Cracking the Complexity Barrier: Towards Exact Data Aggregation for High-Performance Energy System Models (Part 2)

DTU PES Summer School 2026

Sonja Wogrin
Institute of Electricity Economics and Energy Innovation
Graz University of Technology

May 18-22, 2026

Funded by the European Union
European Research Council
Established by the European Commission

Agenda

I	Formulate Full-Scale Optimization Model
II	Aggregated Optimization Models & Time Series Aggregation
III	Challenge!
IV	Solution (and Chocolate)

Source: Photo by Estée Janssens on Unsplash

Investment Optimization Model

Full-Scale (without storage)

Objective Function (Total system cost)

$$\min_{z=\{x,p,d\}} = \sum_{g} C^{\mathrm{inv},g}\, x^{g} + \sum_{g,t} C^{\mathrm{op},g}\, p_t^{g} \Delta + \sum_{t} C^{\mathrm{nse}}\, e_t^{\mathrm{ns}}$$

- Energy Balance Constraint

$$\sum_{g} p_t^{g} \Delta + e_t^{\mathrm{ns}} = D_t \ \forall t$$

- Upper Bound on Production:

$$p_t^{g} \le CF_t^{g} x^{g} \quad \forall g, t$$

- Non-Negativity Constraints:

$$p_t^{g} \ge 0 \ \ \forall g, t$$
$$e_t^{\mathrm{ns}} \ge 0 \ \ \forall t$$
$$x^{g} \ge 0 \ \ \forall g$$

Sets

g	Set of generators.
t	**Set of time periods.**

Variables

x^{g}	Installed capacity of generator g. [MW]
p_t^{g}	Power produced of g in period t. [MW]
e_t^{ns}	Non-supplied demand in period t. [MWh]

Data

$C^{\mathrm{inv},g}$	Investment cost of g. [€/MW]
$C^{\mathrm{op},g}$	Operating cost of generator g. [€/MWh]
C^{nse}	Cost of non-supplied energy. [€/MWh]
D_t	Demand in time period t. [MWh]
CF_t^{g}	Capacity factor of g in period t.
Δ	Duration of time period t. [h] (=1 here)

Investment Optimization Model

Full-Scale (adding storage)

Objective Function (Total system cost)

$$\min_{\{x,p,d\}} \dots + C^{\text{inv},s} x^s + \sum_{t} C^{c,s} p_t^c \Delta + \sum_{t} C^{d,s} p_t^d \Delta$$

- Energy Balance Constraint

$$\left(\sum_g p_t^g + p_t^d - p_t^c \right) \Delta + e_t^{\text{ns}} = D_t \quad \forall t$$

- Upper Bounds Storage:

$$e_t^s \le x^s \tau \quad \forall t$$

$$p_t^c \le x^s \quad \forall t$$

$$p_t^d \le x^s \quad \forall t$$

- Energy Storage Dynamics:

$$e_{t+1}^s = e_t^s + \left(\eta^{c,s} p_t^c - \frac{p_t^d}{\eta^{d,s}} \right) \Delta \quad \forall t \setminus \mathrm{T}$$

$$e_T^s = e_1^s \,, e_1^s = 0.$$

- Non-Negativity Constraints:

$$p_t^c \ge 0 \quad \forall t$$

$$p_t^d \ge 0 \quad \forall t$$

$$e_t^s \ge 0 \, \forall t$$

$$x^s \ge 0$$

Sets

g	Set of generators.
t	**Set of time periods**.

Variables

x^s	Installed capacity of storage. [MW]
$p_t^{c/d}$	Power dis-/charged of storage in period t. [MW]
e_t^s	Energy stored at period t. [MWh]

Data

$C^{\text{inv},s}$	Investment cost of storage. [€/MW]
$C^{\text{c/d},s}$	Storage dis-/chargin cost. [€/MWh]
$\eta^{c/d,s}$	Dis-/charging efficiency of storage. [p.u.]
τ	Discharge duration. [h]

Agenda

- **I** Formulate Full-Scale Optimization Model
- **II** Aggregated Optimization Models & Time Series Aggregation
- **III** Challenge!
- **IV** Solution (and Chocolate)

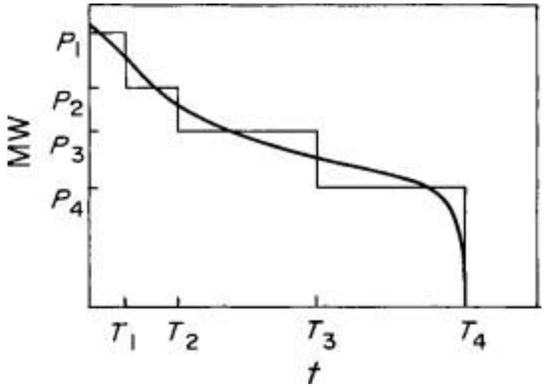
Source: Photo by Estée Janssens on Unsplash

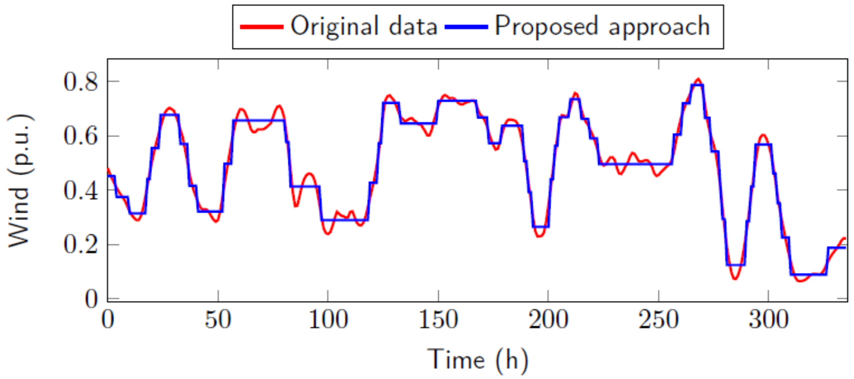
How to represent time in energy system models?

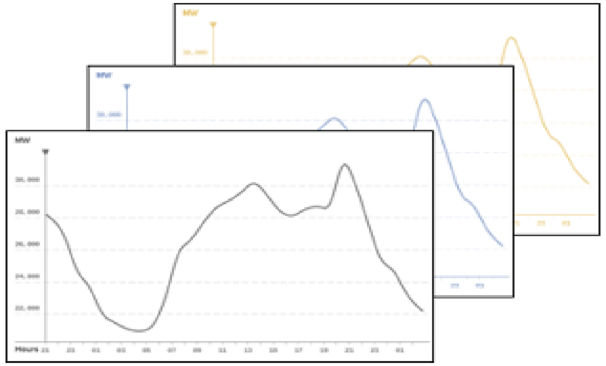
How to represent time in energy system models?
Especially in long-term problems

How to reduce temporal information in tactical and strategic planning models:

1 - Load Periods & Extended Versions	2 - Chronological Time Periods	3 - Representative Periods
Time slices, load duration curve, system states, etc.	Time periods with additional condition of chronology	E.g., days or weeks







How to formulate energy storage systems depending on temporal representation?

1 - Load Periods & Extended Versions

2 - Chronological Time Periods

3 - Representative Periods

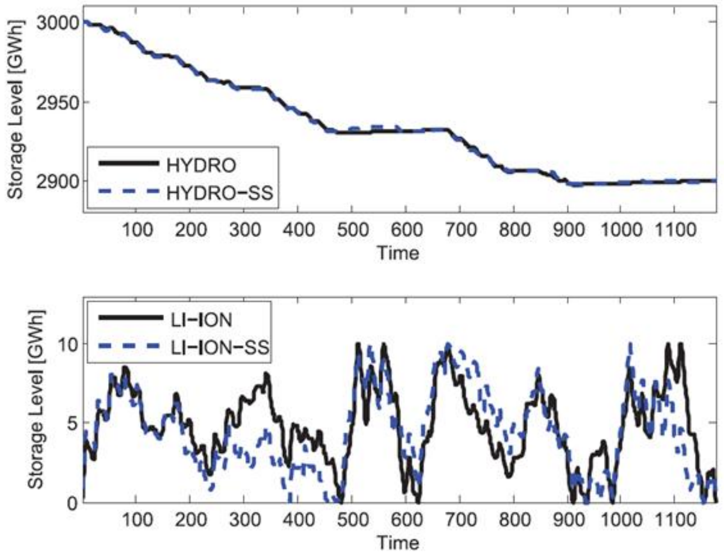
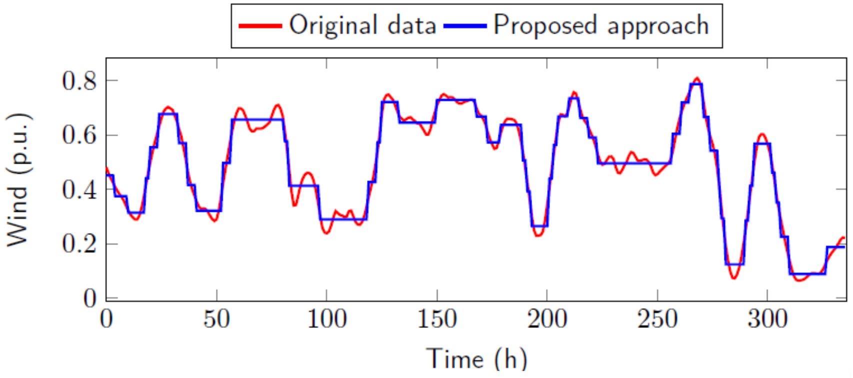
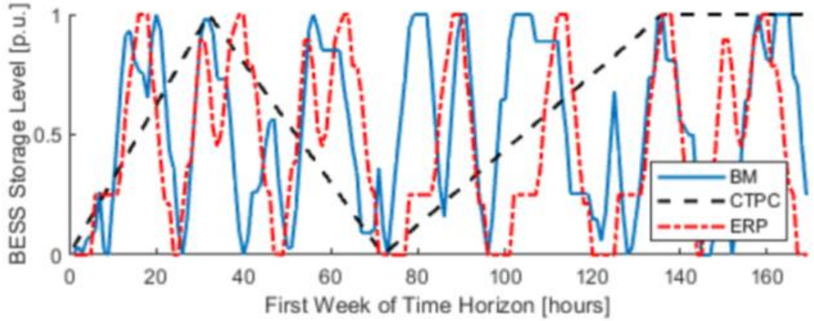


Fig. 3. Evolution of storage level over time for hydro reservoir and Li-ion battery of hourly (continuous line) and system states (discontinuous line) model.

