Chronological Time-Period Clustering for Optimal Capacity Expansion Planning With Storage

EURO 2018

S. Pineda   J. M. Morales

OASYS group, University of Málaga (Spain)

Funded by the Spanish Ministry of Economy, Industry and Competitiveness through projects ENE2016-80638-R and ENE2017-83775-P

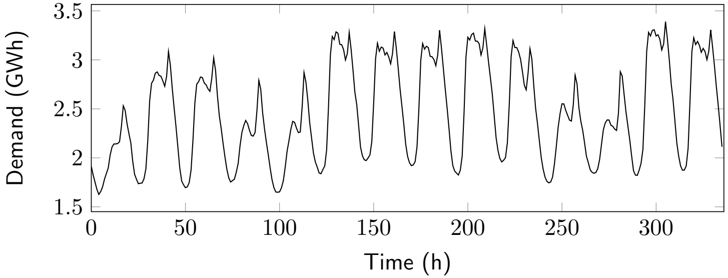
July 10, 2018

Which are the main problems in power systems?

	OPERATION	PLANNING
Horizon	1 second - 1 week	1 year - 20 years
Decisions	Generation dispach Power flows	Generation investments Line investments
Objective	Min production cost	Min prod. + inv. cost
Constraints	Generation = Demand Unit technical limits Line technical limits	Generation = Demand Unit technical limits Line technical limits
Comput. burden	Medium	Very high

How are planning problems usually solved?

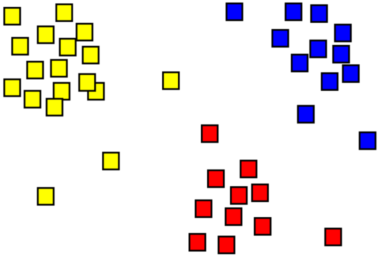
- Taking advantage of the fact that the electrical demand shows strong daily, weekly and annual patterns.



- Using statistical learning techniques such as clustering to group time periods and reduce the computational cost of the planning problem.

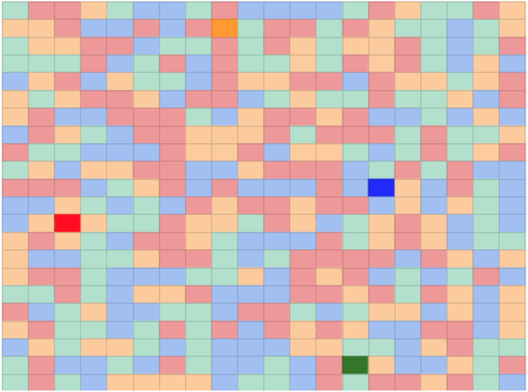
what is clustering?

- The clustering consists in the task of grouping a set of objects in such a way that the members of the same group (called cluster) are more similar, in one way or another.



How is clustering used in planning problems?

- The most common approach is to group the days of the time horizon into a small number of clusters and solve the optimization problem considering only the representative days.



How do I know if two days are similar or not?

- Each day is characterized by normalized demand, wind and solar

$$x_1 = [\underbrace{0.4, 0.6, \ldots, 0.5}_{\text{24 demand values}}, \underbrace{0.1, 0.2, \ldots, 0.4}_{\text{24 wind values}}, \underbrace{0.1, 0.2, \ldots, 0.1}_{\text{24 solar values}}]$$

$$x_2 = [\underbrace{0.6, 0.1, \ldots, 0.3}_{\text{24 demand values}}, \underbrace{0.4, 0.5, \ldots, 0.2}_{\text{24 wind values}}, \underbrace{0.3, 0.4, \ldots, 0.2}_{\text{24 solar values}}]$$

- The similarity between two days is computed using a norm

$$d(x_1, x_2) = ||x_1 - x_2||_2$$

How does the clustering algorithm work?

