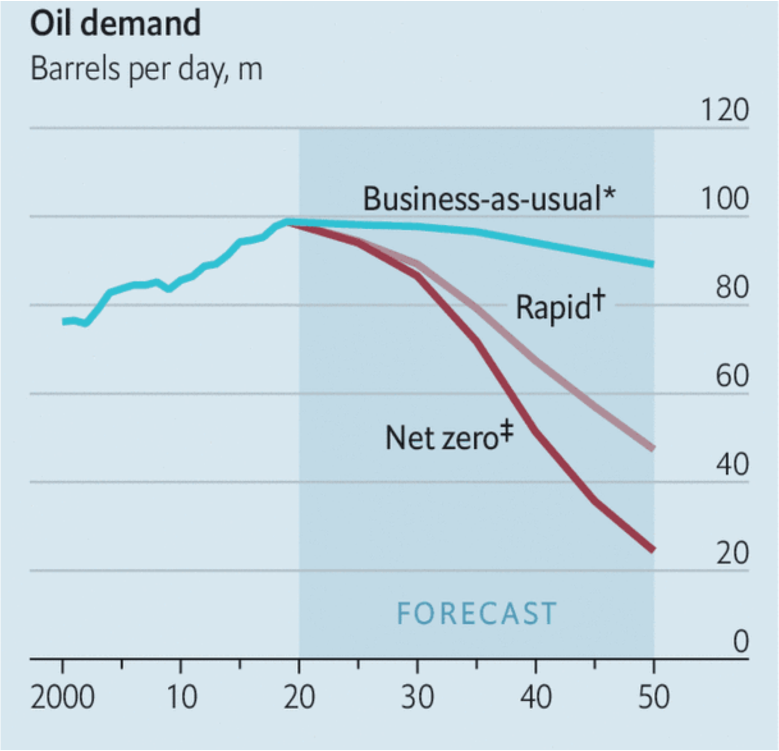
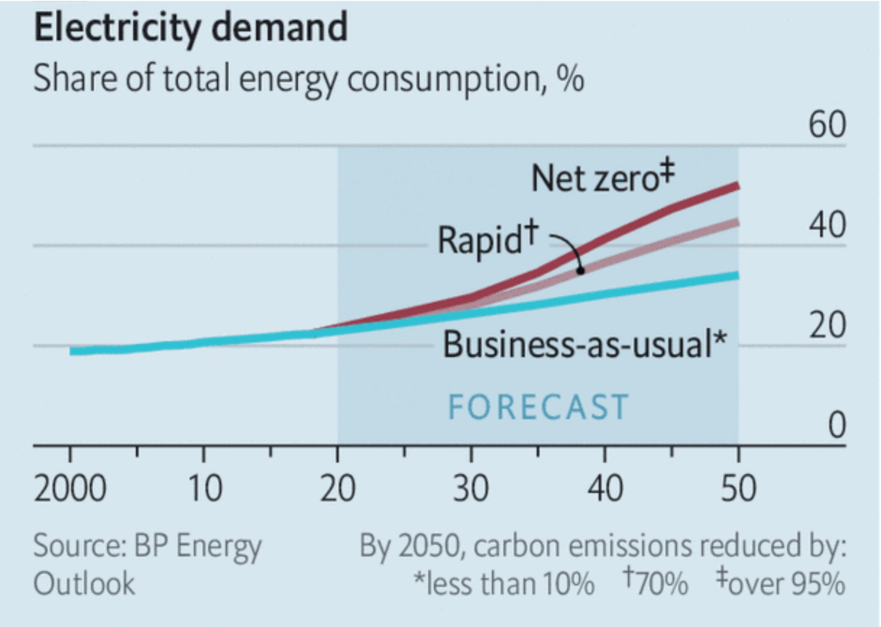
Capacity Expansion Planning of Power Systems under High Renewables Penetration

Can Li
Fifth year PhD candidate
Advised by Prof. Ignacio Grossmann

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



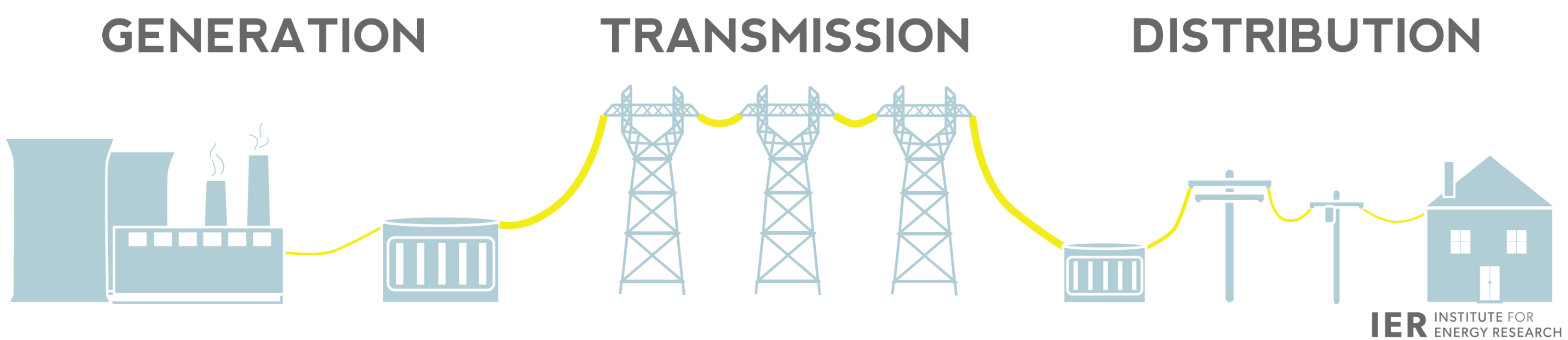


The Economist

BP Energy Outlook 2020

Power Industry

- Electricity is generated at **power plants** and moves through a complex system, sometimes called the **grid**, of electricity substations, transformers, and power lines that connect electricity **producers** and **consumers**.



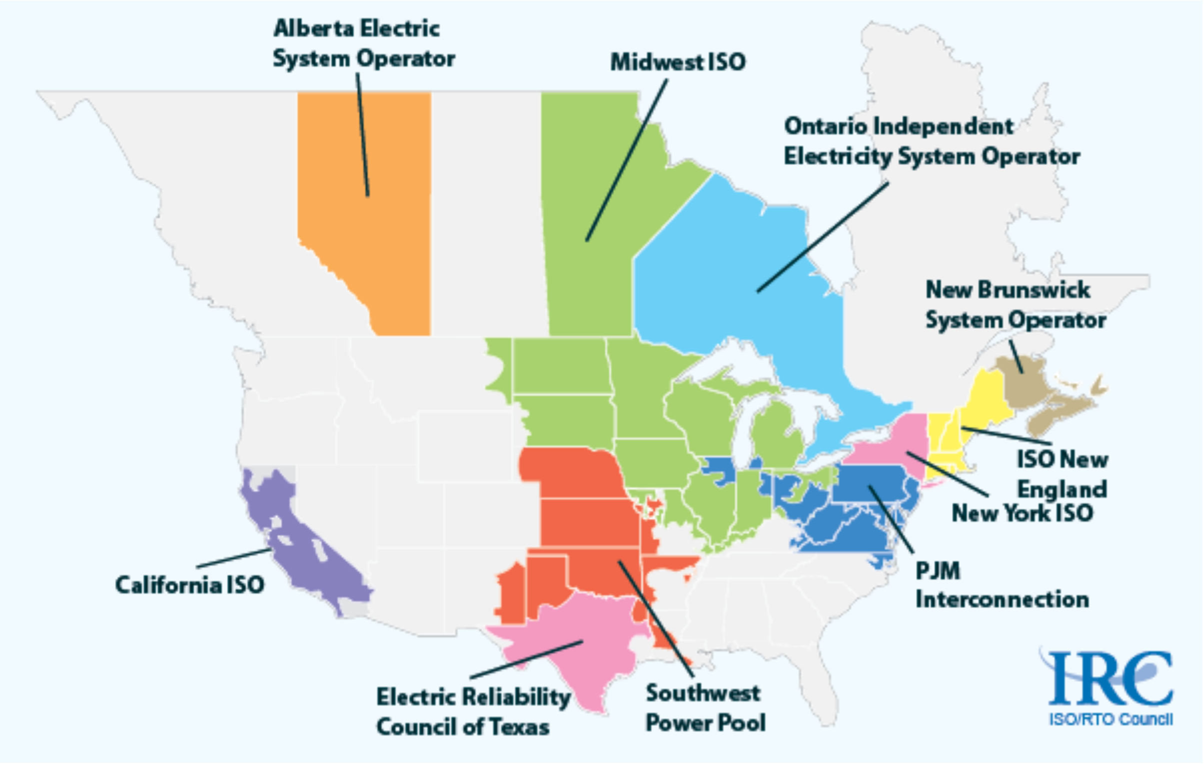
Traditional utilities & Independent merchant generators

Independent system operators

Utilities

Electricity Market in the US

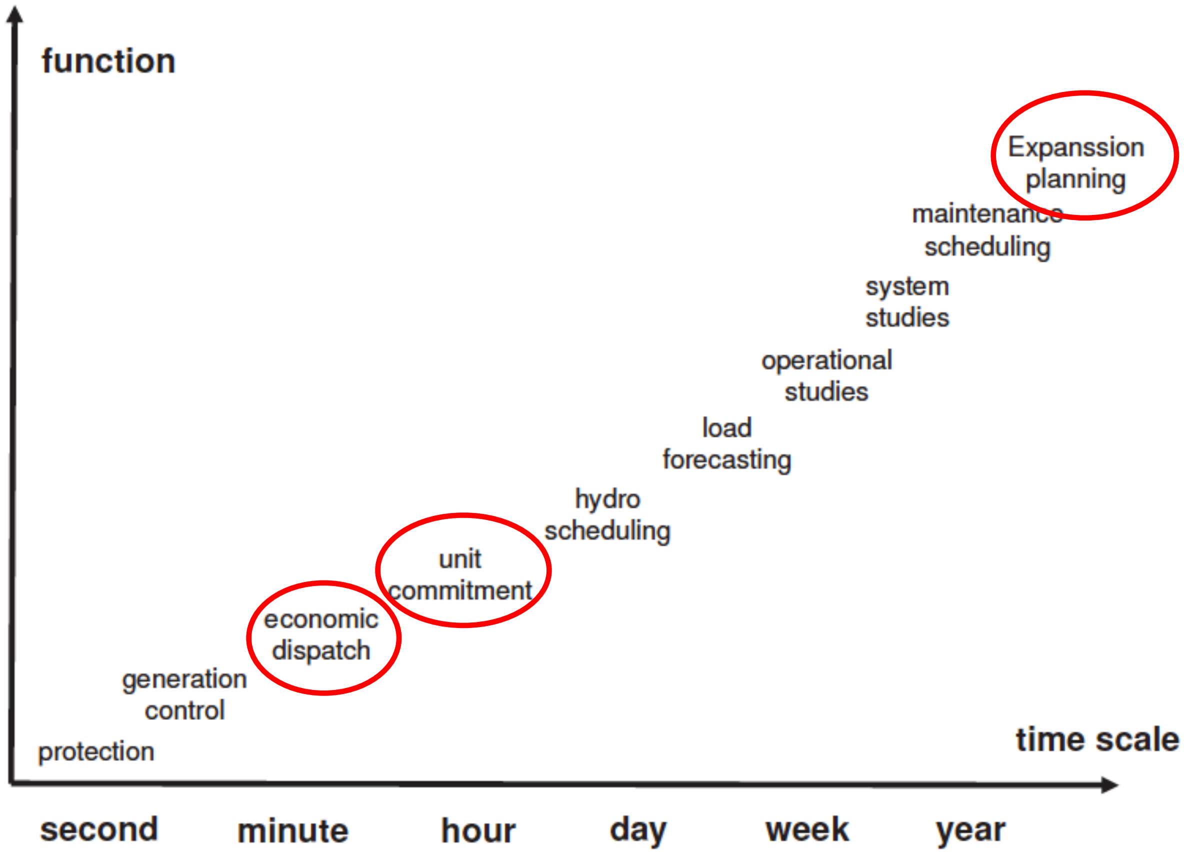
- The electricity transmission network is controlled by **Independent System Operators** (ISOs). An ISO coordinates, controls, and monitors a multi-state electric grid.
- Create **a competitive wholesale electricity market** where all generators can compete on an equal basis and have equal access to the grid.



ISOs in North America

Optimization Problems Involved

- Wide-range applications in terms of the time scale.
- From long term planning to short term control/scheduling

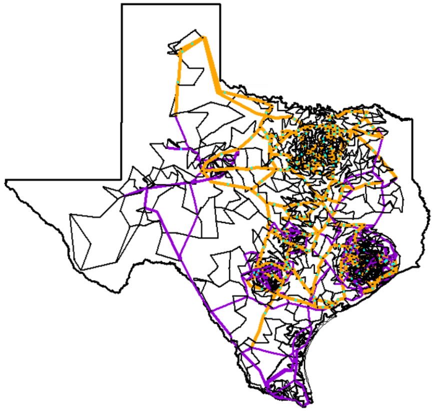


Arriaga et al. (2008)

Economic Dispatch/Optimal Power Flow

- Economic dispatch is the short-term determination of the **optimal output** of a number of electricity generation facilities, to meet the **system load**, at the **lowest possible cost**, subject to **transmission and operational** constraints

$$
\begin{aligned}
\min \quad & \sum_{i \in \mathbf{G}} C_i\left(P_i^{\mathrm{G}}\right), \\
\text{s.t.} \quad & P_i(V, \delta) = P_i^{\mathrm{G}} - P_i^{\mathrm{L}} \quad \forall\, i \in \mathbf{N}, \\
& Q_i(V, \delta) = Q_i^{\mathrm{G}} - Q_i^{\mathrm{L}} \quad \forall\, i \in \mathbf{N}, \\
& P_i^{\mathrm{G,min}} \le P_i^{\mathrm{G}} \le P_i^{\mathrm{G,max}} \quad \forall\, i \in \mathbf{G}, \\
& Q_i^{\mathrm{G,min}} \le Q_i^{\mathrm{G}} \le Q_i^{\mathrm{G,max}} \quad \forall\, i \in \mathbf{G}, \\
& V_i^{\min} \le V_i \le V_i^{\max} \quad \forall\, i \in \mathbf{N}, \\
& \delta_i^{\min} \le \delta_i \le \delta_i^{\max} \quad \forall\, i \in \mathbf{N}.
\end{aligned}
$$



