times of such models reasonable. For this reason, we formulate the GEPM as a linear optimization problem, as is typical of large scale ESOMs.

Of the GEPM formulations compared in this paper, one does not allow any arbitrage between periods (*URD*, Section 3.3), one does so but in a crude fashion (*LRD*, Section 3.4), while the rest represent the state of charge of storage over the entire year in some form (*ERH*, *ERD*, *CTPC-GEP*, Sections 3.5–3.7 respectively). All of these models share the same base structure outlined in Section 3.1 and a summary of these models is presented in Table 2.

Naturally, these GEPMs formulations employ the outputs of the TSA methods, namely $\mathcal{R} = \{i \in \mathcal{P} \mid u_i = 1\}$, W_j and V_{ij}, which are capitalized here to indicate that they are parameters in the GEPMs. Not all TSA methods are compatible with all GEPMs, hence Table 2 outlines which combinations are possible. Only one combination per GEPM was used for the results presented in Section 6.2, and these are highlighted in bold in Table 2.

3.1. Base model

All GEPMs formulations share the following objective function which minimizes the sum of annualized fixed (investment) costs, $C_g^{fix} \cdot k_g$ and $C_h^{fix} \cdot ec_h$, variable (operational) costs, $C_g^{var} \cdot q_{gjt}$ and $C_h^{var} \cdot (c_{hjt} + d_{hjt})$, and load shedding, $C^{shed} \cdot ls_{jt}$. Operational costs and load shedding

are weighted both by the TSA weights W_j and by the length of a time step $\tau_{j,t}$.

$$
\begin{aligned}
\sum_{g \in \mathcal{G}} & C_g^{fix} \cdot k_g + \sum_{h \in \mathcal{H}} C_h^{fix} \cdot ec_h \\
& + \sum_{j \in \mathcal{R}} W_j \cdot \tau_{j,t} \sum_{t \in \mathcal{T}} \left(\sum_{g \in \mathcal{G}} C_g^{var} \cdot q_{gjt} \right. \\
& \left. + \sum_{h \in \mathcal{H}} C_h^{var} \cdot (c_{hjt} + d_{hjt}) + C^{shed} \cdot ls_{jt} \right)
\end{aligned}
\tag{15}
$$

A power balance ensures that demand D_{jt} plus storage charging c_{hjt} is balanced by the sum of generation q_{gjt}, storage discharging d_{hjt} and load shedding ls_{jt}:

$$
\sum_{g \in \mathcal{G}} q_{gjt} + \sum_{h \in \mathcal{H}} (d_{hjt} - c_{hjt}) = D_{jt} - ls_{jt} \quad j \in \mathcal{R}, \, t \in \mathcal{T}
\tag{16}
$$

A VRES penetration target ensures that VRES technologies $g \in \mathcal{G}_R$ generate at least ϕ of the total demand:

$$
\sum_{g \in \mathcal{G}_R, j \in \mathcal{R}, t \in \mathcal{T}} W_j \cdot \tau_{j,t} \cdot q_{gjt} \geq \phi \cdot \sum_{j \in \mathcal{R}, t \in \mathcal{T}} D_{jt}
\tag{17}
$$

The generation, charging and discharging variables are limited by the installed capacities of generation and storage technologies k_g and ec_h multiplied by the availability factors AF_{gjt} and the fixed power to energy ratio PE_h respectively:

$$
0 \leq q_{gjt} \leq AF_{gjt} \cdot k_g \quad g \in \mathcal{G}, \, j \in \mathcal{R}, \, t \in \mathcal{T}
\tag{18}
$$

$$
0 \leq c_{hjt} \leq PE_h \cdot ec_h \quad h \in \mathcal{H}, \, j \in \mathcal{R}, \, t \in \mathcal{T}
\tag{19}
$$

$$
0 \leq d_{hjt} \leq PE_h \cdot ec_h \quad h \in \mathcal{H}, \, j \in \mathcal{R}, \, t \in \mathcal{T}
\tag{20}
$$

The GEPMs all share the same representation of intra-period storage, which must satisfy its own energy balance within a representative period:

$$
e_{hjt} = e_{hj}^0 + \tau_{j,t} \cdot \Delta e_{hjt} \quad h \in \mathcal{H}, \, j \in \mathcal{R}, \, t = \mathcal{T}_1
\tag{21}
$$

$$
e_{hjt} = e_{hjt-1} + \tau_{j,t} \cdot \Delta e_{hjt} \quad h \in \mathcal{H}, \, j \in \mathcal{R},
$$
$$
t \in \mathcal{T}_{2:end}
\tag{22}
$$
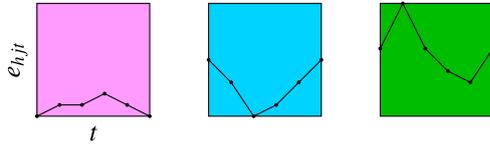
Fig. 5. Illustration of storage operation in *URD* for 3 representative periods. No arbitrage is possible between representative periods and the state of charge at the end of a period is constrained to be greater than or equal to that at the beginning of the period.

Fig. 6. Illustration of storage operation in *LRD* for 3 representative periods. Inter-period arbitrage is possible since the state of charge at the end of one period is equal to the state of charge at the beginning of the next.

where Δe_{hjt} is defined as follows with η_h the roundtrip efficiency of storage:

$$\Delta e_{hjt} = (\sqrt{\eta_h} \cdot c_{hjt} - d_{hjt}/\sqrt{\eta_h})$$

Finally the state of charge of storage is constrained to never exceed the energy capacity of the storage technology:

$$0 \le e^0_{hj} \le ec_h \quad h \in \mathcal{H}, \ j \in \mathcal{R} \tag{23}$$

$$0 \le e_{hjt} \le ec_h \quad h \in \mathcal{H}, \ j \in \mathcal{R}, \ t \in \mathcal{T} \tag{24}$$

3.2. Full year (FY-GEP)

This GEPM is used as the reference case to which the reduced GEPMs are compared. The formulation employs a continuous hourly time series of length $|\mathcal{T}| = 8760$ hours, implying that $|\mathcal{R}| = 1$ and $W_j = 1 \ \forall j \in \mathcal{R}$. In addition a cyclic constraint is defined for all storage technologies to ensure that there is no "free" energy:

$$e_{hjt_{end}} \ge e^0_{hj} \quad h \in \mathcal{H}, \ j \in \mathcal{R} \tag{25}$$

The optimization problem is therefore written as:

min (15)

s.t.

(16)–(25)

3.3. Unlinked Representative Days (URD)

Unlinked Representative Days (URD) does not allow inter-period arbitrage, and hence storage can only arbitrage over the length of a period, in this case a day, as illustrated in Fig. 5. This model has commonly been used to benchmark other storage formulations in the past, see for example [26]. It is the same as the *FY-GEP* model except that representative periods are used instead of the full-year time series, i.e. $|\mathcal{R}| > 1$ and $W_j \ge 1 \ \forall j \in \mathcal{R}$. It is written as:

min (15)

s.t.

(16)–(25)

3.4. Linked Representative Days (LRD)

The Linked Representative Days formulation allows inter-period storage by simply linking the state of charge from one period to the next as illustrated in Fig. 6.

