



DATA DRIVE POWER SYSTEM ANALYSIS: POWER FLOW CALCULATION, TOPOLOGY IDENTIFICATION, AND OPERATION MODE ANALYSIS

Ning Zhang

Associate Professor, Tsinghua University

2020.04 | Beijing China



Energy Intelligence Lab



EI Lab
Energy Intelligence Laboratory
智慧能源实验室





➤ Ning Zhang

- Department of Electrical Engineering, Tsinghua University

- BS (2007), Ph.D(2012) from Tsinghua University, China.
- Post Doc. 2012-2014, Tsinghua university, China.
- Research assistant, 2010.10~2011.07, University of Manchester, UK
- Research assistant. 2013.12~2014.03, Harvard University, US
- Assistant Professor. 2014.09~ 2017.01, Tsinghua University, China.
- Associate Professor. 2017.01~, Tsinghua University, China.

□ Main research interests:

- ✓ Renewable energy analytics and grid integration
- ✓ Multiple energy systems
- ✓ Power system planning
- ✓ Power system uncertainty analysis
- ✓ Power system data driven analysis

- **Website:** <http://www.ningzhang.net/>





Yuxiao Liu, Ph.D. student in Tsinghua University. Academic visit to MIT from 2019 to 2020



Qingchun Hou, Ph.D. student in Tsinghua University. Academic visit to University of Washington from 2019 to 2020



Yi Wang, received the B.S. degree from Huazhong University of Science and Technology (HUST) in 2014 and the Ph.D. degree at Tsinghua University in 2019. He is currently a postdoctoral researcher in ETH Zurich.



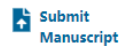
Jiawei Zhang, Ph.D. student in Tsinghua University.



Background

- Machine learning shows increasing successful applications in power systems









IEEE Transactions on Smart Grid











Home Popular **Early Access** Current Issue All Issues About Journal

Early Access Articles

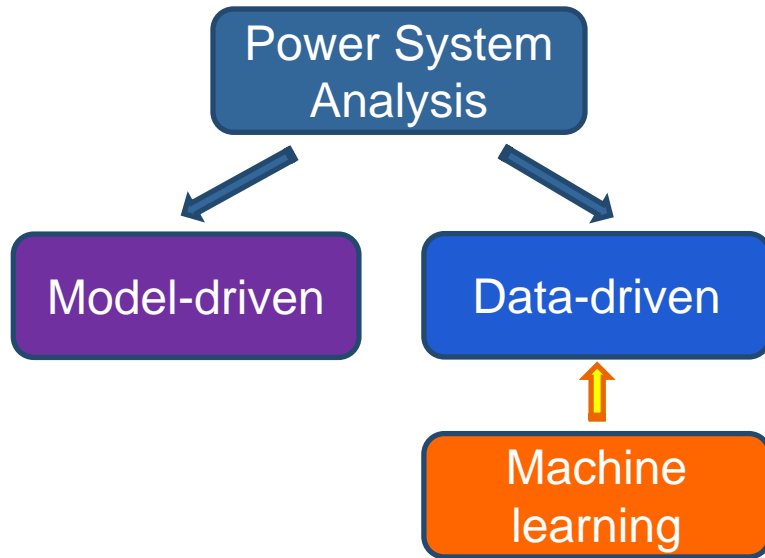
Early Access articles are made available in advance of the final electronic or print versions. Early Access articles are peer reviewed but may not be fully edited. They are fully citable from the moment they appear in IEEE *Xplore*.

- ☐ **A Supervised-Learning-Based Strategy for Optimal Demand Response of an HVAC System in a Multi-Zone Office Building**
 Young-Jin Kim
 Publication Year: 2020, Page(s): 1 - 1
 ▶ Abstract  (2341 Kb) 
- ☐ **Synchrophasor Missing Data Recovery via Data-Driven Filtering**
 Stavros Konstantinopoulos ; Genevieve M. De Mijolla ; Joe H. Chow ; Hanoch Lev-Ari ; Meng Wang
 Publication Year: 2020, Page(s): 1 - 1
 ▶ Abstract  (5302 Kb) 
- ☐ **Deep Reinforcement Learning Based Energy Storage Arbitrage With Accurate Lithium-ion Battery Degradation Model**
 Jun Cao ; Dan Harrold ; Zhong Fan ; Thomas Morstyn ; David Healey ; Kang Li
 Publication Year: 2020, Page(s): 1 - 1
 ▶ Abstract  (11271 Kb) 
- ☐ **Model-free Data Authentication for Cyber Security in Power Systems**
 Shengyuan Liu ; Shutang You ; He Yin ; Zhenzhi Lin ; Yilu Liu ; Wenxuan Yao ; Lakshmi Sundaresh
 Publication Year: 2020, Page(s): 1 - 1
 ▶ Abstract  (933 Kb) 