1) Set the initial number of clusters n to the number of days N.
2) Determine the centroid $\overline{\mathbf{x}}_I$ of each cluster I as

$$\overline{\mathbf{x}}_I = \frac{1}{|I|} \sum_{i \in I} \mathbf{x}_i$$

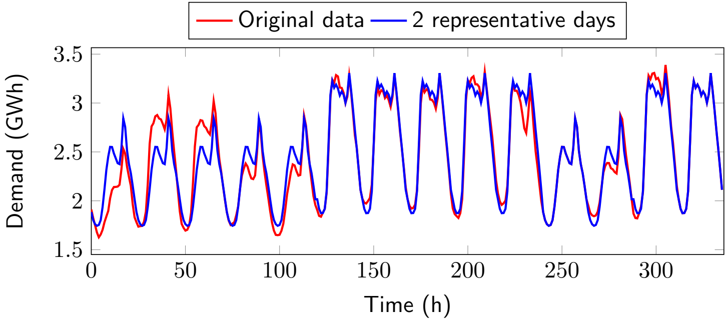
3) Compute the dissimilarity between each pair of clusters I, J according to Ward's method as follows

$$D(I,J) = \frac{2|I||J|}{|I| + |J|} ||\overline{\mathbf{x}}_I - \overline{\mathbf{x}}_J||^2$$

4) Merge the two closest clusters (I', J') according to the dissimilarity matrix, i.e., $(I', J') \in \operatorname{argmin} D(I,J)$ s.t. $I \neq J$.
5) Update $n \leftarrow n - 1$.
6) If $n = N'$ go to step 7). Otherwise go to step 2).
7) Determine the representative days as the elements with minimum dissimilarity to the rest of elements in each cluster (medoid).
8) The number of days belonging to each cluster determines the weight factor of each representative day.

How do the representative days approach work?

- As the figure below shows, using representative days works quite well



- Instead of 14 days (336 hours), we use 2 representative days (48 hours) to reduce the computational burden.

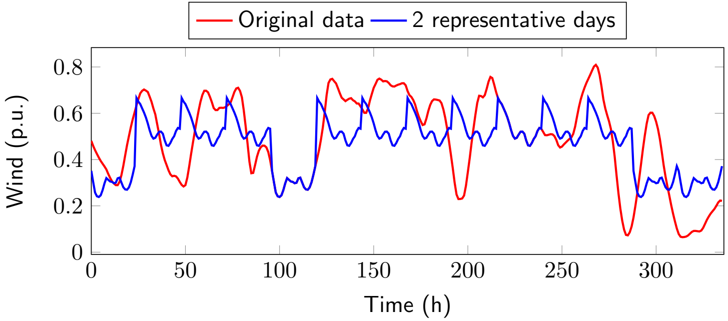
And what about current power systems?

- Current power systems have new actors

- Renewable generation is free and reduce CO2 emissions
- Renewable generation may happen at the wrong time
- Storage energy systems are the perfect partner of renewables:
  - If the wind blows and the demand is low, the battery stores energy
  - If the wind does not blow and the demand is high, we use the energy of the battery

Can we still use representative days?

- Some renewables do not present a strong daily partern



- The energy stored by some storage energy systems can be used several days later.

What do we propose?

- Instead of using representative days, we propose a new clustering methodology to group **consecutive hours** and maintain chronology.
- By doing so we can capture the **longer dynamics** of power generation from renewable sources such as wind
- In addition, we can model the operation of the batteries more accurately since we maintain the **chronology** of the data

How do I know if two consecutive hours are similar or not?

- Each hour is characterized by normalized demand, wind and solar

$$x_1 = [\underbrace{0.4}_{\text{demand}}, \underbrace{0.1}_{\text{wind}}, \underbrace{0.2}_{\text{solar}}]$$

$$x_2 = [\underbrace{0.6}_{\text{demand}}, \underbrace{0.4}_{\text{wind}}, \underbrace{0.3}_{\text{solar}}]$$

- The similarity between two consecutive hours is computed using a norm

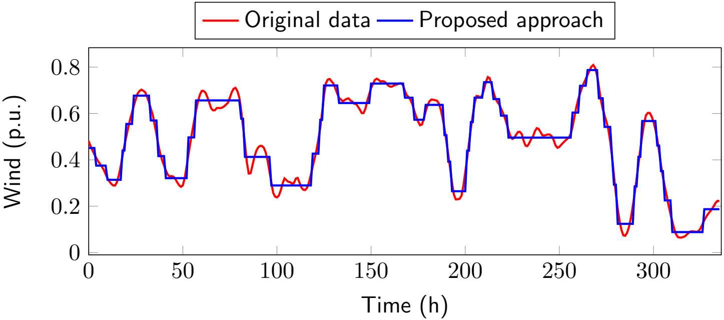
$$d(x_1, x_2) = ||x_1 - x_2||_2$$

How does the proposed clustering algorithm work?