Source: Wogrin, Sonja, et al. "A new approach to model load levels in electric power systems with high renewable penetration." *IEEE Transactions on Power Systems* 29.5 (2014): 2210-2218.





Source: S. Pineda and J. M. Morales. "Chronological Time-Period Clustering for Optimal Capacity Expansion Planning With Storage." in IEEE Transactions on Power Systems. vol. 33. no. 6. pp. 7162-7170. Nov. 2018.

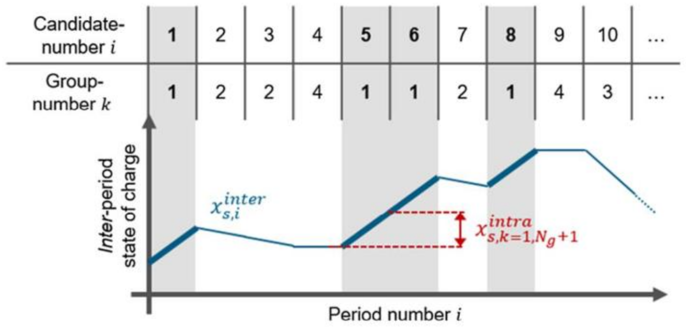


Fig. 2. Sketched high layer inter-period state x_i^{inter} based on the sequence of appearance of the representative periods k. This is highlighted for period or group number 1.

Sources:
Kotzur, Leander, et al. "Time series aggregation for energy system design: Modeling seasonal storage." *Applied energy* 213 (2018): 123-135.

Tejada-Arango, D.A., Domeshek, M., Wogrin, S. and Centeno, E., 2018. Enhanced representative days and system states modeling for energy storage investment analysis. IEEE Transactions on Power Systems, 33(6), pp.6534-6544.

Investment Optimization Model (without storage)

Aggregated assuming **Chronological Time Periods**

Objective Function (Aggregated total system cost)

$$\min_{z=\{\bar{x},\bar{p},\bar{d}\}} = \sum_{g} C^{\text{inv},g}\,\bar{x}^{g} + \sum_{g,k} W_k\, C^{\text{op},g}\,\bar{p}_k^{g}\Delta + \sum_{k} W_k\, C^{\text{nse}}\,\bar{e}_k^{\text{ns}}$$

- Energy Balance Constraint

$$\sum_{g} \bar{p}_k^{g}\Delta + \bar{e}_k^{\text{ns}} = \bar{D}_k \quad \forall k$$

- Upper Bounds on Production:

$$\bar{p}_k^{g} \le \overline{CF}_k^{g}\bar{x}^{g} \quad \forall g, k$$

- Non-Negativity Constraints:

$$\bar{p}_k^{g} \ge 0 \quad \forall g, k$$
$$\bar{e}_k^{\text{ns}} \ge 0 \quad \forall k$$
$$\bar{x}^{g} \ge 0 \quad \forall g$$

Sets

g	Set of generators.
k	Set of **aggr. chron.** time periods. (|k|<<|t|)

Variables

$\bar{x}^{g}$	Installed capacity of generator g. [MW]
$\bar{p}_k^{g}$	**Aggregated** power of g in period k. [MW]
$\bar{e}_k^{\text{ns}}$	**Aggregated** non-supp. demand in k.[MWh]

Data

$C^{\text{inv},g}$	Investment cost of g. [€/MW]
$C^{\text{op},g}$	Operating cost of generator g. [€/MWh]
C^{nse}	Cost of non-supplied energy. [€/MWh]
$\bar{D}_k$	**Aggregated** demand in k. [MWh]
$\overline{CF}_k^{g}$	**Aggregated** capacity factor of g in k.
W_k	Weight of representative period. [h]

Investment Optimization Model (adding storage)

Aggregated assuming **Chronological Time Periods**

Objective Function (Aggregated total system cost)

$$\min_{z=\{\bar{x},\bar{p},\bar{d}\}} = \ldots + C^{\mathrm{inv},s}\,\bar{x}^s + \sum_k W_k\, C^{\mathrm{c},s}\,\bar{p}^c_k \Delta + \sum_k W_k\, C^{\mathrm{d},s}\,\bar{p}^d_k \Delta$$

- Energy Balance Constraint

$$\left(\sum_g \bar{p}^g_k + \bar{p}^d_k - \bar{p}^c_k\right)\Delta + \bar{e}^{\mathrm{ns}}_k = \bar{D}_k \;\; \forall k$$

- Upper Bounds Storage:

$$\bar{e}^s_k \le x^s \tau \quad \forall k$$

$$\bar{p}^c_k \le \bar{x}^s \;\; \forall k$$

$$\bar{p}^d_k \le \bar{x}^s \;\; \forall k$$

- Energy Storage Dynamics:

$$\bar{e}^s_{k+1} = \bar{e}^s_k + \left(\eta^{c,s}\bar{p}^c_k - \frac{\bar{p}^d_k}{\eta^{d,s}}\right) W_k \Delta \quad \forall k \setminus \mathrm{K}$$

$$\bar{e}^s_K = \bar{e}^s_1, \bar{e}^s_1 = 0.$$

- Non-Negativity Constraints:

$$\bar{p}^c_k \ge 0 \;\; \forall k \qquad \bar{e}^s_k \ge 0 \;\; \forall k$$

$$\bar{p}^d_k \ge 0 \;\; \forall k \qquad \bar{x}^s \ge 0$$

Sets

g	Set of generators.
k	Set of **aggregated** time periods. (|k|<<|t|)

Variables

$\bar{x}^s$	Installed capacity of storage. [MW]
$\bar{p}^{c/d}_k$	**Aggregated** dis-/charged power in k. [MW]
$\bar{e}^s_k$	Energy stored over **aggregated peridos** k. [MWh]

Data

$C^{\mathrm{inv},s}$	Investment cost of storage. [€/MW]
$C^{\mathrm{c/d},s}$	Storage dis-/chargin cost. [€/MWh]
$\eta^{c/d,s}$	Dis-/charging efficiency of storage. [p.u.]
τ	Discharge duration. [h]

How to formulate energy storage systems depending on temporal representation?

1 - Load Periods & Extended Versions

2 - Chronological Time Periods

3 - Representative Periods

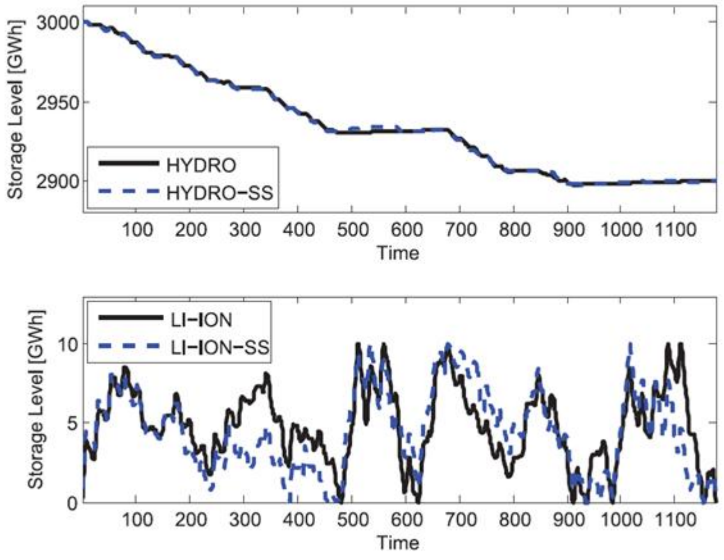
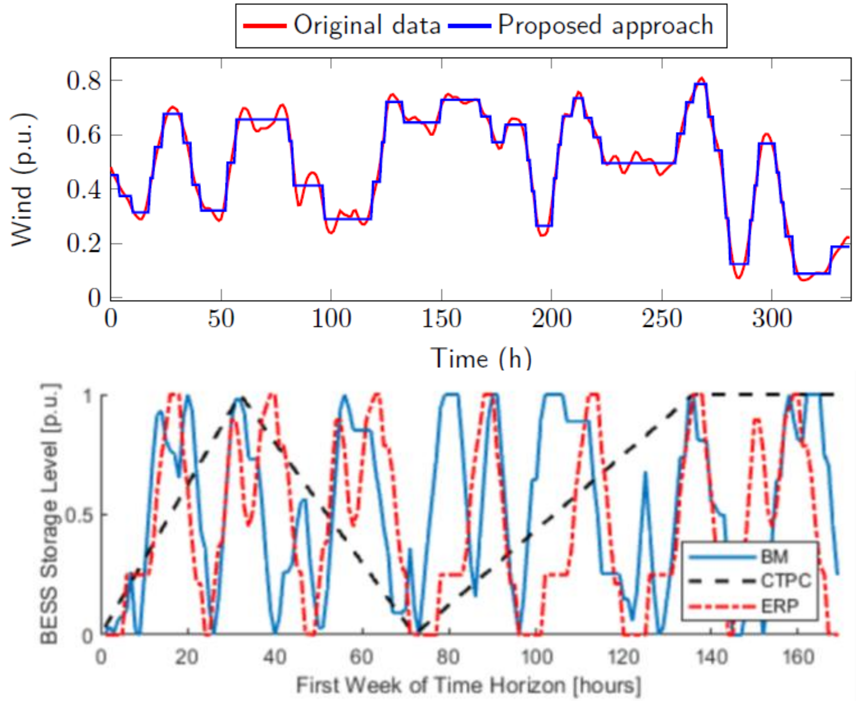


Fig. 3. Evolution of storage level over time for hydro reservoir and Li-ion battery of hourly (continuous line) and system states (discontinuous line) model.