The initial state of charge e^0_{hj} for inter-period storage $h \in \mathcal{H}$ is therefore defined as follows:

$$e^0_{hj} = e_{hj-1t_{end}} \quad h \in \mathcal{H}, \ j \in \mathcal{R}_{2:end} \tag{26}$$

The cyclic condition is defined as:

$$\sum_{j \in \mathcal{R}, t \in \mathcal{T}} \Delta e_{hjt} \cdot W_j \cdot \tau_{j,t} \ge 0 \quad h \in \mathcal{H} \tag{27}$$
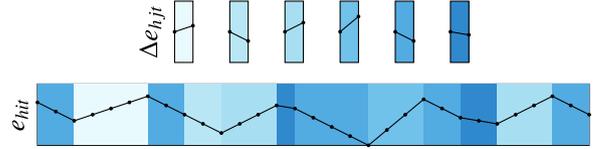
Fig. 7. Illustration of storage representation used in *ERH*. Here 6 representative periods are shown for a year containing 30 periods in total, with colors mapped to representative periods. Inter-period arbitrage is possible since the state of charge is defined for the entire year. Top row: change in the state of charge over a representative period. Bottom row: the resulting evolution of the state of charge.

The model is written as:

min (15)

s.t.

(16)–(24), (26)–(27)

3.5. Enhanced Representative Hours (ERH)

First proposed by [25], the Enhanced Representative Hours (ERH) GEPM has since been picked up and modified by [26,27,29]. In this formulation, the charging and discharging decisions of *non-representative* days are kept the same as those of their *representative* counterparts.[8] This is illustrated in Fig. 7.

This formulation uses a mapping of *non*-representative variables to representative ones, denoted by $V(i) \longrightarrow j$. This requires that V_{ij} contains only ones and zeros and each $\sum_{j \in \mathcal{R}} V_{ij} = 1$.[9]

For inter-period storage, the state of charge dynamics are extended to cover the entire year using the previously mentioned mapping:

$$e_{hit_1} = e^0_{hi} + \tau_{j,t} \cdot \Delta e_{hV(i)t} \quad h \in \mathcal{H}_L, \ i \in \mathcal{P} \setminus \mathcal{R} \tag{28}$$

$$e_{hit} = e_{hit-1} + \tau_{j,t} \cdot \Delta e_{hV(i)t} \quad h \in \mathcal{H}_L, \ i \in \mathcal{P} \setminus \mathcal{R},$$
$$t \in \mathcal{T}_{2:end} \tag{29}$$

$$e^0_{hi} = e_{hi-1t_{end}} \quad h \in \mathcal{H}_L, \ i \in \mathcal{P}_{2:end} \tag{30}$$

The state of charge must be within bounds at all times[10]:

$$0 \le e_{hit} \le ec_h \quad h \in \mathcal{H}_L, \ i \in \mathcal{P}, \ t \in \mathcal{T} \tag{31}$$

A cyclic constraint is then imposed over the entire year:

$$e_{hi_{end},t_{end}} \ge e^0_{hi_1} \quad h \in \mathcal{H}_L \tag{32}$$

The model is written as:

min (15)

---

[8] This was the motivation for ordering representative periods since this provides a mapping from non-representative periods to their representative counterparts.

[9] A theoretically equivalent alternative to this mapping would be to take the dot product with the ordering parameter V_{ij}.

[10] [40] proposed a variant on this in which the state of charge is only checked at certain intervals. In a later publication however, the authors write that this method leads to unrealistic short-term storage operation compare to the Enhanced Representative Days (ERD) formulation [27].
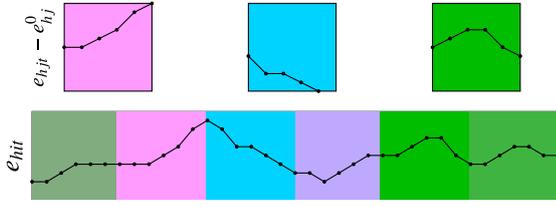
Fig. 8. Storage representation in *ERD* for a fictitious year of 6 periods of which 3 are representative. Top row: evolution in the state of charge over a representative day relative to the state of charge at the beginning of that day. Bottom row: evolution in the state of charge over full-year*. In the depiction above some non-representative days are linear combinations of representative days, hence the evolution in the state of charge is also a linear combination of these. * e_{hit} is not in fact a variable in *ERD*, but it can be computed ex post as $e_{hit} = e_{hi_0}^0 + \sum_{i \in \mathcal{P}, j \in \mathcal{R}, t \in \mathcal{T}} V_{ij} \cdot \Delta e_{hjt}$.

s.t.

(16)–(24), (28)–(32)

3.6. Enhanced Representative Days (ERD)

Recall that in *ERH* the state of charge is defined for every hour or time step in the year. In contrast, Enhanced Representative Days (ERD) reduces the number of variables and constraints such that the problem scales with the number of days in the year, i.e. the state of charge is checked for each day and not each time step of the year. The advantage of this is that the linear combination of representative days produced by the *RR* TSA can be used, which proved intractable in the case of *ERH*. A possible disadvantage is that when used in conjunction with TSA methods which order hours in the year the problem size increases in comparison to the full-year model. The functioning of *ERD* is illustrated in Fig. 8 (see [29] for additional illustrations.).

The energy change over a representative period is written as:

$$\Delta e_{hj} = \sum_{t \in \mathcal{T}} \Delta e_{hjt} \quad h \in \mathcal{H}_L, \ j \in \mathcal{R} \tag{33}$$

The base state of charge e_{hi}^0 for each period in the year can then be defined:

$$e_{hi}^0 = e_{hi-1}^0 + \sum_{j \in \mathcal{R}} V_{ij} \cdot \Delta e_{hj} \quad h \in \mathcal{H}_L, \ i \in \mathcal{P}_{2:end} \tag{34}$$

The maximum positive and negative deviations from the base state of charge, Δe_{hi}^{max} and Δe_{hi}^{min} respectively, are defined for representative periods:

$$0 \leq \Delta e_{hj}^{max} \geq e_{hjt} - e_{hj}^0 \quad h \in \mathcal{H}_L, \ j \in \mathcal{R}, t \in \mathcal{T} \tag{35}$$

$$0 \leq \Delta e_{hj}^{min} \geq e_{hj}^0 - e_{hjt} \quad h \in \mathcal{H}_L, \ j \in \mathcal{R}, t \in \mathcal{T} \tag{36}$$

This allows definition of the maximum positive and negative deviations for *non*-representative periods:

$$\Delta e_{hi}^{max} = \sum_{j \in \mathcal{R}} V_{ij} \cdot \Delta e_{hj}^{max} \quad h \in \mathcal{H}_L, \ i \in \mathcal{P} \setminus \mathcal{R} \tag{37}$$

$$\Delta e_{hi}^{min} = \sum_{j \in \mathcal{R}} V_{ij} \cdot \Delta e_{hj}^{min} \quad h \in \mathcal{H}_L, \ i \in \mathcal{P} \setminus \mathcal{R} \tag{38}$$

With this the state of charge limits can be imposed for all periods:

$$e_{hi}^0 + \Delta e_{hi}^{max} \leq ec_h \quad h \in \mathcal{H}_L, \ i \in \mathcal{P} \tag{39}$$

$$e_{hi}^0 - \Delta e_{hi}^{min} \geq 0 \quad h \in \mathcal{H}_L, \ i \in \mathcal{P} \tag{40}$$

A cyclic constraint is imposed.

$$e_{hi_0}^0 \geq e_{hi_{end}}^0 \sum_{j \in \mathcal{R}, t \in \mathcal{T}} V_{ij} \cdot \Delta e_{hjt} \tag{41}$$

The model is written as:

min (15)

s.t.

(16)–(24), (33)–(41)

3.7. Chronological Time Period Clustering (CTPC-GEP)

CTPC-GEP is the GEPM compatible with the *CTPC* TSA. It can be formulated as a special case of the *FY-GEP* model in which $|\mathcal{R}| = 1$ and the value of $\tau_{j,t}$ in the GEPM is equal to the weight of the representative period w_i as obtained by the TSA method (the weight W_j in the GEPM is set to 1). This illustrates how *CTPC* can be seen as a sophisticated temporal resolution reduction method.