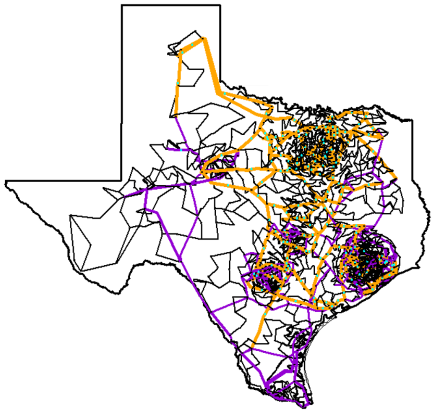
Economic Dispatch/Optimal Power Flow

- Economic dispatch is the short-term determination of the **optimal output** of a number of electricity generation facilities, to meet the **system load**, at the **lowest possible cost**, subject to **transmission and operational** constraints

min — Minimize cost

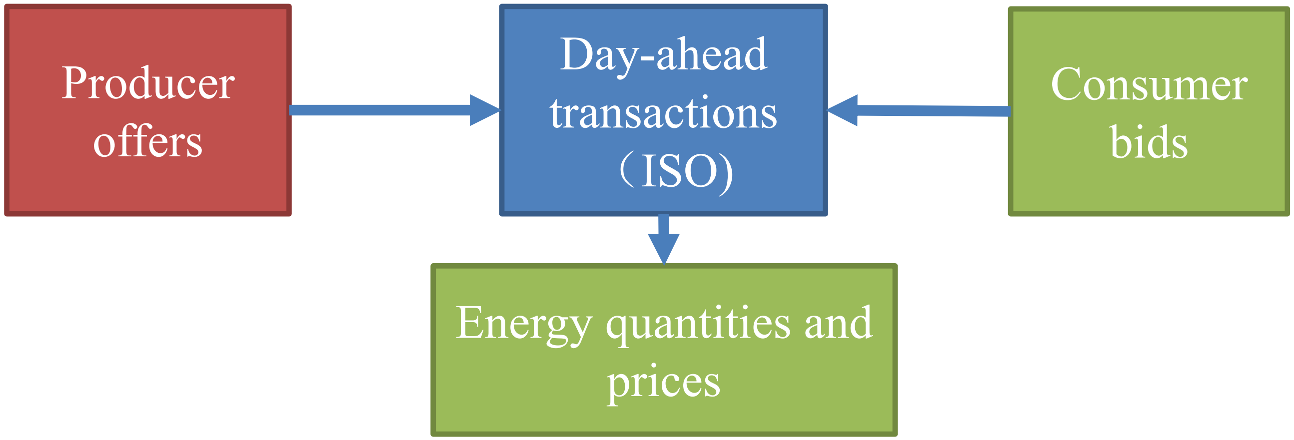
s.t. — Power flow equation at each node, i.e., power injected = generation - load

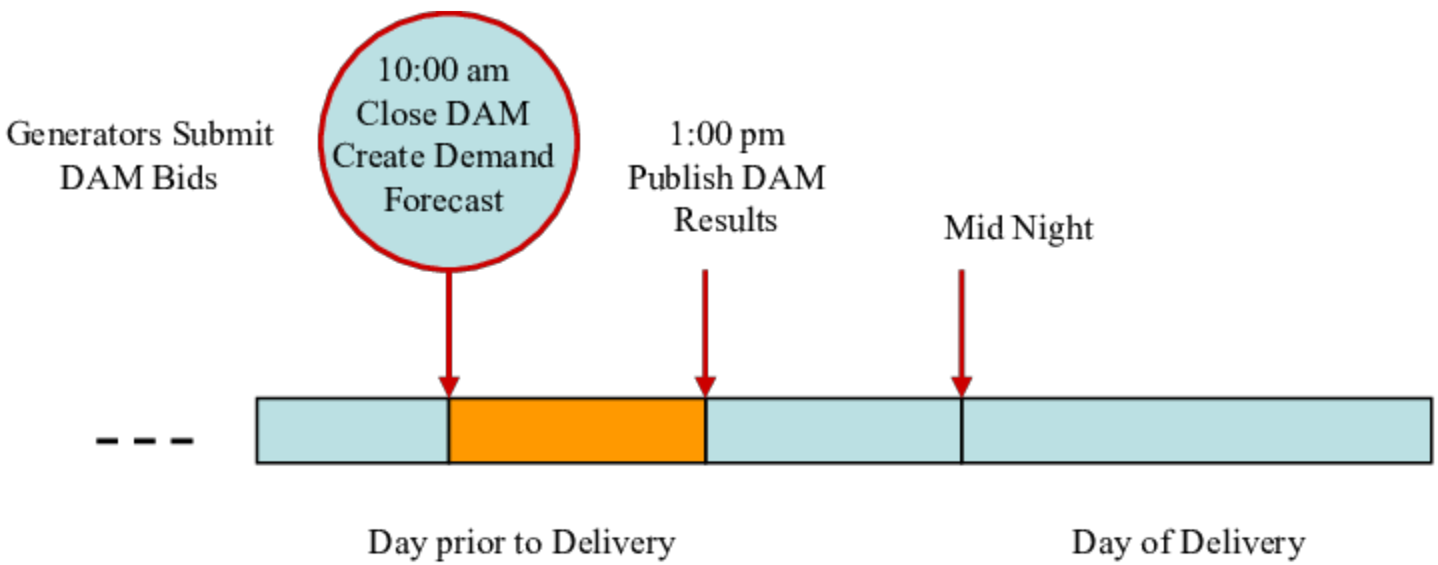
Variable bounds for voltage, phase angle, real, reactive power



Unit Commitment (Day-ahead Market)

- The Day-ahead market lets market participants **commit to buy or sell** wholesale electricity one day before the operating day, to help avoid price volatility





Unit Commitment

- Mixed-integer linear programming (MILP) model
  - Binary variables: generator on/off status
  - Continuous variable: power generation, power flow

$$\min \sum_{\forall g,t} C_g^{su} v_{gt} + C_g^{nl} u_{gt} + C_g^{\text{var}} p_{gt},$$

$$\begin{aligned}
\text{s.t.}\quad & v_{gt} \geq u_{gt} - u_{g,t-1} && \forall g,t, \\
& \sum_{i=t-UT_g+1}^{t} v_{gi} \leq u_{gt} && \forall g,t, \\
& \sum_{i=t+1}^{t+DT_g} v_{gi} \leq 1 - u_{gt} && \forall g,t, \\
& P_g^{\min} u_{gt} \leq p_{gt} \leq P_g^{\max} u_{gt} && \forall g,t, \\
& p_{gt} - p_{g,t-1} \leq R_g^{hr} u_{g,t-1} + R_g^{su} v_{gt} && \forall g,t, \\
& p_{g,t-1} - p_{gt} \leq R_g^{hr} u_{gt} + R_g^{sd} \left(v_{gt} - u_{gt} + u_{g,t-1} \right) && \\
& && \forall g,t,
\end{aligned}$$

$$\begin{aligned}
& e_{nt} = \sum_{g \in G(n)} p_{gt} - D_{nt} && \forall n,t, \\
& -F_l \leq \sum_{\forall n} W_{nl} e_{nt} \leq F_l && \forall l,t, \\
& \sum_{\forall n} e_{nt} = 0 && \forall t, \\
& u_{gt} \in \{0,1\}, 0 \leq v_{gt} \leq 1 && \forall g,t.
\end{aligned}$$

Unit Commitment

- Mixed-integer linear programming (MILP) model
  - Binary variables: generator on/off status
  - Continuous variable: power generation, power flow

min Minimize cost

s.t.

Generator on/off status
Minimum up and down time

Power generation bounds

Generation ramping constraints

Power flow equations

Binary constraints

Research Communities Involved

- Electrical engineers (traditionally)
  - IEEE Transactions on Power Systems
- Increasing interest in industrial engineering
  - Operations Research, INFORMS Journal on Computing, Mathematical Programming

Strong SOCP Relaxations for the Optimal Power Flow Problem

Burak Kocuk, Santanu S. Dey, X. Andy Sun
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Nonconvex NLP

Learning to Solve Large-Scale Security-Constrained Unit Commitment Problems

Álinson S. Xavier,[a] Feng Qiu,[a] Shabbir Ahmed[b]

Large scale MILP

A model and approach to the challenge posed by optimal power systems planning

Richard P. O'Neill · Eric A. Krall · Kory W. Hedman · Shmuel S. Oren

Large scale MILP

Project Motivation

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

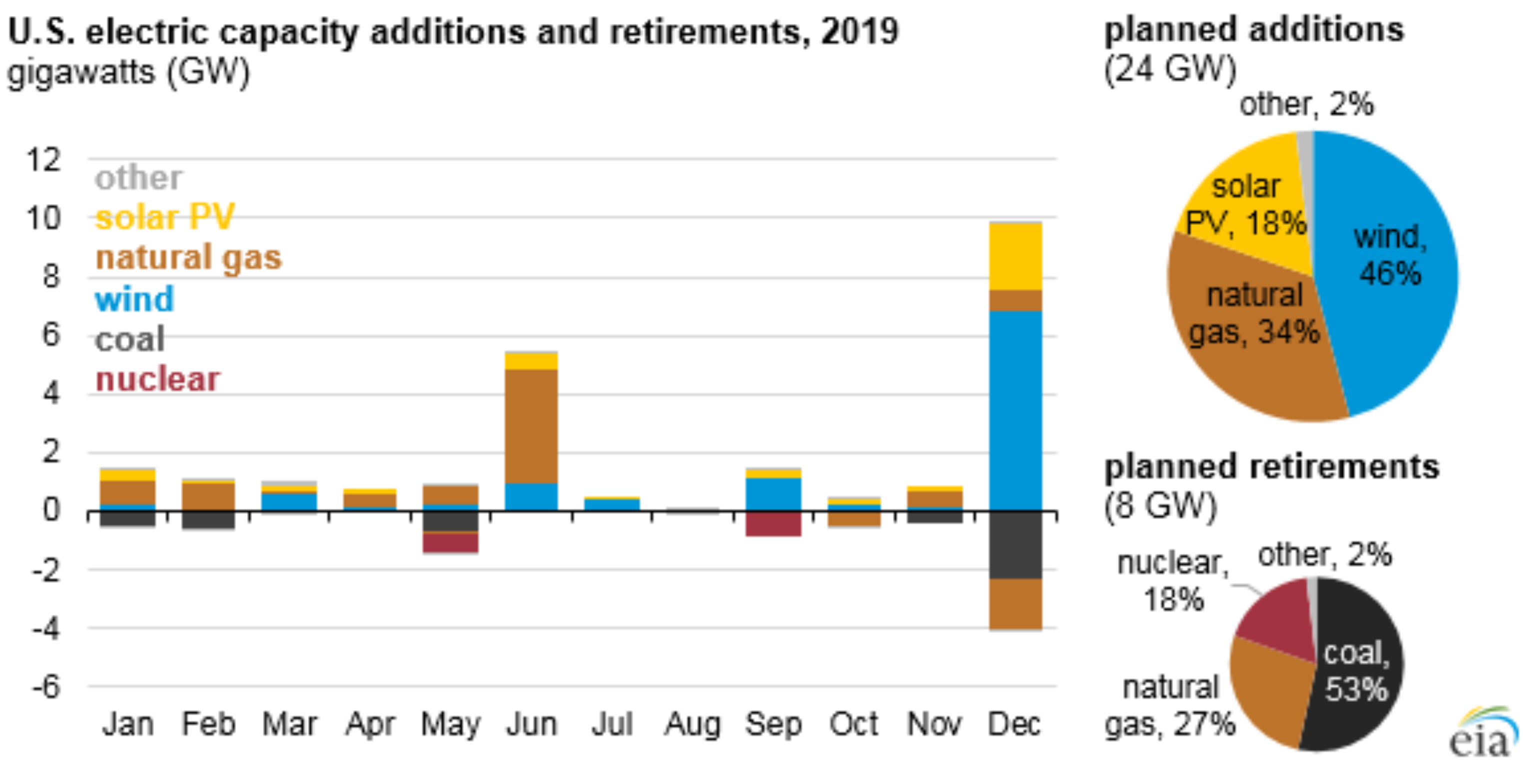
Consider major generation sources:

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

Emphasis: Long term Planning to Minimize Total Cost

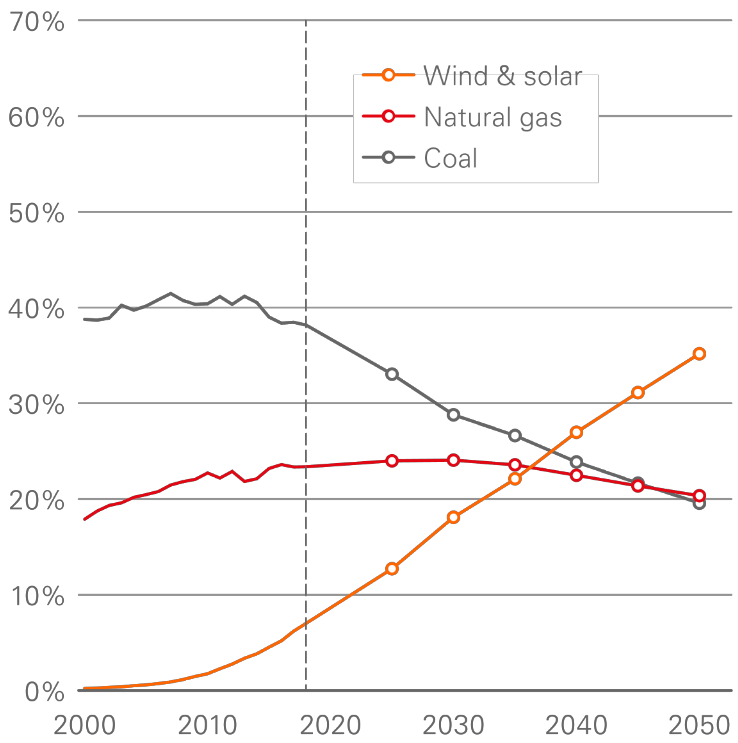
Current Capacity Additions

- Most electric capacity additions come from **renewables**
  - In 2019, 64% capacity additions in the US are from renewables. 34% from natural gas