- ☐ **Virtual Inertia from Smart Loads**
 Tong Chen ; Jinrui Guo ; Balarko Chaudhuri ; Ron S. Y. Hui
 Publication Year: 2020, Page(s): 1 - 1
 ▶ Abstract  (1014 Kb) 
- ☐ **Graph-based Faulted Line Identification Using Micro-PMU Data in Distribution Systems**
 Ying Zhang ; Jianhui Wang ; Mohammad E. Khodayar
 Publication Year: 2020, Page(s): 1 - 1
 ▶ Abstract  (1801 Kb) 
- ☐ **Two-stage WECC Composite Load Modeling: A Double Deep Q-Learning Networks Approach**
 Xinan Wang ; Yishen Wang ; Di Shi ; Jianhui Wang ; Zhiwei Wang
 Publication Year: 2020, Page(s): 1 - 1
 ▶ Abstract  (2515 Kb) 
- ☐ **Adaptive Distributionally Robust Optimization for Electricity and Electrified Transportation Planning**
 Ali Hajebrahimi ; Innocent Kamwa ; Erick Delage ; Morad Abdelaziz
 Publication Year: 2020, Page(s): 1 - 1
 ▶ Abstract  (5901 Kb) 



Brief of the presentation



➤ 1) Power Flow Calculation

- How to calculate power flow using a data-driven model.

➤ 2) Topology Identification

- How to identify the topology and estimate line parameters of distribution network without the measurement of voltage phase angles.

➤ 3) Operation Mode Analysis

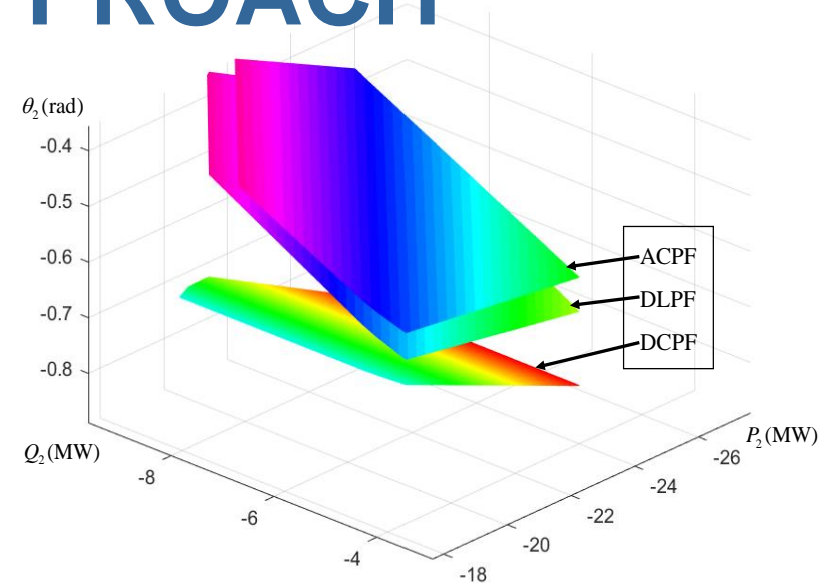
- How to cluster the operation mode of power system under high renewable energy penetration.





DATA-DRIVEN POWER FLOW LINEARIZATION

-A REGRESSION APPROACH



Yuxiao Liu, Ning Zhang, Yi Wang, Jingwei Yang and Chongqing Kang. Data-driven power flow linearization: a regression approach, IEEE Transactions on Smart Grid, 2019, 10(3): 2569 - 2580.



Purpose and background



Why power flow function linearization?



$$P_i = V_i \sum_{j \in i} V_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}) \quad i = 1, 2, \dots, N$$

$$Q_i = V_i \sum_{j \in i} V_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij}) \quad i = 1, 2, \dots, N$$

- Basis of optimization and control
 - Unit commitment, voltage control, OPF, planning optimization...
- Basis of analytical analysis
 - Stability analysis, reliability analysis, contingency analysis, LMP prices



Purpose and background



Power Flow Linearization

Yang Z, Zhong H, Xia Q, et al. A novel network model for optimal power flow with reactive power and network losses[J]. *Electric Power Systems Research*, 2017, 144:63-71.