Source: Wogrin, Sonja, et al. "A new approach to model load levels in electric power systems with high renewable penetration." *IEEE Transactions on Power Systems* 29.5 (2014): 2210-2218.



Source: S. Pineda and J. M. Morales. "Chronological Time-Period Clustering for Optimal Capacity Expansion Planning With Storage." in IEEE Transactions on Power Systems. vol. 33. no. 6. pp. 7162-7170. Nov. 2018.

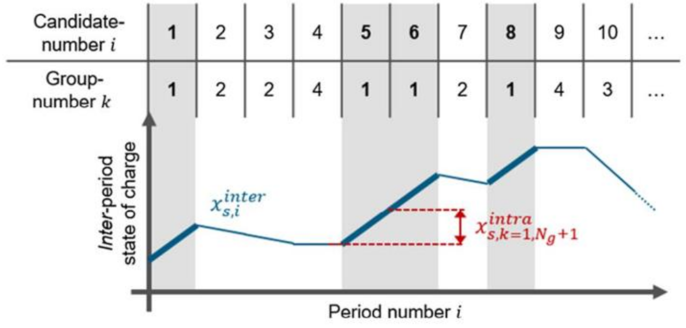


Fig. 2. Sketched high layer inter-period state x_i^{inter} based on the sequence of appearance of the representative periods k. This is highlighted for period or group number 1.

Sources:
Kotzur, Leander, et al. "Time series aggregation for energy system design: Modeling seasonal storage." *Applied energy* 213 (2018): 123-135.

Tejada-Arango, D.A., Domeshek, M., Wogrin, S. and Centeno, E., 2018. Enhanced representative days and system states modeling for energy storage investment analysis. IEEE Transactions on Power Systems, 33(6), pp.6534-6544.

Investment Optimization Model (without storage)

Aggregated assuming **Representative Days** (Kotzur et al.)

Objective Function (Aggregated total system cost)

$$\min_{z=\{\bar{x},\bar{p},\bar{d}\}} = \sum_{g} C^{\mathrm{inv},g}\,\bar{x}^{g} + \sum_{g,d} W_{d} \sum_{r} \left(C^{\mathrm{op},g} \bar{p}^{g}_{d,r} \Delta + C^{nse} \bar{e}^{\mathrm{ns}}_{d,r} \right)$$

- Energy Balance Constraint

$$\sum_{g} \bar{p}^{g}_{d,r} \Delta + \bar{e}^{\mathrm{ns}}_{d,r} = \bar{D}_{d,r} \;\; \forall d, r$$

- Upper Bounds on Production:

$$\bar{p}^{g}_{d,r} \leq \overline{CF}^{g}_{d,r} \bar{x}^{g} \quad \forall g, d, r$$

- Non-Negativity Constraints:

$$\bar{p}^{g}_{d,r} \geq 0 \;\; \forall g, d, r$$
$$\bar{e}^{\mathrm{ns}}_{d,r} \geq 0 \;\; \forall d, r$$
$$\bar{x}^{g} \geq 0 \;\; \forall g$$

Sets

d	Representative days.
n	Original time horizon days.
r	Representative time periods within d.
$f(n)$	Mapping from n to d.

n_1					n_2					n_3					…	n_N				
d_3					d_2					d_3					…	d_1				
1	2	3	…	R	1	2	3	…	R	1	2	3	…	R	…	1	2	3	…	R

Investment Optimization Model (adding storage)

Aggregated assuming **Chronological Time Periods**

Objective Function (Aggregated total system cost)

$$\min_{z=\{\bar{x},\bar{p},\bar{d}\}} = \ldots + C^{\text{inv},s}\,\bar{x}^s + \sum_{d,r} W_d\,(C^{c,s}\,\bar{p}^c_{d,r}\Delta + C^{d,s}\,\bar{p}^d_{d,r}\Delta)$$

- Energy Balance Constraint

$$\left(\sum_g \bar{p}^g_{d,r} + \bar{p}^d_{d,r} - \bar{p}^c_{d,r}\right)\Delta + \bar{e}^{\text{ns}}_{d,r} = \bar{D}_{d,r}\ \forall d,r$$

- Upper Bounds Storage:

$$\bar{e}^s_n \le x^s\tau \quad \forall n$$

$$\bar{p}^c_{d,r} \le \bar{x}^s\ \forall d,r$$

$$\bar{p}^d_{d,r} \le \bar{x}^s\ \forall d,r$$

- Intra- and Inter-day Energy Storage Dynamics:

$$\delta\bar{e}^s_{d,r+1} = \delta\bar{e}^s_{d,r} + \left(\eta^{c,s}\bar{p}^c_{d,r} - \frac{\bar{p}^d_{d,r}}{\eta^{d,s}}\right)\Delta \quad \forall d,r$$

$$\bar{e}^s_{n+1} = \bar{e}^s_n + \delta\bar{e}^s_{d,R} + \left(\eta^{c,s}\bar{p}^c_{d=f(n),R} - \frac{\bar{p}^d_{d=f(n),R}}{\eta^{d,s}}\right)\Delta \quad \forall n\backslash N$$

- Non-Negativity Constraints:

$$\bar{p}^c_{d,r} \ge 0\ \ \forall d,r$$

$$\bar{p}^d_{d,r} \ge 0\ \ \forall d,r$$

$$\bar{e}^s_n \ge 0\ \forall n$$

$$\bar{x}^s \ge 0$$

Sets

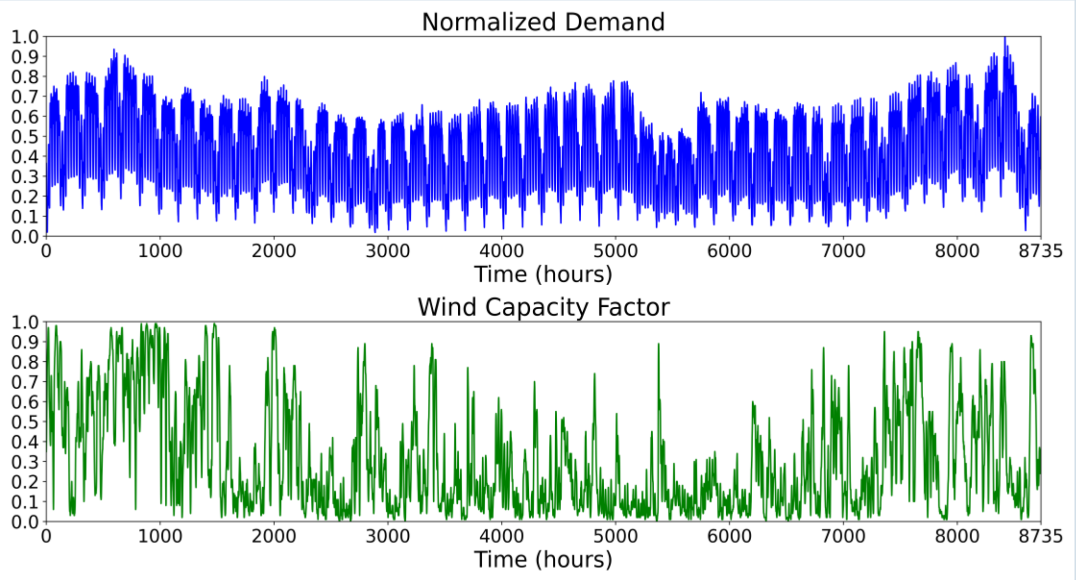
d	Representative days.
n	Original time horizon days.
r	Representative time periods within d.
$f(n)$	Mapping from n to d.

Variables

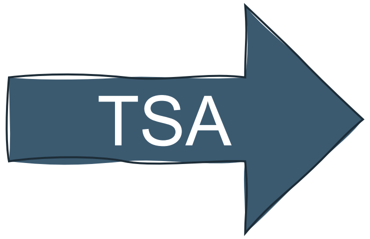
$\bar{x}^s$	Installed capacity of storage. [MW]
$\delta\bar{e}^s_{d,r}$	Difference in state of charge in time period r of repr. day d. [MW]
$\bar{e}^s_n$	Energy stored at the end of day n. [MWh]

Time Series Aggregation (TSA)

Original Data Time Series

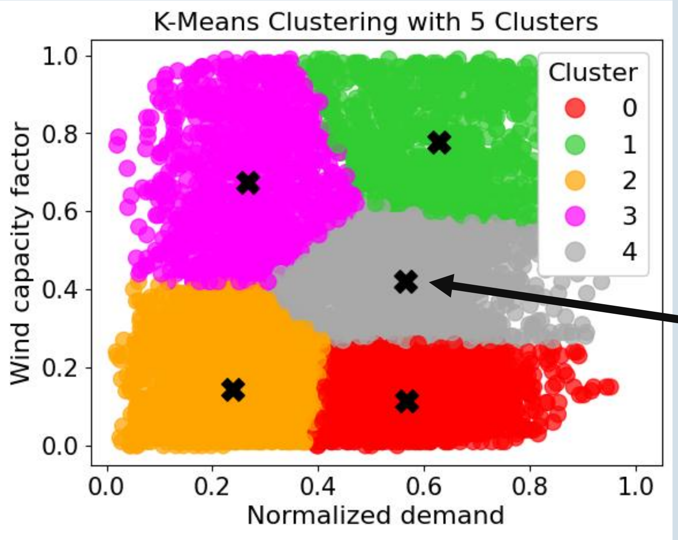


Focus on reducing dimensionality of input data.



