

UC Berkeley

Proceedings of the PowerUp Conference

Title

Unlocking the Informational Value of Marginal Costs for Exact Time Series Aggregation in Generation Expansion Planning

Permalink

<https://escholarship.org/uc/item/1w01v8vb>

Journal

Proceedings of the PowerUp Conference, 1(1)

Authors

Santosuosso, Luca

Wogrin, Sonja

Publication Date

2026-09-09

DOI

None/powerup_proceedings.69117

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nd/4.0/>

Peer reviewed

PowerUp 2026

BOULDER, COLORADO
SEPTEMBER 9 – 11, 2026

Unlocking the Informational Value of Marginal Costs for Exact Time Series Aggregation in Generation Expansion Planning

Luca Santosuosso, Sonja Wogrin
Graz University of Technology

Suggested Citation

L. Santosuosso and S. Wogrin, "Unlocking the Informational Value of Marginal Costs for Exact Time Series Aggregation in Generation Expansion Planning," in *Proceedings of the PowerUp Conference*, 2026.

BibTeX Entry

```
@inproceedings{santosuosso2026unlockinginformation,  
  author = {Santosuosso, Luca and Wogrin, Sonja},  
  title = {Unlocking the Informational Value of Marginal Costs for Exact Time Series  
    Aggregation in Generation Expansion Planning},  
  booktitle = {Proceedings of the PowerUp Conference},  
  year = {2026},  
}
```

Unlocking the Informational Value of Marginal Costs for Exact Time Series Aggregation in Generation Expansion Planning

Luca Santosuosso, Sonja Wogrin

Abstract—This paper addresses the generation expansion planning (GEP) problem, formulated as a mixed-integer linear programming model with intertemporal storage constraints. Being generally NP-hard, the problem’s computational complexity grows sharply with the planning horizon and the number of binary variables. While previous research has tackled this challenge using heuristic time series aggregation (TSA) methods, we propose a theoretically grounded marginal-cost-based TSA, designed to construct an aggregated model that preserves the active constraints of its full-scale counterpart, thereby explicitly targeting exact temporal aggregation. This TSA method is embedded within solution algorithms that iteratively refine theoretically validated bounds on the maximum error introduced by the temporal aggregation relative to conventional full-scale optimization, thus offering a formal performance guarantee to the decision-maker. Numerical results highlight the computational advantages of the proposed algorithms, which notably recover tractability whereas full-scale optimization proves intractable.

Index Terms—Exact time series aggregation, generation expansion planning, energy storage, mixed-integer programming

I. INTRODUCTION

A. Literature Review

Generation expansion planning (GEP) is a cornerstone problem in energy system optimization, aiming to determine the optimal mix and sizing of power generation units required to meet future electricity demand while minimizing both capital investment and operational costs [1]. Recently, the large-scale integration of variable renewable energy sources (vRES), the growing deployment of flexible technologies (e.g., energy storage systems), and the evolution of regulatory and market designs have significantly increased the complexity of GEP models [2]. Modern GEP formulations must therefore capture nonlinear system dynamics, alongside intricate intertemporal coupling constraints, over long-term planning horizons [3]. This has driven a steady evolution of GEP models from basic linear programming (LP) formulations [4] toward increasingly complex nonconvex formulations [5], most notably mixed-integer linear programming (MILP) models [6]. Critically, MILP formulations of the GEP problem are strongly NP-hard in general [7], making the trade-off between modeling fidelity and computational tractability a central challenge [8].

To address this trade-off, time series aggregation (TSA) methods are commonly employed [9]. By leveraging clustering techniques, these methods condense the input time series of the GEP problem, such as energy demand profiles or vRES generation forecasts, into a temporally aggregated counterpart

of the original full-scale GEP model, defined over a reduced set of representative time periods (or clusters) [10].

Traditional *a priori* TSA methods rely on standard clustering techniques, such as k-means [11], k-medoids [12], and hierarchical clustering [13], to select representative periods based exclusively on the statistical properties of the input time series. Despite their widespread adoption in GEP [14], the accuracy of these methods is highly case-dependent and must therefore be validated individually for each application [15]. This limitation has motivated growing interest in establishing performance guarantees for aggregated models obtained via TSA [16], typically formulated as bounds on the approximation error relative to the original full-scale model [17]. Although such bounds are valuable for enabling decision-makers to assess the accuracy of an aggregated model [18], they do not allow *a priori* TSA-based models to systematically attain a prescribed level of accuracy. This shortcoming reflects a fundamental limitation of *a priori* methods: they implicitly assume that an accurate statistical representation of the full-scale model’s input space yields a correspondingly accurate approximation of its output solution space. In practice, however, this assumption often fails, as the model outputs depend not only on the input time series but also critically on the underlying mathematical structure of the full-scale model [19].

The recognition of this limitation has led to the recent emergence of the *a posteriori* TSA paradigm [20]. The key distinction of this paradigm from conventional *a priori* TSA is its use of clustering features derived directly from the full-scale model, rather than relying solely on statistical features extracted from the input time series [21]. This paradigm has proven highly effective for GEP [22], particularly in the presence of intertemporal coupling constraints (e.g., due to storage or ramping) [23], which significantly complicate TSA by requiring preservation of temporal chronology to maintain consistency with the full-scale model dynamics [24]. Existing studies have explored *a posteriori* methods based on clustering features derived from approximations of the full-scale primal solution [25] or the dual solution [26]. Nevertheless, *a posteriori* TSA remains underdeveloped, as the current research has yet to conclusively determine which clustering features most effectively lead to accurate aggregated models.

In this regard, the theoretical result of [27] offers a strong conceptual foundation by demonstrating that, if the decision-maker had prior knowledge of the full-scale model’s active constraints, TSA could be performed **exactly**, with the aggregated formulation reproducing both the optimal solution space and the optimal objective function value of its full-scale counterpart. Notably, [28] further extends this result to optimization models featuring intertemporal storage constraints.

While perfect knowledge of the full-scale model’s active constraint sets during TSA is clearly unattainable, the theoret-

Luca Santosuosso and Sonja Wogrin are with the Institute of Electricity Economics and Energy Innovation, Graz University of Technology, and the Research Center ENERGETIC, 8010 Graz, Austria. Emails: luca.santosuosso@tugraz.at; wogrin@tugraz.at.

ical result of [27] provides a valuable insight: any clustering feature that reliably captures constraint activation in the full-scale model can be exploited within an a posteriori TSA method to achieve exact temporal aggregation. This highlights a significant, yet largely untapped, opportunity to dramatically reduce the computational complexity of nonconvex GEP models, without compromising solution accuracy.

B. Research Gaps and Contributions

The literature review reveals the following research gaps:

- Traditional *a priori* TSA methods rely on the frequently violated assumption that an accurate representation of the full-scale model's input space is sufficient to accurately capture its solution space [19]. In contrast, *a posteriori* TSA methods offer greater potential for constructing high-fidelity aggregated models by exploiting information extracted directly from the full-scale model to enrich the clustering feature space [20]. Although prior studies have proposed deriving clustering features from approximations of the full-scale primal [25] and dual [26] solutions, it remains unclear which full-scale model information should be leveraged, and how clustering features should be systematically constructed, to improve the accuracy of temporally aggregated models. This challenge is particularly pronounced for MILP optimization models, where dual information is generally ill-defined.
- Recent analyses [27], [28] demonstrate that exact temporal aggregation can be achieved when the active constraints of the full-scale model are preserved in the aggregated model. These results provide clear theoretical guidance on which aspects of the full-scale model are most critical for ensuring aggregation accuracy and, therefore, should be captured by TSA methods. However, because the set of active constraints is unknown during the TSA process, it remains unclear how this theoretical insight can be translated into practical methods that effectively reduce GEP models while maintaining high solution accuracy and formal performance guarantees.