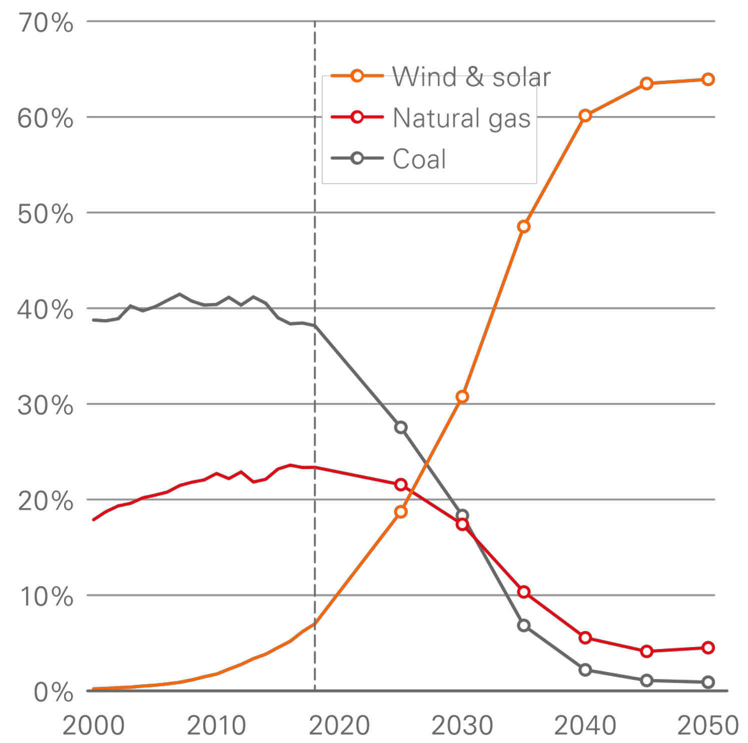


Renewable Generation

- Share of global power generation from **wind&solar** is expected to **increase**



business as usual

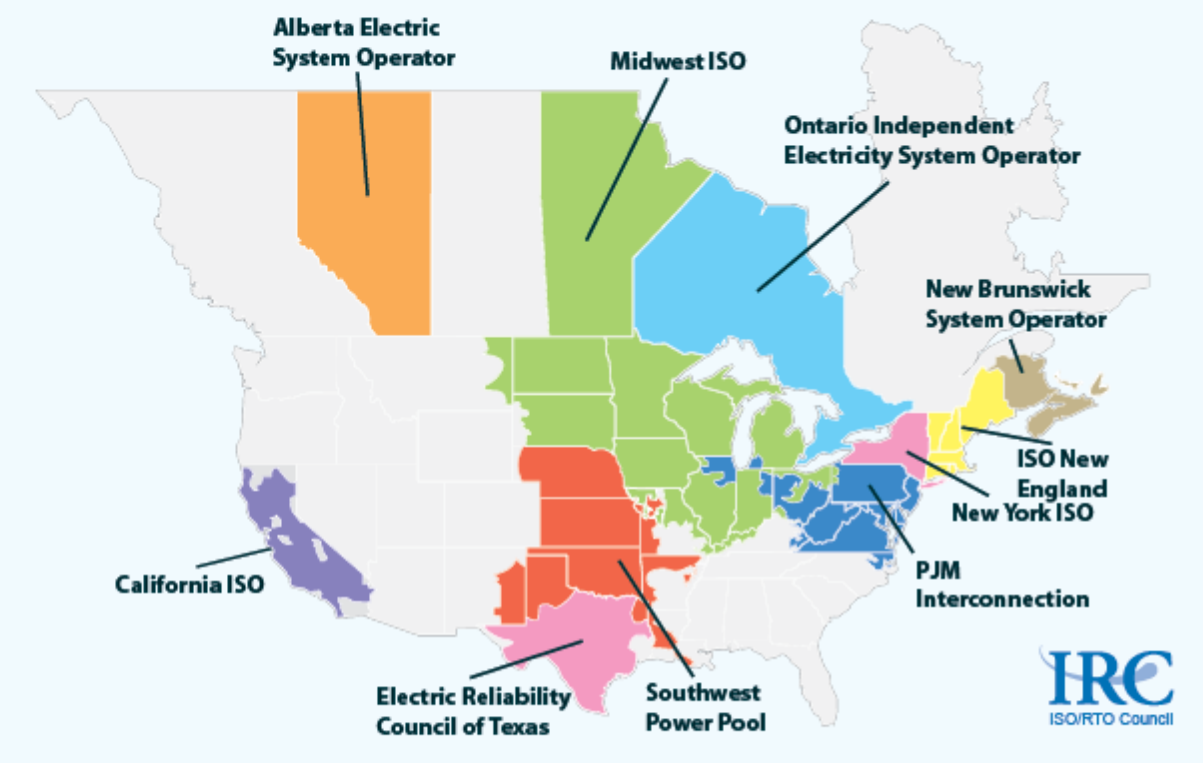


net zero

BP Energy Outlook 2020

Problem Addressed in This Presentation

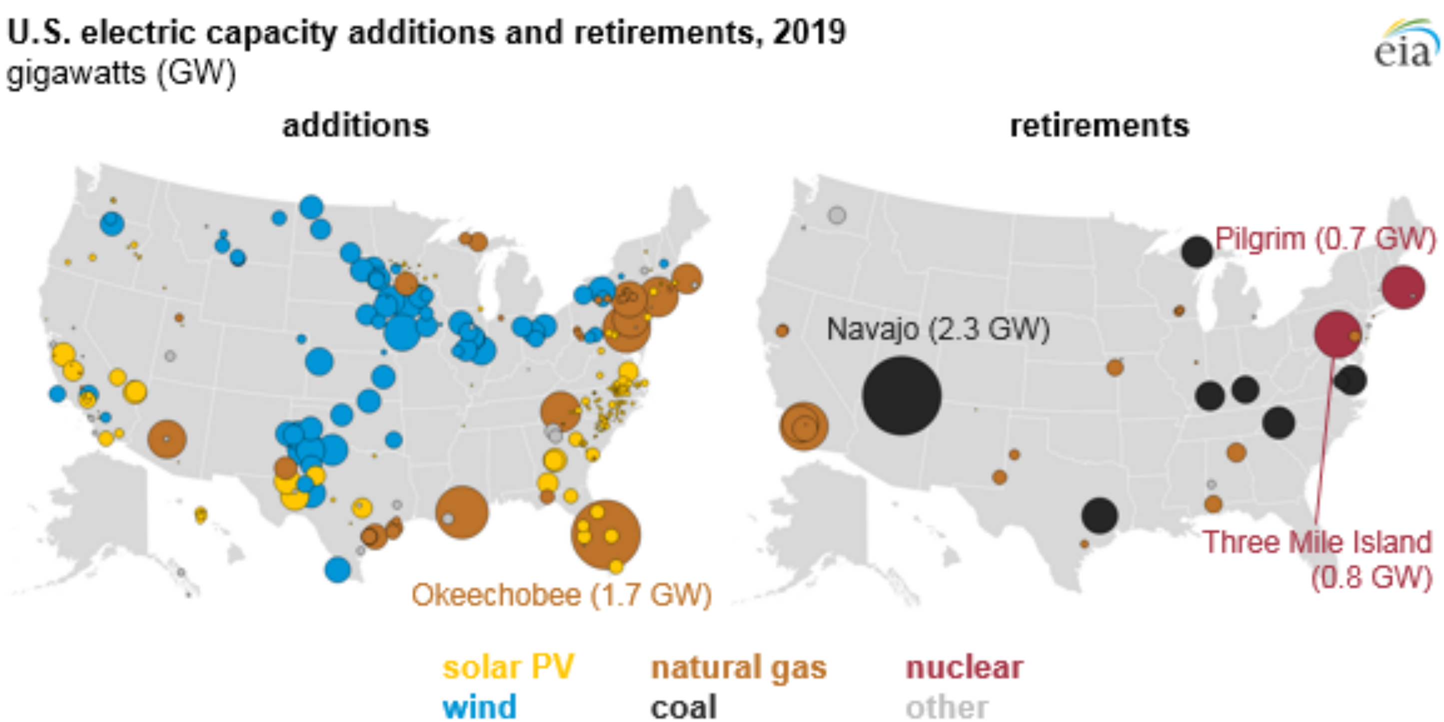
- We take the role of a **central planner** on the **capacity expansion of generating units and transmission lines** to satisfy the increase in demand within a geographical region, like a region corresponding to an Independent System Operator (ISO)



ISOs in North America

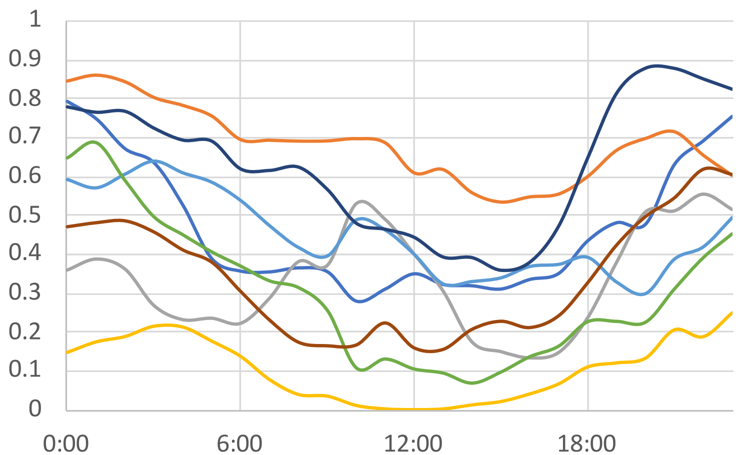
Research Challenges in Transitioning to Renewables

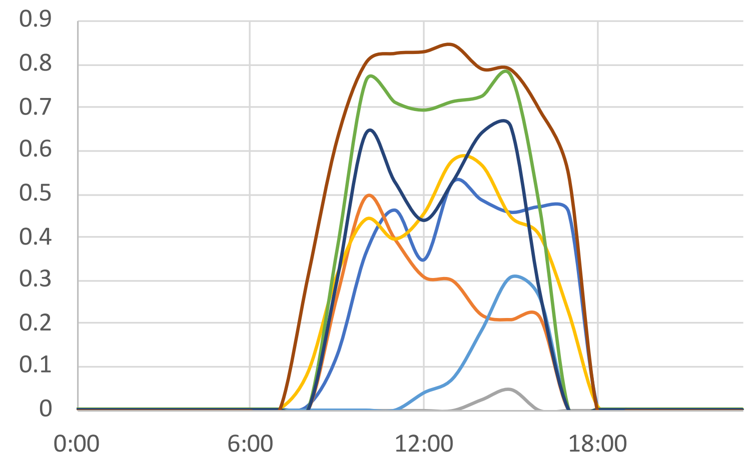
- Renewables concentrate in **remote areas** not well connected to load demand. The model needs to **coordinate transmission and generation** expansion.



Research Challenges in Transitioning to Renewables

- Power systems need to be able to adjust to the **volatile** power generation from renewables. The model has to capture the **hourly** variations.





Hourly wind and solar generator output in 8 days

Overview of our work on expansion planning

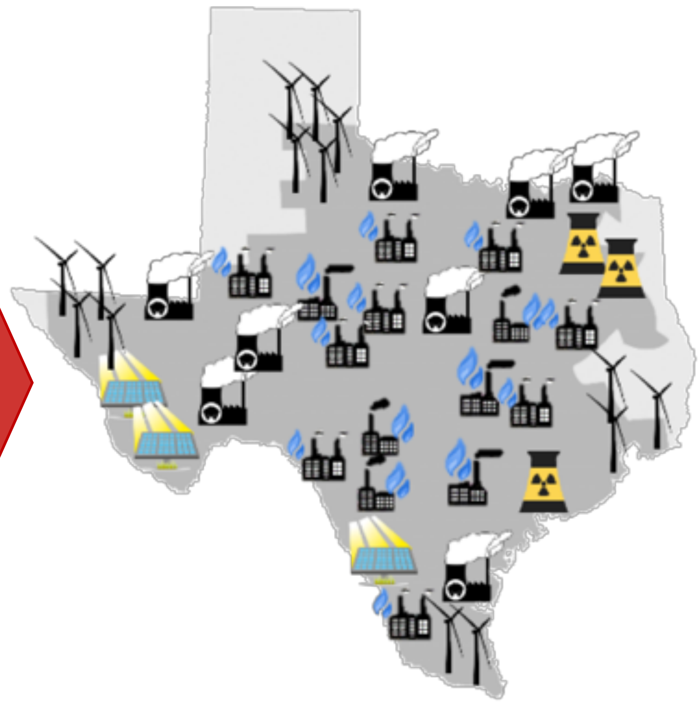
- **Generation Expansion Planning (GEP) models and algorithm**
  - Lara, C. L., Mallapragada, D. S., Papageorgiou, D. J., Venkatesh, A., & Grossmann, I. E. (2018). Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. *European Journal of Operational Research*, 271(3), 1037-1054.
- **Representative day selection in Generation Expansion Planning**
  - Mallapragada, D. S., Papageorgiou, D. J., Venkatesh, A., Lara, C. L., & Grossmann, I. E. (2018). Impact of model resolution on scenario outcomes for electricity sector system expansion. *Energy*, 163, 1231-1244.
  - Li, C., A.J. Conejo, J.D. Siirola, I.E. Grossmann. On representative day selection for capacity expansion planning of power systems under extreme events. *Under Review in Energy.*
- **Generation Expansion Planning under Uncertainty**
  - Lara, C. L., Siirola, J. D., & Grossmann, I. E. (2019). Electric power infrastructure planning under uncertainty: stochastic dual dynamic integer programming (SDDiP) and parallelization scheme. *Optimization and Engineering*, 1-39.
- **Integrated Generation and Transmission Expansion (GTEP) Planning**
  - Li, C., A.J. Conejo, P. Liu, B.P. Omell, J.D. Siirola, I.E. Grossmann. Mixed-integer Linear Programming Models and Algorithms for Generation and Transmission Expansion Planning of Power Systems. *Under Review in European Journal of Operational Research.*