4. Performance evaluation

4.1. Investigation procedure

A contribution of this paper is to distinguish between errors due to the TSA method and the GEPM formulation. For models which do not consider any inter-temporal constraints, there is no distinction to be made, but this is not the case here since storage technologies have a state of charge which evolves over the entire optimization horizon. To distinguish between these errors, full-year models (Section 3.2) are run on synthetic time series obtained from the TSA methods. These results can then be compared with the results from running the GEPM models with a reduced temporal representation.

In summary, the investigate procedure followed in this paper is as follows:

1. Select and order representative periods (days or hours) using the TSA methods described in Section 2.
2. Construct synthetic time series from the ordered periods and run a full-year GEPM on them. Comparing these results with the results obtained from a full-year GEPM run on the original time series yields the error introduced purely from the TSA methods.
3. Run the reduced GEPM models described in Section 3. Comparing these results with the results obtained from a full-year GEPM run on the original time series yields the total error arising from the TSA method and the GEPM model.

4.2. Creating synthetic time series

To obtain a synthetic time series, we take the product of an ordering V_{ij} obtained from an aggregation method a, V_{ij}^a, and the original time series:

$$\tilde{X}_{sit}^a = \sum_{j \in \mathcal{R}} V_{ij}^a \cdot X_{sjt} \tag{42}$$

For an illustration of the output of this process see Fig. 3 rows 3 and 4 as well as Figs. 4a and 4b.

4.3. Performance metrics

Since Section 6 presents results for different test systems, the metrics here have superscripts y to indicate that these are calculated per test system.

To compare time series aggregation methods, we use the mean time series error $MTSE$ of a TSA method a and test system y:

$$MTSE^{a,y} = \sum_{s \in S, i \in \mathcal{P}, t \in \mathcal{T}} \frac{W_s}{|\mathcal{T}| \cdot |\mathcal{P}|} \cdot (\tilde{X}_{sit}^a - X_{sit}) \tag{43}$$

Very different power systems can have very similar total costs, hence we introduce the normalized distance error $L1$ (adapted from [19]) as a more informative alternative to the total cost error. If the objective function of an optimization problem m is given by $\vec{a} \cdot \vec{z}^m$ where $\vec{a}$ and
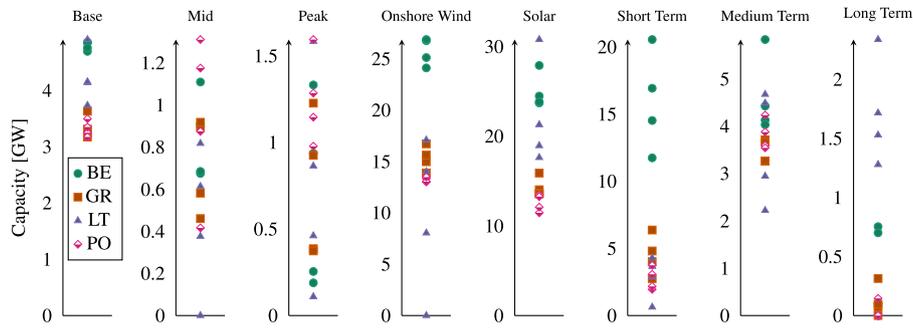
Fig. 9. Installed capacities for *FY-GEP* model. These are grouped by country: Belgium (BE), Greece (GR), Lithuania (LT) and Portugal (PT). Capacities are scaled such that each test system has a peak demand of 10 GW.

Table 3
Summary of generation and storage technology parameters. Short Term, Medium Term and Long Term characteristics are inspired by battery, pumped hydro and power to gas technologies respectively.

Technology	C^{fix} [€/kW]	C^{var} [€/MWh]	AF_{gjt} [–]
Base	82.120	54	0.85
Mid	50.910	90	0.85
Peak	46.400	112	0.85
Onshore Wind	66.960	0	–
Solar	42.270	0	–

	C^{fix} [€/ kWh]	C^{var} [€/ MWh]	PE_h [1/h]	η_h
Short Term	4.44	0	1	0.9
Medium Term	15.52	5	16	0.75
Long Term	0.176	0	1000	0.3

$\vec{z}$ are the cost and decision variable vectors respectively, then $L1^{m,y}$ is defined as:

$$L1^{m,y} = \frac{|\vec{a}^y \cdot (\vec{z}^{m,y} - \vec{z}^{FY,y})|}{TC^{FY,y}} \tag{44}$$

where the superscript *FY* denotes the full-year solution.

This metric can be computed for a full-year GEPM run on a synthetic time series or a reduced GEPM, denoted by $L1^{a,y}$ and $L1^{m,y}$ respectively. To compute $L1^{m,y}$, the reduced GEPM models were rerun with capacities fixed to ensure that m and FY have the same number of decision variables. Since doing this could make the results sensitive to the amount of load shedding, similarly to [21] it was assumed that any load shedding was covered by installing peaking capacity.

Capacity mixes are compared using the normalized capacity mix error $NCME$ of a technology g, GEPM m and test system y:

$$NCME_g^{m,y} = \frac{k_g^{m,y} - k_g^{FY,y}}{k_g^{FY,y}} \tag{45}$$

$$NCME_g^{m,y} = PE_h \cdot \frac{ec_h^{m,y} - ec_h^{FY,y}}{ec_h^{FY,y}} \tag{46}$$

Note that the above metric is not symmetric: if 1 GW of a technology is installed in the full-year GEPM, then 1 GW in the reduced GEPM gives a value of 0; 2 GW a value of 1; and 0 GW a value of −1. If $k_g^{FY,y}$ or $ec_h^{FY,y}$ was 0, then $NCME_g^{m,y}$ was either 0 or infinity. Median values, along with maxima and minima, were reported to avoid these cases skewing results.

5. Test systems and data

Unless stated otherwise, the following parameters were used. The length of a time step was set to one hour and a number of representative periods equivalent to 5, 10, 32, and 128 days were chosen. 5 and 10 representative days is typical of large ESOMs while 128 should be enough days to accurately represent the full-year, with 32 days lying between the two extremes. For the optimization based TSA methods,

equal weights were given to both the time series and duration curve error, i.e. $W^{err} = 0.5$. Time series were equally weighted, i.e. $W_s = 1 \ \forall s \in S$. A VRES penetration target of 90% was set, to provide conditions favorable to long-term storage investment [24]. The cost of load shedding C^{shed} was set to 10,000 €/MWh. The investment problems are greenfield with all technology capacities as decision variables. The values of important technical and cost-related parameters can be found in Table 3.

[20] notes that the performance of temporal aggregation techniques can differ significantly from one system to another. To mitigate this, load, solar, and onshore wind time series for Belgium, Greece, Lithuania, and Portugal (chosen for small size and data availability) for 4 years between 2015 and 2019 were used to increase the robustness of the results. These were obtained from the ENTSO-E transparency platform, with VRES availability factors calculated by normalizing generation by installed capacity [41].

The optimal installed capacities when using the full-year time series are shown in Fig. 9, where capacities have been scaled such that each test system has a peak demand of 10 GW. The spread in the results for different weather years can be contrasted with the capacity mix errors in Section 6. In 5 test systems Long Term storage was not invested in while in one test system both Solar and Mid were not invested in. This has consequences for the interpretation of the $NCME$ metric.

All code and data can be found at https://gitlab.kuleuven.be/u0128861/rep-days-4-storage.

