On representative day selection for capacity expansion planning of power systems under extreme events

Can Li[a], Antonio J. Conejo[b], John D. Siirola[c], Ignacio E. Grossmann[a]

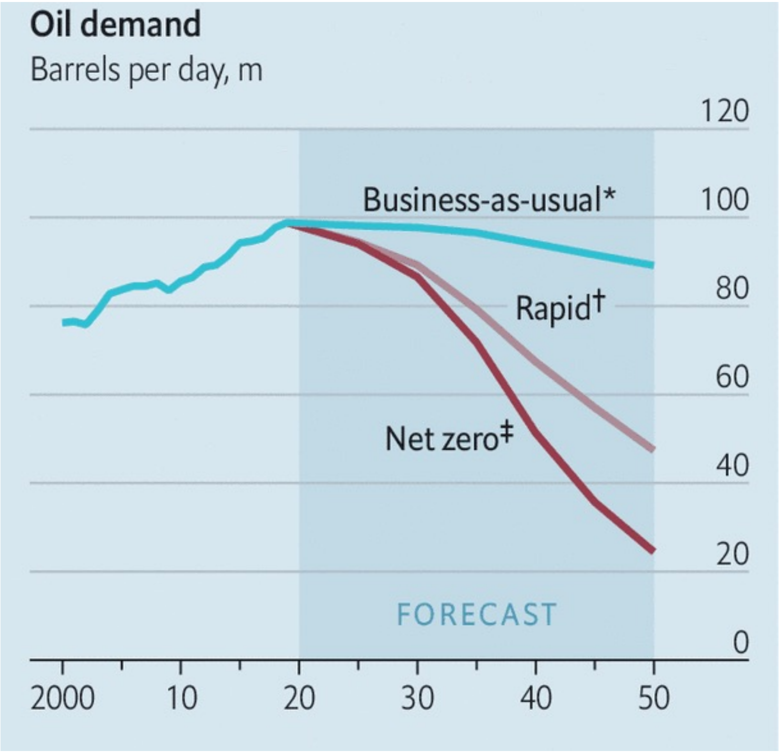
a Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

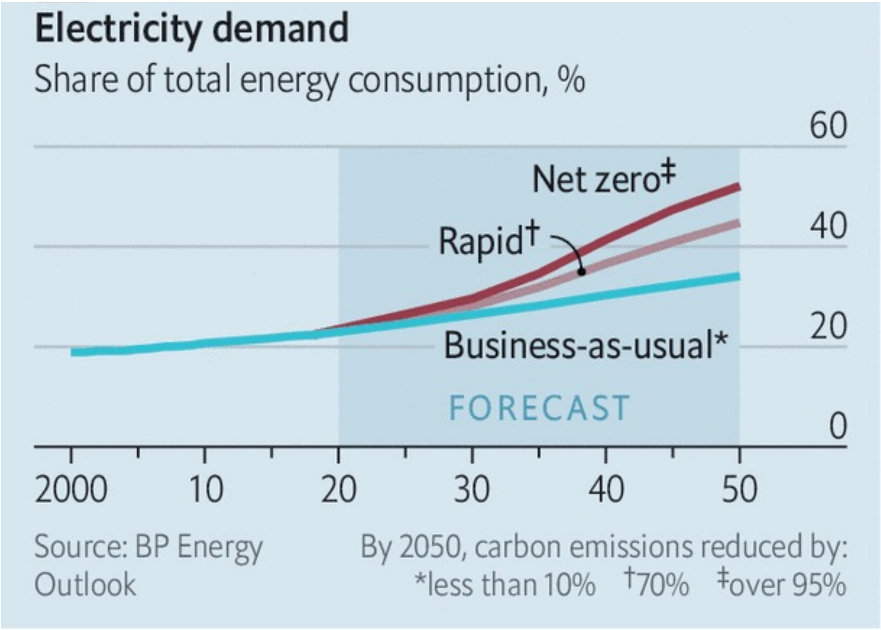
b Department of Integrated Systems Engineering, The Ohio State University, 1971 Neil Avenue, Columbus, OH 43210,USA

c Center for Computing Research, Sandia National Laboratories, P.O. 5800, Albuquerque, NM, 87185, USA

Energy Transition from Oil to Electricity

- **Electricity** demand would account for over 50% of total energy demand if we were to achieve **net zero carbon** emission in 2050





The Economist

BP Energy Outlook 2020

Project Motivation

Goal: Develop Optimization Models for Power Generation and Transmission Expansion Planning *(multiperiod MILP)*

Consider major generation sources:

- coal
- natural gas (simple and combined cycle)
- nuclear
- wind
- solar

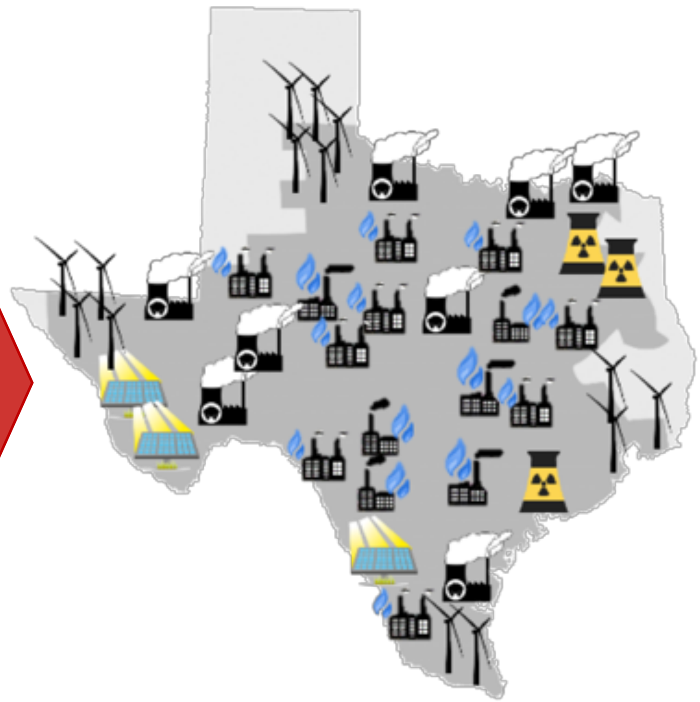
Emphasis: Long term Planning to Minimize Total Cost

Generation Transmission Expansion Planning + Unit Commitment

Minimize the **net present cost** (operating, investment, and environmental).

INPUT

- Energy source (coal, natural gas, nuclear, solar, wind*);
- **Generation** and **storage** technology;
- Location of existing generators;
- Nameplate capacity;
- Age and expected lifetime
- Potential transmission lines
- Emissions
- Operating and investment costs
- **Ramping rates, operating limits, maximum operating reserve**.
- Renewable generation profile.
- Load demand



OUTPUT

- **Location**, **year**, **type** and **number** of **generators, transmission lines** and **storage** units to install;
- When to retire them;
- Whether or not to extend their lifetime;
- Approximate power flow between locations;
- Approximate operating schedule