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

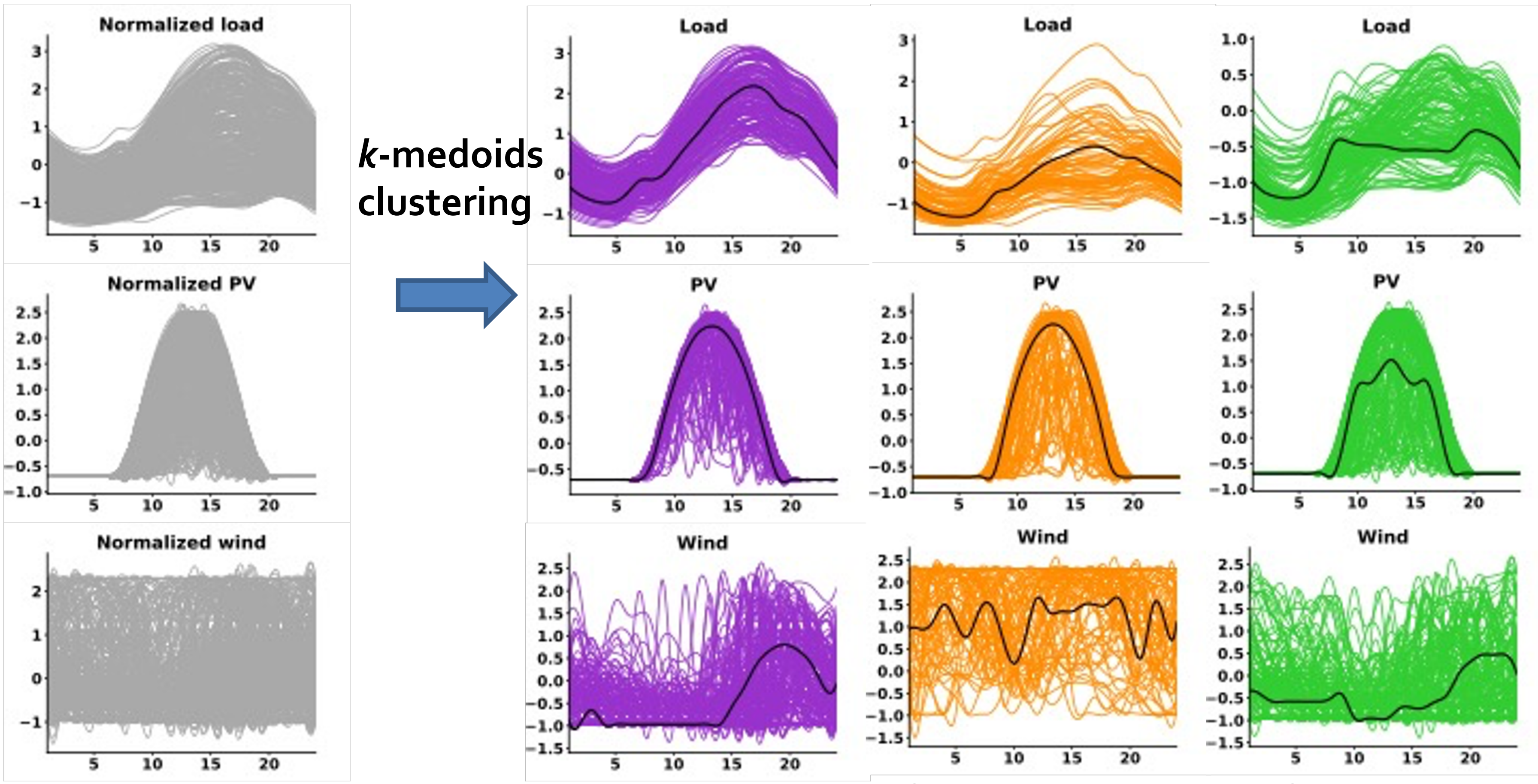
Research Challenges

- **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

Temporal Aggregation

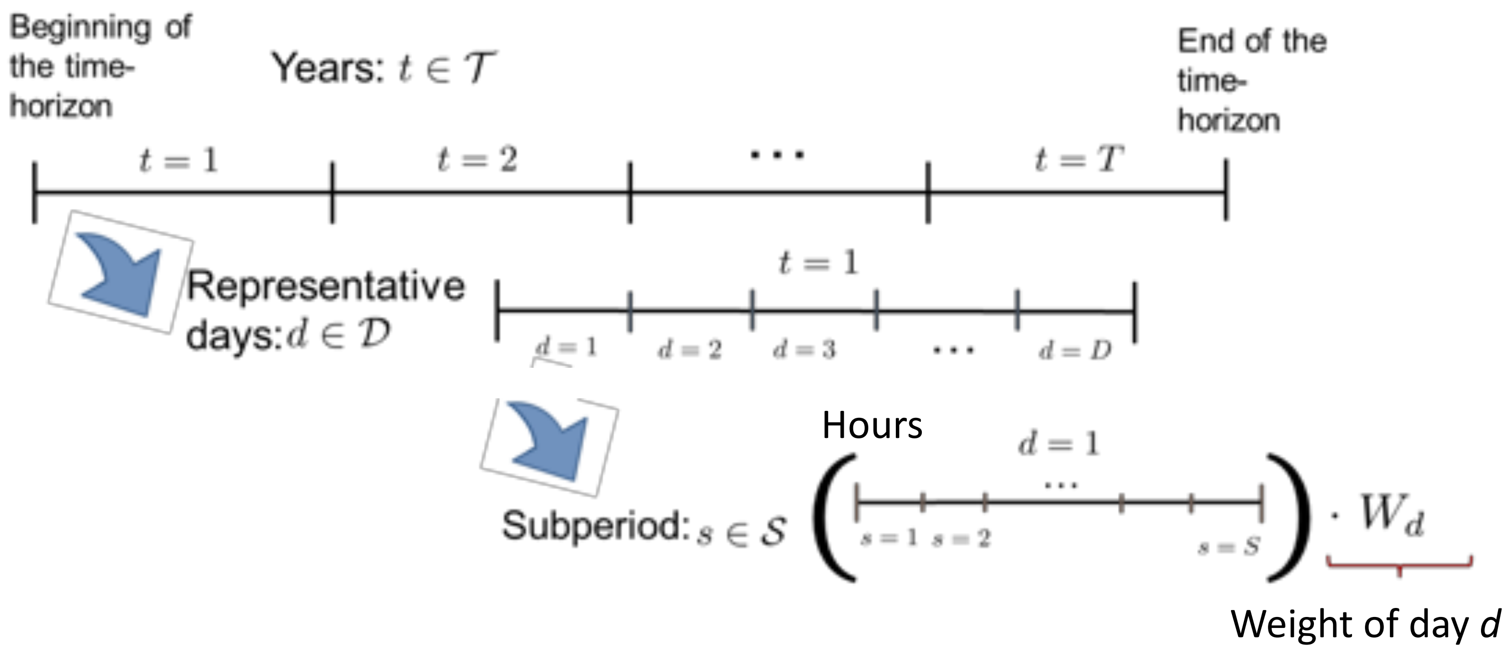
- Aggregate the days with **similar load and renewable** output time series using **machine learning-based clustering algorithms.**



Li, C., A.J. Conejo, J.D. Siirola, I.E. Grossmann. On representative day selection for capacity expansion planning of power systems under extreme events. Working paper.

Temporal Aggregation

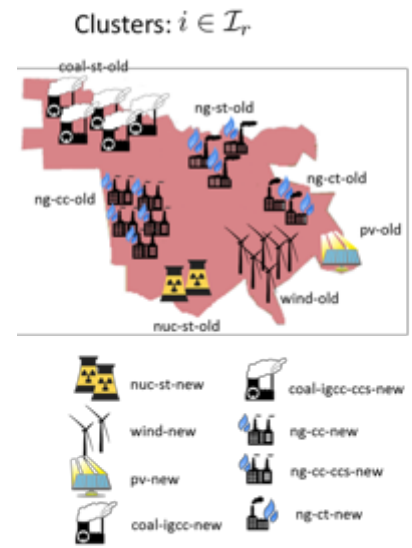
- d **representative days** per year to account for unit commitment and power flow in the **hourly** level

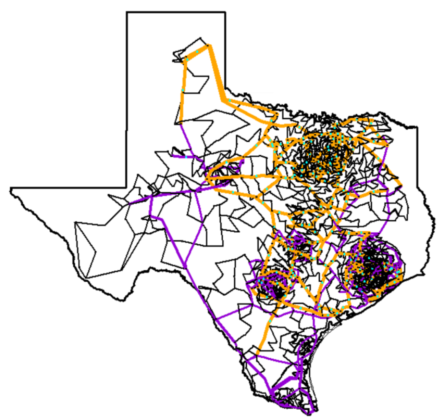


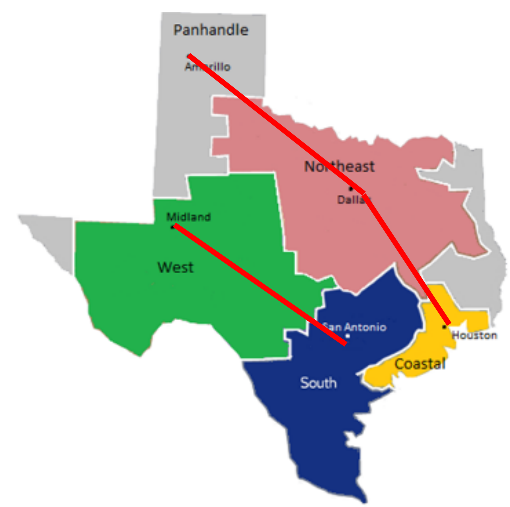
Spatial Aggregation

Region and cluster representation

- Area represented by a few **zones**
- Potential locations are the **midpoint** in each zone
- Center for each region: Panhandle (Amarillo), West (Midland), South (San Antonio), Coastal (Houston), Northeast (Dallas).
- Clustering of generators and storage units
- Only consider the **tielines** that connect the centers of two neighboring regions







Overview of Mixed-integer Linear Programming (MILP) Model

Continuous variables:	**Discrete variables:**
• Power output at sub-period *s*	• no. of generators installed at period *t*
• Curtailment generation slack at *s*	• no. of generators built at *t*
• Power flow between regions at *s*	• no. of generators retired at *t*
• Deficit from renewable quota at *t*	• no. of generators with life extended at *t*
• Spinning reserve at *s*	• whether transmission line *l* is installed at *t*
• Quick-start reserve at *s*	• whether transmission line *l* exists at *t*
• Voltage angle of region *r* at *s*	• no. of generators ON at sub-period *s*
• Power level and power charged or discharged at storage cluster *j*	• no. of generators starting up at *s*
	• no. of generators shutting down at *s*

<u>Objective function</u>:

Minimization of the **net present cost** over the planning horizon comprising:

- Variable operating cost
- Fixed operating cost
- Startup costs
- Cost of investments in new generators, transmission lines and storage units
- Cost to extend the life of generators that achieved their expected lifetime
- Fuel consumption
- Carbon tax for CO_2 emission
- Penalty for not meeting the minimum renewable annual energy production requirement

Lara, C. L., Mallapragada, D. S., Papageorgiou, D. J., Venkatesh, A., & Grossmann, I. E. (2018). Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. European Journal of Operational Research, 271(3), 1037-1054.
Li, C., A.J. Conejo, P. Liu, B.P. Omell, J.D. Siirola, I.E. Grossmann. Mixed-integer Linear Programming Models and Algorithms for Generation and Transmission Expansion Planning of Power Systems. Under review in European Journal of Operations Research.

Overview of Mixed-integer Linear Programming (MILP) Model

Summary of constraints:

- **Energy balance** in each region *r*.
- **DC power flow** calculate the power flow between any two nodes at each subperiod *s*
- **Capacity factor** of renewable generators .
- **Unit commitment constraints** to compute the startup and shutdown, operating limits and ramping rates for thermal generators.
- **Operating reserve constraints** to determine the maximum contribution per thermal generator for spinning and quick-start reserves, and the minimum total operating reserves.
- **Investment constraints** to ensure that the planning reserve and renewable energy contribution requirements are satisfied, and to limit the yearly installation per generation type.
- **Balance of generators** to define the number of generators that are **operational**, **built**, **retired**, and have their life **extended** in each time period *t*.

DC v.s. AC Power Flow Equations

DC power flow

$$P_i = \sum_{k=1}^{N} B_{ik}(\delta_i - \delta_k) \quad \forall i \in N$$

Real power only
Linear equations

AC power flow

$$\begin{aligned} P_i(V,\delta) &= V_i \sum_{k=1}^{N} V_k \big(G_{ik} \cos(\delta_i - \delta_k) \\ &\quad + B_{ik} \sin(\delta_i - \delta_k) \big) \quad \forall\, i \in \mathbf{N}, \\ Q_i(V,\delta) &= V_i \sum_{k=1}^{N} V_k \big(G_{ik} \sin(\delta_i - \delta_k) \\ &\quad - B_{ik} \cos(\delta_i - \delta_k) \big) \quad \forall\, i \in \mathbf{N}. \end{aligned}$$

Real and reactive power
nonlinear equations
(trigonometric functions)

DC is a good approximation for AC if

1) All system branch resistances are approximately zero
2) The differences between adjacent bus voltage angles are small
3) The system bus voltages are approximately equal to the 1.0 per unit
4) Reactive power flow is neglected

Comparison of Formulations of Transmission Expansion

Generalized Disjunctive Programming

Grossmann, I.E. and F. Trespalacios, "Systematic Modeling of Discrete-Continuous Optimization Models through Generalized Disjunctive Programming," *AIChE J*. **59**, 3276-3295 (2013).

$$\begin{bmatrix} NTE_{l,t} \\ p^{\text{flow}}_{l,t,d,s} = B_l(\theta_{sr(l),t,d,s} - \theta_{er(l),t,d,s}) \\ -F_l^{\max} \le p^{\text{flow}}_{l,t,d,s} \le F_l^{\max} \end{bmatrix} \vee \begin{bmatrix} \neg NTE_{l,t} \\ p^{\text{flow}}_{l,t,d,s} = 0 \end{bmatrix} \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

Big M reformulation

$$-(1-nte_{l,t})M \le p^{\text{flow}}_{l,t,d,s} - B_l(\theta_{sr(l),t,d,s} - \theta_{er(l),t,d,s}) \le (1-nte_{l,t})M \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$-F_l^{\max} nte_{l,t} \le p^{\text{flow}}_{l,t,d,s} \le F_l^{\max} nte_{l,t} \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

Hull reformulation

$$p^{\text{flow}}_{l,t,d,s} = B_l \Delta\theta^1_{l,t,d,s} \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$\theta_{sr(l),t,d,s} - \theta_{er(l),t,d,s} = \Delta\theta^1_{l,t,d,s} + \Delta\theta^2_{l,t,d,s} \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$-\pi \cdot nte_{l,t} \le \Delta\theta^1_{l,t,d,s} \le \pi \cdot nte_{l,t} \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$-\pi(1 - nte_{l,t}) \le \Delta\theta^2_{l,t,d,s} \le \pi(1 - nte_{l,t}) \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

Tighter formulation, also has more variables

Comparison of Formulations of Transmission Expansion

Alternative big M formulation

Bahiense, L., Oliveira, G. C., Pereira, M., & Granville, S. (2001). A mixed integer disjunctive model for transmission network expansion. *IEEE Transactions on Power Systems*, *16*(3), 560-565.

$$p^{\text{flow}+}_{l,t,d,s} - B_l \Delta\theta^{+}_{l,t,d,s} \leq 0 \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$p^{\text{flow}-}_{l,t,d,s} - B_l \Delta\theta^{-}_{l,t,d,s} \leq 0 \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$p^{\text{flow}+}_{l,t,d,s} - B_l \Delta\theta^{+}_{l,t,d,s} \geq -M_l(1 - nte_{l,t}) \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$p^{\text{flow}-}_{l,t,d,s} - B_l \Delta\theta^{-}_{l,t,d,s} \geq -M_l(1 - nte_{l,t}) \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$p^{\text{flow}}_{l,t,d,s} = p^{\text{flow}+}_{l,t,d,s} - p^{\text{flow}-}_{l,t,d,s} \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$\theta_{sr(l),t,d,s} - \theta_{er(l),t,d,s} = \Delta\theta^{+}_{l,t,d,s} - \Delta\theta^{-}_{l,t,d,s} \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$p^{\text{flow}+}_{l,t,d,s} \leq F^{\max}_l nte_{l,t} \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$p^{\text{flow}-}_{l,t,d,s} \leq F^{\max}_l nte_{l,t} \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

$$p^{\text{flow}+}_{l,t,d,s}, p^{\text{flow}-}_{l,t,d,s}, \Delta\theta^{+}_{l,t,d,s}, \Delta\theta^{-}_{l,t,d,s} \geq 0 \quad \forall l \in \mathcal{L}^{new}, t, d, s$$