Times Series Aggregation e.g. using k-means with 5 clusters

You obtain:



5 cluster **centroids**. Values for $\overline{D}_k$ & $\overline{CF}_k$

A **mapping** that matches every original time period to its cluster.

{t1:0,t2:0,…,t37:1 ,…, t8736:3}

This also yields the **weights** W_k of each cluster.

Sources:
Teichgraeber, H. and A.R. Brandt. "Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities." Renewable and Sustainable Energy Reviews (2022)
Li, C. et al. "On representative day selection for capacity expansion planning of power systems under extreme operating conditions." International Journal of Electrical Power & Energy Systems (2022)
Hoffmann, M. et al. "A review on time series aggregation methods for energy system models." Energies (2020)
Hilbers, A.P. et al. "Importance subsampling: improving power system planning under climate-based uncertainty." Applied Energy (2019).

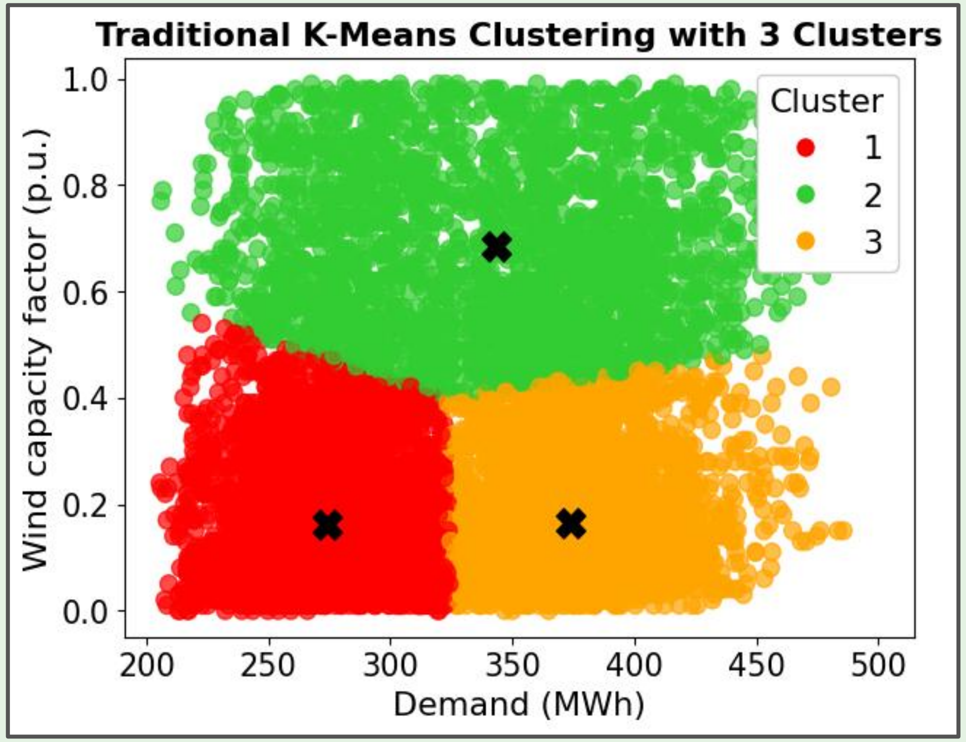
K-means

K-means is an algorithm that **partitions the data into K clusters** by iteratively assigning each data point to the nearest cluster centroid and updating cluster centroids based on the mean of the assigned points. It minimizes overall distance of cluster points to their centroid.

$$\min J = \min \sum_{i=1}^{K} \sum_{\boldsymbol{x} \in S_i} ||\boldsymbol{x} - \boldsymbol{\mu}_i||^2$$

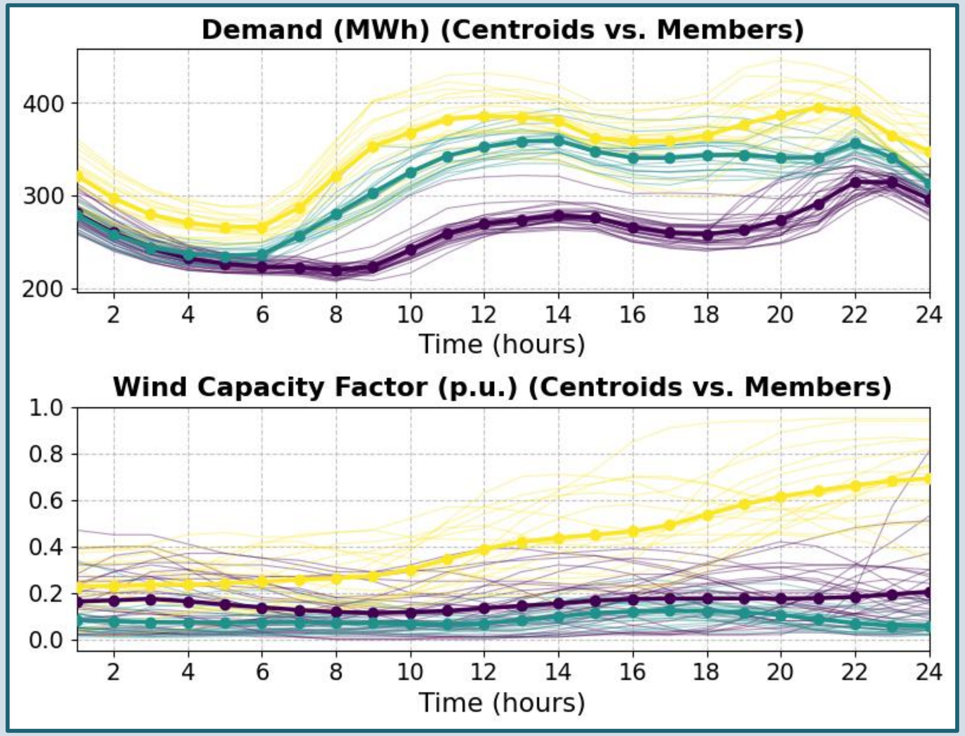
Representative hours

Approximate 8760 hours of original time series using only **3 representative hours** with K-means clustering.
Warning: Does not preserve **chronology** among hours!



Representative days

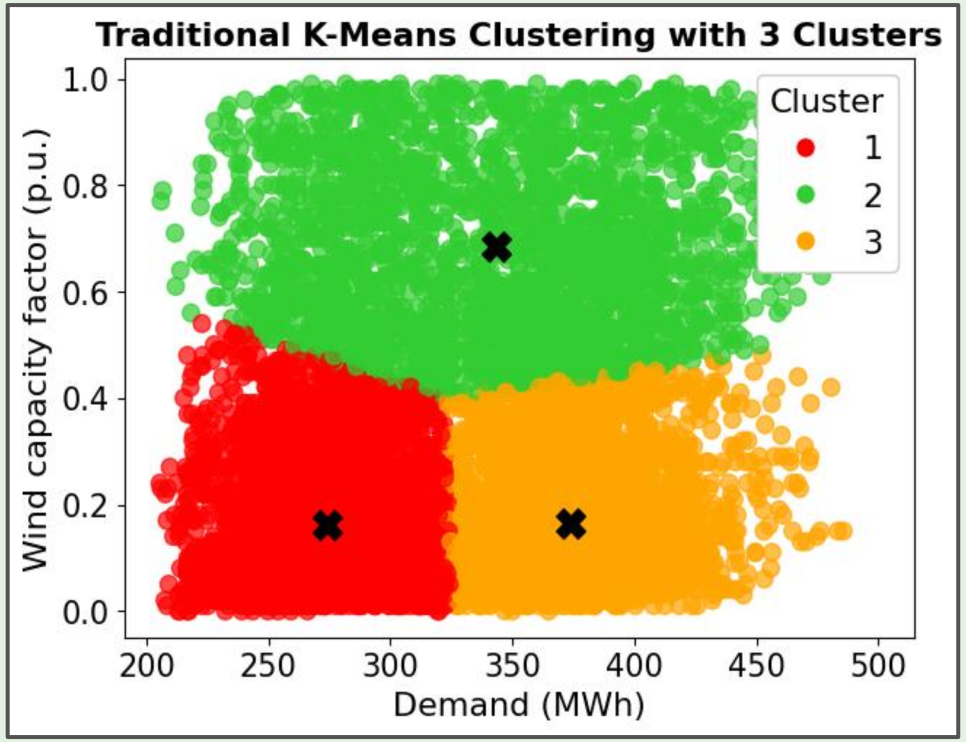
Use **3 representative days** for the whole year.
Warning: Chronology between days not preserved!
Requires **strong daily patterns** to be effective!



Chronological mapping

How to make K-means preserve chronology?