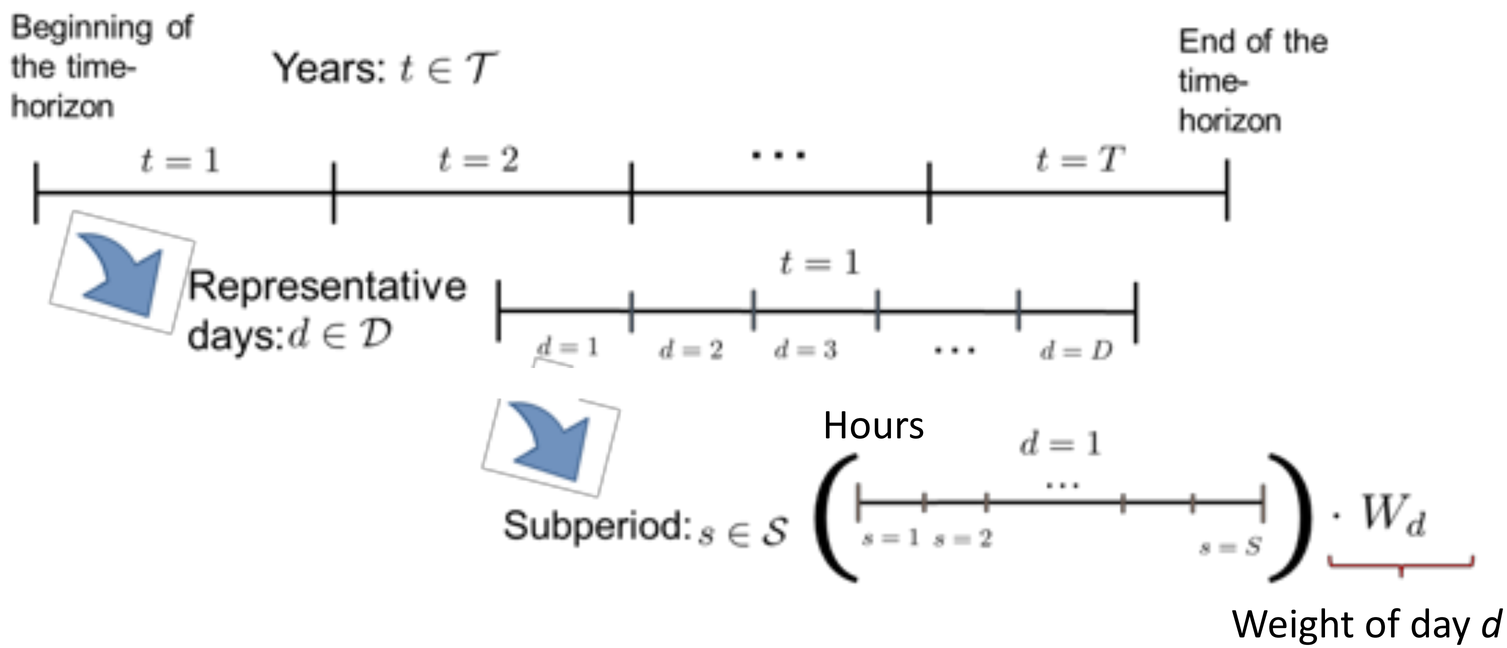
Research Challenges

- **Temporal** complexity: 20 years × 365days × 24hours=175,200 hours
- **Spatial** complexity: Around 500-2,000 individual generators depending on the region
- Complexity of the **optimization problem** with hourly decisions can be easily over **1 billion** variables.

Intractable. Need simplification

Representative Day Selection

- **Motivation:** Expansion planning decisions sensitive to the selection of representative days
  - Algorithms to select the representative days
  - Estimation of "optimality gap"



Fullspace model and Reduced model

$$(FD) \quad OBJ_{FD} = \min \sum_{t \in \mathcal{T}} \left(c_t^\top x_t + \sum_{d \in \mathcal{D}} \frac{365}{|\mathcal{D}|} f_t^\top y_{t,d} \right)$$

$$\text{s.t.} \quad A_{t,d} x_t + B_t y_{t,d} \leq b_{t,d} \quad \forall t \in \mathcal{T}, d \in \mathcal{D}$$

$$C_{t-1} x_{t-1} + D_t x_t \leq g_t \quad t = 2, 3, \ldots, |\mathcal{T}|$$

$$x_t \in X_t, \quad \forall t \in \mathcal{T}, \quad y_{t,d} \in Y_t, \quad \forall t \in \mathcal{T}, d \in \mathcal{D}$$

The whole dataset

Investment decisions for year *t*

operating decisions for year *t* day *d*

$$(RD) \quad OBJ_{RD} = \min \sum_{t \in \mathcal{T}} \left(c_t^\top x_t + \sum_{k \in \mathcal{K}} w_k f_t^\top y_{t,k} \right)$$

$$\text{s.t.} \quad A_{t,k} x_t + B_t y_{t,k} \leq b_{t,k} \quad \forall t \in \mathcal{T}, k \in \mathcal{K}$$

$$C_{t-1} x_{t-1} + D_t x_t \leq g_t \quad t = 2, 3, \ldots, |\mathcal{T}|$$

$$x_t \in X_t, \quad \forall t \in \mathcal{T}, \quad y_{t,k} \in \tilde{Y}_t, \quad \forall t \in \mathcal{T}, k \in \mathcal{K}$$

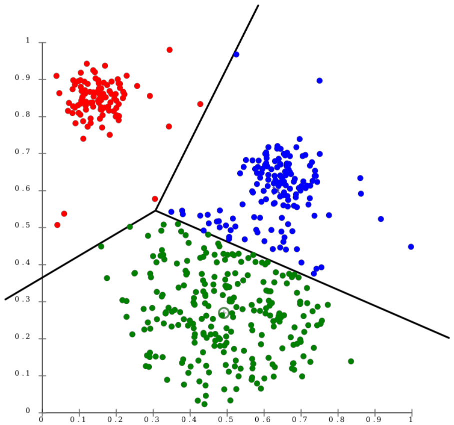
The set of representative days

operating decisions for year *t* representative day *k*

Relaxed integrality constraints

K-means clustering

- **Objective:** minimize the within cluster variance.



$$\mathbf{S}^* = \underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2$$

MINLP formulation:

$$\min_{\mathbf{c},\mathbf{d},\mathbf{y}} \sum_{i=1}^{n} d_i$$

$$d_i \geq \left(\sum_{j=1}^{D} (x_{ij} - c_{lj})^2 \right) - M_i(1 - y_{il}) \quad \forall i \in \{1, \ldots, n\}, l \in \{1, \ldots, k\}$$