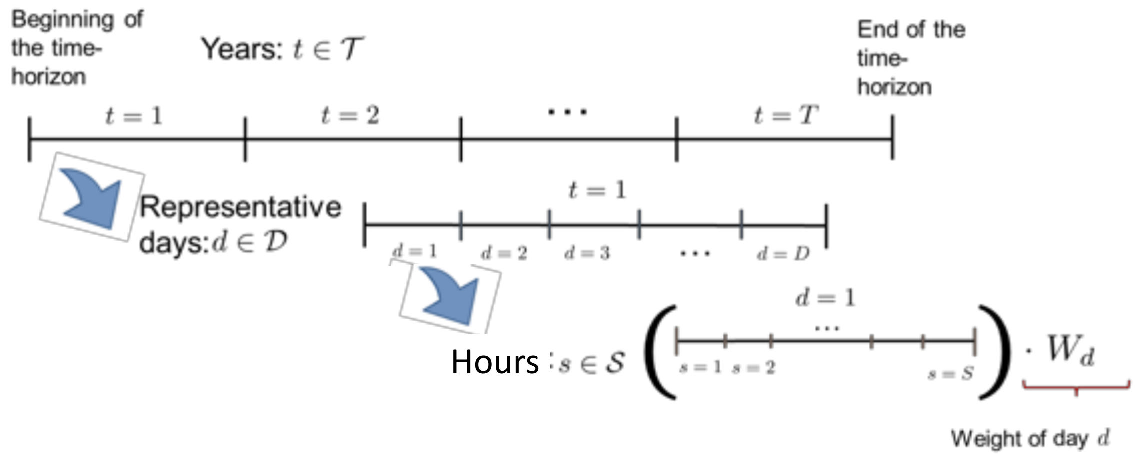
The authors claim that alternative big M formulation is **tighter** than big M formulation

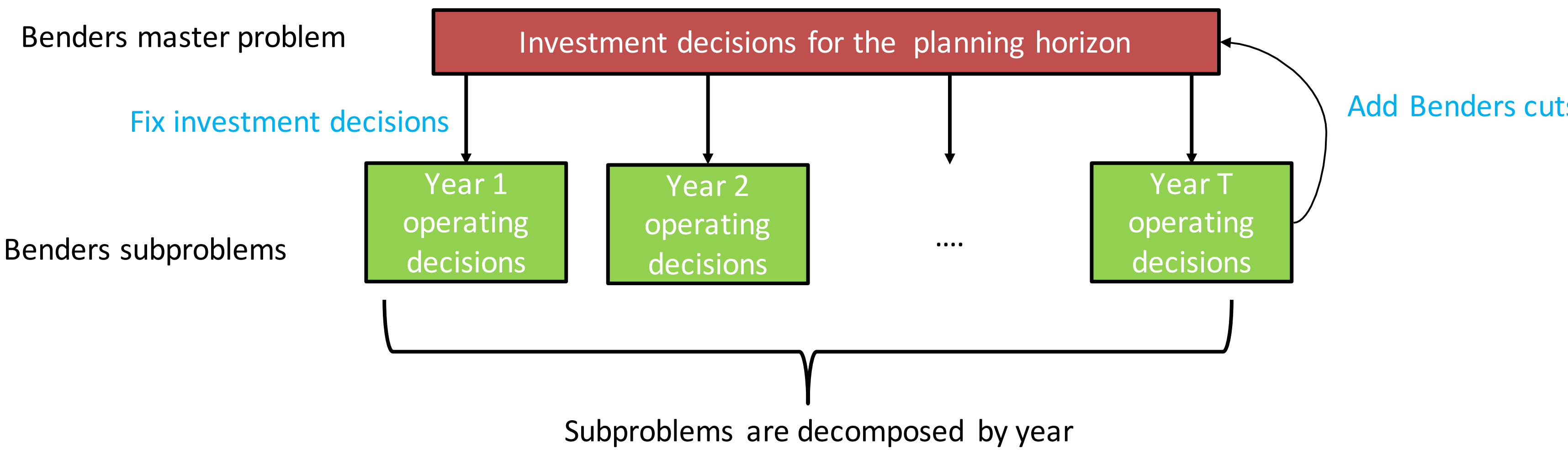
Theorem: The two formulations have **the same feasible region** when project on the original variable space

Solution Techniques-Benders Decomposition

Time scale approach:

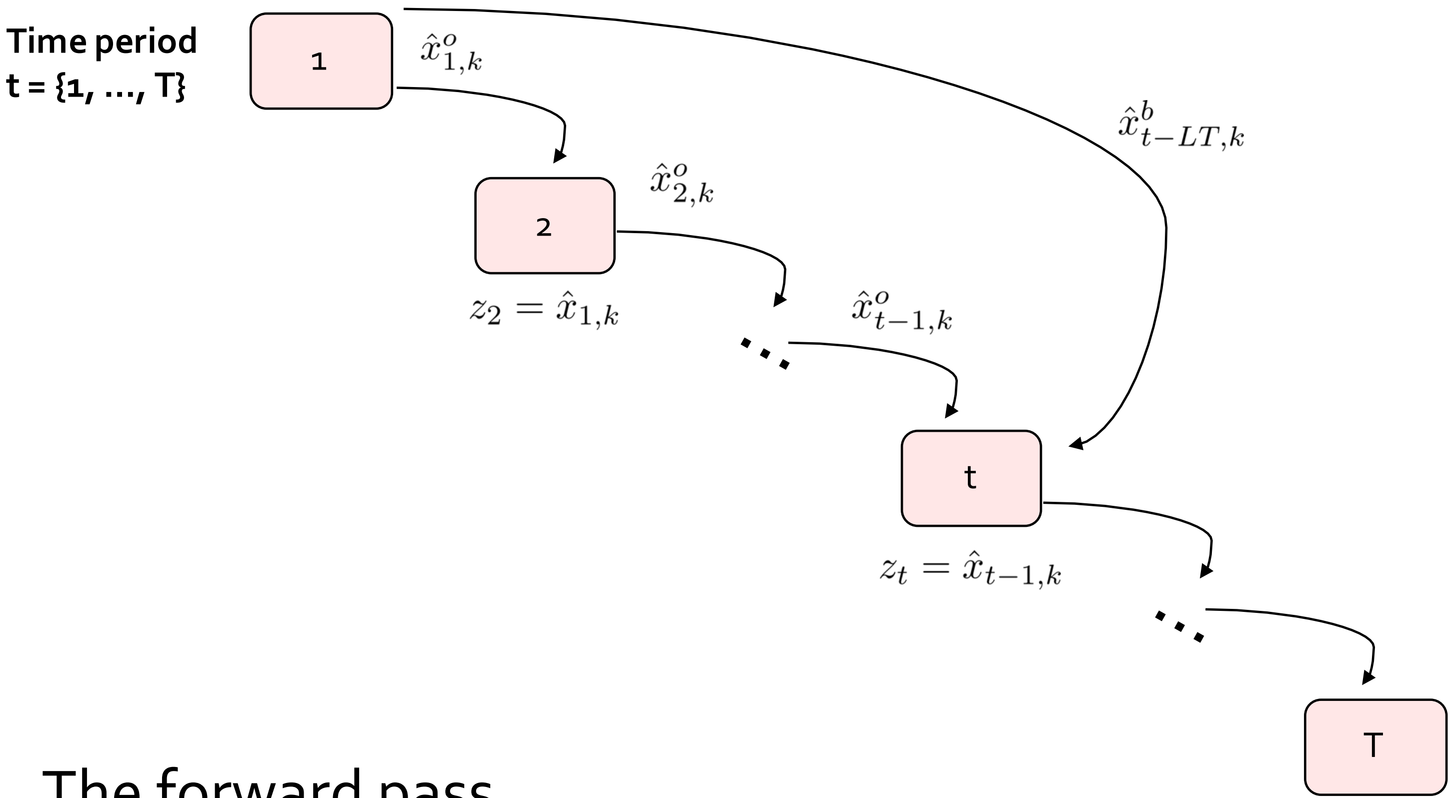
d **representative days** per year to account for unit commitment in the **hourly** level





Li, C., A.J. Conejo, P. Liu, B.P. Omell, J.D. Siirola, I.E. Grossmann. Mixed-integer Linear Programming Models and Algorithms for Generation and Transmission Expansion Planning of Power Systems. Under review in European Journal of Operations Research.

Solution Techniques-Nested Benders Decomposition

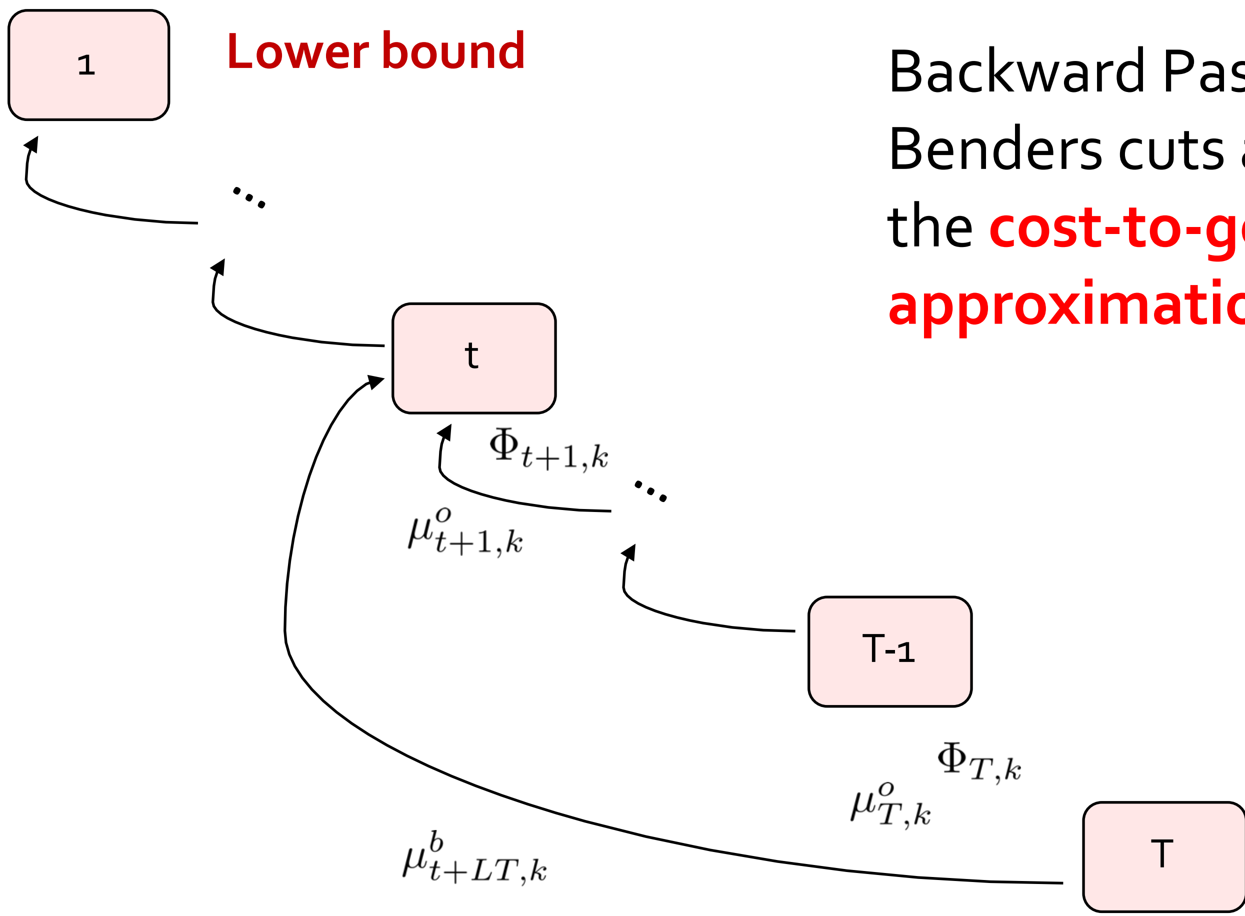


Upper bound

The forward pass solves the model in a **myopic fashion.**

Lara, C.L. et al., "Electric Power Infrastructure Planning: Mixed-Integer Programming Model and Nested Decomposition Algorithm," *European Journal of Operational Research* **271**, 1037–1054 (2018).
Birge, J. R. (1985). Decomposition and partitioning methods for multistage stochastic linear programs. Operations research, 33(5), 989-1007.

Solution Techniques-Nested Benders Decomposition

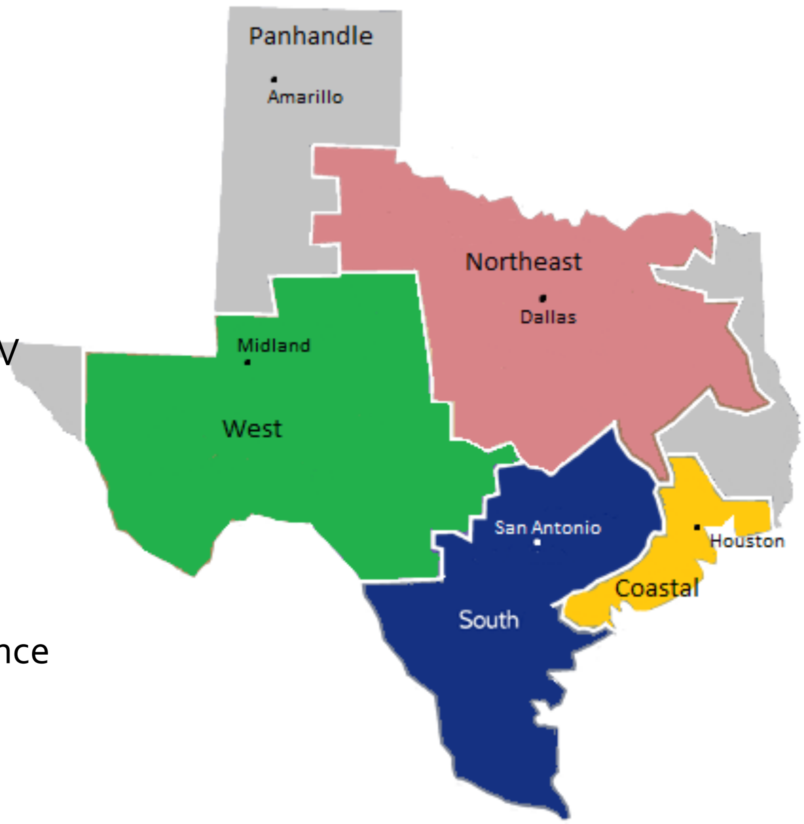


Backward Pass generates Benders cuts and improves the **cost-to-go approximation**

Time period
t = {T, …, 1}

ERCOT Case Study

- 20 year time horizon (1st year is 2019)
- **Load** Data from ERCOT database
- **Solar and wind capacity factor** data from NREL
- **Generator cost** information from NREL (Annual Technology Baseline (ATB)
- **Storage data** from Schmidt et al. (2017) Nature Energy.
- **Transmission line** data from Texas Synthetic Grid. Only 500 kV tielines between two neighboring regions are considered
- All costs in 2019 USD
- Regions: Northeast, West, Coastal, South, Panhandle
- **Fuel price** data from EIA Annual Energy Outlook 2016 (reference case)
- **Carbon tax** is zero in the first year and grows linearly across years to $0.325/kg CO_2.