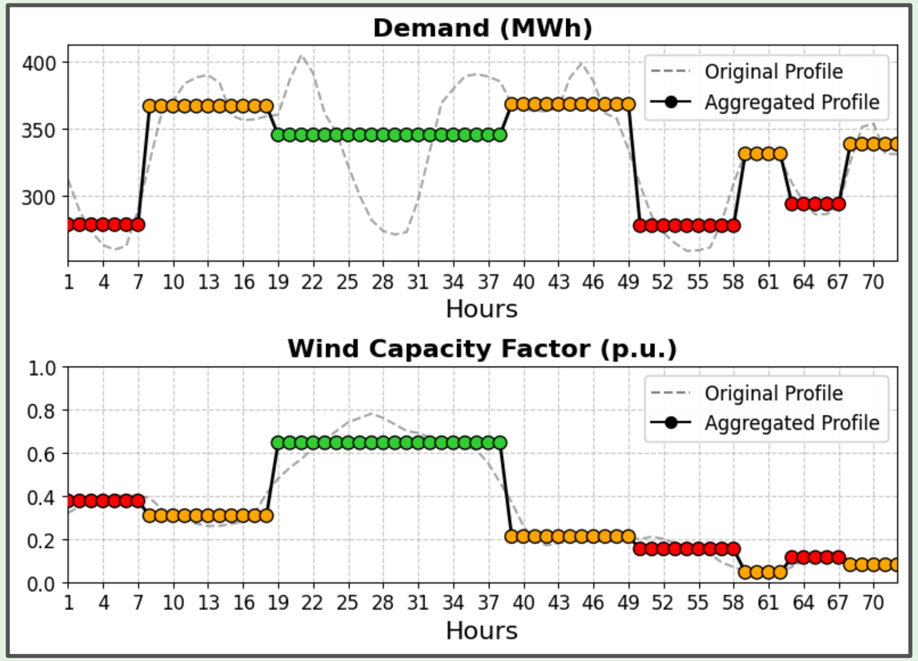
Traditional K-means



1. Take **Mapping**:
   {t1:1,t2:1,…, t37:2,…}
2. **Aggregate** consecutive hours with the same cluster

This yields 720 chronological representative timesteps

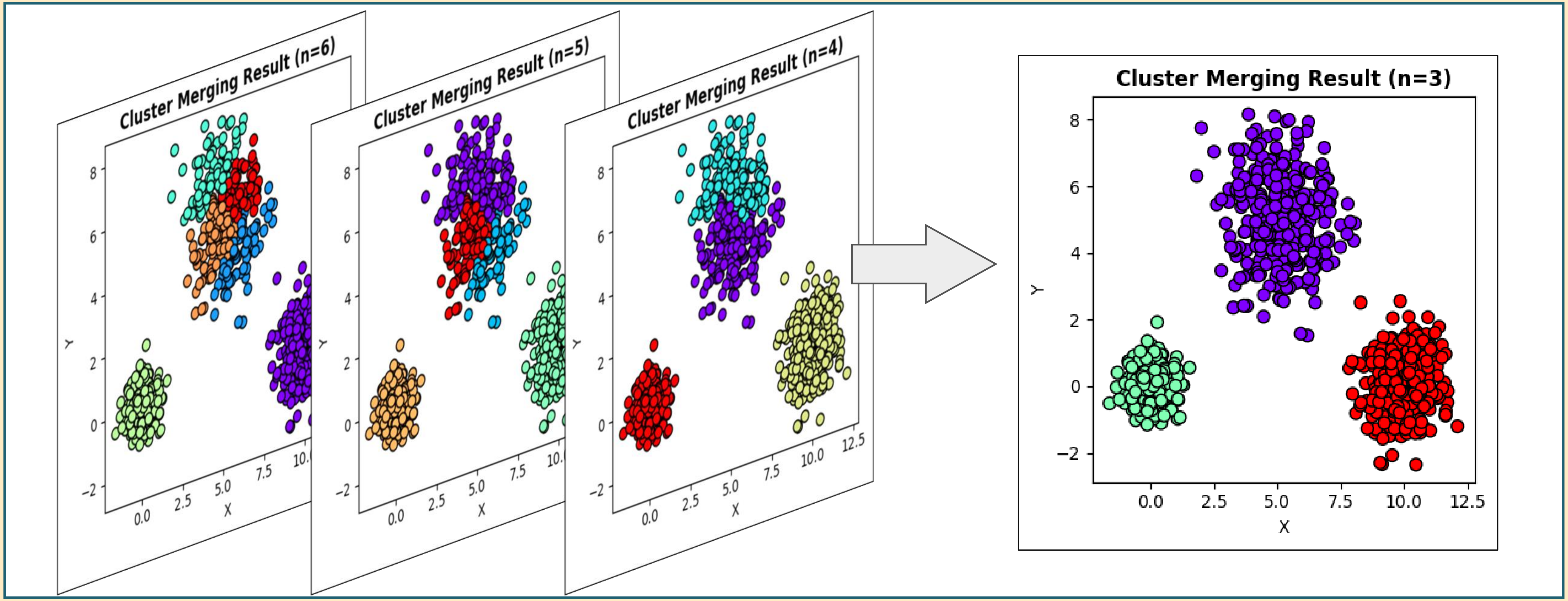
Chronological K-means



This gives very limited control over the aggregated model size!

Hierarchical clustering

Hierarchical clustering is an **unsupervised learning framework** that builds a hierarchy of clusters by **iteratively grouping** data points based on their similarity. Similarity is defined by **the distance between the centroids** of the clusters.



Chronological Hierarchical clustering

Algorithm 1: Chronological Hierarchical clustering

Input: T (Initial number of periods), K (Reduced number of periods), Time series data.

Output: Representative centroids x_k, Cluster weights W_k.

Initialization:;

Set the initial number of clusters $n \leftarrow T$;

Assign each hour to its own cluster $k \in \{1, \ldots, T\}$;

while $n > K$ **do**

- Determine the centroid x_k of each cluster k using by calculating mean $\overline{x}_k$;
- Compute the dissimilarity $D(k, l)$ between each pair of **adjacent** clusters k, l according to Ward's method:

$$D(k,l) = \frac{2|k||l|}{|k| + |l|} ||\overline{x}_k - \overline{x}_l||^2$$

- Identify the two closest adjacent clusters (k', l') such that:

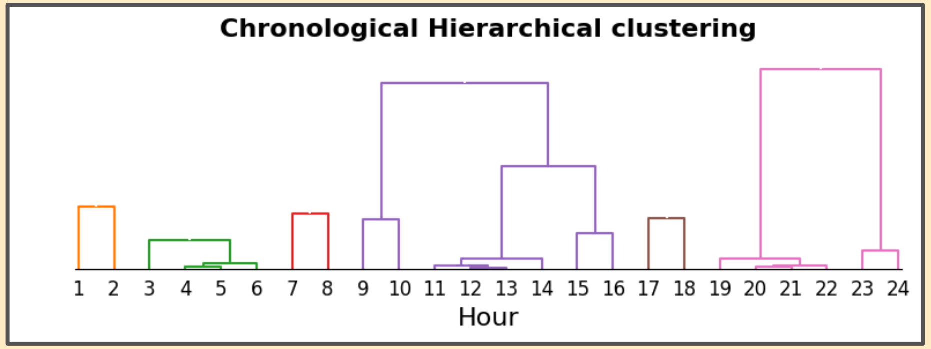
$$(k', l') \in \text{argmin } D(k,l) \quad \text{s.t. } l \in A(k)$$

  where $A(k)$ is the set of clusters containing hours consecutive to those in k;
- Merge clusters k' and l' into a single cluster;
- Update the cluster count: $n \leftarrow n - 1$;

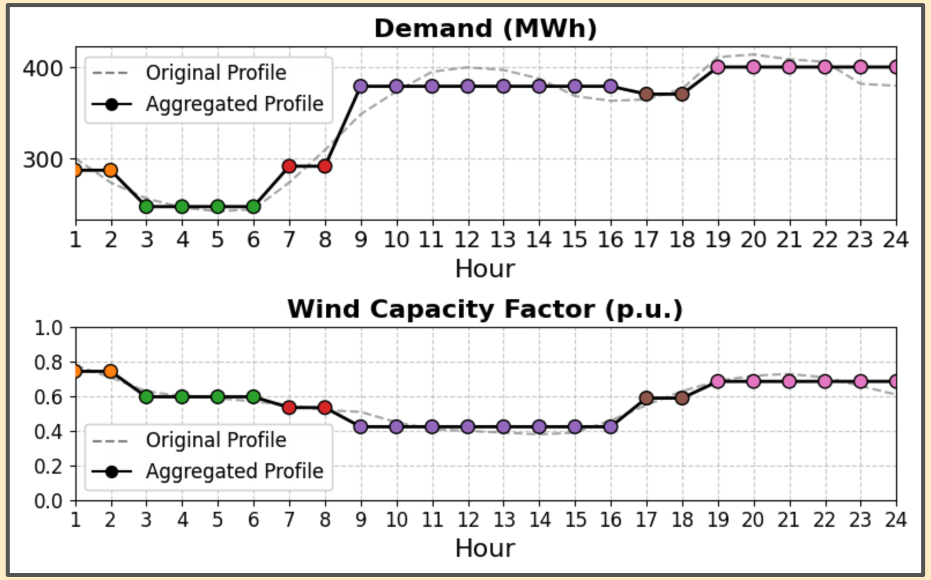
Calculate the time-period duration W_k as the number of hours belonging to each cluster k;

Source:
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.

Algorithm **iteratively merges two closest** timesteps until reaches **target number** of representative hours.