$$\sum_{l=1}^{k} y_{il} = 1 \quad \forall i \in \{1, \ldots, n\}$$

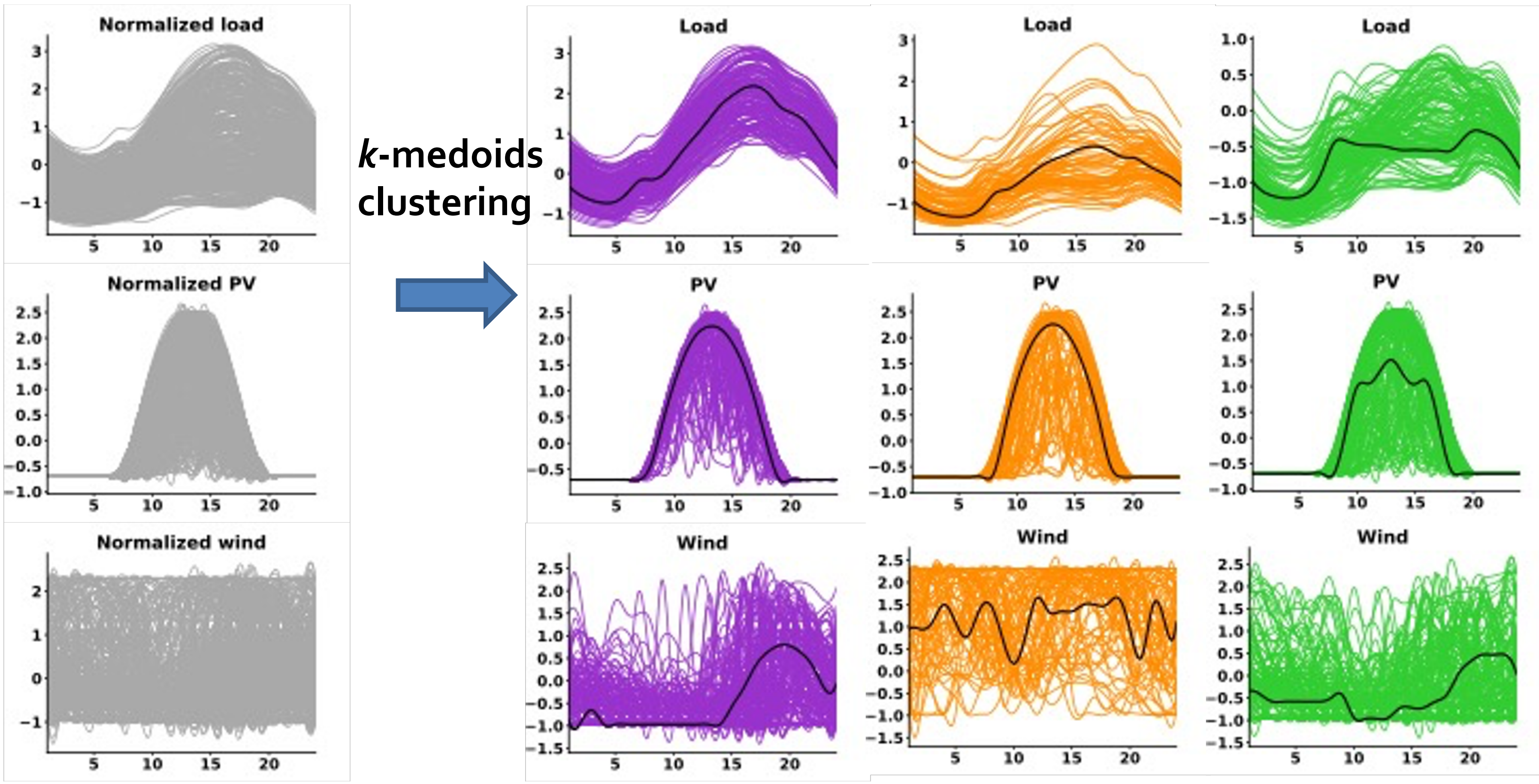
$$\mathbf{c}_l \in \mathbb{R}^D \quad \forall l \in \{1, \ldots, k\}$$

$$d_i \in \mathbb{R}_+ \quad \forall i \in \{1, \ldots, n\}$$

$$y_{il} \in \{0, 1\} \quad \forall i \in \{1, \ldots, n\}, l \in \{1, \ldots, k\}$$

Input-based method

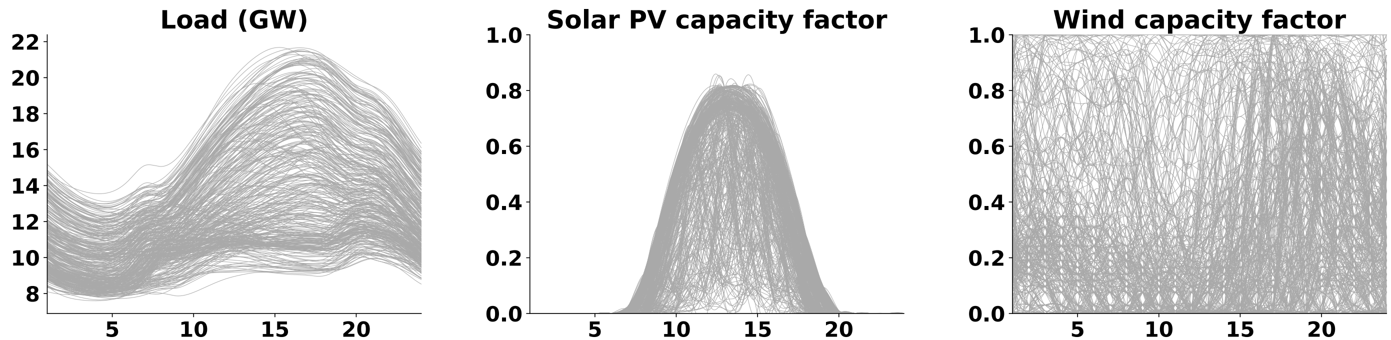
- Clustering is performed directly on the input data (load, capacity factors)