1) Set the initial number of clusters n to the total number of hours N.
2) Determine the centroid of each cluster as $\overline{\mathbf{x}}_I = \frac{1}{|I|}\sum_{i \in I}\mathbf{x}_i$
3) Compute the dissimilarity between each pair of adjacent clusters I, J according to Ward's method as $D(I,J) = \frac{2|I||J|}{|I|+|J|}||\overline{\mathbf{x}}_I - \overline{\mathbf{x}}_J||^2$
4) Merge the two closest *adjacent* clusters (I', J') according to the dissimilarity matrix, i.e., $(I', J') \in$ argmin $D(I, J)$ s.t. $J \in \mathcal{A}(I)$, where $\mathcal{A}(I)$ is the set of clusters adjacent to cluster I. Two clusters I and J are said to be adjacent if I contains an hour that is consecutive to an hour in J, or vice versa, according to the original time series.
5) Update $n \leftarrow n - 1$.
6) If $n = N'$ go to step 7). Otherwise go to step 2).
7) Determine the representative periods as the clusters' centroids $\overline{\mathbf{x}}_I$.
8) The number of hours belonging to each cluster corresponds to the value of the time-period duration.

Does the proposed clustering work?

- With only 48 time periods (2 days) we managed to represent the wind much better



How does the aggregation affect the optimization model?

$$\min \sum_{gnt} \tau_t w_t c_g p_{gnt} + \sum_{nt} \tau_t w_t s c_n \left(\overline{d}_{nt} - d_{nt}\right) + \sum_{gn} \frac{i_g^G}{y_g^G} \overline{p}_{gn} + \sum_{n} \frac{i^H}{y^H} \overline{h}_n + \sum_{sn} \frac{i_s^S}{y_s^S} \overline{b}_{sn} + \sum_{nm} \frac{i^F l_{nm}}{y^F} \overline{f}_{nm}$$

s.t.

$$\sum_{g \in \mathcal{G}^r nt} \tau_t w_t p_{gnt} \geqslant \kappa \sum_{gnt} \tau_t w_t p_{gnt} \qquad 0 \leqslant \overline{p}_{gn} \leqslant \hat{p}_{gn}, \quad \forall g \in \mathcal{G}^r, n$$

$$0 \leqslant \overline{h}_n \leqslant \hat{h}_n, \quad \forall n \qquad 0 \leqslant \overline{f}_{nm} \leqslant \hat{f}_{nm}, \quad \forall n, m$$

$$\sum_g p_{gnt} + \sum_s \left(b_{snt}^- - b_{snt}^+\right) + h_{nt} = d_{nt} + \sum_m f_{nmt}, \forall n, t$$

$$0 \leqslant p_{gnt} \leqslant \rho_{gnt} \left(\overline{p}_{gn}^0 + \overline{p}_{gn}\right), \quad \forall g, n, t$$

$$-r_g^- \left(\overline{p}_{gn}^0 + \overline{p}_{gn}\right) \leqslant p_{gnt} - p_{gnt-1} \leqslant r_g^+ \left(\overline{p}_{gn}^0 + \overline{p}_{gn}\right), \quad \forall g, n, t \notin T_1$$

$$0 \leqslant h_{nt} \leqslant \overline{h}_n^0 + \overline{h}_n, \quad \forall n, t$$

$$\sum_t \tau_t w_t h_{nt} \leqslant a_n \sum_t \tau_t w_t \left(\overline{h}_n^0 + \overline{h}_n\right), \quad \forall n$$