This paper seeks to address these research gaps through the following **key contributions**:

- We formulate the full-scale GEP model as a MILP problem encompassing multiple generation technologies, including both vRES and thermal power, and subject to intertemporal storage constraints. We then construct the corresponding temporally aggregated model and show that the full-scale model's marginal costs serve as a reliable proxy for identifying the sets of active constraints in the full-scale model, enabling their use as clustering features to achieve exact temporal aggregation.
- Acknowledging that system marginal costs are not directly available during temporal aggregation, we introduce a practical a posteriori TSA method that leverages marginal cost estimates obtained from the dual space of the full-scale model's linear relaxation to guide the construction of the temporally aggregated model.
- We develop two iterative solution algorithms that systematically reduce the temporal dimensionality of the

full-scale MILP GEP model while providing a formal performance guarantee in the form of theoretically validated bounds on the approximation error introduced by temporal aggregation. Notably, both algorithms deliver feasible solutions for the full-scale model at each iteration, and the aggregated model is progressively refined toward exact temporal aggregation through the proposed marginal-costs-based a posteriori TSA method.

Finally, the proposed algorithms are evaluated using real-world data from the ENTSO-E Transparency Platform [29].

The remainder of the paper is organized as follows: Section II details the proposed methodology, Section III presents the numerical results, and Section IV concludes the study.

II. METHODOLOGY

This section presents the proposed methodology. Subsection II-A introduces the full-scale GEP model, while Subsection II-B presents its temporally aggregated counterpart. Subsection II-C details the proposed marginal-cost-based a posteriori TSA method, which is employed in the algorithms of Subsection II-D to iteratively refine bounds on the optimal objective function value of the original full-scale model.

In the following, sets and vectors are denoted in boldface (e.g., $\mathbf{z}$). The notation $|\cdot|$ denotes the cardinality of a set.

A. The Full-Scale Model

The goal is to determine the optimal mix and sizing of generation units that minimize capital investment and operational costs while meeting the energy demand. Specifically, we consider an asset portfolio comprising vRES, namely wind and solar power plants, thermal units, and energy storage systems. In the following formulation, we omit network constraints, consistent with the operation of generation portfolios that do not participate in the grid management, such as commercial virtual power plants or generation companies [1]. Future work will extend the formulation to incorporate these constraints.

Let $\mathbf{G}$, $\mathbf{N}$, and $\mathbf{T}$ denote the sets of generators (indexed by g), energy storage systems (indexed by n), and time steps (indexed by t), respectively. The operational costs of generator g and non-supplied energy are denoted by C_g^{pg} and C^{ms} (€/MWh), respectively. For each storage unit n , C_n^{pc} and C_n^{pd} denote the charging and discharging costs (€/MWh). The investment costs (€/MW) are $C_g^{\text{g,inv}}$ for generators and $C_n^{\text{s,inv}}$ for storage units. The input time series comprise the generator capacity factors $F_{g,t}$ and the energy demand D_t (MWh).

We denote by x_g^{g} and x_n^{s} the installed capacities (MW) of generator g and storage unit n , respectively. The binary variables b_g^{g} and b_n^{s} enforce that installed capacities are either zero or lie within $[\underline{X}_g^{\text{g}}, \overline{X}_g^{\text{g}}]$ and $[\underline{X}_n^{\text{s}}, \overline{X}_n^{\text{s}}]$, respectively.

The operational variables include the power output $p_{g,t}^{\text{g}}$ (MW) of generator g at time t , the non-supplied energy e_t^{ns} (MWh), the charging and discharging power of storage unit n , denoted by $p_{n,t}^{\text{c}}$ and $p_{n,t}^{\text{d}}$ (MW), respectively, and its stored energy $e_{n,t}^{\text{s}}$ (MWh) at time t . The charging and discharging efficiencies of storage unit n are denoted by η_n^{c} and η_n^{d} , respectively, ρ_n denotes its energy-to-power ratio (h), while E_n^0 denotes its initial stored energy (MWh).

We group the decision variables of the full-scale model in the set $\mathbf{z}$, defined as:

$$\mathbf{z} := \{x_g^g, x_n^s, b_g^g, b_n^s, p_{g,t}^g, e_{n,t}^s, p_{n,t}^c, p_{n,t}^d, e_t^{\text{ns}}\}_{g \in G, n \in N, t \in T}.$$

The objective function of the GEP problem is defined as

$$J(\mathbf{z}) := \sum_{g \in G} C_g^{\text{g,inv}} x_g^g + \sum_{n \in N} C_n^{\text{s,inv}} x_n^s + \sum_{t \in T} \sum_{g \in G} C_g^{\text{pg}} p_{g,t}^g \Delta + \sum_{t \in T} \left(\sum_{n \in N} (C_n^{\text{pc}} p_{n,t}^c + C_n^{\text{pd}} p_{n,t}^d) \Delta + C^{\text{ns}} e_t^{\text{ns}} \right), \quad (1)$$

which represents the sum of both capital investment and operational costs over the planning horizon T .

The **full-scale GEP model**, formulated as a MILP problem, is defined over T with sampling time Δ (h) as follows:

$$\min_{\mathbf{z}} J(\mathbf{z}) \quad (2a)$$

$$\text{s.t.} \quad \sum_{g \in G} p_{g,t}^g \Delta + \sum_{n \in N} (p_{n,t}^d - p_{n,t}^c) \Delta + e_t^{\text{ns}} = D_t, \quad \forall t, \quad (2b)$$

$$e_{n,t+1}^s = e_{n,t}^s + (\eta_n^c p_{n,t}^c - \eta_n^d p_{n,t}^d) \Delta, \quad \forall n, \forall t \in T \setminus \{|T| - 1\}, \quad (2c)$$

$$e_{n,0}^s = E_n^0, \quad \forall n, \quad (2d)$$

$$0 \leq p_{g,t}^g \leq F_{g,t} x_g^g, \quad \forall g, \forall t, \quad (2e)$$

$$0 \leq e_{n,t}^s \leq x_n^s \rho_n, \quad \forall n, \forall t, \quad (2f)$$

$$0 \leq p_{n,t}^c \leq x_n^s, \quad \forall n, \forall t, \quad (2g)$$

$$0 \leq p_{n,t}^d \leq x_n^s, \quad \forall n, \forall t, \quad (2h)$$

$$b_n^s \underline{X}_n^s \leq x_n^s \leq b_n^s \overline{X}_n^s, \quad \forall n, \quad (2i)$$

$$b_g^g \underline{X}_g^g \leq x_g^g \leq b_g^g \overline{X}_g^g, \quad \forall g, \quad (2j)$$

$$e_t^{\text{ns}} \geq 0, \quad \forall t, \quad (2k)$$

$$b_g^g \in \{0, 1\}, \quad \forall g, \quad (2l)$$

$$b_n^s \in \{0, 1\}, \quad \forall n. \quad (2m)$$

In (2), the constraints (2b) enforce the energy balance; (2c) and (2d) characterize the storage dynamics; (2i) and (2j) impose capacity investment limits; and (2e)–(2h) together with (2k)–(2m) define the bounds on the decision variables.