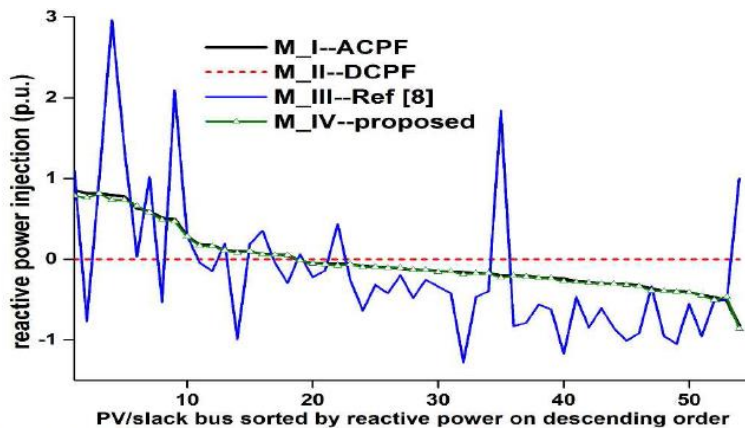


Fig. 3. PV Bus reactive power injections for the IEEE 118-bus system.

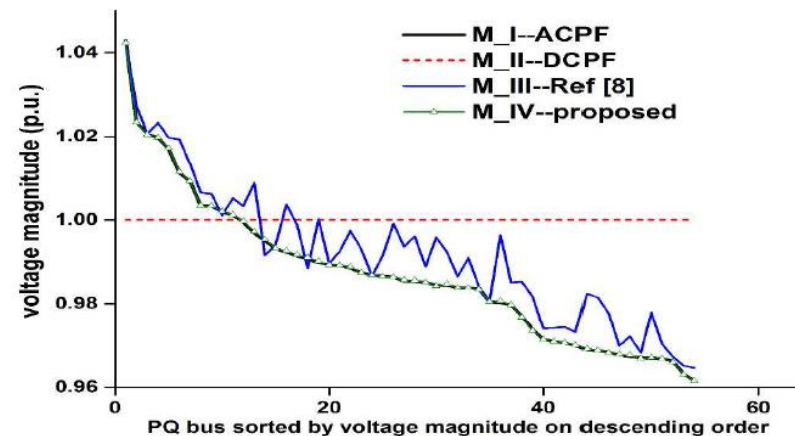
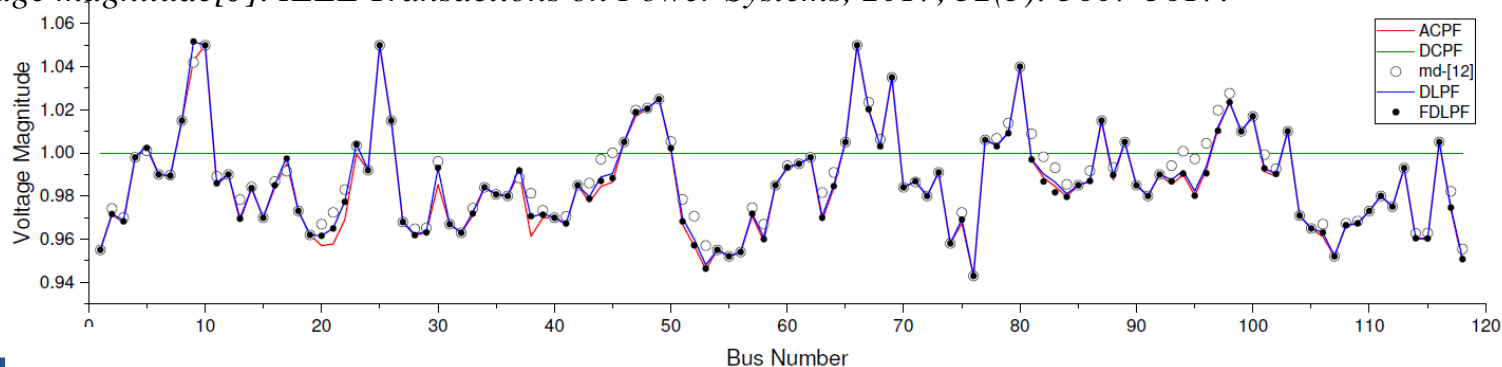


Fig. 2. PQ Bus voltage magnitudes for the IEEE 118-bus system.

Yang J, Zhang N, Kang C, et al. A state-independent linear power flow model with accurate estimation of voltage magnitude[J]. *IEEE Transactions on Power Systems*, 2017, 32(5): 3607-3617.