$$0 \leqslant b_{snt}^+ \leqslant \overline{b}_{sn}^0 + \overline{b}_{sn}, \quad \forall s, n, t \qquad 0 \leqslant b_{snt}^- \leqslant \overline{b}_{sn}^0 + \overline{b}_{sn}, \quad \forall s, n, t$$

$$b_{snt} = b_{snt-1} - \tau_t b_{snt}^- + \tau_t \xi_s b_{snt}^+, \quad \forall s, n, t \notin T_1$$

$$0 \leqslant b_{snt} \leqslant \eta_s \left(\overline{b}_{sn}^0 + \overline{b}_{sn}\right), \quad \forall s, n, t$$

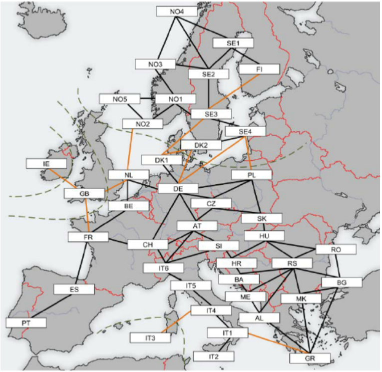
$$b_{snt} = b_{snt'}, \quad \forall s, n, (t, t') \in T_2$$

$$0 \leqslant d_{nt} \leqslant \overline{d}_{nt}, \quad \forall n, t$$

$$-\overline{f}_{nm}^0 - \overline{f}_{nm} \leqslant f_{nmt} \leqslant \overline{f}_{nm}^0 + \overline{f}_{nm}, \quad \forall n, m, t$$

Have you tried in a realistic case study?

- Electric power system (28 countries) for 2030 (single target year)
- Investments in conventional and renewable generation, transmission lines and two storage technologies (intraday and interday).
- Greenfield approach (no initial capacities)
- Given renewable penetration target



Have you tried in a realistic case study?

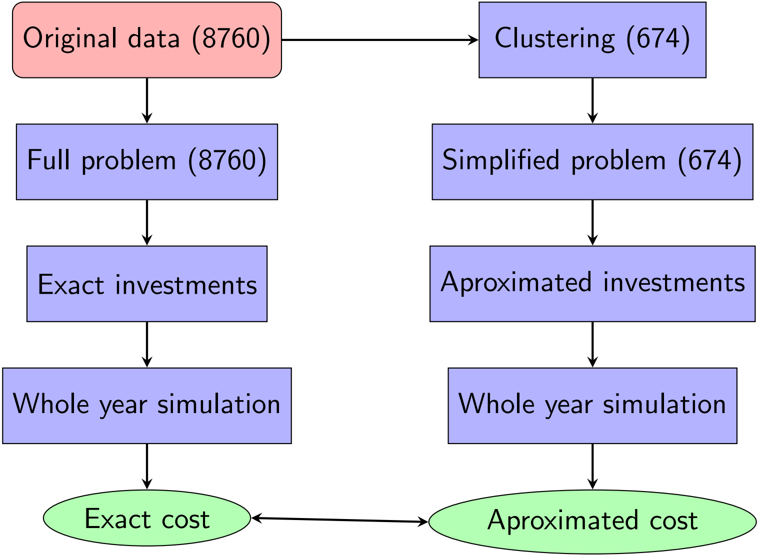
Table: Generation technology data

Technology	i_g^G (€/MW)	y_g^G (years)	c_g (€/MWh)	r_g^+/r_g^- (p.u.)
Base	$4 \cdot 10^6$	60	10	0.1
Peak	$1.5 \cdot 10^6$	40	40	1.0
Wind	$1.5 \cdot 10^6$	25	-	-
Solar	$1 \cdot 10^6$	25	-	-

Table: Storage technology data

Storage	η_s (h)	ξ_s (p.u.)	i_s^S (€/MW)	y_s (years)
intraday	6	0.8	$1.5 \cdot 10^6$	80
interday	48	0.7	$2 \cdot 10^6$	60

What about the results?



What about the results?