B. The Temporally Aggregated Model

Mixed-integer GEP problems are known to be strongly NP-hard [7]. As the problem dimension increases with the number of binary variables and time steps, (2) may become computationally intractable. To mitigate this, TSA can be employed to construct an aggregated counterpart of (2), defined over a reduced set of representative periods, denoted by $\mathbf{K}$ and indexed by k . When $|\mathbf{K}| \ll |T|$, the aggregated model offers a significant computational advantage over the full-scale model.

Let T_k denote the set of consecutive time steps $t \in T$ assigned to the k -th cluster via TSA. We define $K := |\mathbf{K}|$ and $T_k := |T_k|$. The aggregated counterparts of $F_{g,t}$ and D_t are computed as

$$\hat{F}_{g,k} := \sum_{t \in T_k} \frac{F_{g,t}}{T_k}, \quad \forall g, \forall k, \quad (3) \quad \hat{D}_k := \sum_{t \in T_k} \frac{D_t}{T_k}, \quad \forall k. \quad (4)$$

We group the variables of the aggregated model as follows:

$$\hat{\mathbf{z}} := \{\hat{x}_g^g, \hat{x}_n^s, \hat{b}_g^g, \hat{b}_n^s, \hat{p}_{g,k}^g, \hat{e}_{n,k}^s, \hat{p}_{n,k}^c, \hat{p}_{n,k}^d, \hat{e}_k^{\text{ns}}\}_{g \in G, n \in N, k \in \mathbf{K}}.$$

The aggregated counterpart of $J(\mathbf{z})$ in (1) is

$$\hat{J}(\hat{\mathbf{z}}) := \sum_{g \in G} C_g^{\text{g,inv}} \hat{x}_g^g + \sum_{n \in N} C_n^{\text{s,inv}} \hat{x}_n^s + \sum_{k \in \mathbf{K}} \sum_{g \in G} C_g^{\text{pg}} \hat{p}_{g,k}^g T_k \Delta + \sum_{k \in \mathbf{K}} \left(\sum_{n \in N} (C_n^{\text{pc}} \hat{p}_{n,k}^c + C_n^{\text{pd}} \hat{p}_{n,k}^d) \Delta + C^{\text{ns}} \hat{e}_k^{\text{ns}} \right) T_k. \quad (5)$$

The **temporally aggregated GEP model**, formulated as a MILP problem, is defined over $\mathbf{K}$ as follows:

$$\min_{\hat{\mathbf{z}}} \hat{J}(\hat{\mathbf{z}}) \quad (6a)$$

$$\text{s.t.} \quad \sum_{g \in G} \hat{p}_{g,k}^g \Delta + \sum_{n \in N} (\hat{p}_{n,k}^d - \hat{p}_{n,k}^c) \Delta + \hat{e}_k^{\text{ns}} = \hat{D}_k, \quad \forall k, \quad (6b)$$

$$\hat{e}_{n,k+1}^s = \hat{e}_{n,k}^s + (\eta_n^c \hat{p}_{n,k}^c - \eta_n^d \hat{p}_{n,k}^d) T_k \Delta, \quad \forall n, \forall k \in \mathbf{K} \setminus \{K - 1\}, \quad (6c)$$

$$\hat{e}_{n,0}^s = E_n^0, \quad \forall n, \quad (6d)$$

$$0 \leq \hat{p}_{g,k}^g \leq \hat{F}_{g,k} \hat{x}_g^g, \quad \forall g, \forall k, \quad (6e)$$

$$0 \leq \hat{e}_{n,k}^s \leq \hat{x}_n^s \rho_n, \quad \forall n, \forall k, \quad (6f)$$

$$0 \leq \hat{p}_{n,k}^c \leq \hat{x}_n^s, \quad \forall n, \forall k, \quad (6g)$$

$$0 \leq \hat{p}_{n,k}^d \leq \hat{x}_n^s, \quad \forall n, \forall k, \quad (6h)$$

$$\hat{b}_n^s \underline{X}_n^s \leq \hat{x}_n^s \leq \hat{b}_n^s \overline{X}_n^s, \quad \forall n, \quad (6i)$$

$$\hat{b}_g^g \underline{X}_g^g \leq \hat{x}_g^g \leq \hat{b}_g^g \overline{X}_g^g, \quad \forall g, \quad (6j)$$

$$\hat{e}_k^{\text{ns}} \geq 0, \quad \forall k, \quad (6k)$$

$$\hat{b}_g^g \in \{0, 1\}, \quad \forall g, \quad (6l)$$

$$\hat{b}_n^s \in \{0, 1\}, \quad \forall n. \quad (6m)$$

In (6), the constraints (6b)–(6m) represent the temporally aggregated counterparts of (2b)–(2m), respectively.

C. Clustering Based on Marginal Costs

The construction of the aggregated model (6) requires clustering to identify representative periods within the planning horizon T . The intertemporal constraints in (2) render standard methods (e.g., k-means) unsuitable, as they neglect temporal chronology; we therefore use a *sliding-window clustering* to group consecutive time steps by clustering feature similarity.

Let $\{\mathbf{a}_t\}_{t \in T}$ be the set of clustering features. We evaluate the similarity between consecutive elements of $\{\mathbf{a}_t\}$, assigning t to cluster k if

$$\|\mathbf{a}_t - \boldsymbol{\mu}_k\|_2 \leq \zeta, \quad (7)$$

where $\boldsymbol{\mu}_k$ denotes the centroid of cluster k , defined as $\boldsymbol{\mu}_k := \frac{1}{T_k} \sum_{t \in T_k} \mathbf{a}_t$, and ζ is a user-defined similarity threshold. If (7) is not met for $\mathbf{a}_t$, a new cluster is initialized starting at t .

The selection of clustering features is a key determinant of the quality of the aggregated model (6). Conventional a priori TSA methods construct these features exclusively from the input time series of the GEP problem. Although this strategy preserves the problem input space, it offers no guarantee that the aggregated model will accurately reproduce the outputs

Algorithm 1 Marginal Cost–Based Time Series Aggregation with Bounded Objective Function Error

Input: Parameters of (2), optimality threshold ϵ^{thr} , maximum number of iterations $I := |\mathbf{I}|$, and similarity threshold ζ .
Output: Objective function bounds J^{UB^*} and J^{LB^*} .

```

1:  $i \leftarrow 1$ ;  $\epsilon^i \leftarrow +\infty$ ;  $J^{\text{LB}^i} \leftarrow -\infty$ ;  $J^{\text{UB}^i} \leftarrow +\infty$ ;
2: while  $\epsilon^i > \epsilon^{\text{thr}}$  and  $i \leq I$  do
3:   Estimate marginal costs according to Steps I and II of Subsection II-D;
4:    $\{\mathbf{T}_k^i\}_{k \in \mathbf{K}^i} \leftarrow$  Sliding-window clustering of Subsection II-C using threshold  $\zeta$  and estimated marginal costs as features;
5:    $\hat{\mathbf{z}}^* \leftarrow$  Solve the aggregated model (6) for  $\{\mathbf{T}_k^i\}_{k \in \mathbf{K}^i}$ ;
6:    $J^{\text{LB}^{i+1}} \leftarrow \max(\hat{J}(\hat{\mathbf{z}}^*), J^{\text{LB}^i})$ ;
7:    $\tilde{J} \leftarrow$  Solve the full-scale model (2) with investment variable values fixed to those in  $\hat{\mathbf{z}}^*$  (LP model);
8:    $J^{\text{UB}^{i+1}} \leftarrow \min(\tilde{J}, J^{\text{UB}^i})$ ;
9:    $\epsilon^{i+1} \leftarrow 100 \frac{J^{\text{UB}^{i+1}} - J^{\text{LB}^{i+1}}}{J^{\text{UB}^{i+1}}}$ ;
10:   $i \leftarrow i + 1$ ;
11: end while
12:  $J^{\text{LB}^*} \leftarrow J^{\text{LB}^i}$  and  $J^{\text{UB}^*} \leftarrow J^{\text{UB}^i}$ ;
  
```

(e.g., the optimal objective function value or the optimal solution space) of the full-scale model [19]. In general, a decision-maker is primarily concerned with preserving accuracy in the GEP model's output space rather than its input space.