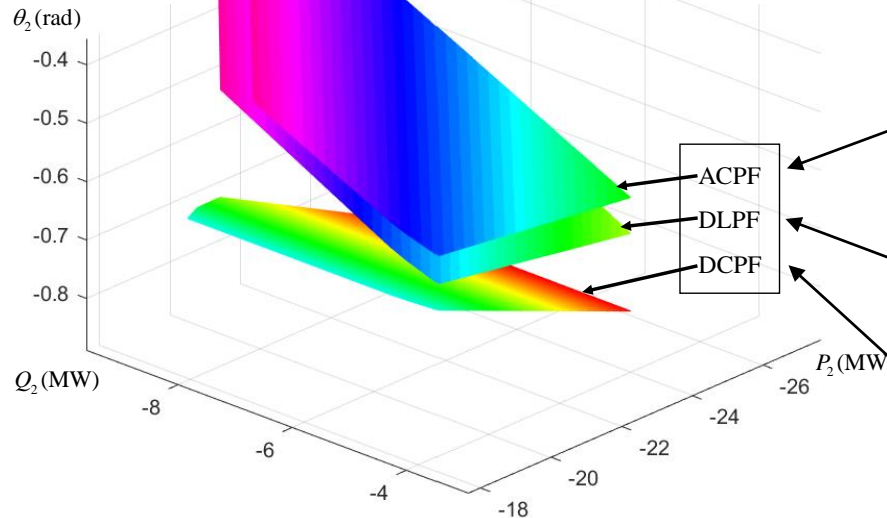
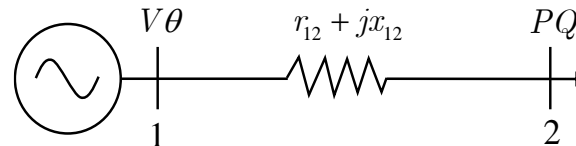


Purpose and background



PF Linearization

An illustrative two-bus system:



$$P_i = V_i \sum_{j \in i} V_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij})$$

$$Q_i = V_i \sum_{j \in i} V_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij})$$

$$\begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} = - \begin{bmatrix} \mathbf{B}' & -\mathbf{G} \\ \mathbf{G} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{V} \end{bmatrix}$$

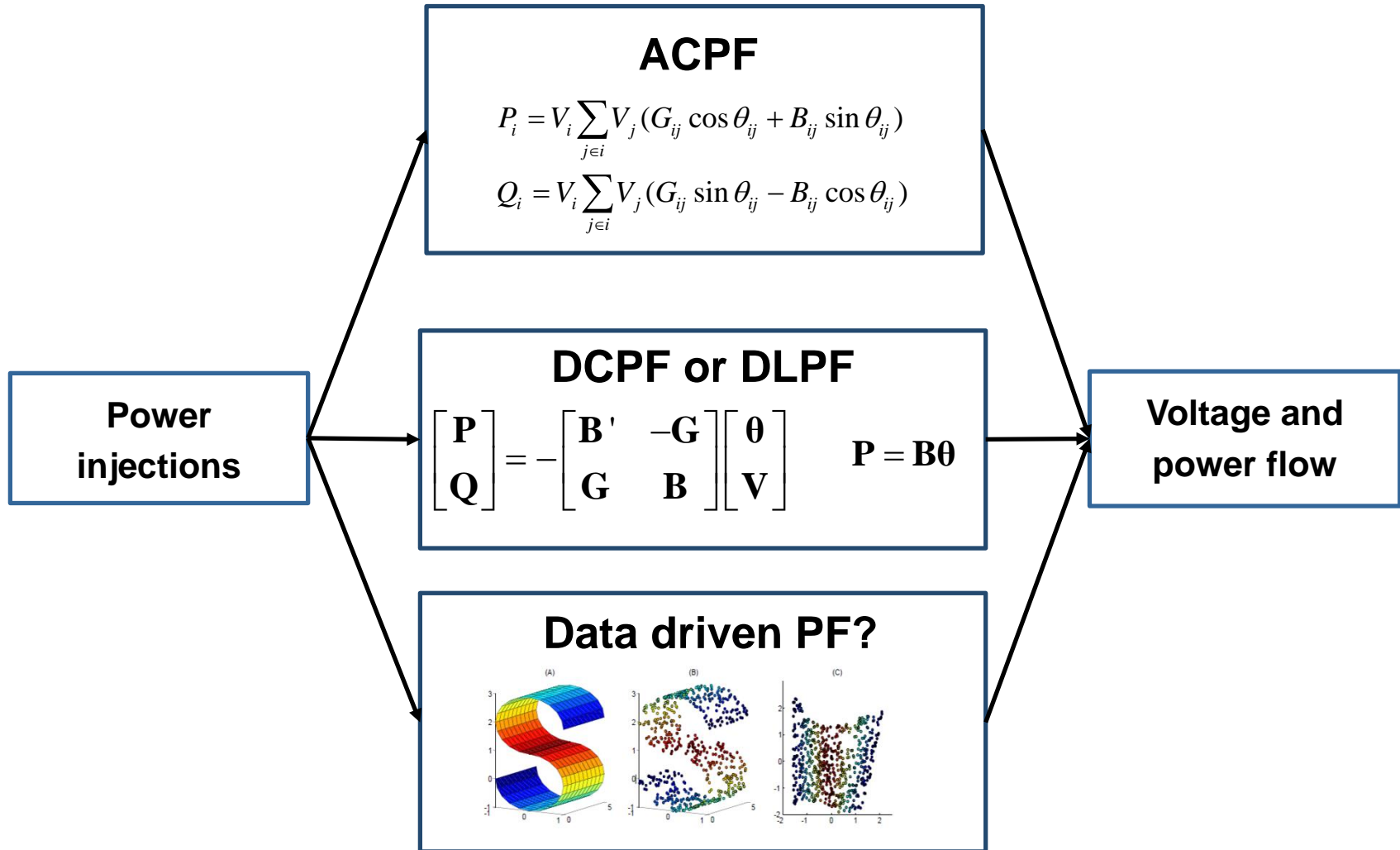
$$\mathbf{P} = \mathbf{B}\boldsymbol{\theta}$$