Cost-based method

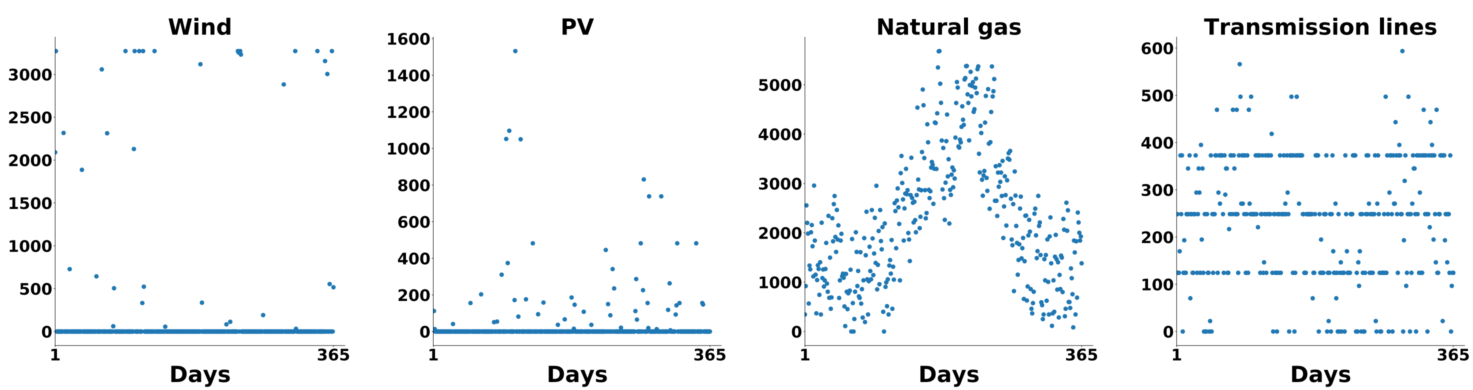
➢ **Hypothesis:** The days with similar **optimal investment decisions**, i.e., the days that need similar generators, transmission lines, and storage units, are similar and should be assigned to the same cluster

Raw data



Solve CEP for each day in the full dataset individually & Dimension reduction

Investment cost breakdown after reduction (million dollars)