To address this issue, we build on the theoretical results of [27] and [28], which establish that an aggregated model reproduces the exact outputs of its full-scale counterpart when it captures the active constraints of the full-scale model at optimality. In the full-scale model (2), the constraints (2e)–(2h) become active when a generator reaches its operational limits, as dictated by (2i) and (2j), thereby triggering a transition in the marginal generator that sets the system's marginal cost, which represents the sensitivity of the optimal total system cost to a marginal change in the energy demand. Hence, marginal costs provide an effective proxy for constraint activation and can be used as clustering features to guide the construction of (6) toward exact temporal aggregation.

D. Practical Solution Algorithms

While Subsection II-C highlighted the potential of marginal cost-based clustering for exact TSA, two fundamental challenges remain. First, the nonconvexity of the full-scale model (2) precludes the computation of marginal costs via duality theory. Second, even if marginal costs were obtainable, their evaluation would require solving the full-scale model itself, thereby negating the computational benefits of TSA.

To overcome these, we propose the following marginal cost estimation procedure, executed over $i \in \mathbf{I}$ iterations.

Step I. For each month in the GEP horizon, randomly sample i days. Next, solve a surrogate linear model obtained by relaxing the binary variables in (2) to continuous variables in the interval $[0, 1]$. This surrogate model is defined at full-scale temporal resolution but only over the set of sampled

Algorithm 2 Adaptive Marginal Cost–Based Time Series Aggregation with Bounded Objective Function Error

Input: Parameters of (2), optimality threshold ϵ^{thr} , maximum number of iterations $I := |\mathbf{I}|$, and similarity threshold ζ .
Output: Objective function bounds J^{UB^*} and J^{LB^*} .

```

1:  $i \leftarrow 1$ ;  $\epsilon^i \leftarrow +\infty$ ;  $J^{\text{LB}^i} \leftarrow -\infty$ ;  $J^{\text{UB}^i} \leftarrow +\infty$ ;
2: while  $\epsilon^i > \epsilon^{\text{thr}}$  and  $i \leq I$  do
3:   if  $i = 1$  then
4:      $\hat{\lambda}^i \leftarrow$  Estimate marginal costs according to Steps I and II of Subsection II-D;
5:   else
6:      $\hat{\lambda}^i \leftarrow$  Estimate marginal costs based on the discrepancies between  $\hat{\lambda}^{i-1}$  and  $\hat{\lambda}^{i-1}$ , as detailed in Subsection II-D;
7:   end if
8:    $\{\mathbf{T}_k^i\}_{k \in \mathbf{K}^i} \leftarrow$  Sliding-window clustering of Subsection II-C using threshold  $\zeta$  and  $\hat{\lambda}^i$  as features;
9:    $\hat{\mathbf{z}}^* \leftarrow$  Solve the aggregated model (6) for  $\{\mathbf{T}_k^i\}_{k \in \mathbf{K}^i}$ ;
10:   $J^{\text{LB}^{i+1}} \leftarrow \max(\hat{J}(\hat{\mathbf{z}}^*), J^{\text{LB}^i})$ ;
11:   $\{\tilde{\lambda}^i, \tilde{J}\} \leftarrow$  Solve the full-scale model (2) with investment variable values fixed to those in  $\hat{\mathbf{z}}^*$  (LP model);
12:   $J^{\text{UB}^{i+1}} \leftarrow \min(\tilde{J}, J^{\text{UB}^i})$ ;
13:   $\epsilon^{i+1} \leftarrow 100 \frac{J^{\text{UB}^{i+1}} - J^{\text{LB}^{i+1}}}{J^{\text{UB}^{i+1}}}$ ;
14:   $i \leftarrow i + 1$ ;
15: end while
16:  $J^{\text{LB}^*} \leftarrow J^{\text{LB}^i}$  and  $J^{\text{UB}^*} \leftarrow J^{\text{UB}^i}$ ;
  
```

days. For instance, if $i = 2$ and the full-scale model is formulated at hourly resolution, the surrogate model spans 576 time steps (i.e., 2 sampled days per month over 12 months, with 24 hours per day). This optimization step yields an estimate of the marginal cost trajectory for each sampled day.

Step II. For each month, all non-sampled days are assigned, with equal probability, to one of the i sampled days of that month. Each non-sampled day is assumed to share the same marginal cost trajectory as the sampled day to which it is assigned. For example, if a particular day is assigned to a sampled day d , the hourly marginal costs of d are used as a proxy for the marginal costs of that day. This step yields a marginal cost estimate over the full-scale horizon $\mathbf{T}$, which is then used as a feature in the sliding-window clustering of Subsection II-C to construct the aggregated model (6).

Importantly, irrespective of the clustering technique employed, the aggregated model (6) provides, by construction, a lower bound on the optimal objective function value of the full-scale model (2) [30]. Fixing the investment variables in (2) to those computed via (6) reduces (2) to a linear program, whose solution yields an upper bound on the full-scale optimal objective. At each iteration $i \in \mathbf{I}$, the relative difference between the upper and lower bounds, denoted J^{UB^i} and J^{LB^i} , respectively, defines the optimality gap ϵ^i , indicating that the MILP GEP solution derived via TSA is at most ϵ^i suboptimal. Convergence is achieved when ϵ^i falls below a prescribed threshold ϵ^{thr} . This iterative procedure, which

combines marginal cost-based a posteriori TSA with bound-based error evaluation, constitutes the proposed Algorithm 1.

Algorithm 1 refines the aggregated model by progressively increasing the number of sampled days used to estimate marginal costs. However, the optimization steps employed to compute the objective function bounds also generate additional, underexploited information. Specifically, solving (2) with fixed investment variables to obtain an upper bound on the objective function produces short-run marginal cost estimates, which can be used to enhance the TSA as follows.

Let $\tilde{\lambda}^i$ denote the marginal costs obtained from the optimization step yielding J^{UB^i} (i.e., the short-run marginal costs), and let $\hat{\lambda}^i$ denote those used to construct the aggregated model at iteration i (i.e., the long-run marginal costs). Algorithm 2 introduces an adaptive alternative to Algorithm 1 executing the above Steps I and II at iteration $i + 1$ over the days exhibiting the largest average hourly deviation between $\tilde{\lambda}^i$ and $\hat{\lambda}^i$, rather than over randomly sampled days as in Algorithm 1, adapting TSA to the observed optimality gap.

Critically, both Algorithms 1 and 2 exhibit the following key properties. First, they are accompanied by a formal performance guarantee, in the form of theoretically validated bounds on the objective function error incurred relative to the original full-scale GEP model (2). Second, at each iteration, they yield a feasible solution for (2), i.e., the decision variable values employed to compute the objective function upper bound. Finally, they offer two complementary and practical strategies to exploit the theoretical results of [27] and [28], thereby guiding the construction of the aggregated model toward exact temporal aggregation based on marginal cost estimates.

III. NUMERICAL RESULTS

This section presents the numerical results. Subsection III-A describes the case study, Subsection III-B illustrates how marginal-cost-based clustering enables exact TSA, and Subsection III-C compares the proposed solution algorithms.