4 representative days, 15 years results

Fullspace mixed-integer linear programming (MILP) models

formulation	Integer Var	Binary Var	Continuous Var	Constraints	UB	LB	Wall time
big-M	274,920	2,800	564,826	1,543,966	-	21.13	36,000
alternative big M	274,920	2,800	1,102,426	2,081,566	-	21.13	36,000
hull	274,920	2,800	833,626	2,081,566	-	281.73	36,000

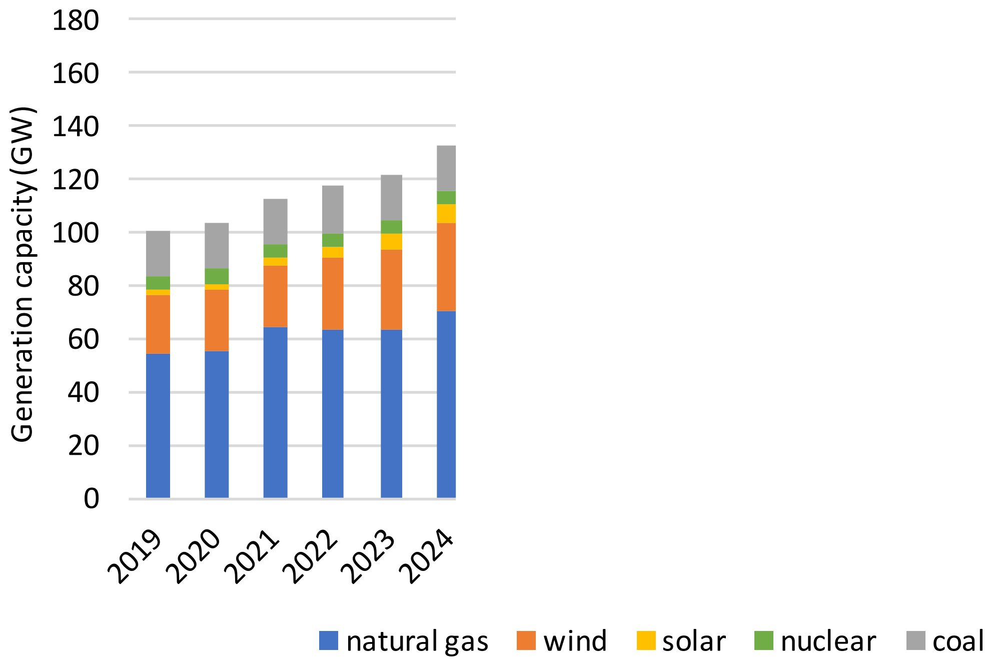
All the problems are solved with Cplex v 12.9.0.0 from Pyomo. **The fullspace model cannot be solved directly. No feasible solution can be found within 10 hours**

Decomposition algorithms

algorithm	formulation	ub	lb	gap	Wall time (secs)
Benders	big-M	283.7	282.6	0.38%	5,115
Benders	**alternative big M**	**283.9**	**281.6**	**0.82%**	**3,693**
Benders	hull	282.6	280.6	0.71%	8,418
nested Benders	big-M	295.7	268.9	9.98%	53,682
nested Benders	alternative big M	294.2	265.5	10.81%	43,389
nested Benders	hull	288.0	269.3	6.97%	37,577

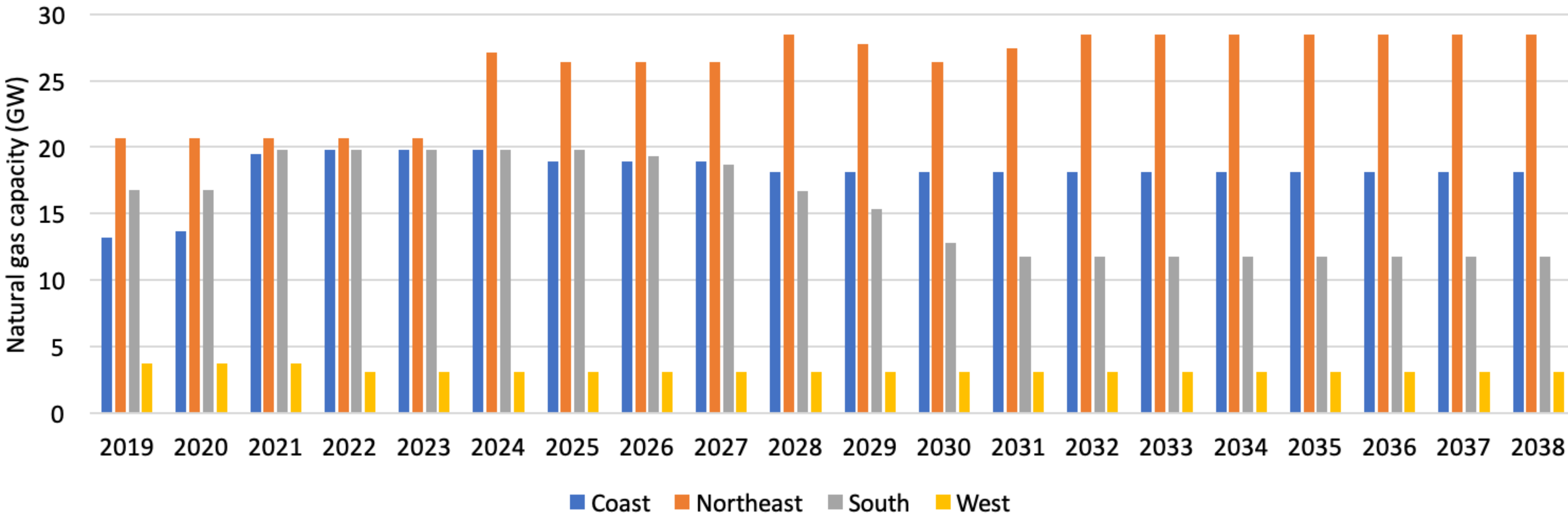
The **Benders decomposition algorithm with the alternative big-M** formulation has the best computational performance

20-year Generation Expansion



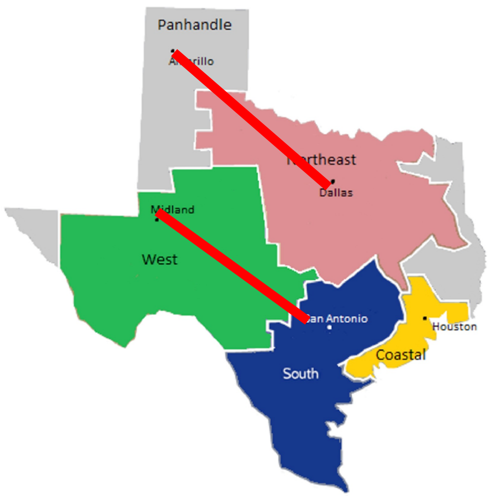
- Natural gas capacity increases in the beginning and then decreases due to the increase in carbon tax
- Most projected capacity expansion is in wind and solar. **27-fold** increase in **solar** and **87%** increase in **wind**.

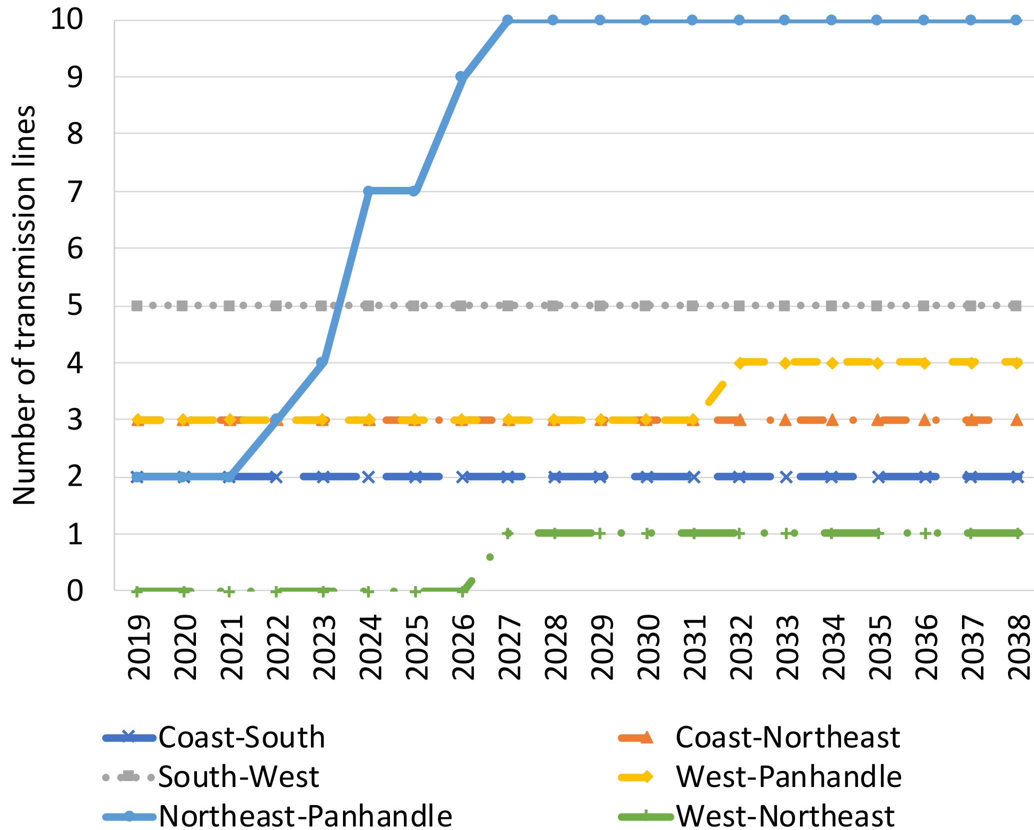
Geographical Distribution of Natural Gas Capacity



- Most natural gas expansions are expected to take place in the **Northeast** and **Coast** regions where the absolute increase in **load** is **high** and capacity factors for **renewables** are relatively **low**.

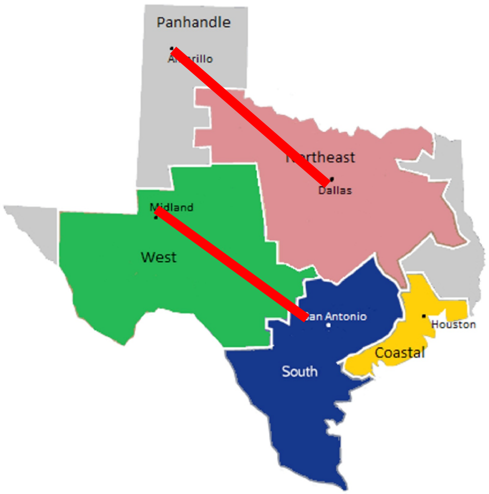
Transmission Expansion



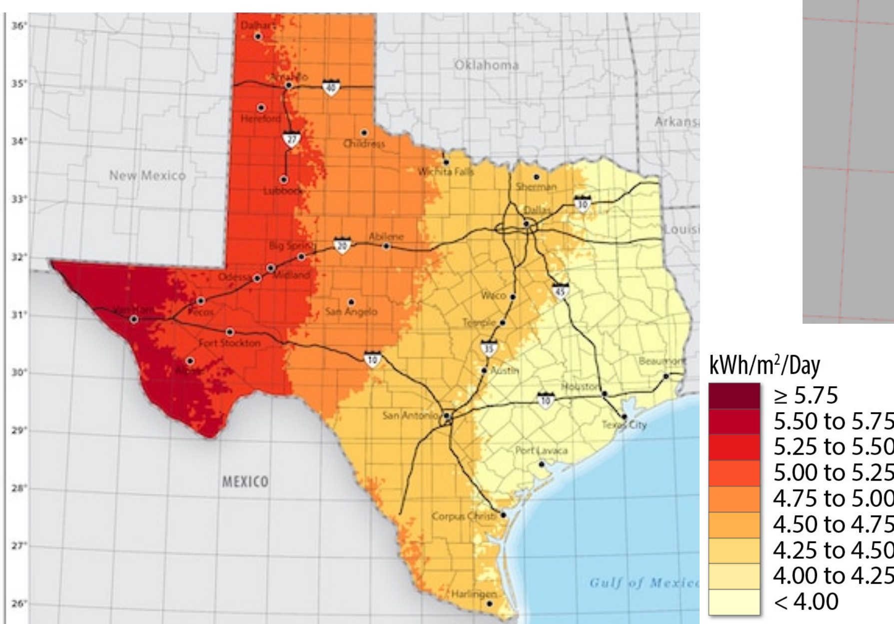


- Most of the transmission lines are built for **Northeast-Panhandle** and **South-West** in order to transfer the power generated by the **renewables** in West and Panhandle to other regions