24 hours be approximated by **6 consecutive representative timesteps**



Agenda

I	Formulate Full-Scale Optimization Model
II	Aggregated Optimization Models & Time Series Aggregation
III	Challenge!
IV	Solution (and Chocolate)

Source: Photo by Estée Janssens on Unsplash

7.1. Challenge!

Create your own **mapping** (Maximum of 500 time steps)

Open Colab

https://colab.research.google.com/drive/1b5bc4P-H-8XP_VyYKzUwbkfFY2kppMME?usp=sharing

7.1: How to verify your solution?

Submit 7.1. Results

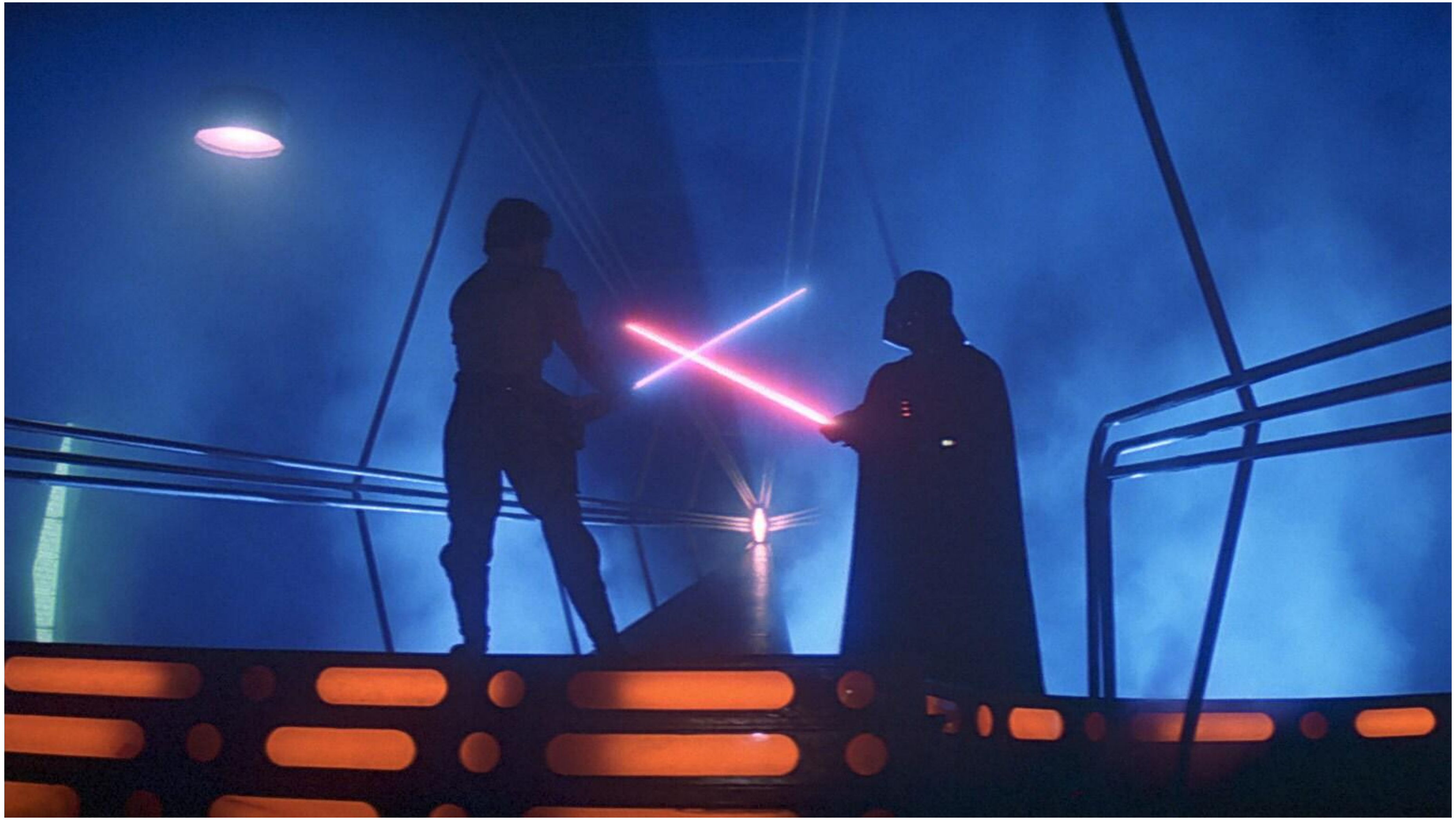
Please submit your output error here:

https://forms.gle/2P4dJJzJMSuLW2CMA

7.2. More Challenge?

Time series aggregation with **performance guarantees**

Same Colab 7.2.



https://colab.research.google.com/drive/1b5bc4P-H-8XP_VyYKzUwbkfFY2kppMME?usp=sharing

7.2: How to verify your solution?

Submit 7.2 Results

Please submit your output error here:

https://forms.gle/UWf7vt7ZjzSCpkdc6

Agenda

I Formulate Full-Scale & Aggregated Investment Optimization Model

II Time Series Aggregation & Aggregated Problem

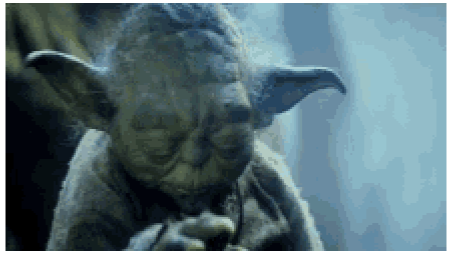
III Challenge!

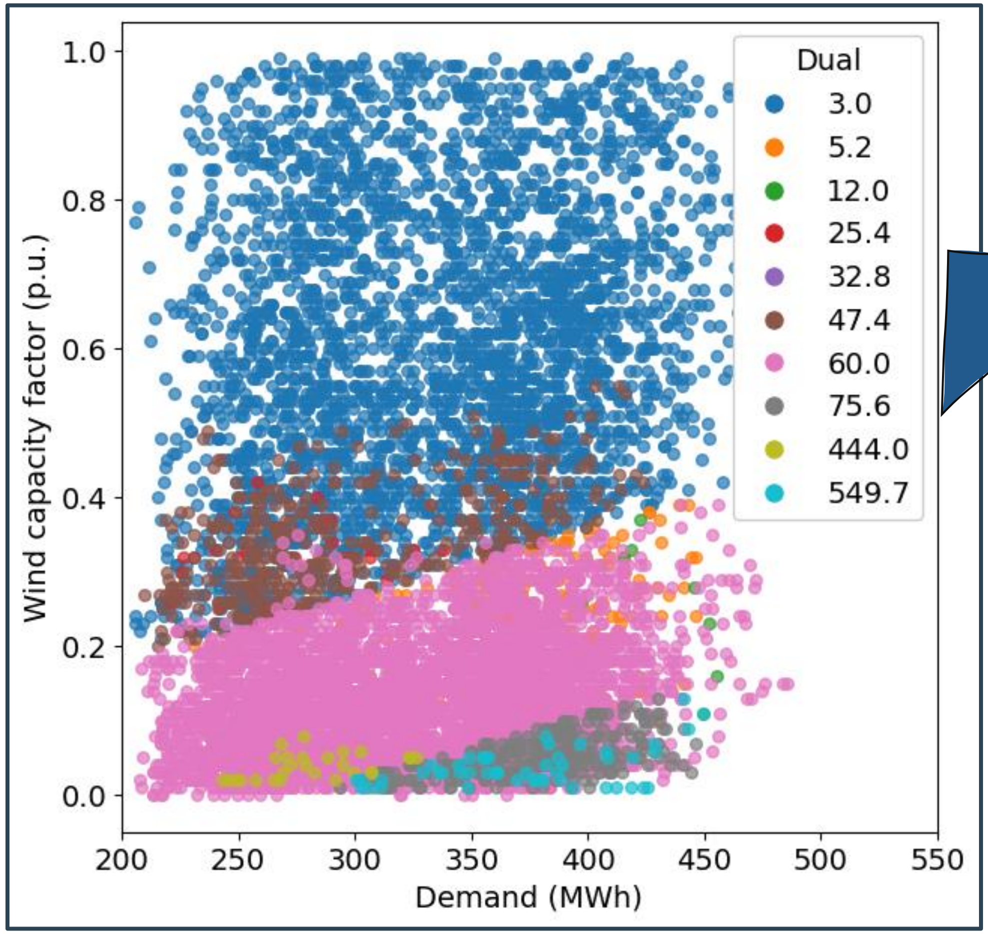
IV Solution (and Chocolate … finally)

Source: Photo by Estée Janssens on Unsplash

Challenge 7.1. Solution

Leverage dual variable values you must





Leverage the **dual variable values** of the optimization model (in particular, of the energy balance constraints).

1. Retrieve the dual variable values for the energy balance constraints (i.e., system marginal costs (MCs))
2. We observe only **10 different MC values**.
3. Group **consecutive** time steps with same MC into the same cluster. This yields **409 consecutive time steps**.

This approach achieves **exact aggregation** (i.e., output error = 0), reducing the number of representative periods **from 8736 to just 409**.

Challenge 7.1. Solution

Where do these 10 different values come from?