A. Case Study Description

To ensure transparency and reproducibility, we generate the problem parameters from prescribed distributions as follows.

For each storage unit, the charging and discharging efficiencies are set to 0.9 and 1/0.9, respectively, while the associated operational costs are drawn from a uniform distribution over $[5, 15]$ €/MWh. Capacity investment costs are sampled uniformly from $[4.5 \times 10^5, 5.5 \times 10^5]$ €/MW, and storage capacity investments are either zero or constrained to $[0.25, 1]$ MW. The stored energy is initialized to 0 MWh, with the energy-to-power ratio fixed at 2 h. Among the generators, 20% are thermal units, 40% wind, and 40% solar. Capacity investments in thermal and vRES units are either zero or constrained to $[0.5, 1]$ MW and $[0.2, 1]$ MW, respectively. The investment costs are sampled uniformly from $[3 \times 10^6, 4 \times 10^6]$ €/MW for thermal units and $[5 \times 10^5, 6 \times 10^5]$ €/MW for vRES. The operational costs are set to 50 €/MWh for thermal units and 1 €/MWh for vRES, while the cost of non-supplied energy is 10^5 €/MWh. The GEP horizon is one year with hourly

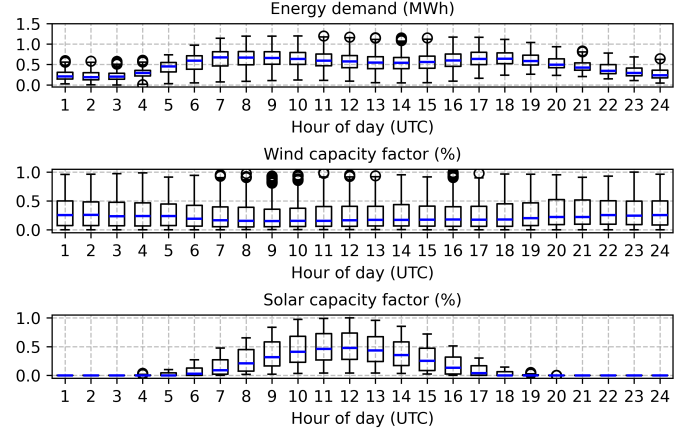


Fig. 1. Boxplots of the hourly input data. Each boxplot characterizes the distribution of values observed at a given hour: the box spans the interquartile range, the blue line indicates the median, the whiskers extend to the 10th and 90th percentiles, and the outliers are shown as individual points.

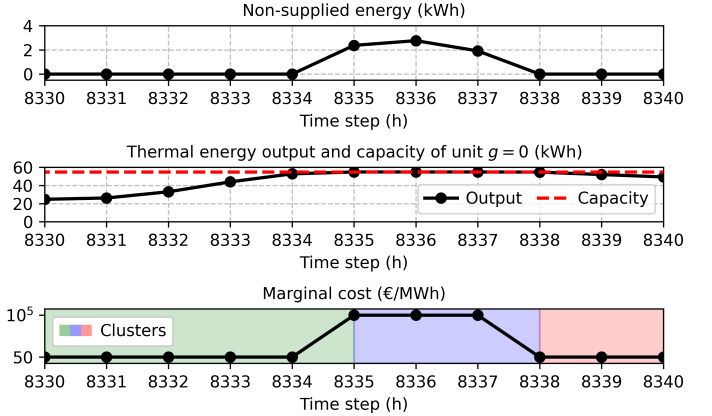


Fig. 2. Example of marginal-cost-based time series aggregation.

resolution ($|T| = 8760$). In the proposed algorithms, we set $I = 25$, $\zeta = 10$ and $\epsilon^{\text{thr}} = 0.01\%$.

The capacity factor of thermal units is fixed at 1, whereas vRES capacity factors and the energy demand time series are derived by scaling real-world observations from the Austrian power grid in 2024, available on the ENTSO-E Transparency Platform [29], and illustrated in Fig. 1. The baseline demand profile shown in Fig. 1 is scaled by a factor $0.05 (|G| + |N|)$ in order to reflect the aggregate system size. To generate heterogeneous production profiles across vRES units, uniform multiplicative noise in the range $[0.85, 1.15]$ is applied to the baseline wind and solar capacity factors depicted in Fig. 1.

The optimization models are implemented on an Intel i7 processor with 32 GB of RAM, using Gurobi 12.0.1.

B. Using Marginal Costs for Exact Time Series Aggregation

Fig. 2 illustrates how marginal costs can be used to identify the active constraints in the LP relaxation of the full-scale GEP model (2). Specifically, the results shown in the figure are obtained by setting $|G| = 5$ and $|N| = 0$, thereby limiting the available sources of flexibility in the considered power system. This configuration is deliberately chosen in this example to

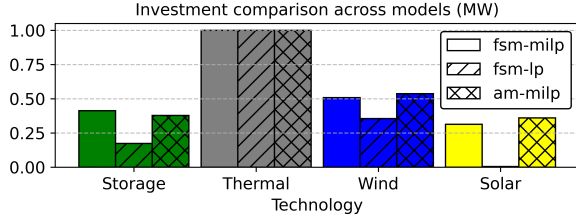


Fig. 3. Comparison of investment decisions obtained with the full-scale MILP model (fsm-milp), its LP relaxation (fsm-lp), and the proposed aggregated MILP model (am-milp).

induce the activation of the power generation constraints (2e) in the full-scale GEP model (2), resulting in the occurrence of non-supplied energy.

As shown in the figure, when the marginal generator (the only available thermal unit) reaches its maximum capacity (at $t = 8335$), the marginal cost increases from 50€ (its operating cost) to 10^5€ , corresponding to the non-supplied energy penalty. Conversely, as demand decreases and the generator operates strictly within its feasible region (starting from $t = 8338$), the marginal cost decreases from 10^5€ back to 50€ . Hence, marginal costs alone constitute a sufficient statistic to identify the active constraint sets of the GEP problem, specifically whether the upper limit on thermal power generation or the lower limit on non-supplied energy is binding. This example illustrates that using marginal costs as clustering features for TSA yields an aggregated model that preserves the full-scale active constraints, thereby resulting in exact temporal aggregation, as demonstrated in [27]. In this sense, the informational value embedded in the system's marginal costs is effectively unlocked, enabling exact TSA for the GEP problem under study.

When addressing the MILP GEP model (2), marginal costs obtained from its LP relaxation are used as proxies for the true marginal costs within the proposed solution algorithms (see Subsection II-D). Clearly, the LP relaxation could, in principle, serve directly as an approximation of the original MILP GEP model rather than solely for estimating marginal costs. However, Fig. 3 shows that, for an illustrative stylized system comprising both storage ($|N| = 1$) and generation ($|G| = 10$) units, the combination of TSA based on marginal cost estimates from the LP relaxation with the construction of an aggregated model that preserves the nonconvex investment constraints, as proposed, provides a substantially more accurate approximation of the MILP GEP model's optimal investment decisions than the conventional LP relaxation.

C. Comparative Performance of the Proposed Algorithms

To illustrate the performance of the proposed solution algorithms, Fig. 4 reports the convergence of their objective function bounds for a GEP model with $|G| = 100$ and $|N| = 10$. Specifically, in the first two subplots of the figure, the lower and upper bounds on the optimal objective function value of the full-scale GEP model (2), obtained at each iteration of the proposed solution algorithms, are represented by triangles and circles, respectively.