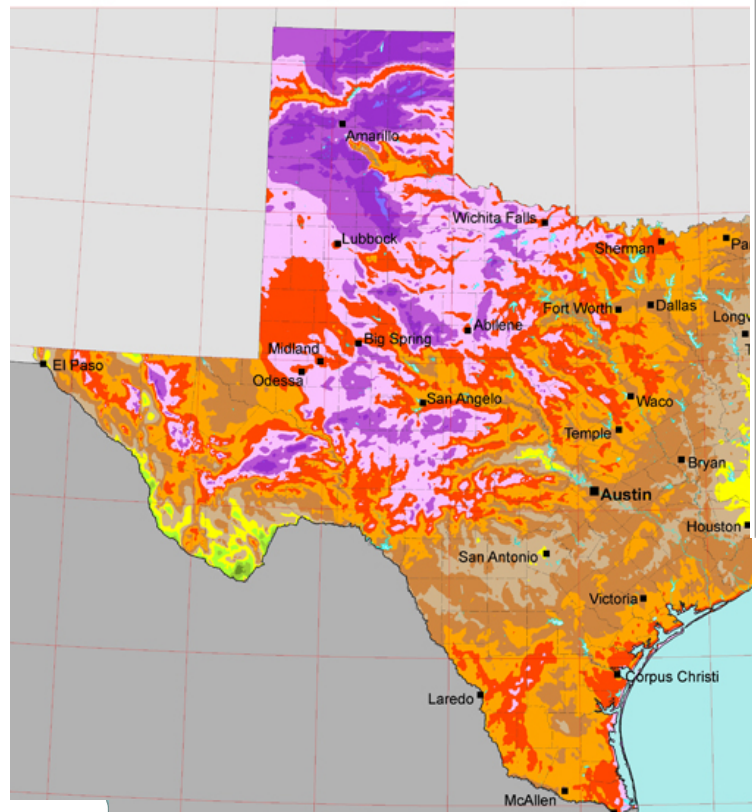
Transmission Expansion

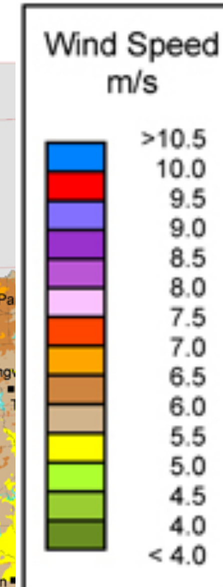


Average annual solar irradiance



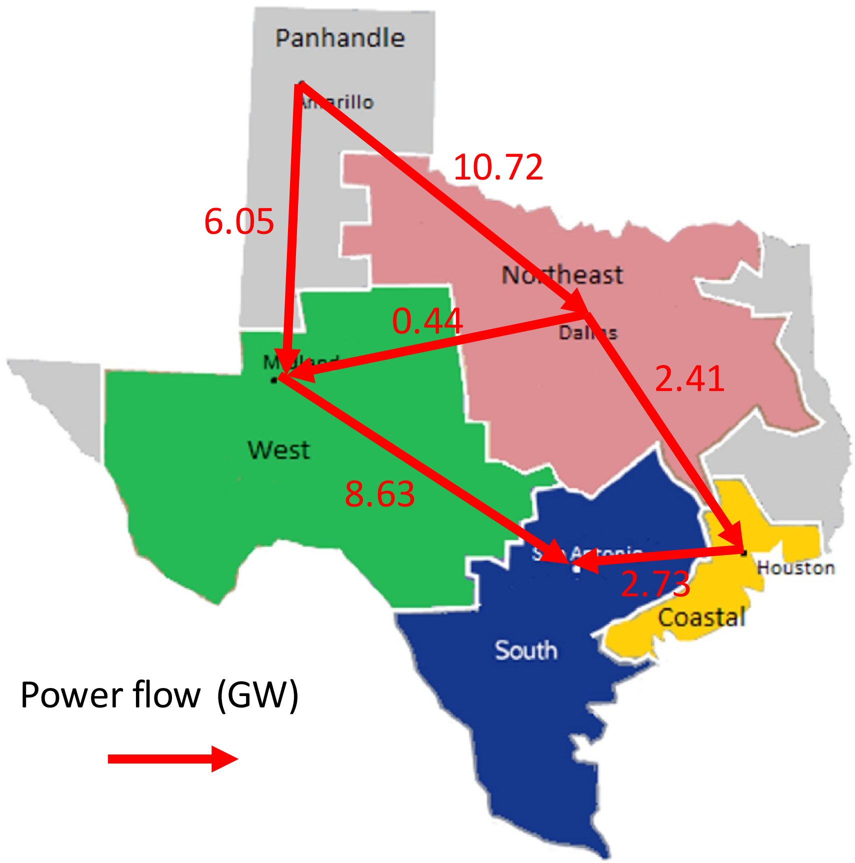
Average annual wind speed at 80 meter





Data source: NREL

Power Flow in ERCOT



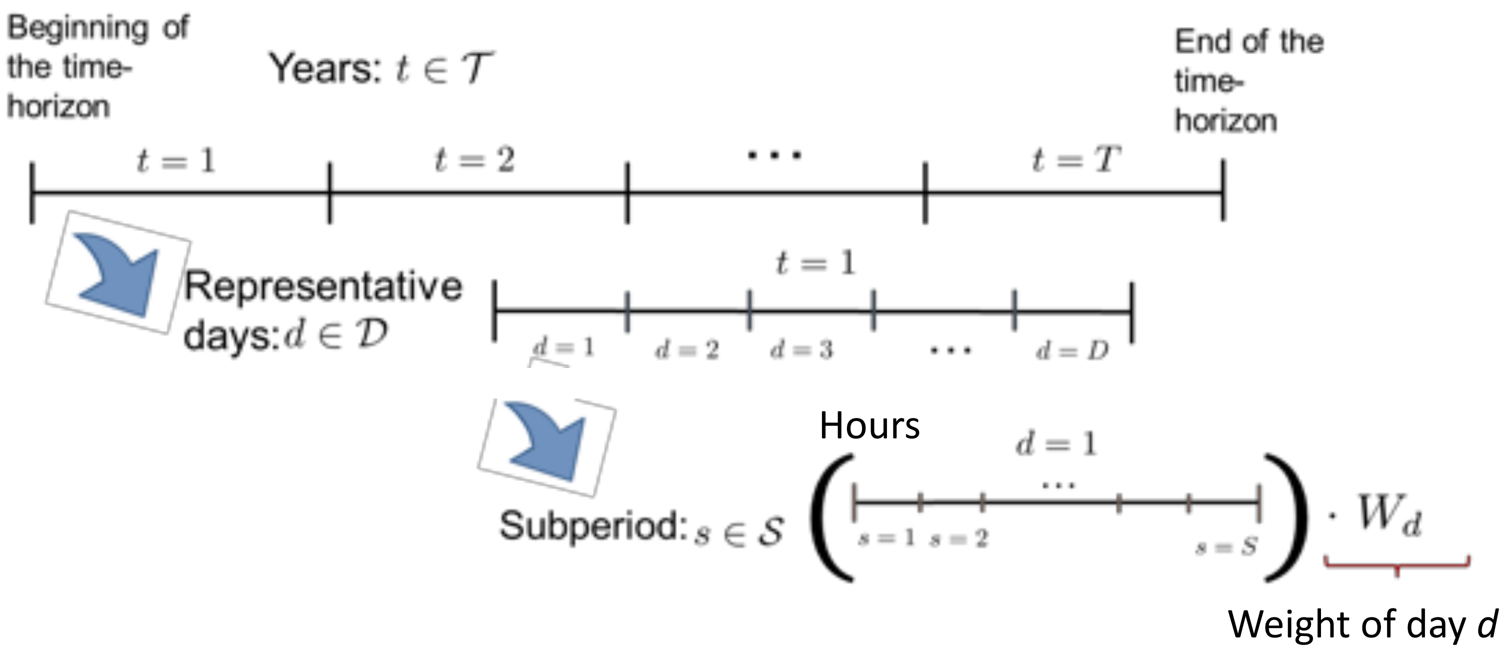
The largest power flow magnitudes are **Panhandle-Northeast, West-South** due the surplus of their renewable energy generation

There are potential benefits in **integrating generation** and **transmission** expansion

Year 20 (2038), representative day 15, 11pm

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

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

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

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

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

Investment decisions for year *t*

The whole dataset

operating decisions for year *t* day *d*

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

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

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

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

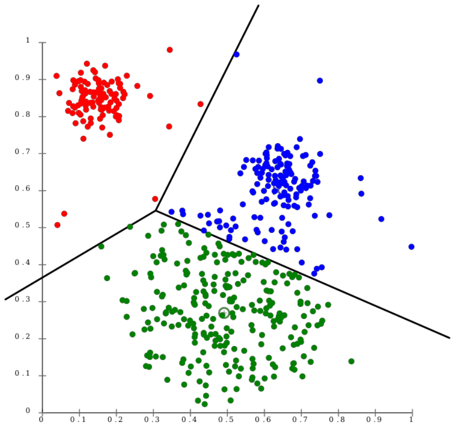
The set of representative days

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

Relaxed integrality constraints

K-means clustering

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



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

MINLP formulation:

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

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

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

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

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

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

K-medoids clustering

- The center μ_i has to be a data point. Centroid v.s. medoid

MILP formulation:

$$\min_{\mathbf{z},\mathbf{y}} \sum_{ij} d_{ij} z_{ij}$$

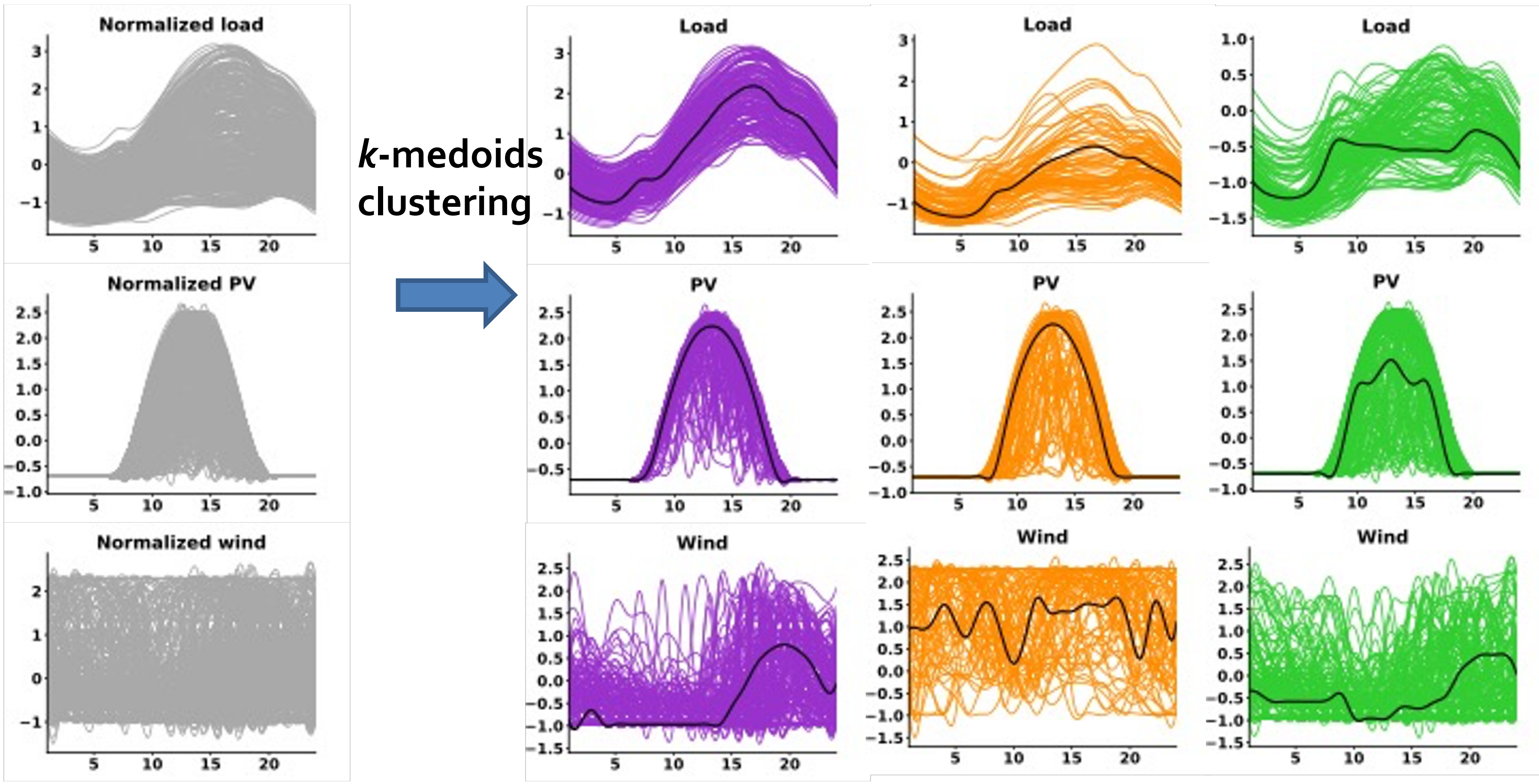
$$\sum_{j=1}^{n} z_{ij} = 1 \quad \forall i = 1, 2, \ldots, n$$

$$z_{ij} \leq y_j \quad \forall i = 1, 2, \ldots, n, j = 1, 2, \ldots, n$$

$$\sum_{i=1}^{n} y_i = k$$

Input-based method

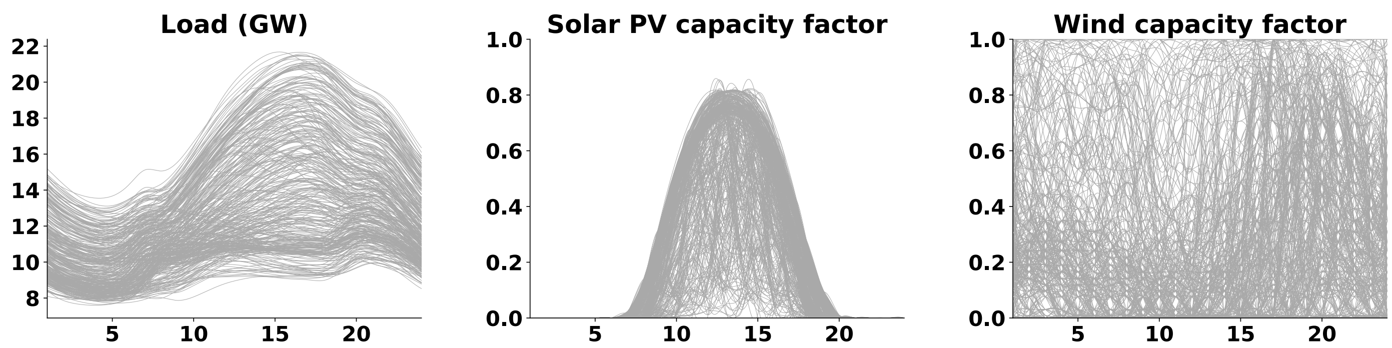
- Clustering is performed directly on the input data (load, capacity factors)



Cost-based method

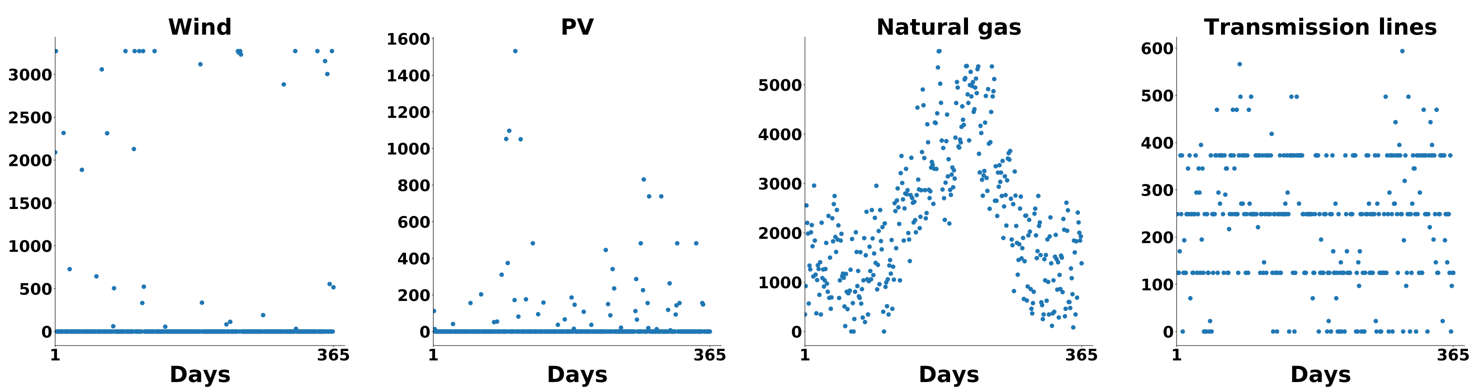
- **Hypothesis:** The days with similar **optimal investment decisions**, i.e., the days that need similar generators, transmission lines, and storage units, are similar and should be assigned to the same cluster