These 10 clusters correspond to the different active constraint sets or system states (operating and investment states)

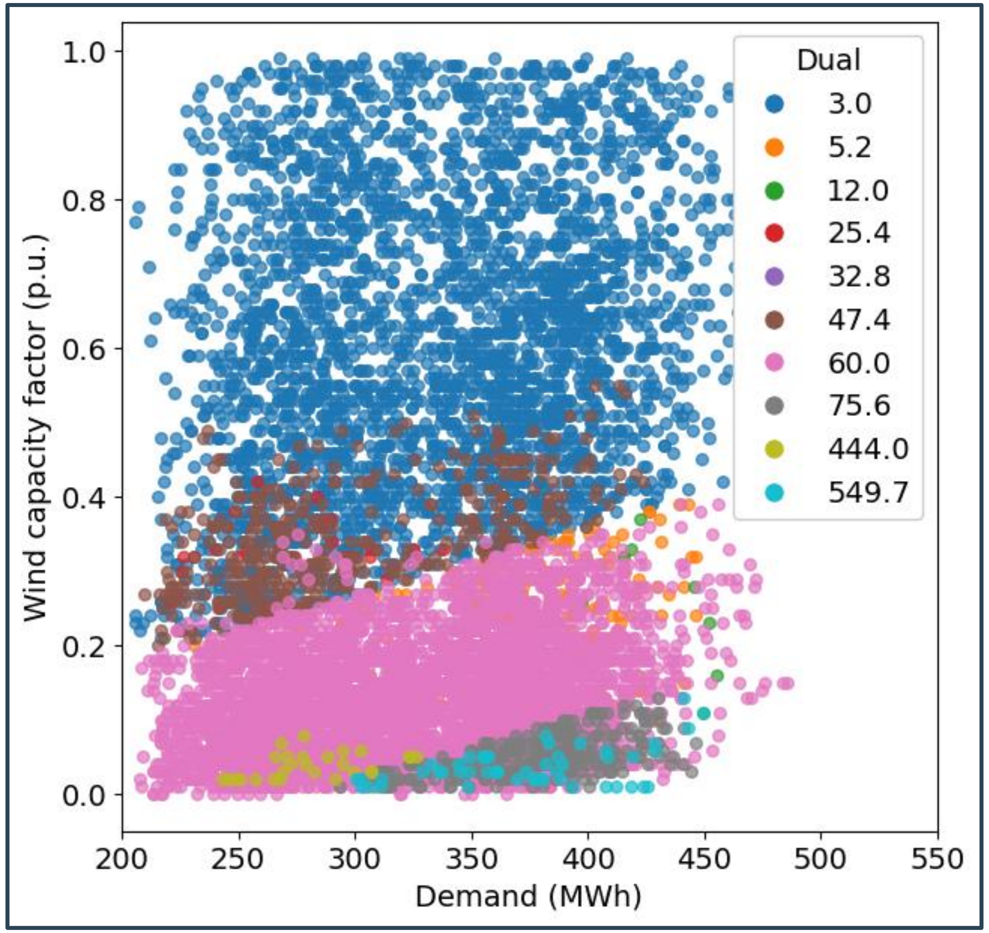
Operating states:

1. **Wind** marginal operator (MC = 3, wind variable cost)
7. **Thermal** marginal operator (MC=60, thermal VC)
6. **Storage charging** (MC = 47.4, if charged with wind)
2. **Storage discharging** (MC = 5.2 if charged with wind)
8. **Storage discharging** (MC = 75.6 charged with thermal)

Investment states:

Storage and wind investment (3., 4. and 5.) (MC = 12.0, MC = 25.4 and 32.8, includes all operating, charging/discharging & investment cost)

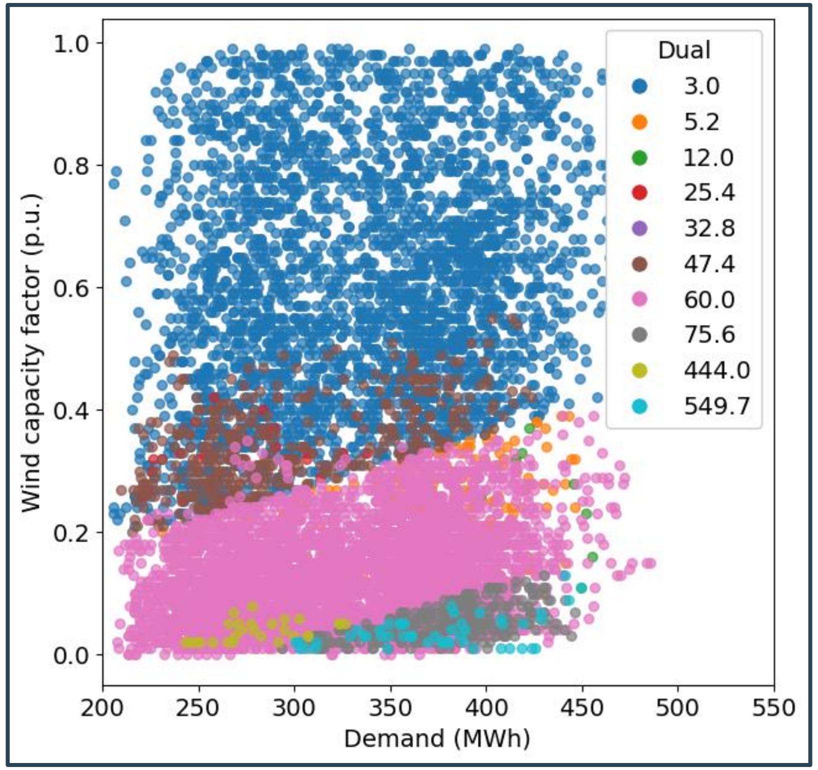
Storage and thermal investment (9. and 10.) (MC = 444.0 and 549.7, thermal operating & investment costs)



How does this work?

It‘s math

Theorem. *For each* $i = 1, \ldots, I$ *consider the following LPs* (E_i): $\min c^T x$ *s.t.* $Ax = b_i$, *where* $x \in \mathbb{R}^n, A \in \mathbb{R}^{mxn}, c \in \mathbb{R}^n, b_i \in \mathbb{R}^m$ *and the LPs only differ in the RHS values* b_i. *Then,* B *is also an optimal basis for the problem* $(\overline{E})$: $\min c^T x$ *s.t.* $Ax = \mathbb{E}(b_i)$, *where* $\mathbb{E}(b_i) = \sum_i \frac{b_i}{I}$.



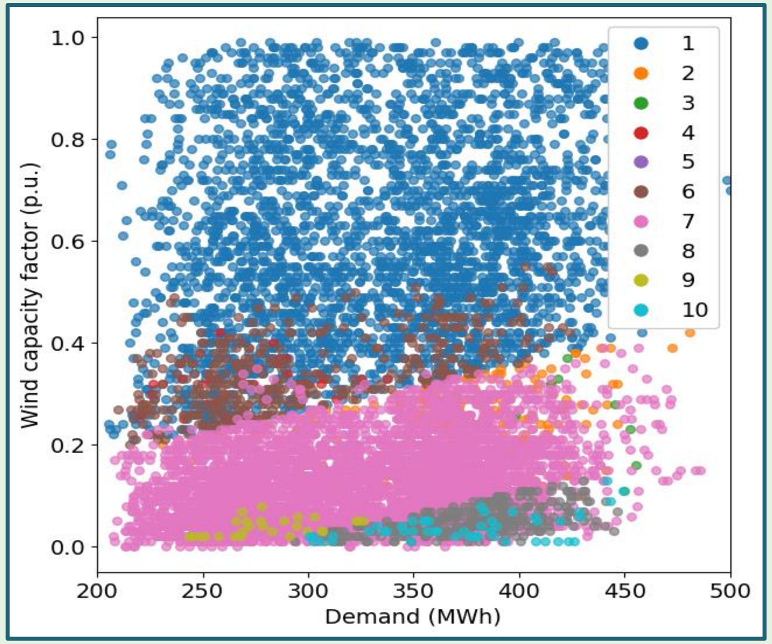
Take-aways

- Time periods with same color have the same **active constraint set**.
- The **Theorem** states that those hours can be approximated by its average hour, without incurring an error in the outputs (on average).
- This problem has **10** unique hourly **active constraint sets** (simplex bases).

Comparison Challenge 7.1.

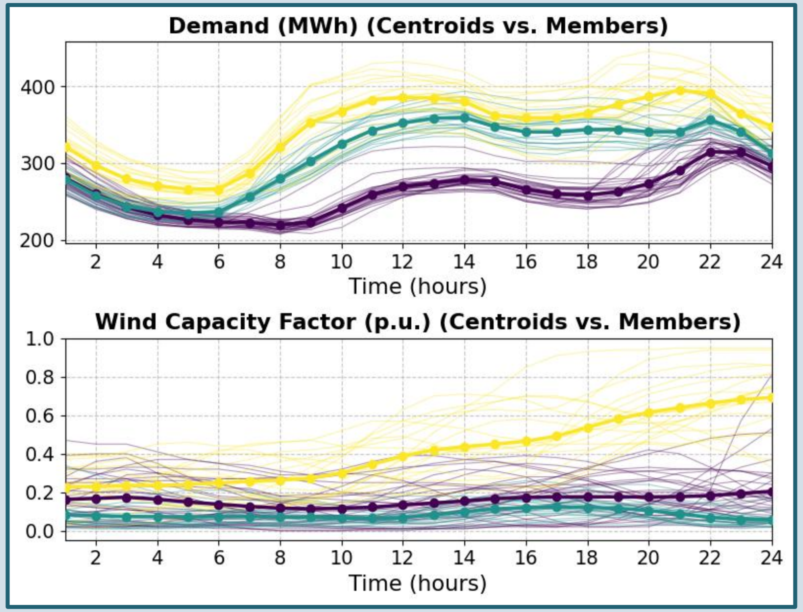
A-posteriori Method vs a-priori Methods

Proposed Solution



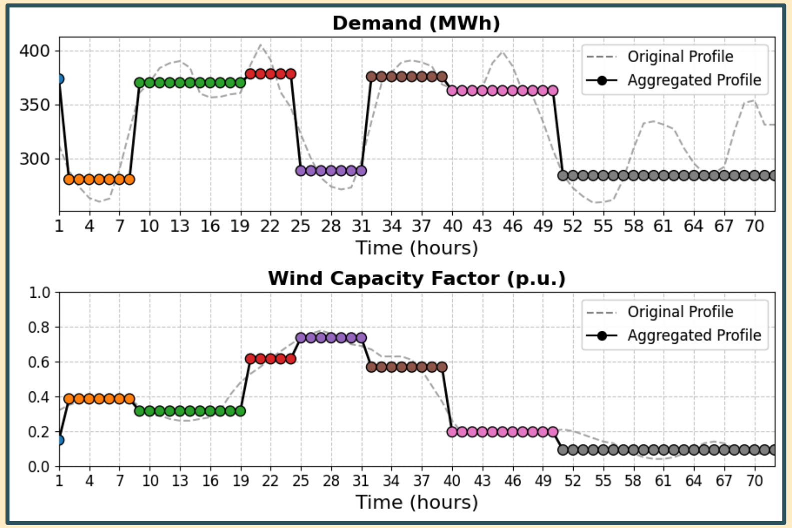
Variable	Value
OFV (M€)	119.22
Thermal (GWh)	863
Renewable (GWh)	2043
Storage Inv. (MW)	539
Renewable Inv. (MW)	1115
Thermal Inv. (MW)	277