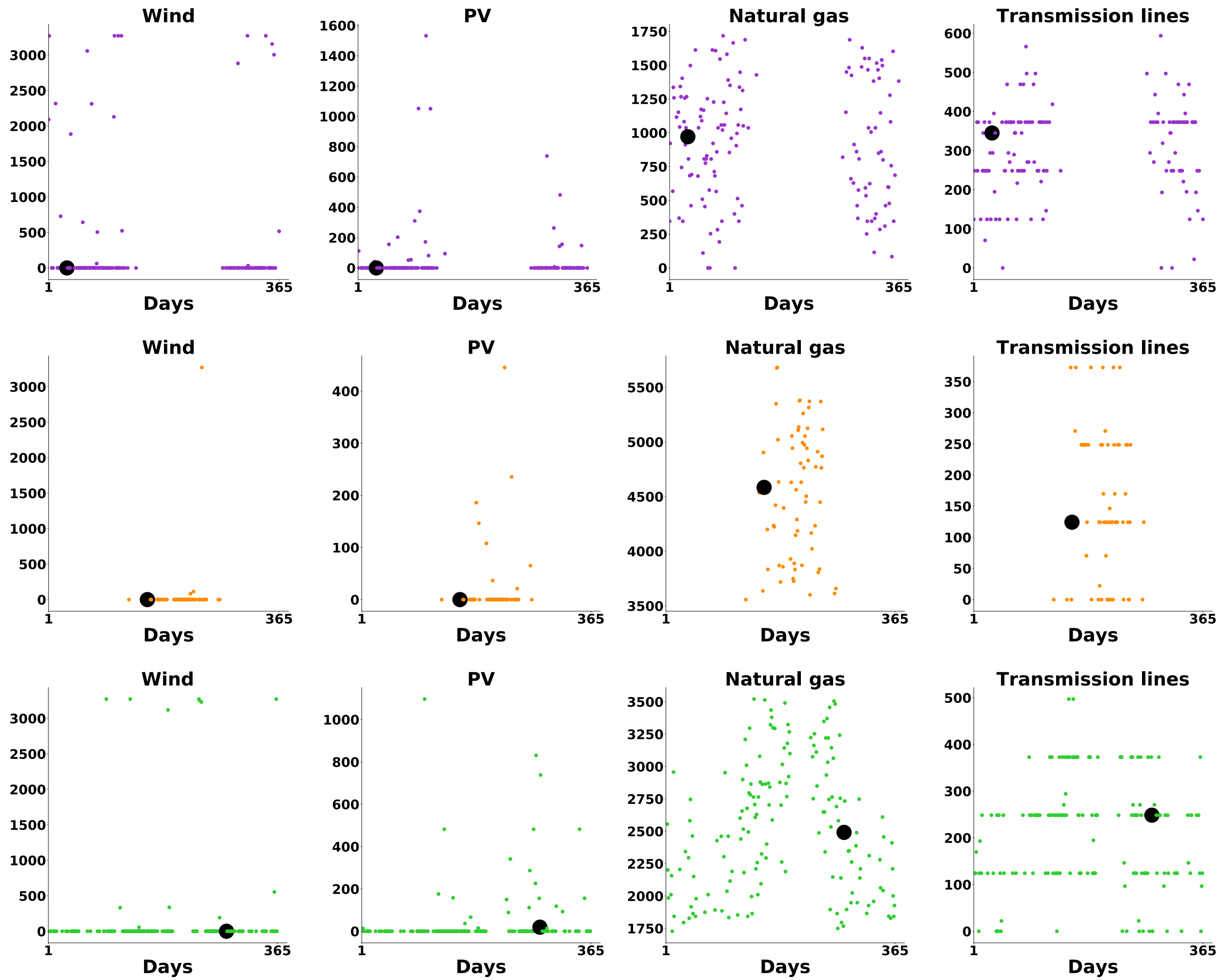
Cost-based method

K-medoids clustering

Cluster 1, w=141

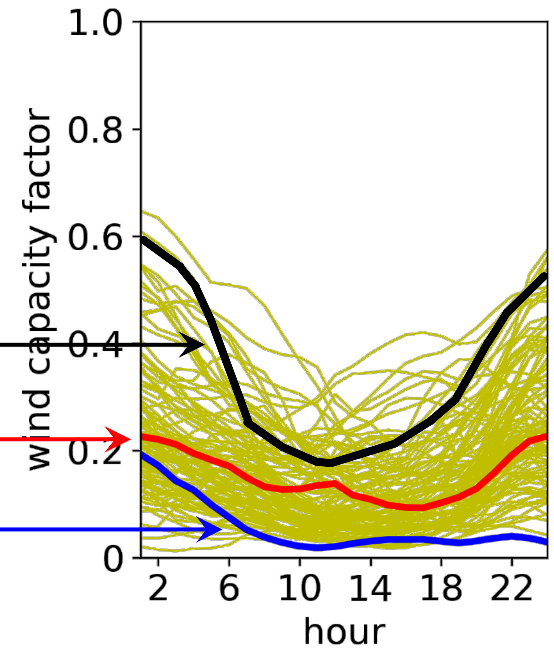
Cluster 2, w=65

Cluster 3, w=159



Failures of the Representative Day Approach

- Extreme events, such as highest ramp and lowest generation, are not captured by the representative days.
- The investment decisions from (RD) are usually **infeasible** for (FD).
- Solution: adding days with **extreme events**
- Option 1: adding extreme days based on some predefined characteristics, e.g., peak load day.
- Alternative strategy?



Extreme Events Selection

- **Load shedding cost**

Energy balance at each node

Min Load shedding

$$\sum_{i}(p_{i,r,t,d,s}) + \sum_{l|r(l)=r} p^{\text{flow}}_{l,t,d,s} - \sum_{l|s(l)=r} p^{\text{flow}}_{l,t,d,s} + \sum_{j} p^{\text{discharge}}_{j,r,t,d,s} - \sum_{j} p^{\text{charge}}_{j,r,t,d,s} = L_{r,t,d,s}$$

Power generation $\pm$ power flow in/out $\pm$ power discharge/charge = Load

Power generation $\pm$ power flow in/out $\pm$ power discharge/charge = Load – **Load shedding**

1) Fix the investment decisions from (RD)
2) Solve the operating problem corresponding to each day in our dataset
3) Find the infeasible day with the highest load shedding cost

Extreme Events Selection

- **Highest cost**
  - In the cost-based approach, we have obtained the total cost (operating + investment) for each day in our dataset
  - Select the day with the highest cost as our extreme day

Optimality Gap

- Motivation: Provide upper and lower bound for the fullspace problem (FD)
- **Upper bound:** Fix the optimal investment decisions from the reduced model, solve each day in the fullspace model.

$$OBJ_{FD}(\mathbf{x}^{RD}) \geq OBJ_{FD}(\mathbf{x}^{FD}) = OBJ_{FD}$$

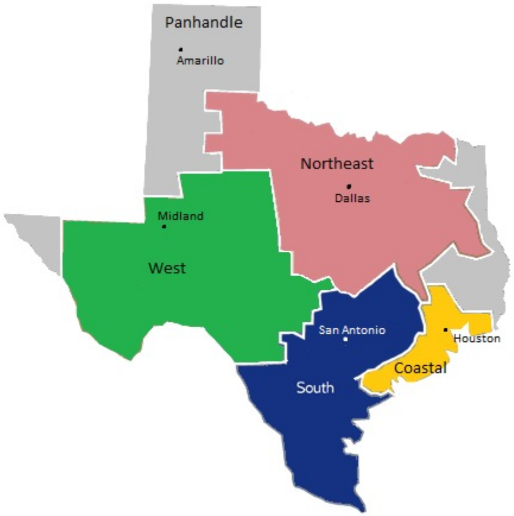
- **Lower bound:** Reduced model provides lower bound under certain assumptions.