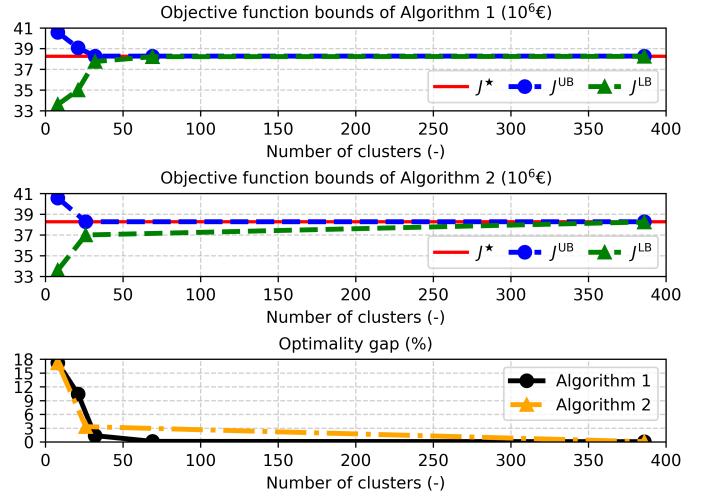


Fig. 4. Performance comparison of the proposed Algorithms 1 and 2.

TABLE I
COMPARISON BETWEEN THE RUNTIMES OF THE FULL-SCALE MODEL, ALGORITHM 1, AND ALGORITHM 2 AS $|N|$ AND $|G|$ INCREASE. RELATIVE DIFFERENCES WITH RESPECT TO THE FULL-SCALE MODEL ARE REPORTED IN BRACKETS. THE SYMBOL ∞ INDICATES INTRACTABLE INSTANCES OF THE FULL-SCALE MODEL.

$ N $	$ G $	Runtime (min)		
		Full-scale model	Algorithm 1	Algorithm 2
50	50	2.1	2.9 (+38%)	2.4 (+14%)
250	250	51.9	22.4 (−57%)	19.6 (−62%)
750	750	432.8	211.1 (−51%)	102.3 (−76%)
1500	1500	∞	291.9 (− ∞)	144.2 (− ∞)

As detailed in Subsection II-D, both algorithms execute the same TSA and optimization steps during the first iteration. From the second iteration onward, a key distinction emerges: Algorithm 2 leverages the upper bound information obtained in the previous iteration to refine marginal cost estimates, whereas Algorithm 1 does not. This difference is evident in Fig. 4, where Algorithm 2 achieves a $\sim 3\%$ optimality gap by the second iteration, while Algorithm 1 exhibits a $\sim 10\%$ gap. Notably, both algorithms ultimately converge to over 99.99% accuracy using 386 clusters, corresponding to a reduction of more than 95% in the temporal dimension of the GEP model. However, Algorithm 1 requires five iterations to reach convergence, whereas the adaptive Algorithm 2 converges in only three iterations, achieving the same dimensionality reduction with lower computational effort.

These results indicate that the computational advantage observed for Algorithm 2 over Algorithm 1 does not arise from achieving a more accurate temporal aggregation for a given number of clusters. Instead, it results from exploiting the information embedded in the upper bound computation, as described in Subsection II-D, to accelerate the identification of the minimum number of clusters required to achieve the prescribed optimality gap.

Table I reports the computational performance of the proposed algorithms in comparison with conventional full-scale optimization. As anticipated, Algorithm 2 consistently exhibits faster runtimes than Algorithm 1. While for small instances of

TABLE II
PERCENTAGE OF TOTAL RUNTIME DEVOTED TO THE LOWER BOUND (LB)
AND UPPER BOUND (UB) COMPUTATIONS IN ALGORITHMS 1 AND 2. THE
LB COMPUTATION INCLUDES TSA AND THE SOLUTION OF THE
TEMPORALLY AGGREGATED MODEL, WHILE THE UB COMPUTATION
CONSISTS OF SOLVING THE FULL-SCALE MODEL WITH FIXED
INVESTMENT DECISIONS, AS DETAILED IN SECTION II.

N	G	Percentage of total runtime			
		Algorithm 1		Algorithm 2	
		LB	UB	LB	UB
50	50	19%	81%	19%	81%
250	250	31%	69%	29%	71%
750	750	33%	67%	32%	68%
1500	1500	38%	62%	34%	66%

the GEP problem the TSA-based algorithms may incur higher computational costs relative to full-scale optimization, their efficiency increases markedly as the number of generation and storage units, and consequently the number of binary variables, grows. Notably, both algorithms maintain tractability, whereas conventional full-scale optimization proves intractable, i.e., it fails to solve within the prescribed time of 24 hours.

Finally, Table II reports the percentage of the total runtime devoted to computing the objective function lower and upper bounds in the proposed solution algorithms. The results indicate that, for small instances of the GEP problem, the upper bound computation accounts for the majority of the total runtime in both algorithms, representing more than 80% of the overall computational effort. This behavior is expected, since the upper bound computation requires solving the full-scale GEP model with investment decisions fixed to those obtained from the temporally aggregated model. Consequently, despite being linear, the resulting optimization problem remains computationally demanding due to its large temporal dimensionality.

As the size of the GEP problem increases, the relative computational burden gradually shifts toward the lower bound computation. Specifically, the share of total runtime devoted to this algorithmic step increases from 19% for the smallest GEP problem to approximately 36–38% for the largest problem considered. This behavior reflects the different scalability characteristics of the two computational tasks: while the upper bound computation requires solving a linear optimization model, the lower bound computation involves solving a temporally aggregated MILP model whose computational complexity grows more rapidly with the size of the GEP problem.

Notably, the growth in computational effort associated with the lower bound derivation is less pronounced for Algorithm 2 than for Algorithm 1. By leveraging the marginal cost information obtained during the objective function upper bound computation, Algorithm 2 can accurately infer, after only a limited number of iterations, the number of representative time periods required to attain the prescribed optimality gap. Consequently, several intermediate iterations that would otherwise be performed by Algorithm 1 can be avoided, thereby reducing the number of temporally aggregated MILP models solved for refining the objective function lower bound during

the execution of Algorithm 2.

IV. CONCLUSION AND FUTURE WORK

This paper addresses the GEP problem, formulated as a MILP model incorporating intertemporal storage constraints. Being NP-hard, the problem's computational complexity increases sharply with the planning horizon and the number of generators. To address this challenge, we propose a novel a posteriori TSA method that leverages marginal cost estimates to ensure the aggregated model retains the active constraints of the full-scale model, thereby targeting exact temporal aggregation. The TSA method is embedded within solution algorithms that iteratively refine theoretically validated bounds on the maximum error introduced by the temporal aggregation, while producing a feasible GEP solution at each iteration.

Numerical results show the potential of marginal-cost-based clustering for exact TSA and the efficiency of the proposed algorithms, which restore tractability where the full-scale model proves intractable.

Future work will focus on refining the marginal cost estimation procedure, incorporating additional nonconvex dynamics into the GEP formulation, and explicitly accounting for network constraints in larger-scale instances of the GEP problem. In particular, extending the GEP model to include network constraints would require reformulating the proposed marginal-cost-based clustering technique of Subsection II-C into a clustering approach based on locational marginal prices. In this context, the prediction of locational marginal prices would depend not only on the system-wide marginal generation cost, as considered in the present study, but also on network congestion and transmission losses. To this end, machine learning-based prediction methods, recently initiated in our previous work [31], will be further investigated to assess whether they can preserve a favourable balance between prediction accuracy and computational efficiency. Moreover, the marginal cost estimation procedure proposed in Subsection II-D could be further investigated by replacing the current random sampling of days within the GEP horizon with a feature-informed sampling strategy. Specifically, selecting days based on relevant system characteristics, such as demand profiles and vRES capacity factors, may improve the quality of the marginal cost estimates and potentially enhance the convergence of the proposed solution algorithms.