6. Results

We first compare the TSA methods using synthetic time series in Section 6.1. We then build on these results in order to compare GEPM formulations in Section 6.2 while distinguishing between TSA and GEPM errors.

6.1. Comparison of Time Series Aggregation methods

Fig. 10 shows the $L1^{a,y}$ error plotted against $MTSE^{a,y}$, which measure the performance of the TSA methods in approximating the full-year GEPM solution and the time series values respectively. The top right portion of Fig. 10 is a scatter plot of all the results, while the bottom right and top left split the $MTSE^{a,y}$ and $L1^{a,y}$ results by number of representative days selected using the median, maximum and minimum values for each TSA across all test systems.

It is clear from the top right plot that these metrics are correlated, though there is a significant spread in the results, particularly for a low number of representative days. In some cases, for example in large scale ESOMs, the time series error or related metrics are the only indications a modeler may have on how well their TSA method performs, but Fig. 10 clearly illustrates that it is difficult to determine a priori how well a TSA performs without running a planning model on the resulting time series. This is why improving TSA methods is challenging, and it was the motivation for the importance subsampling
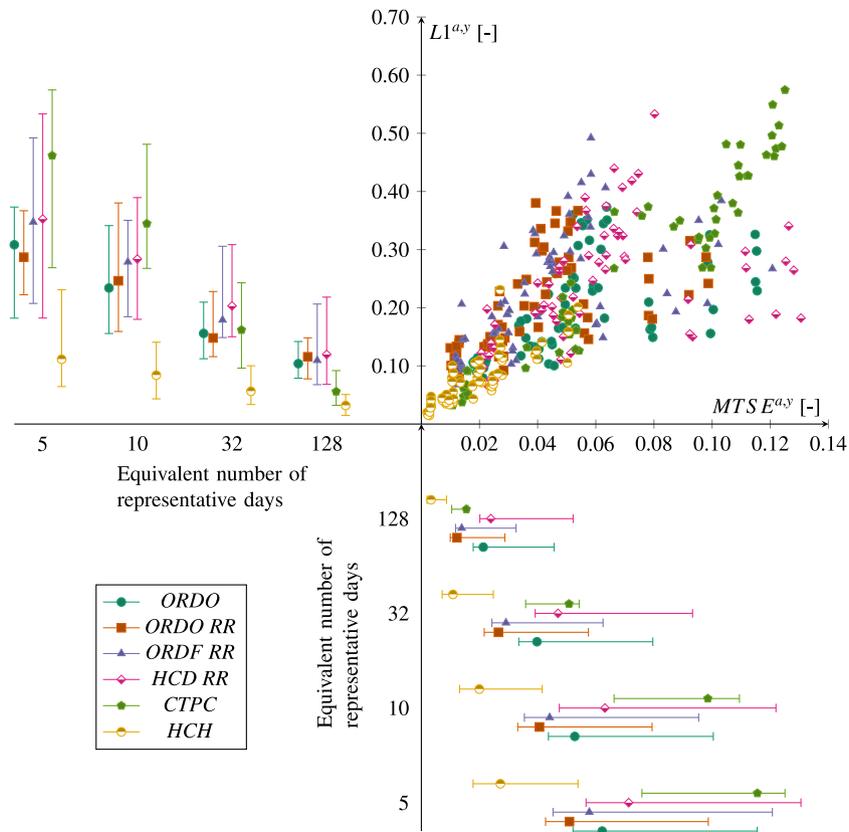
Fig. 10. $L1^{m,y}$ error obtained from running a full-year GEPM on synthetic time series versus the mean time series error $MTSE^{a,y}$ of TSA methods. The dots shown are median values with error bars indicating maximum and minimum values. The trend suggests that time series error is only partially correlated with GEPM error, illustrating the difficulty in improving TSA methods based solely on the original time series.
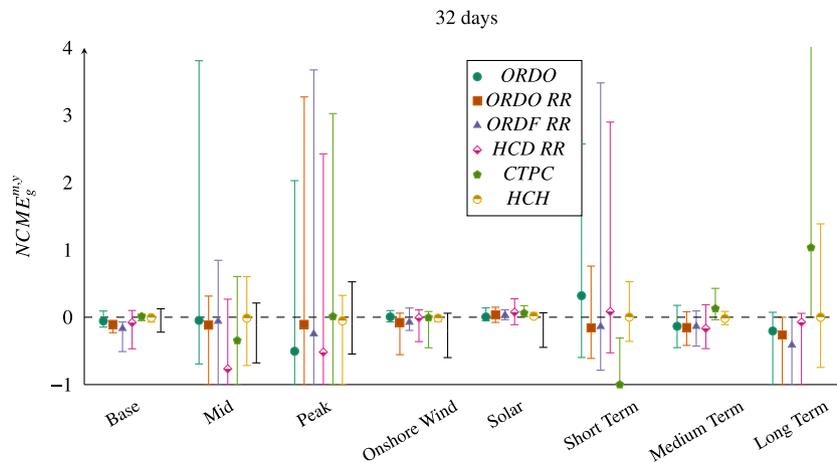
Fig. 11. Capacity mix errors for 32 equivalent number of representative days for different TSA methods. Dots shown are median values with error bars indicating maximum and minimum values. Solid black error bars indicate results across all TSA methods for when storage investment was not allowed. It can be seen that when storage is included, deviations by a factor of 2 or more from the benchmark capacities are possible in the case of Mid, Peak, Short and Long Term storage.

based approach of [21] which is not readily applicable to models which include long-term storage.

The spread in results in Fig. 10 also highlights the importance of testing a TSA on multiple test systems, since it suggests that relying on only a single test system could have lead to erroneous conclusions of one method performing better than another. As it happens, the top left quadrant of Fig. 10 suggests TSAs which select days perform similarly despite the difference in approaches. Comparing *ORDO* and *ORDO-RR* also reveals that the additional post-processing step *RR* has little influence on the results.