Theorem 1. *For both cost-based and input-based approaches, if k-means clustering is used, (RD) provides a lower bound for the optimal objective value of (FD), i.e.,* $OBJ_{RD} \leq OBJ_{FD}$. *This lower bound holds before and after adding extreme days.*

$$\texttt{Gap} = \frac{OBJ_{FD}(\mathbf{x}^{RD}) - OBJ_{RD}}{OBJ_{FD}(\mathbf{x}^{RD})} \times 100\%$$

Case Study

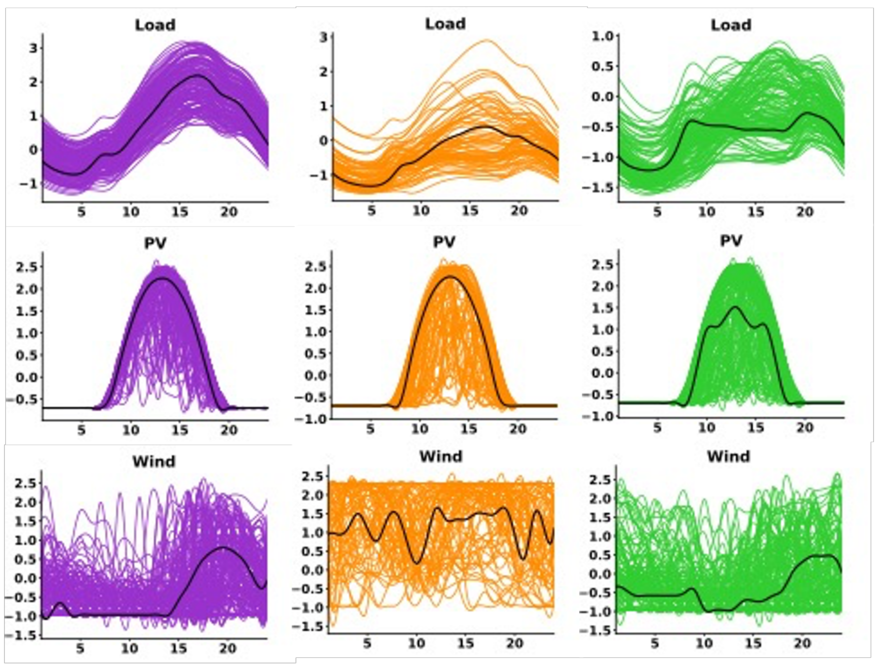
- ERCOT region, 5 years planning problem
- The whole dataset D has 365 days that consists of load and capacity factor data



Algorithm option	Data	Clustering Algorithm	Extreme Day Method
1	Input	k-means	load shedding cost
2	Input	k-medoids	load shedding cost
3	Cost	k-medoids	highest cost
4	Cost	k-medoids	load shedding cost
5	Cost	k-means	highest cost
6	Cost	k-means	load shedding cost

Infeasibility without the Extreme Days

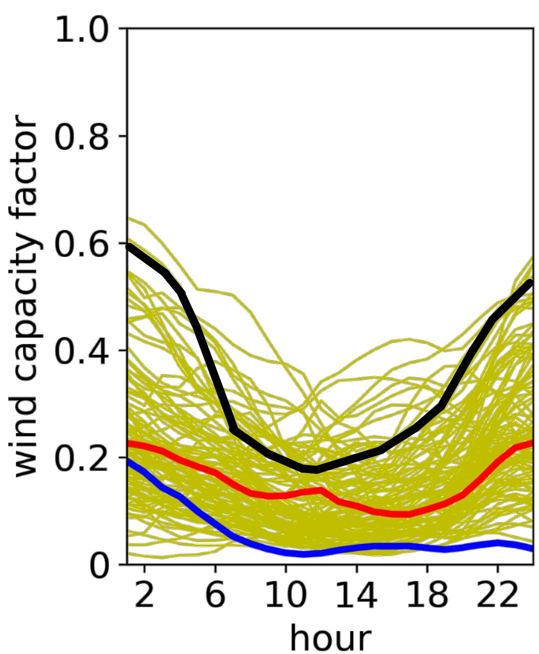
- Only using the representative days from centroids/medoids of the clustering algorithms **cannot guarantee feasibility**
- Cost-based approach has fewer infeasible days when *k* is large