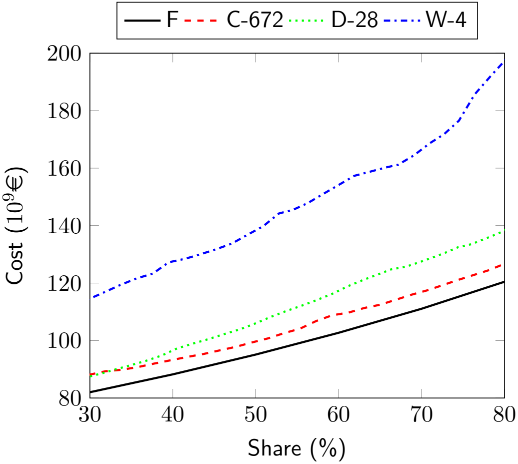
These are the investment results for a 50% renewable target

	F	C-672	D-28	W-4
Base (GW)	208	206	235	207
Peak (GW)	20	16	41	0
Wind (GW)	772	747	692	790
Solar (GW)	217	255	276	155
Hydro (GW)	160	160	160	160
Intraday (GW)	48	31	135	100
Interday (GW)	144	151	0	69
Network (GW)	23	18	34	19
Cost (10^9€)	95.13	99.31	102.17	133.72
Share (%)	50	49.5	46.1	47.1
Shed (%)	0	0.1	0.2	1.1
Spil (%)	2	3.3	3.4	3.3

- W-4 underinvest in peak, highest cost because of load shedding
- D-28 underinvest in wind+interday and overinvest in solar+intraday
- C-672 closest to full model, balance between intraday and interday storage

What about the results?

We plot now the cost of the 4 approaches for different renewable targets



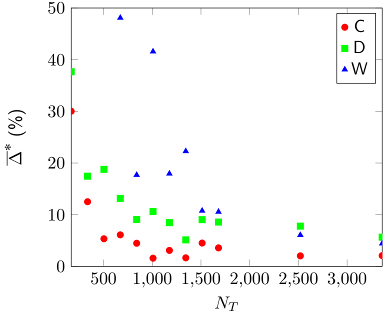
What about the results?

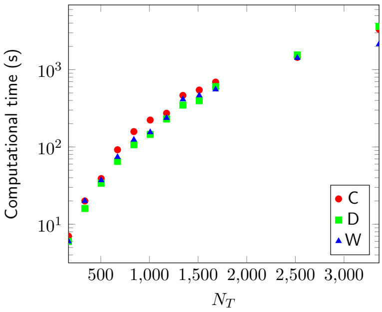
- Below we provide the average cost increase over all renewable penetration levels.

Approach	Number periods	Av. cost increase	Time
F	8760	0 %	~ 10 h
D-28	$28 \times 24 = 672$	13.1 %	~ 100 s
W-4	$4 \times 168 = 672$	48.1 %	~ 100 s
C-672	672	6.1 %	~ 100 s

- The proposed approach has the same computational burden as existing methodologies but reduces significantly the cost increase with respect to the benchmark model.

We also compare the proposed approach with existing ones for different number of time periods





What about the results?

Finally, we evaluate the proposed method in different scenarios

Scenario	C-672	D-28	W-4
Base	6.1%	13.1%	48.1%
No_solar	8.6%	18.3%	31.5%
No_wind	9.0%	7.2%	32.8%
No_hydro	2.3%	13.2%	60.3%
No_storage	11.1%	7.2%	6.3%

- If wind or storage investments are not possible, the performance of the proposed method is worse than the representative days

Conclusions

- Existing models to reduce the computational burden of planning problems do not properly capture the mid-term dynamics of renewable generation and fail to model storage operation.
- We propose a new time-period clustering technique that retains the chronology of the time-dependent parameters throughout the whole planning horizon.
- The proposed method determines capacity expansion plans that take into account the economic value of using interday storage to handle prolonged periods of high or low renewable power generation.
- Numerical results show the superior performance of our method, which determines more efficient capacity expansion plans without increasing the computational burden.

Thanks for the attention!

Questions?

OASYS

Optimization and Analytics for Sustainable energY Systems

More info: oasys.uma.es