K-Means Rep Days



Variable	Value	% Error
OFV (M€)	113.45	4.84
Thermal (GWh)	736	14.68
Renewable (GWh)	2145	4.99
Storage Inv. (MW)	76	85.82
Renewable Inv. (MW)	1321	18.45
Thermal Inv. (MW)	228.6	17.35

Chron. Hierarchical



Variable	Value	% Error
OFV (M€)	117.47	1.47
Thermal (GWh)	788	8.74
Renewable (GWh)	2105	3.03
Storage Inv. (MW)	470	12.84
Renewable Inv. (MW)	1208	8.27
Thermal Inv. (MW)	268	2.95

7.2. Challenge solution

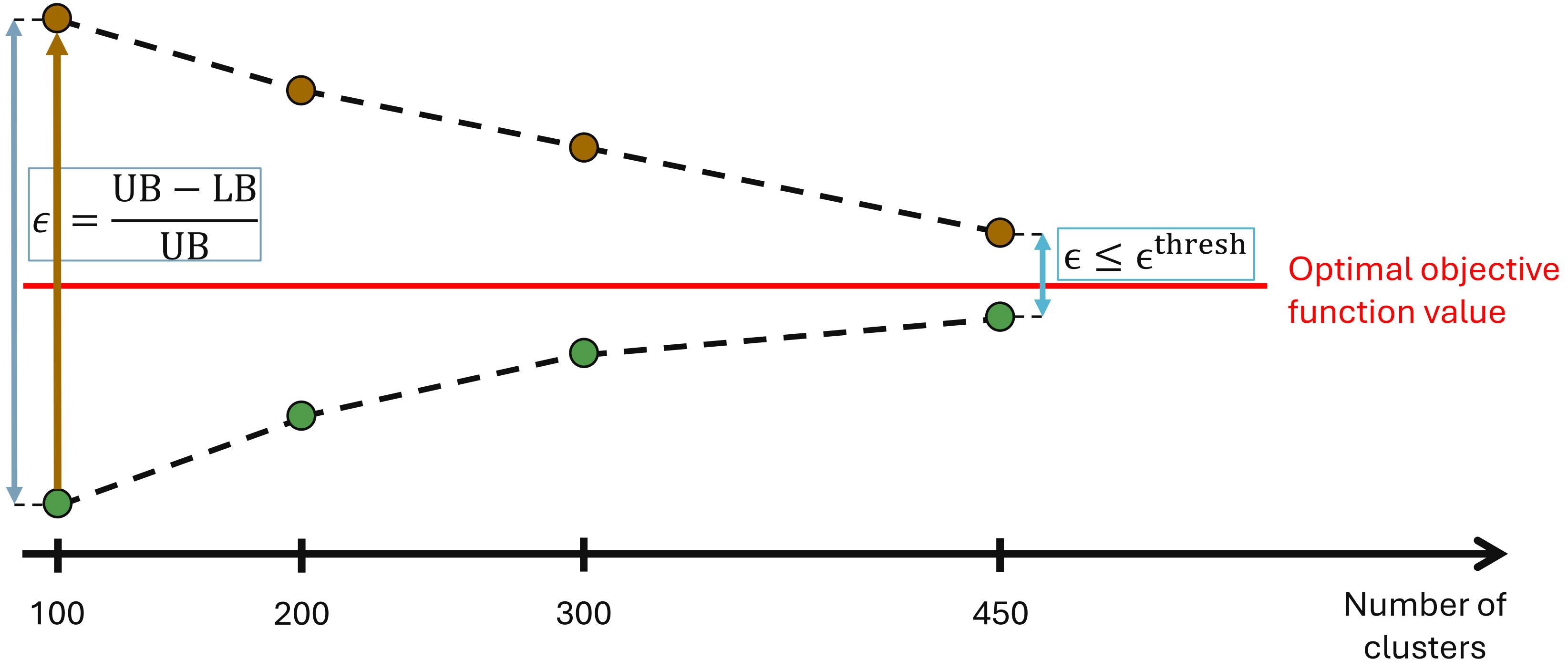
Upper and lower bound estimation

Solve **aggregated model** to **obtain lower bound** and investment decisions.

Fix investments in full-scale model. Solve resulting operational problem. The solution yields the **upper bound** for a true solution.

Calculate resulting **optimality gap** and refine aggregated model with larger number of clusters.

Repeat until **convergence**, i.e., desired optimality gap.



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

7.2. Challenge solution

Marginal cost-based time series aggregation

Predict **marginal costs** for whole time horizon

Merge together only time steps that have common marginal cost with exclusion of extreme time steps

Cluster by using **chronological hierarchical** clustering technique with exclusion of extreme time steps and Marginal cost-based Ward's distance

Build and solve aggregated model to obtain **lower bound**

Build and solve full-scale model with fixed capacity variables from aggregated model to obtain **upper bound**

Marginal costs prediction algorithm

Algorithm 1: Machine Learning Framework for Marginal Cost Prediction

Input: Input time series $\Omega^{\mathcal{T}} := \{F_{g,t}, D_t, \pi_t\}_{g\in\mathcal{G}, t\in\mathcal{T}}$, Number of time steps in submodel K.

Output: Predicted marginal costs $\bar{\boldsymbol{\mu}} := \{\bar{\mu}_t\}_{t\in\mathcal{T}}$.

1 $\bar{P}\left(\Omega^{\mathcal{T}}\right) \leftarrow$ Estimate input data distribution via Kernel Density Estimation $P\left(\Omega^{\mathcal{T}}\right) \approx \bar{P}\left(\Omega^{\mathcal{T}}\right)$;

2 $\Omega^{\mathcal{K}} \leftarrow$ Draw K samples from $\bar{P}\left(\Omega^{\mathcal{T}}\right)$ to generate a reduced time series $\Omega^{\mathcal{K}} := \left\{\bar{P}(\Omega^{\mathcal{T}})_k\right\}_{k\in\mathcal{K}}$, where $K \ll |\mathcal{T}|$;

3 $\tilde{\boldsymbol{\mu}} \leftarrow$ Solve the GEP model (1) for each sample in $\Omega^{\mathcal{K}}$ to obtain marginal cost estimates $\{\tilde{\mu}_k\}_{k\in\mathcal{K}}$;

4 $\mathcal{C} \leftarrow$ Train a Random Forest classifier $\mathcal{C} : \Omega^{\mathcal{T}} \rightarrow \bar{\boldsymbol{\mu}}$ using the labeled pairs $\left(\Omega^{\mathcal{K}}, \tilde{\boldsymbol{\mu}}\right)$ as training data;

5 $\bar{\boldsymbol{\mu}} \leftarrow$ Perform inference for the full planning horizon $\mathcal{T}$ using the trained classifier $\mathcal{C}$ s.t. $\bar{\mu}_t = \mathcal{C}(\Omega^{\mathcal{T}}_t)$ for all $t \in \mathcal{T}$;

With solution of submodel preserving input data distribution, you can train ML classifier (e.g. Random Forest) to predict marginal costs for whole time horizon.

Congrats to Challenge winners!

Sonja Wogrin
Univ.-Prof. Dipl.-Ing. Dr. MSc.

Graz University of Technology

Institute of Electricity Economics and Energy Innovation
Inffeldgasse 18
8010 Graz
Austria

E-Mail: wogrin@tugraz.at
Web: iee.tugraz.at

linkedin.com/company/iee-tugraz

instagram.com/iee.tugraz

NETZERO-OPT

Funded by the European Union (ERC, NetZero-Opt, 101116212). *Views and opinions expressed are however those of the author(s) 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.*

DTU Summer School Tutorial

All links

- **Unsolved tutorial link:**
https://colab.research.google.com/drive/1b5bc4P-H-8XP_VyYKzUwbkfFY2kppMME?usp=sharing

- **Solved tutorial link:**
https://colab.research.google.com/drive/1jfFNuntrZ_fhc9Xo_YKu-EIEXwLngmQT?usp=sharing

- **Submission for challenge 7.1. link:**
https://forms.gle/2P4dJJzJMSuLW2CMA

- **Submission for challenge 7.2. link:**
https://forms.gle/UWf7vt7ZjzSCpkdc6

QR for tutorial

QR for submission challenge 7.1

QR for submission challenge 7.2