It is perhaps surprising to see such a high sensitivity of GEPM solutions to the temporal scope. This is addressed in [19], who notes that the optimal solution to linear GEPMs such as the one used here are sensitive to peak pricing or scarcity hours, in which the peaking technology (here Peak) recuperates its fixed costs. The availability of VRES and storage during these hours can have a large impact on the optimal capacity, hence the variance observed in Fig. 12. An equivalent explanation is that linear GEPMs can be sensitive to changes in input as discussed in [21] for example.
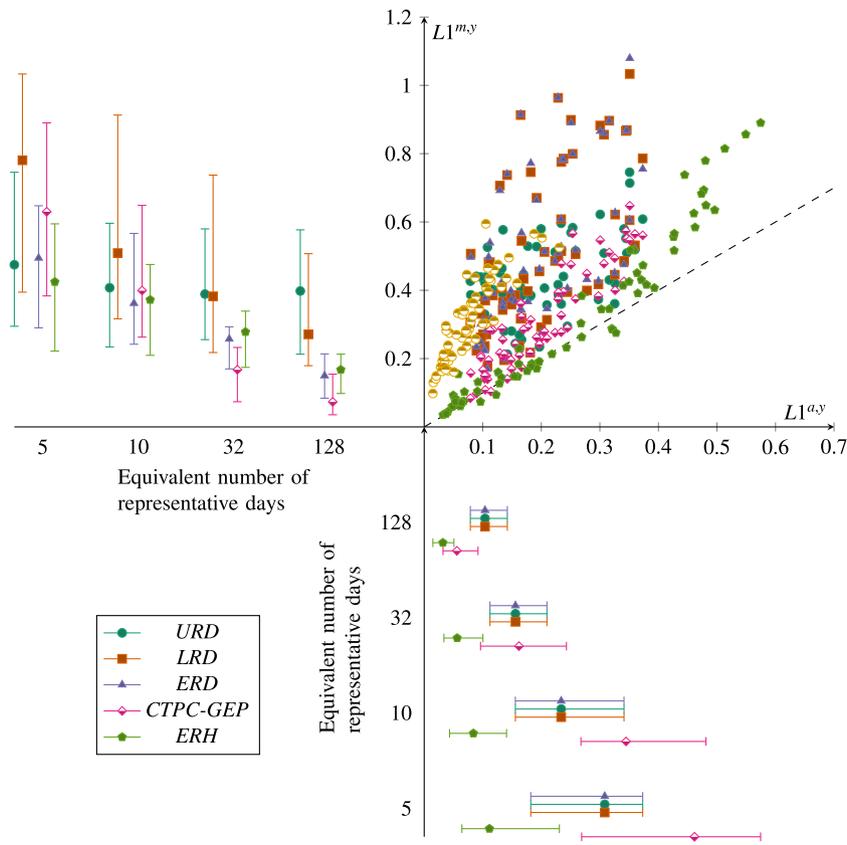
Fig. 12. TSA method vs GEPM L1 norm error. For the latter, GEPM models were rerun with additional peak capacity to cover any load shedding events. Dots shown are median values with error bars indicating maximum and minimum values. It appears that the reduced GEPM formulations exacerbate errors in approximating the full-year GEPM solution, since most data points lie above the dashed 45 degree line.
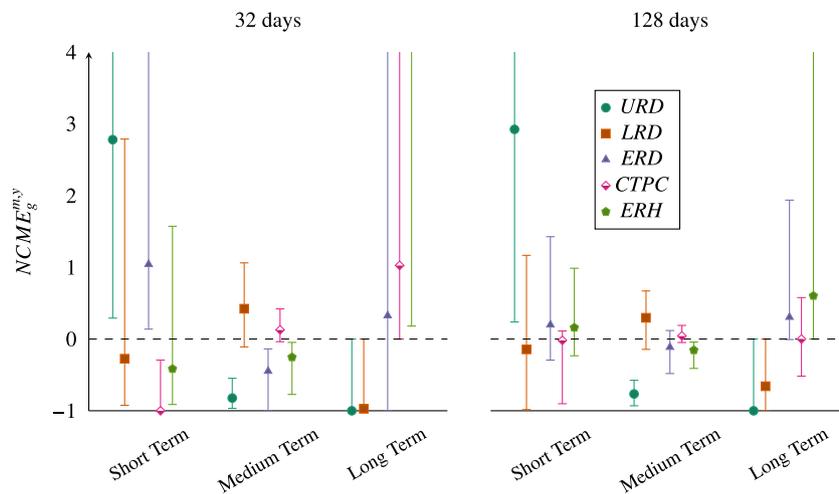
Fig. 13. Storage capacity mix errors for 32 and 128 equivalent number of representative days for different GEPM formulations. Dots shown are median values with error bars indicating maximum and minimum values. Though only storage technologies are shown here, all technologies could be invested in. Significant over or underinvestment in storage is clearly possible for 32 days irrespective of the models used.

Of the TSA methods compared, the *HCH* and *CTPC* stand out in Fig. 12. Since these select hours they can perform significantly better than TSA methods which select days. This is particularly true of *HCH*. This result is in contrast with [13], since methods which select hours performed worse in their comparison. Our results suggest that this is due to the GEPM and not the TSA itself (see Section 6.2)

With regards to the *CTPC* formulation, Fig. 12 shows that this method performs poorly for 5 and 10 representative days, but it improves drastically as the number of days increases. This is unsurprising since for 5 equivalent days, e.g., the whole year would be represented by 120 continuous time steps which leads to significant temporal smoothing of VRES availability. Previous literature has shown that this leads to underestimating the difficulty of VRES integration [20].

Fig. 11 shows the technology capacity mix errors for 32 representative days with and without the possibility of storage investment. Dots represent median values with error bars denoting the minimum and maximum capacity mix error observed across the 16 test systems. Though the extreme minimum and maximum errors may be due to the choice of error metric, it can be seen from this figure that allowing storage investment leads to greater capacity mix errors for conventional technologies while partially reducing it for VRES.

In addition Fig. 11 further highlights the sensitivity of GEPMs to the temporal scope. While median values across test systems are generally close to 0, particularly for Base, Onshore Wind, Solar and Medium Term, for individual test systems these can be off by factors of 2 or more.

Some of the TSAs also exhibit biases. In agreement with [28], the *CTPC* formulation favors Long Term storage over Short Term storage. This is because the method retains long and mid-term dynamics of time series but not the hour to hour variations, hence the synthetic time series does not signal the need for energy-constrained storage which arbitrages over a day.

For TSAs methods which select days, it could be argued that a bias exists against Long Term storage. While the median values of $NCME_g^{a,y}$ shown here are close to 0, this includes cases in which Long Term storage was not invested in. If these are omitted, the median values all lie below -0.5, i.e. underinvestment by a factor of 2.

6.2. Comparison of GEPM formulations

Similarly to Figs. 10, 12 shows the TSA error plotted against the combined TSA and GEPM error. The TSA error is obtained by computing the $L1$ error for a full-year GEPM run on a synthetic time series. The combined error is obtained by running a reduced GEPM on the representative periods selected and ordered by the corresponding TSA method (see bold entries in Table 2). As in Fig. 10, there is a correlation between the two errors but also a significant spread. It might be expected that this spread would be limited to above the 45-degree line, i.e. that TSA and GEPM errors would compound each other, and this is mostly the case.

Similarly to Figs. 11, 13 shows the storage technology capacity mix errors for 32 and 128 representative days. Focusing first on the results for 32 days, it is unsurprising that *URD* does not invest in Long Term storage and overinvests in Short Term storage, since it cannot arbitrage between representative periods which makes Long Term storage unattractive. This is also the case for the *LRD* model, perhaps because the few constraints on inter-period arbitrage lead to Short Term and Medium Term storage filling the role of Long Term storage.

ERD overinvests in both Short Term and Long Term storage and with a high variance on these results. This is particularly true when its corresponding TSA, *ORDO*, is used as a reference.

The *CTPC-GEP* formulation exhibits the same biases as in Section 6.1 since there is no distinction between TSA and GEPM induced error.[11] Interestingly *ERH* performs worse than it's corresponding TSA, *HCH*, in addition to being slower to solve than even than the full-year model in some cases (see Fig. 14). It is thus unable to leverage the better performance of the *HCH* TSA method in comparison to the methods which select days.

Turning to the results for 128 days in Fig. 13, it seems that for more sophisticated GEPMs such as *ERD* and *CTPC* the technology capacities converge to the correct solutions.[12] It is therefore possible to obtain

Fig. 14. Mean optimization times for GEPM formulations recorded on a single core of a Intel Xeon Gold 6140 CPU at 2.30 GHz. All GEPM formulations lead to a reduction in optimization time save for *ERH*.

accurate technology capacities for GEPMs with a reduced temporal scope but a high level of temporal detail is still required, in this case around a third of the full year. Fig. 13 shows that even for such a large amount of days, individual results could however still be incorrect.

For both 32 and 128 days, the median $NCME_g^{m,y}$ values for Medium Term storage are generally closer to 0 than for those of Short Term and Long Term storage. This suggests that storage technologies which have a more balanced power to energy ratio (compared to the extremes of Short Term and Long Term storage) may be less sensitive to the GEPM formulation.