V. AI USAGE DISCLOSURE

AI was utilized for spelling and grammar checking.

ACKNOWLEDGMENTS

Funded by the European Union (ERC, NetZero-Opt, 101116212). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

- [1] A. Baringo, L. Baringo, and J. M. Arroyo, "Robust virtual power plant investment planning," *Sustain. Energy Grids Netw.*, vol. 35, pp. 101105, Sept. 2023.
- [2] N. E. Koltsaklis and A. S. Dagoumas, "State-of-the-art generation expansion planning: A review," *Appl. Energy*, vol. 230, pp. 563–589, Nov. 2018.
- [3] V. N. Motta, M. F. Anjos, and M. Gendreau, "Survey of optimization models for power system operation and expansion planning with demand response," *Eur. J. Oper. Res.*, vol. 312, no. 2, pp. 401–412, Jan. 2024.
- [4] B. F. Hobbs, "Optimization methods for electric utility resource planning," *Eur. J. Oper. Res.*, vol. 83, no. 1, pp. 1–20, May 1995.
- [5] J. F. Franco, M. J. Rider, and R. Romero, "A mixed-integer quadratically-constrained programming model for the distribution system expansion planning," *Int. J. Electr. Power Energy Syst.*, vol. 62, pp. 265–272, Nov. 2014.
- [6] C. L. Lara, 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," *Eur. J. Oper. Res.*, vol. 271, no. 3, pp. 1037–1054, Dec. 2018.
- [7] S. Goderbauer, M. Comis, and F. J. L. Willamowski, "The synthesis problem of decentralized energy systems is strongly NP-hard," *Comput. Chem. Eng.*, vol. 124, pp. 343–349, May 2019.
- [8] C. Li, A. J. Conejo, P. Liu, B. P. Omell, J. D. Sirola, and I. E. Grossmann, "Mixed-integer linear programming models and algorithms for generation and transmission expansion planning of power systems," *Eur. J. Oper. Res.*, vol. 297, no. 3, pp. 1071–1082, Mar. 2022.
- [9] S. Pineda and J. M. Morales, "Chronological time-period clustering for optimal capacity expansion planning with storage," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7162–7170, Nov. 2018.
- [10] R. Granell, C. J. Axon, and D. C. H. Wallom, "Impacts of raw data temporal resolution using selected clustering methods on residential electricity load profiles," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3217–3224, Nov. 2015.
- [11] N. Sarajpoor, L. Rakai, J. Arteaga, N. Amjadi, and H. Zareipour, "Time aggregation in presence of multiple variable energy resources," *IEEE Trans. Power Syst.*, vol. 39, no. 1, pp. 587–601, Jan. 2024.
- [12] T. Schütz, 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," *Renew. Energy*, vol. 129, pp. 570–582, Dec. 2018.
- [13] Y. Liu, R. Sioshansi, and A. J. Conejo, "Hierarchical clustering to find representative operating periods for capacity-expansion modeling," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3029–3039, May 2018.
- [14] F. D. Munoz, B. F. Hobbs, and J.-P. Watson, "New bounding and decomposition approaches for MILP investment problems: Multi-area transmission and generation planning under policy constraints," *Eur. J. Oper. Res.*, vol. 248, no. 3, pp. 888–898, Feb. 2016.
- [15] L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten, "Impact of different time series aggregation methods on optimal energy system design," *Renew. Energy*, vol. 117, pp. 474–487, Mar. 2018.
- [16] L. Santosuosso and S. Wogrin, "Optimal virtual power plant investment planning via time series aggregation with bounded error," in *2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, Valletta, Malta, 2025, pp. 1–5.
- [17] D. F. Rogers, R. D. Plante, R. T. Wong, and J. R. Evans, "Aggregation and disaggregation techniques and methodology in optimization," *Oper. Res.*, vol. 39, no. 4, pp. 553–582, Aug. 1991.
- [18] W. W. Tso, 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," *Appl. Energy*, vol. 270, pp. 115190, July 2020.
- [19] M. Sun, F. Teng, X. Zhang, G. Strbac, and D. Pudjianto, "Data-driven representative day selection for investment decisions: A cost-oriented approach," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2925–2936, July 2019.
- [20] M. Hoffmann, L. Kotzur, D. Stolten, and M. Robinius, "A review on time series aggregation methods for energy system models," *Energies*, vol. 13, no. 3, pp. 641, Feb. 2020.
- [21] C. Li, A. J. Conejo, J. D. Sirola, and I. E. Grossmann, "On representative day selection for capacity expansion planning of power systems under extreme operating conditions," *Int. J. Electr. Power Energy Syst.*, vol. 137, pp. 107697, May 2022.
- [22] K. Poncelet, 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 Trans. Power Syst.*, vol. 32, no. 3, pp. 1936–1948, May 2017.
- [23] D. A. Tejada-Arango, M. Domeshek, S. Wogrin, and E. Centeno, "Enhanced representative days and system states modeling for energy storage investment analysis," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6534–6544, Nov. 2018.
- [24] L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten, "Time series aggregation for energy system design: Modeling seasonal storage," *Appl. Energy*, vol. 213, pp. 123–135, Mar. 2018.
- [25] Y. Zhang, V. Cheng, D. S. Mallapragada, J. Song, and G. He, "A model-adaptive clustering-based time aggregation method for low-carbon energy system optimization," *IEEE Trans. Sustain. Energy*, vol. 14, no. 1, pp. 55–64, Jan. 2023.
- [26] D. Cardona-Vasquez, T. Klatzer, B. Klinz, and S. Wogrin, "Enhancing time series aggregation for power system optimization models: Incorporating network and ramping constraints," *Elect. Power Syst. Res.*, vol. 230, pp. 110267, May 2024.
- [27] S. Wogrin, "Time series aggregation for optimization: One-size-fits-all?," *IEEE Trans. Smart Grid*, vol. 14, no. 3, pp. 2489–2492, May 2023.
- [28] T. Klatzer, D. Cardona-Vasquez, L. Santosuosso, and S. Wogrin, "Towards exact temporal aggregation of time-coupled energy storage models via active constraint set identification and machine learning," *arXiv preprint*, 2025. [Online]. Available: <https://arxiv.org/abs/2510.14451>.
- [29] L. Hirth, J. Mühlenpfordt, and M. Bulkeley, "The ENTSO-E Transparency Platform—A review of Europe's most ambitious electricity data platform," *Appl. Energy*, vol. 225, pp. 1054–1067, Sept. 2018.
- [30] L. Santosuosso, B. Klinz, and S. Wogrin, "What are we clustering for? Establishing performance guarantees for time series aggregation in generation expansion planning," *arXiv preprint*, 2025. [Online]. Available: <https://arxiv.org/abs/2510.09357>.
- [31] J. Rybka, L. Santosuosso, T. Klatzer, and S. Wogrin, "Machine learning for exact time series aggregation in generation expansion planning with energy storage," *arXiv preprint*, 2026. [Online]. Available: <https://arxiv.org/abs/2605.28356>.

I. REVIEW #146A

A. Paper summary

The authors propose a marginal cost based approach for time series aggregation, focusing on generation expansion planning. The relaxed MILP is used to generate marginals for a subset of representative days, which are used to randomly assign marginal values to the remaining days. Numerical results are presented for the two algorithm variants, at four scales of generation and storage investment.