Raw data



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

Investment cost breakdown after reduction (million dollars)



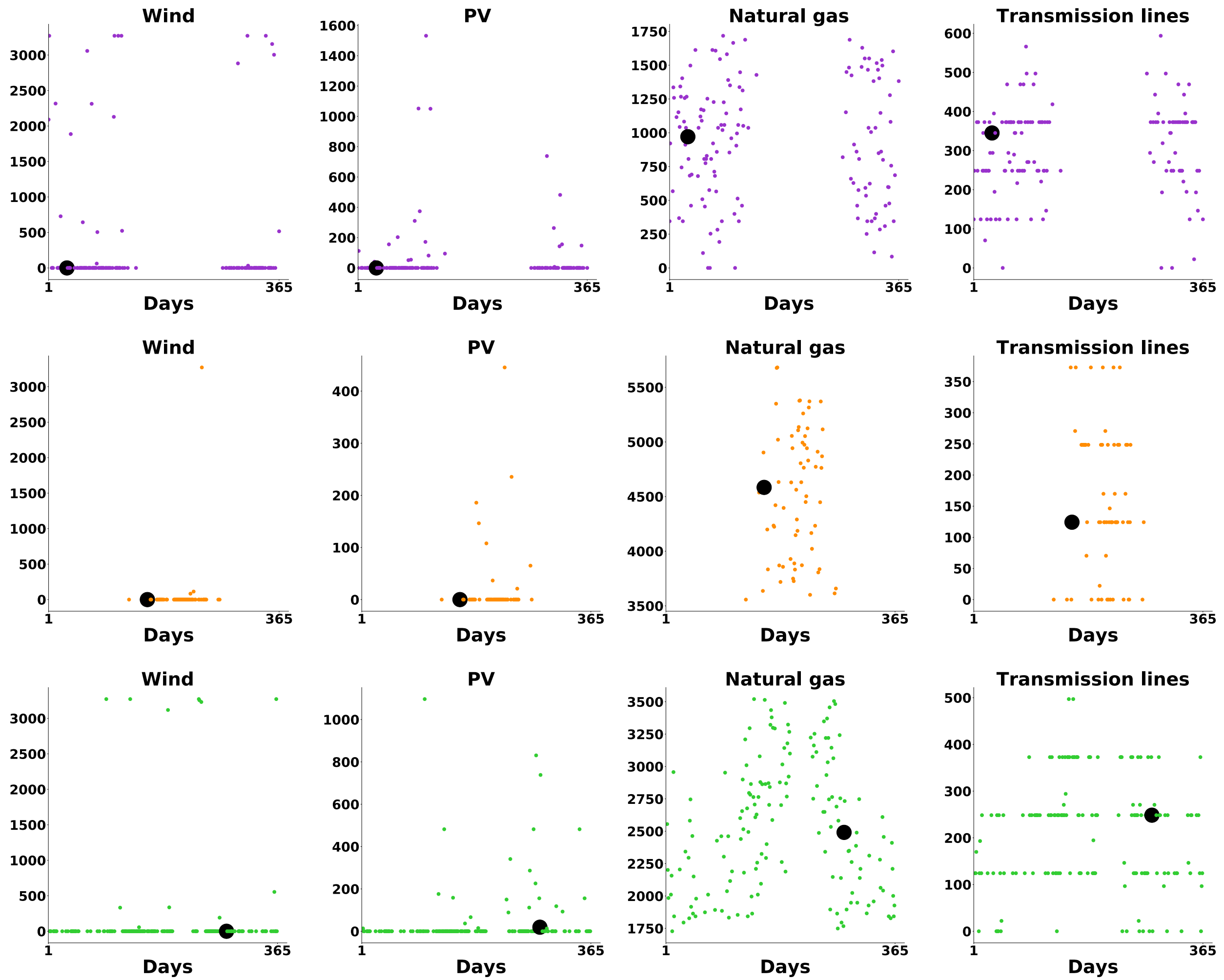
Cost-based method

K-medoids clustering

Cluster 1, w=141

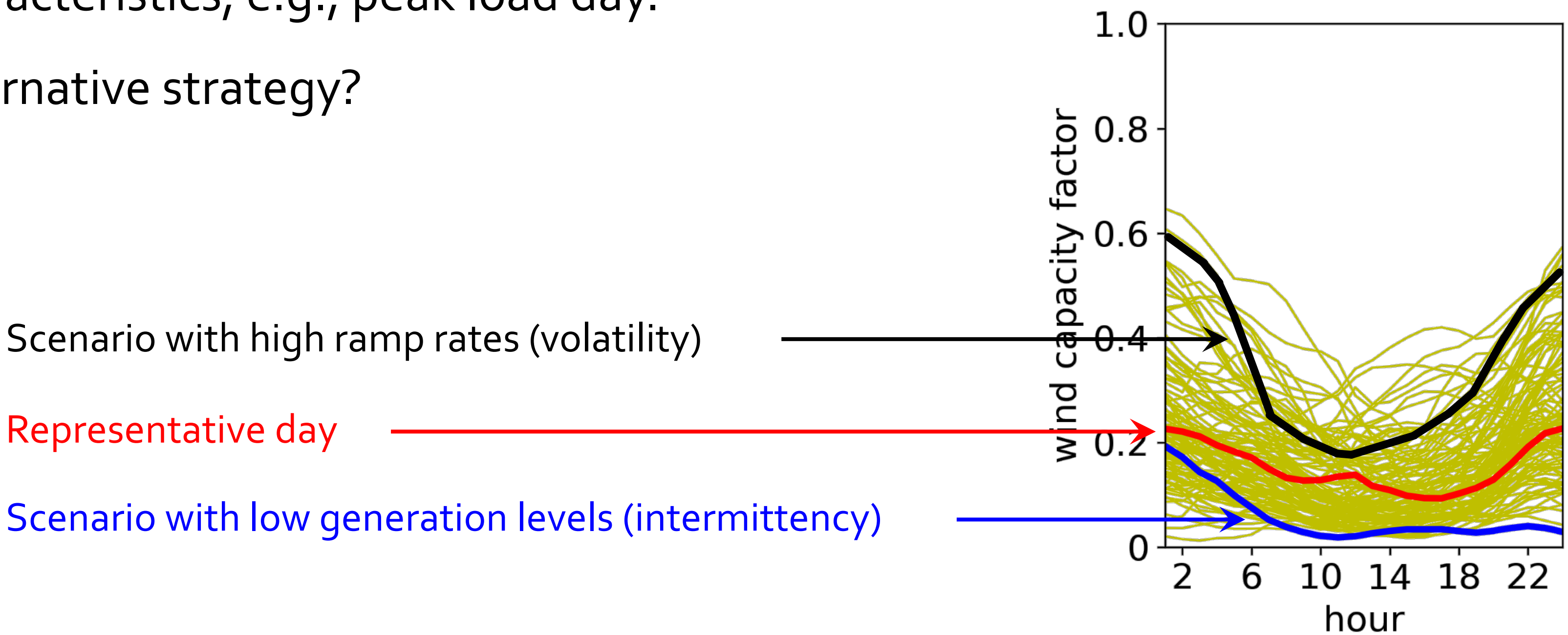
Cluster 2, w=65

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

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

Power generation ± power flow in/out ± 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.

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

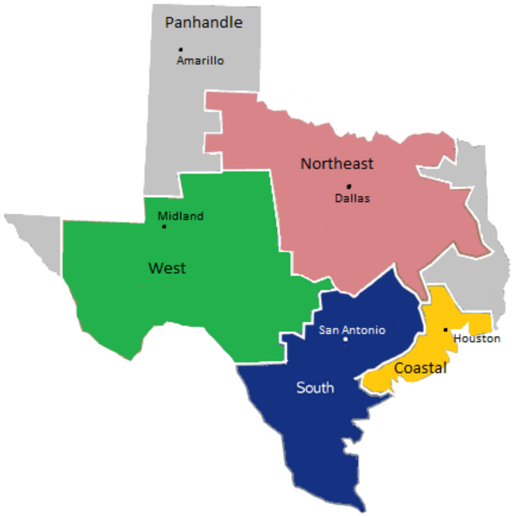
- **Lower bound:** Reduced model provides lower bound under certain assumptions.

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

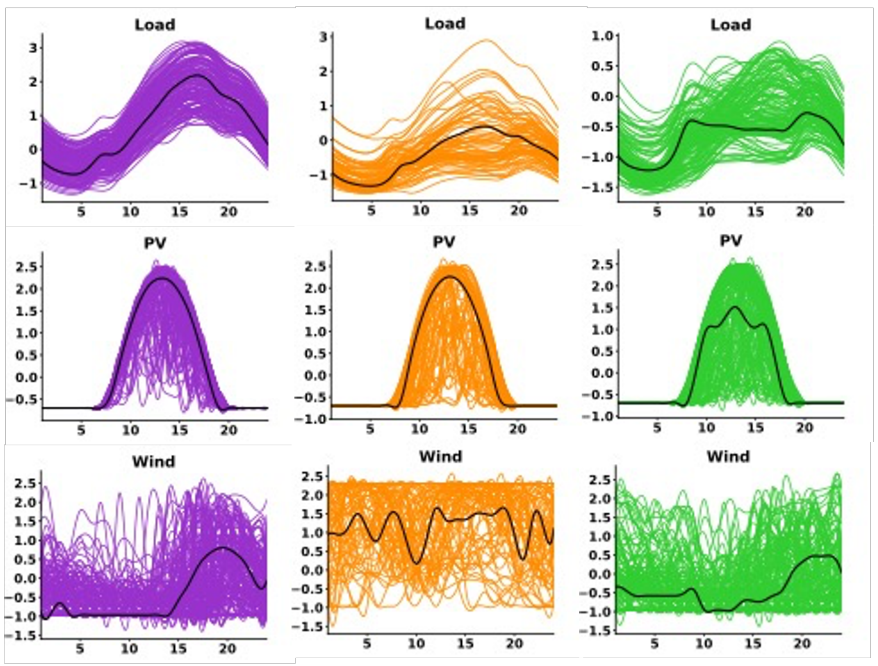
- ERCOT region, 5 years planning problem
- The whole dataset *D* has 365 days that consists of load and capacity factor data



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

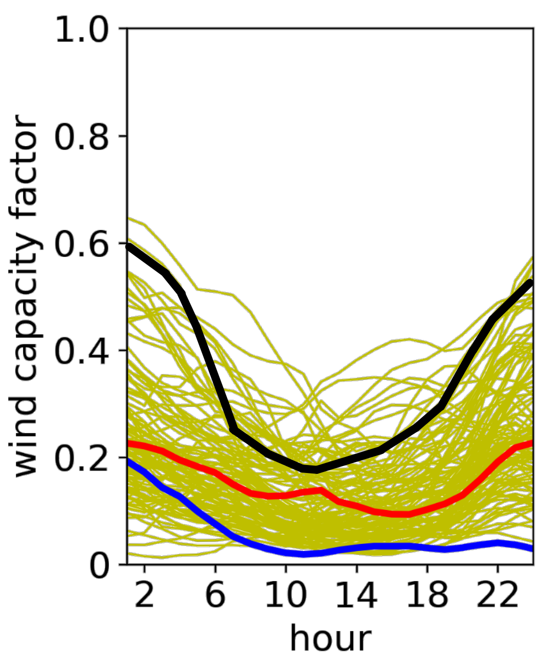
- Only using the representative days from centroids/medoids of the clustering algorithms **cannot guarantee feasibility**
- Cost-based approach has fewer infeasible days when *k* is large



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

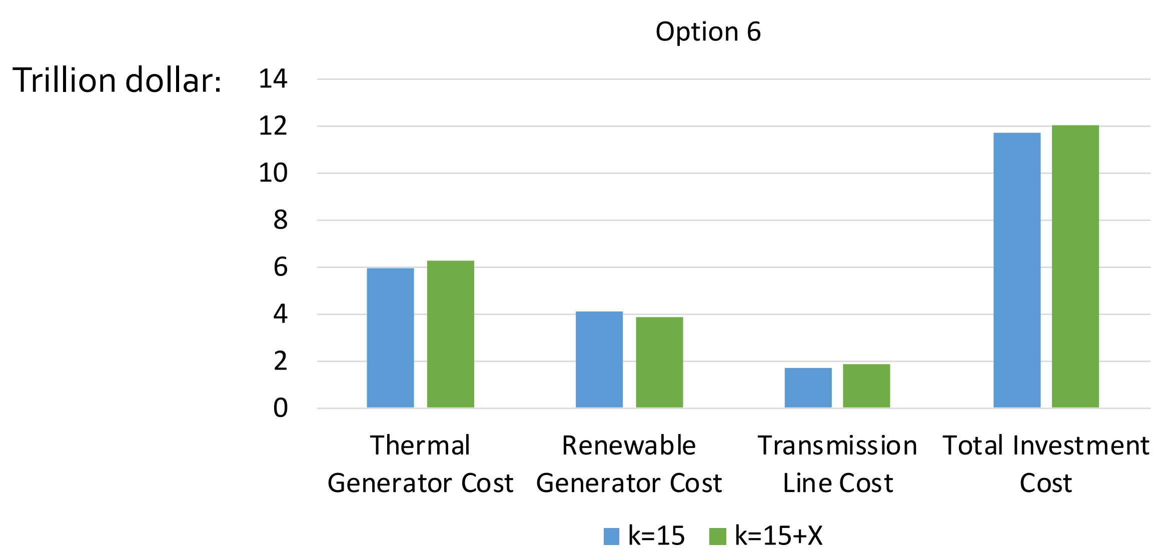
Optimality Gap

- "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.**

Acknowledgment

- Advisor: Prof. Ignacio E. Grossmann
- Collaborators:
  - Prof. Antonio Conejo (OSU)
  - Dr. John Siirola (Sandia)
  - Dr. Benjamin Omell (NETL)
  - Dr. Peng Liu (NETL)
- Acknowledgment
  - Dr. Cristiana Lara
  - Dr. Dimitri Papageorgiou (ExxonMobil)
- Funding: DOE/NETL IDAES Project

IDAES
Institute for the Design of Advanced Energy Systems