Fig. 14 shows the mean optimization times for the various GEPM formulations. All formulations save *ERH* reduce this time by a factor of 10–1000 with respect to the full-year depending on the number of days. It seems that using a period length of an hour for the *ERH* formulation is counterproductive in terms of computational efficiency rendering it largely useless, though the results of [25] indicate that the computational time would likely be similar to *ERD* if days instead of hours were used.

7. Discussion

7.1. Limitations of the study

There are several limitations to the present study which means the findings may not be universally applicable. The first is the sensitivity of the GEPM models used. Due to their linear nature and few constraints, these are particularly sensitive to any changes in input parameters.[13] As noted by [19], this sensitivity comes from scarcity time step(s), i.e. time steps where a unit of extra demand requires investment in a unit of capacity. Since VRES have a time-dependent availability factor, the high VRES penetration target adds to this sensitivity because scarcity hours are highly uncoupled from the load. For a lower penetration target and more constrained model, e.g., brownfield with limits on total installed capacity, the variance of the results could be less pronounced.

The second and related issue is that in this work the driver for installing Long Term storage is to minimize electricity system costs. This may not be the case in practice, since other drivers could exist, for example the coupling of energy sectors such as heating or industry could instead drive investments in power to gas.

7.2. Recommendations to modelers

In light of these results, what is an ESOM[14] modeler to do? Ideally, the entire year's worth of time series should be used, given the possibility of introducing errors or biases in resulting capacity mixes. Previous

---

[11] The alert reader will have noticed that this is not entirely true, since the *CTPC-GEP* results do not lie on the 45 degree line in Fig. 12. This is because $L1^{m,y}$ was obtained by running an operational model on the capacity mixes resulting from each GEPM.

[12] *LRD* should also theoretically converge to the full-year solution as the number of periods increases.

[13] This is well known and has motivated the use of Methods to Generate Alternatives, as used by, e.g., [42].

[14] Recall that the term ESOMs is used here to denote both ESOMs and GEPMs unless noted otherwise.

research e.g from [17,18] suggests that this should take precedent over other modeling details such as technical constraints on thermal generators. If that is not possible, then both *CTPC-GEP* and *ERD* are viable ways of reducing the temporal scope of a GEPM while preserving the value of long-term storage.[15] If long-term storage is not included, then *URD* is adequate, as has been shown elsewhere [26].

Two questions remain, namely how many representative periods to include and which method to choose. For the former, the simple answer is "as much as is computationally tractable (given that, e.g., multiple scenarios are to be investigated)". As we have shown, omitting temporal detail while still accurately valuing generation and storage technologies is a non-trivial task for high VRES scenarios, and so including temporal detail should be a priority.

The level of temporal detail also affects the choice of *CTPC-GEP* or *ERD*, with the former being particularly unsuitable for a low number of representative time steps. It is however simpler to implement in comparison to *ERD*, particularly for models which can handle variable duration time series. The focus of the study could also inform this decision — if it is solely on longer-term storage technologies then *CTPC-GEP* might be preferred.

There are cases where it may not be feasible to include enough temporal detail to make these GEPM formulations viable. This is particularly true of ESOMs with a large sectoral scope, and so from henceforth, we will refer specifically to these and not GEPMs. These are often used for energy transition studies, where it is particularly important to avoid errors or biases in installed capacities. Four possible solutions are identified: improving existing TSA methods and GEPM formulations to avoid errors; identifying and characterizing errors and biases in a capacity mix, possibly to correct these; omitting TSA and using decomposition methods to solve the full model; finally using multiple models to capture the effects of including a high level of temporal detail. A non-exhaustive discussion of these is presented below.

With regards to improving existing TSA methods and GEPM formulations, from our experience in attempting this, it appears that any new method still requires a significant amount of temporal detail to be accurate. In other words, slight alterations to any of the formulations presented here are unlikely to prove fruitful, since there will still be errors intrinsic to the reduction of the temporal scope which a storage formulation would have to overcome. This is even more evident when considering the many demand and weather years which should ideally be considered in planning exercises [20,21,32].

Regarding the identification and characterization of capacity mix errors, perhaps the simplest approach is to run an operational model on a power system determined by an ESOM, sometimes referred to as "soft linking" (see e.g. [43,44]). While good practice, this will merely identify whether the power system is able to achieve its reliability or environmental targets while providing little information on biases in the capacity mix. Even if these targets are met, it is not straightforward to determine whether another power system configuration would also have done so but at a lower cost.

Another alternative would be to use Monte Carlo error analysis, in which an ESOM is rerun multiple times with different samples of uncertain parameters (see, e.g., [45]). A modified version of this approach is presented by [37] for the specific case where the uncertain parameters are the representative periods chosen, which could be modified also to include the ordering of said periods. Of the GEPM formulations presented here, only *ERD* appears to be a viable candidate for this, since the other formulations have consistent biases for or against long-term storage investment. This has not been done before however, and so should be the subject of future research.

If one accepts the need for a high level of temporal detail, then decomposition methods could provide a solution. In particular, parallelizing such methods could allow for more temporal detail to be included while keeping computation times acceptably low. Examples of this can be found in [46–49].

The last possibility discussed is adopting a multi-model approach, which can be done in various ways. The previously mentioned soft linking of an operational model to an ESOM could be satisfactory if the shortcomings of the ESOM (such as long-term storage modeling) are taken into account. An example of this is [32], where an ESOM is used to reduce the number of investment decisions in a subsequent GEPM to just peaking units, storage and transmission, thereby allowing for 10 demand and weather years to be included. Other examples include [50,51]. The capacity mixes evaluated by the operational model need not even come from an ESOM - they could be (judiciously) set by the user for example, which is the approach taken by the EnergyPLAN model [52].

8. Conclusion

Reducing the temporal scope of ESOMs is commonly used to reduce the computation time of such models, yet incorporating mid and long-term storage technologies in this context is difficult due to the loss of chronology and inter-temporal constraints. In this paper we compared novel TSA methods and GEPM formulations which allow for long-term storage for the first time in an investment setting, distinguishing between errors that arose from the TSA method or the GEPM formulation using the novel concept of synthetic time series. We compared TSA methods by varying the equivalent number of representative days and then running a full-year GEPM on the aforementioned synthetic time series. The results were compared in terms of time series, L1 and capacity mix error (see Section 4.3) for 16 test systems (see Fig. 1). The GEPM formulations were run with the selected and ordered representative periods and then similarly compared in terms of L1 and capacity mix error. The primary results of this study are:

- TSA methods which selected days all performed similarly in terms of $L1^{a,y}$ despite differing performance in approximating the full-year time series. This highlights the difficulty of improving TSA methods, since approximating the time series is merely correlated with approximating the full-year GEPM solution.
- Methods which selected hours instead of days were however appreciably different - *HCH* consistently outperformed other methods in terms of $L1^{a,y}$ while *CTPC* did so but only for 128 days. For 32 representative days, *CTPC* exhibited a bias against short-term storage and towards long-term storage.
- Regarding GEPM formulations, *URD* and *LRD* were found to underinvest in long-term storage.
- *ERH* was unable to leverage the better performance of its corresponding TSA method, *HCH*, since it was slower to solve than the full-year model for more than 10 representative days. .
- The *CTPC-GEP* and *ERD* models converged to the correct capacities for a 128 representative days while reducing computation time. For 32 days, potential errors and biases, such as *ERD* consistently overinvesting in short and long-term storage, could prove problematic when drawing conclusions on the results.