B. Comments for authors

Major comments:

- It's not clear whether the final investment results permit load shed, and if so, why this isn't penalized as would be expected in a typical capacity expansion model.
- The motivation behind using the marginal cost of the generators to cluster representative days is reasonable. However, our experience with planning models (capacity and transmission), specifically those including storage, is that network congestion is a major investment driver since it limits the amount of load served. Since this formulation does not include congestion, it's not clear how well it can capture the investment decisions on a real system. The congestion challenge is only increased with higher renewable penetration and large load additions.
- Step II of the solution algorithms assigns the non-sampled days with equal probability to one of the i sampled days. What is the rationale to assign via a uniform distribution, rather than clustering on some characteristic of the days (e.g., load level, CF?) to better align the days with realized generator dispatch? This would likely be more important when considering network effects, since the power flow patterns could vary.
- The presentation of results could be revised to improve readability:
 - It's not clear what network the methodology is being tested on, and whether the possible investment decisions are a reduced candidate set on the grid, or on the whole network.
 - Figure 4 could be better explained to clarify that the iterations are annotated by the circle/triangle, and the clustering level is different across the two algorithms at each iteration. Also, while it's clear that Algorithm 2 converges faster, could the authors expand on the comparable error between the two algorithms as a function of the number of clusters?
 - Table 1 lists the cardinality of the candidate set, but does not present the resulting investment decisions – how many generators and storage units are selected for investment? Since these decisions then impact the downstream operational decisions and the size of the resulting problem, reporting these results would provide additional details for the system.

C. Summarize your recommendation in one to two sentences

The topic is well-motivated and an interesting solution approach is proposed. However, a major limitation is the lack of network constraints, as congestion drives significant storage investment decisions in planning models. Additionally, the results are not easy to follow and the discussion could be improved.

— Rabab Haider

II. REVIEW #146B

A. Paper summary

This paper presents the generation expansion planning (GEP) problem as a mixed-integer linear programming model with intertemporal storage constraints. Specifically, it proposes a theoretically grounded marginal-cost-based time series aggregation method using a sliding window clustering technique to group consecutive time steps by clustering feature to construct an aggregated model that preserves active constraints of its full scale counterpart—targeting exact temporal aggregation. The paper presents a strong literature review and clear articulation of the paper's main contributions. It seems strong, but it is also a big claim, so strong evaluation results are needed. The paper does also present clear case studies and results that quantitatively articulate the benefits of the approach.

B. Comments for authors

Overall this is a well-written and well-motivated paper on an interesting topic. The literature review comparison of the proposed method with the state of the art, and the contributions that are sought to be addressed in this work are articulated well. Clearly presented algorithms and pseudocode are given.

The authors develop two iterative solution algorithms that systematically reduce the temporal dimensionality of the full-scale MILP GEP

Some higher level explanations of the need for this method could greatly help position this work.

Also, it would be good if the title was clearly explained in the text "unlocking the informational value of marginal costs" – and/or even if a slightly modified title would be possible. The current title is a bit vague.

The paper does also present clear case studies and results that quantitatively articulate the benefits of the approach. Could more studies be performed? It seems that the potential theoretical value is shown in the algorithms while the experiments are limited. Can the authors comment on future extensions or plans of extensions in this direction?

C. Summarize your recommendation in one to two sentences

This is overall a well-written paper on a GEP formulation the incorporates clustering in a full-scale MILP problem that encompasses multiple generation technologies. Some suggestions include to include a discussion of potential future evaluations/case studies and to strengthen the overall discussion of the motivation for this particular approach, and further showing or quantifying its potential value. A more clear title may also be helpful, if allowed.

— Katherine Davis

III. REVIEW #146C

A. Paper summary

The paper presents two algorithms for time series aggregation for grid expansion planning models that use system marginal cost as a feature to decide upon time-step similarity for aggregation. Both of the algorithms provide performance guarantees (upper and lower objective bounds) for the original (non-aggregated) expansion planning model.

B. Comments for authors

The authors present an insightful and well-written paper. The proposed idea is very interesting, and this reviewer would like to particularly commend the derivation and analysis of the performance guarantees.

The following comments are provided:

- 1) The performance of the algorithm likely depends strongly on the initial selection of days. A stochastic performance analysis might help reveal this dependence. Additionally, it would be interesting to explore whether there are more informed ways to select the initial days rather than relying on random sampling.
- 2) Iteration $i = 1$ appears to perform rather poorly in general. Would it not be more efficient for the algorithm to start at, say, $i = 3$, thereby avoiding a few runs of the full-scale model with fixed investments?
- 3) For clarity and to avoid ambiguity, I recommend writing out the formal definition of the marginal cost.
- 4) It would be interesting to analyze the per-iteration time distribution of the different components of the algorithm. I suspect that solving the full-scale model, even as an LP, could become a bottleneck for larger-scale systems.
- 5) I may be misunderstanding some of the notation, but it seems that Δ should be cluster-dependent, since clusters can have different lengths.

For the final version of this paper, the authors are encouraged to add some discussion on how they expect their method to hold up for more complex transmission-constrained systems.

C. Summarize your recommendation in one to two sentences

Based on the above I recommend acceptance. The paper presents an interesting idea well.

— Anonymous reviewer

IV. BRIEF RESPONSE TO REVIEWERS (OPTIONAL)

We would like to sincerely thank the reviewers for their time and effort in evaluating our work, as well as for their valuable, constructive, and insightful feedback, which we believe has significantly contributed to improving the quality and clarity of the manuscript.

V. BRIEF SUMMARY OF CHANGES MADE TO THE FINAL PAPER

We would like to sincerely thank the reviewers for their time and effort in evaluating our work, as well as for their valuable, constructive, and insightful feedback, which we believe has significantly contributed to improving the quality and clarity of the manuscript. In response to their comments, our revision efforts can be summarized as follows. First, the discussion of the proposed methodology has been revised to better reflect the title of the study, to provide an explicit definition of marginal costs, and to refine the notation used in the generation expansion planning model under consideration. Moreover, the section “Numerical Results” has been revised in response to several comments from the reviewers, as detailed below. In response to Reviewer 1, the discussion accompanying Fig. 2 has been further elaborated to clarify the interpretation of the reported results, particularly with regard to the occurrence of non-supplied energy in the analyzed system. Similarly, the discussion of Fig. 4 has been expanded to clearly distinguish the different iterations of the proposed algorithm and to better articulate the conclusions drawn from the reported results concerning their relative performance and convergence behavior. In response to the comments from Reviewers 2 and 3, we have introduced a new table (Table II), which reports the percentage of total runtime

devoted to the computation of the objective function lower and upper bounds within the proposed solution algorithms, in order to improve the understanding of the relative computational complexity of the various algorithmic steps involved and their contribution to the overall performance of the algorithms. Finally, the discussion in Section “Conclusion and Future Work” has been significantly expanded to better reflect our future research directions, particularly regarding the applicability and effectiveness of the proposed methodology in more comprehensive generation expansion planning models, including network constraints and additional nonconvex dynamics. Furthermore, we now explicitly articulate our interest in developing advanced methods for estimating marginal costs in more complex instances of the generation expansion planning problem, as well as potential refinements of the proposed estimation procedure. In particular, and in line with the reviewers’ suggestions, we consider moving beyond the current random-sampling-based mechanism for selecting days within the generation expansion planning horizon used for estimating marginal costs, toward more informed selection strategies that exploit the input time series of the problem (e.g., energy demand and renewable capacity factors) to identify the most relevant days for this purpose. In this context, we also discuss the potential application of machine learning techniques to support and enhance the estimation procedure, building on recent work that has already shown promising results.