Algorithm option	*k*	#infeasible day
	5	**70**
1	10	**63**
	15	**42**
	5	**35**
2	10	**21**
	15	**40**
	5	**98**
3	10	**13**
	15	**12**
	5	**98**
4	10	**13**
	15	**12**
	5	**34**
5	10	**30**
	15	**29**
	5	**34**
6	10	**30**
	15	**29**

Feasible After Adding Extreme Days

- Adding the **extreme days** makes the investment decisions **feasible** for the fullspace problem. $OBJ_{FD}(\mathbf{x}^{RD}) < +\infty$
- **K-medoids** clustering has lower cost in most cases



Option	k	#Extreme day	$OBJ_{FD}(\mathbf{x}^{RD})$
1	5	**3**	**79.16**
	10	**2**	**79.04**
	15	**2**	**78.81**
2	5	**3**	**78.92**
	10	**2**	**78.72**
	15	**2**	**78.74**
3	5	**5**	**78.83**
	10	**3**	**78.67**
	15	**3**	**78.81**
4	5	**3**	**78.93**
	10	**2**	**78.79**
	15	**1**	**78.75**
5	5	**4**	**78.98**
	10	**6**	**79.09**
	15	**4**	**78.98**
6	5	**3**	**79.12**
	10	**4**	**78.93**
	15	**3**	**78.81**

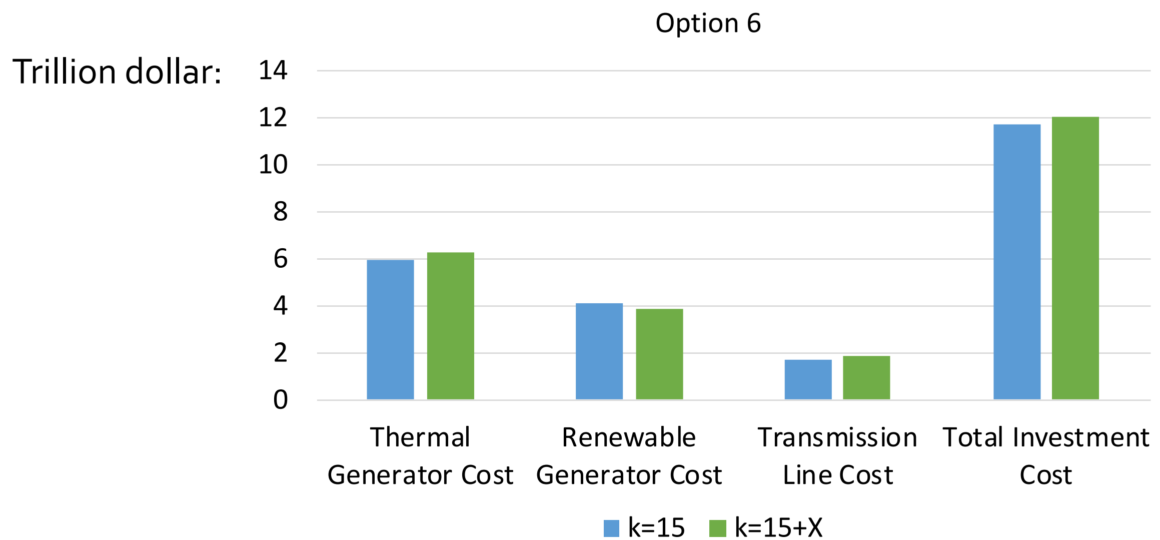
Optimality Gap

- "Optimality gap" can be obtained when k-means clustering is used
- Gap improves as *k* increases

Option	k	$OBJ_{FD}(\mathbf{x}^{RD})$	LB	Gap
1	5	79.16	**76.09**	**4.0%**
	10	79.04	**76.29**	**3.6%**
	15	78.81	**76.58**	**2.9%**
2	5	78.92	-	-
	10	78.72	-	-
	15	78.74	-	-
3	5	78.83	-	-
	10	78.67	-	-
	15	78.81	-	-
4	5	78.93	-	-
	10	78.79	-	-
	15	78.75	-	-
5	5	78.98	**76.16**	**4.2%**
	10	79.09	**76.64**	**3.7%**
	15	78.98	**76.74**	**3.4%**
6	5	79.12	**76.15**	**3.9%**
	10	78.93	**76.63**	**3.0%**
	15	78.81	**76.73**	**2.7%**

Effects of Adding Extreme days

- Comparison of *k*=15, option 6 before and after adding the extreme days
  - Total investment cost +325 million
  - Thermal generator cost +350 million
  - Transmission line cost +186 million
  - Storage investment cost +0.2 million
  - Renewable generator cost -212 million



Conclusion and Future work

- We have developed **models and algorithms** for capacity expansion of **power systems** with high penetration of **renewables**.
- The capability to analyze powers systems enables to **study hybrid energy systems** that have both electricity generators and electricity/heat consumers, such as **chemical plants.**