The present study has two primary limitations. The first is that only one type of GEPM was used for comparison, which, due to the linear formulation of the model, may be particularly sensitive to changes in input data. The second is that the driver for installing long-term storage in our case studies was reducing power system costs, whereas in practice other factors, such as sector integration, may be the primary incentive for such technologies. Taken together, these limitations mean that the results listed above may be case-specific and that in practice fewer representative days may be sufficient to capture the value of long-term storage.

Given our results, we recommend that modelers take care when reducing the temporal scope of a ESOM to avoid potential errors and

---

[15] It is probable that *ERH* is also viable if run using representative days given the results observed in [25].

biases. As for possible solutions and avenues of future research, we suggest either improving the state of the art in TSA methods and ESOM formulations, identifying such errors and biases ex-post, using decomposition methods to reduce computation times while avoiding aggregation or using multiple models to capture the effects of including a high level of temporal detail.

CRediT authorship contribution statement

Sebastian Gonzato: Conceptualization, Methodology, Software, Writing - original draft. **Kenneth Bruninx:** Conceptualization, Methodology, Writing - review & editing. **Erik Delarue:** Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] Loulou R, Goldstein G, Kanudia A, Lettila A, Remme U. Documentation for the TIMES model: PART I. Energy technology systems analysis programme. Technical Report, 2016, p. 151.

[2] Howells Mark, Rogner Holger, Strachan Neil, Heaps Charles, Huntington Hillard, Kypreos Socrates, Hughes Alison, Silveira Semida, DeCarolis Joe, Bazillian Morgan, Roehrl Alexander. OSeMOSYS: The Open Source Energy Modeling System: An introduction to its ethos, structure and development. Energy Policy 2011;39(10):5850–70. http://dx.doi.org/10.1016/j.enpol.2011.06.033, URL: http://www.sciencedirect.com/science/article/pii/S0301421511004897.

[3] Cohen Stuart M, Becker Jonathon, Bielen David A, Brown Maxwell, Cole Wesley J, Eurek Kelly P, Frazier Allister, Frew Bethany A, Gagnon Pieter J, Ho Jonathan L, Jadun Paige, Mai Trieu T, Mowers Matthew, Murphy Caitlin, Reimers Andrew, Richards James, Ryan Nicole, Spyrou Evangelia, Steinberg Daniel C, Sun Yinong, Vincent Nina M, Zwerling Matthew. Regional energy deployment system (ReEDS) model documentation: Version 2018. Technical report, Golden, CO (United States): National Renewable Energy Laboratory (NREL); 2019, http://dx.doi.org/10.2172/1505935.

[4] Nahmmacher Paul, Schmid Eva, Knopf Brigitte. Documentation of LIMES-EU - A long-term electricity system model for Europe. Technical report, 2014, p. 43.

[5] Koppelaar Rembrandt HEM, Keirstead James, Shah Nilay, Woods Jeremy. A review of policy analysis purpose and capabilities of electricity system models. Renew Sustain Energy Rev 2016;59:1531–44. http://dx.doi.org/10.1016/j.rser.2016.01.090.

[6] Haas J, Cebulla F, Cao K, Nowak W, Palma-Behnke R, Rahmann C, Mancarella P. Challenges and trends of energy storage expansion planning for flexibility provision in low-carbon power systems – a review. Renew Sustain Energy Rev 2017;80(June):603–19. http://dx.doi.org/10.1016/j.rser.2017.05.201.

[7] Hoffmann Maximilian, Kotzur Leander, Stolten Detlef, Robinius Martin. A review on time series aggregation methods for energy system models. Energies 2020;13(3). http://dx.doi.org/10.3390/en13030641.

[8] Diaz Gabriel, Inzunza Andrés, Moreno Rodrigo. The importance of time resolution, operational flexibility and risk aversion in quantifying the value of energy storage in long-term energy planning studies. Renew Sustain Energy Rev 2019;112(July 2018):797–812. http://dx.doi.org/10.1016/j.rser.2019.06.002.

[9] Poncelet Kris, Hoschle Hanspeter, Delarue Erik, Virag Ana, D'haeseleer William. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. IEEE Trans Power Syst 2017;32(3):1936–48. http://dx.doi.org/10.1109/TPWRS.2016.2596803.

[10] Teichgraeber Holger, Brandt Adam R. Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. Appl Energy 2019;239(August 2018):1283–93. http://dx.doi.org/10.1016/j.apenergy.2019.02.012.

[11] Scott Ian J, Carvalho Pedro MS, Botterud Audun, Silva Carlos A. Clustering representative days for power systems generation expansion planning: Capturing the effects of variable renewables and energy storage. Appl Energy 2019;253(July):113603. http://dx.doi.org/10.1016/j.apenergy.2019.113603.

[12] Nahmmacher Paul, Schmid Eva, Hirth Lion, Knopf Brigitte. Carpe diem: A novel approach to select representative days for long-term power system modeling. Energy 2016;112:430–42. http://dx.doi.org/10.1016/j.energy.2016.06.081.

[13] Buchholz Stefanie, Gamst Mette, Pisinger David. A comparative study of time aggregation techniques in relation to power capacity expansion modeling. T O P 2019;27(3):353–405. http://dx.doi.org/10.1007/s11750-019-00519-z.

[14] Buchholz Stefanie, Gamst Mette, Pisinger David. Sensitivity analysis of time aggregation techniques applied to capacity expansion energy system models. Appl Energy 2020;269(May):114938. http://dx.doi.org/10.1016/j.apenergy.2020.114938.

[15] Zipkin Paul H. Bounds for row-aggregation in linear programming. Oper Res 1980;28(4):903–16. http://dx.doi.org/10.1287/opre.28.4.903, URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.28.4.903.

[16] Zipkin Paul H. Bounds on the effect of aggregating variables in linear programs. Oper Res 1980;28(2):403–18. http://dx.doi.org/10.1287/opre.28.2.403, URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.28.2.403.

[17] Poncelet Kris, Delarue Erik, Six Daan, Duerinck Jan, D'haeseleer William. Impact of the level of temporal and operational detail in energy-system planning models. Appl Energy 2015;162:631–43. http://dx.doi.org/10.1016/j.apenergy.2015.10.100.

[18] Helistö Niina, Kiviluoma Juha, Holttinen Hannele, Lara Jose Daniel, Hodge Bri-Mathias. Including operational aspects in the planning of power systems with large amounts of variable generation: A review of modeling approaches. Wiley Interdiscip Rev: Energy Environ 2019;8(5). http://dx.doi.org/10.1002/wene.341, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wene.341.

[19] Merrick James H. On representation of temporal variability in electricity capacity planning models. Energy Econ 2016;59:261–74. http://dx.doi.org/10.1016/j.eneco.2016.08.001.

[20] Pfenninger Stefan. 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. Appl Energy 2017;197:1–13. http://dx.doi.org/10.1016/j.apenergy.2017.03.051.

[21] Hilbers Adriaan P, Brayshaw David J, Gandy Axel. Importance subsampling: improving power system planning under climate-based uncertainty. Appl Energy 2019;251(January):113114. http://dx.doi.org/10.1016/j.apenergy.2019.04.110, arXiv:1903.10916.

[22] Sun Mingyang, Teng Fei, Zhang Xi, Strbac Goran, Pudjianto Danny. Data-driven representative day selection for investment decisions: A cost-oriented approach. IEEE Trans Power Syst 2019;34(4):2925–36. http://dx.doi.org/10.1109/TPWRS.2019.2892619.

[23] Teichgraeber Holger, Lindenmeyer Constantin P, Baumgärtner Nils, Kotzur Leander, Stolten Detlef, Robinius Martin, Bardow André, Brandt Adam R. Extreme events in time series aggregation: A case study for optimal residential energy supply systems. Appl Energy 2020;275(May):115223. http://dx.doi.org/10.1016/j.apenergy.2020.115223, arXiv:2002.03059.

[24] Belderbos Andreas. Storage via power-to-gas in future energy systems. Leuven: KU Leuven; 2019.

[25] Gabrielli Paolo, Gazzani Matteo, Martelli Emanuele, Mazzotti Marco. Optimal design of multi-energy systems with seasonal storage. Appl Energy 2018;219(June 2017):408–24. http://dx.doi.org/10.1016/j.apenergy.2017.07.142.

[26] Kotzur Leander, Markewitz Peter, Robinius Martin, Stolten Detlef. Time series aggregation for energy system design: Modeling seasonal storage. Appl Energy 2018;213(October 2017):123–35. http://dx.doi.org/10.1016/j.apenergy.2018.01.023, arXiv:1710.07593.

[27] Tejada-Arango Diego A, Domeshek Maya, Wogrin Sonja, Centeno Efraim. Enhanced representative days and system states modeling for energy storage investment analysis. IEEE Trans Power Syst 2018;33(6):6534–44. http://dx.doi.org/10.1109/TPWRS.2018.2819578.

[28] Pineda Salvador, Morales Juan M. Chronological time-period clustering for optimal capacity expansion planning with storage. IEEE Trans Power Syst 2018;33(6):7162–70. http://dx.doi.org/10.1109/TPWRS.2018.2842093.

[29] van der Heijde Bram, Vandermeulen Annelies, Salenbien Robbe, Helsen Lieve. Representative days selection for district energy system optimisation: a solar district heating system with seasonal storage. Appl Energy 2019;248(February):79–94. http://dx.doi.org/10.1016/j.apenergy.2019.04.030.

[30] Wogrin S, Tejada-Arango DA, Pineda S, Morales JM. 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). IEEE; 2019, p. 1–6. http://dx.doi.org/10.1109/sest.2019.8849027.

[31] Dowling Jacqueline A, Rinaldi Katherine Z, Ruggles Tyler H, Davis Steven J, Yuan Mengyao, Tong Fan, Lewis Nathan S, Caldeira Ken. Role of long-duration energy storage in variable renewable electricity systems. Joule 2020;4(9):1907–28. http://dx.doi.org/10.1016/j.joule.2020.07.007.

[32] Zeyringer Marianne, Price James, Fais Birgit, Li Pei-hao, Sharp Ed. Designing low-carbon power systems for great britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather. Nature Energy 2018;3(May). http://dx.doi.org/10.1038/s41560-018-0128-x.

[33] Kotzur Leander, Markewitz Peter, Robinius Martin, Stolten Detlef. Impact of different time series aggregation methods on optimal energy system design. Renew Energy 2018;117:474–87. http://dx.doi.org/10.1016/j.renene.2017.10.017, arXiv:1708.00420.

[34] LLC Gurobi Optimization. Gurobi optimizer reference manual. 2020.

[35] Nemhauser George, Wolsey Laurence. Integer and combinatorial optimization. 1999.

[36] Ward Joe H. Hierarchical grouping to optimize an objective function. J Amer Statist Assoc 1963;58(301):236–44. http://dx.doi.org/10.1080/01621459.1963.10500845, URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845.

[37] Hilbers Adriaan, Brayshaw David, Gandy Axel. Efficient quantification of the impact of demand and weather uncertainty in power system models. IEEE Trans Power Syst 2020;1. http://dx.doi.org/10.1109/TPWRS.2020.3031187, URL: https://ieeexplore.ieee.org/document/9224169/, arXiv:1912.10326.

[38] Koltsaklis Nikolaos E, Dagoumas Athanasios S. State-of-the-art generation expansion planning: A review. Appl Energy 2018;230(July):563–89. http://dx.doi.org/10.1016/j.apenergy.2018.08.087.

[39] EnergyVille. EnergyVille introduces additional energy system scenarios for electricity provision in Belgium in 2030 and 2050. 2020, https://www.energyville.be/en/news-events/energyville-introduces-additional-energy-system-scenarios-electricity-provision-belgium.

[40] Wogrin Sonja, Galbally David, Reneses Javier. Optimizing storage operations in medium- and long-term power system models. IEEE Trans Power Syst 2016;31(4):3129–38. http://dx.doi.org/10.1109/TPWRS.2015.2471099.

[41] ENTSO-E. ENTSO-E transparency platform. 2019, https://transparency.entsoe.eu/. Accessed: 2020-09-06.

[42] Trutnevyte Evelina. Does cost optimization approximate the real-world energy transition? Energy 2016;106. http://dx.doi.org/10.1016/j.energy.2016.03.038.

[43] Deane JP, Chiodi Alessandro, Gargiulo Maurizio, Ó Gallachóir Brian P. Soft-linking of a power systems model to an energy systems model. Energy 2012;42(1):303–12. http://dx.doi.org/10.1016/j.energy.2012.03.052.

[44] Collins Seán, Deane JP, Ó Gallachóir Brian. Adding value to EU energy policy analysis using a multi-model approach with an EU-28 electricity dispatch model. Energy 2017;130:433–47. http://dx.doi.org/10.1016/j.energy.2017.05.010.

[45] Hart Elaine K, Jacobson Mark Z. A Monte Carlo approach to generator portfolio planning and carbon emissions assessments of systems with large penetrations of variable renewables. Renew Energy 2011;36(8):2278–86. http://dx.doi.org/10.1016/j.renene.2011.01.015.

[46] Watson Jean Paul, Woodruff David L. Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. Comput Manage Sci 2011;8(4):355–70. http://dx.doi.org/10.1007/s10287-010-0125-4.

[47] Liu Yixian, Sioshansi Ramteen, Conejo Antonio J. Multistage stochastic investment planning with multiscale representation of uncertainties and decisions. IEEE Trans Power Syst 2017;33(1):781–91. http://dx.doi.org/10.1109/tpwrs.2017.2694612.

[48] Cao Karl Kiên, Von Krbek Kai, Wetzel Manuel, Cebulla Felix, Schreck Sebastian. Classification and evaluation of concepts for improving the performance of applied energy system optimization models. Energies 2019;12(24). http://dx.doi.org/10.3390/en12244656.

[49] Maluenda Benjamin, Negrete-Pincetic Matias, Olivares Daniel E, Lorca Álvaro. Expansion planning under uncertainty for hydrothermal systems with variable resources. Int J Electr Power Energy Syst 2018;103(March):644–51. http://dx.doi.org/10.1016/j.ijepes.2018.06.008.

[50] Crespo del Granado Pedro, van Nieuwkoop Renger H, Kardakos Evangelos G, Schaffner Christian. Modelling the energy transition: A nexus of energy system and economic models. Energy Strategy Rev 2018;20:229–35. http://dx.doi.org/10.1016/j.esr.2018.03.004.

[51] Trutnevyte Evelina, Barton John, O'Grady Áine, Ogunkunle Damiete, Pudjianto Danny, Robertson Elizabeth. Linking a storyline with multiple models: A cross-scale study of the UK power system transition. Technol Forecast Soc Change 2014;89:26–42. http://dx.doi.org/10.1016/j.techfore.2014.08.018, http://dx.doi.org/10.1016/j.techfore.2014.08.018 https://linkinghub.elsevier.com/retrieve/pii/S0040162514002571.

[52] Lund Henrik, Arler Finn, Østergaard Poul Alberg, Hvelplund Frede, Connolly David, Mathiesen Brian Vad, Karnøe Peter. Simulation versus optimisation: Theoretical positions in energy system modelling. Energies 2017;10(7):1–17. http://dx.doi.org/10.3390/en10070840.