- 1) The non-linear ACPF has a high degree of linearity
- 2) The two model-based linear approximations (DCPF and DLPF) still result in clear errors

Yang J, Zhang N, Kang C, et al. A state-independent linear power flow model with accurate estimation of voltage magnitude[J]. IEEE Transactions on Power Systems, 2017, 32(5): 3607-3617.



Problem formulation

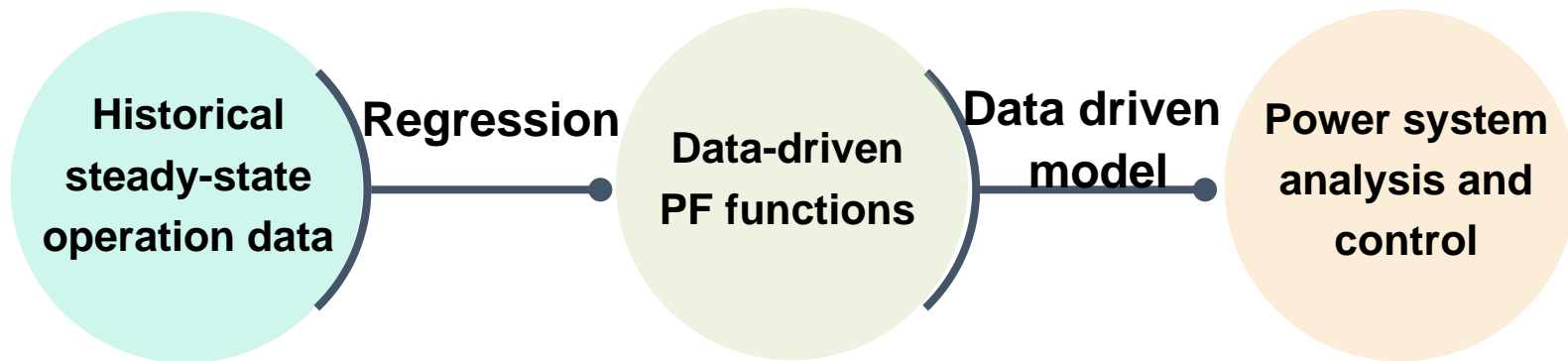




Purpose and background



Why data driven



- Do not require knowledge of the system topologies and parameters
- The exact system **topologies**, element **parameters**, and the **control logic** of active control devices are difficult to model accurately in some **distribution** network.
- Improve the linearization accuracy of PF calculations
- The **measurement data** reflects the operation status more efficiently than equivalent parameters. (e.g. parameters may change due to the **atmospheric condition and aging**)



Technology roadmap



Forward regression (P, Q) as a function of (V, θ)

- Formulation:
$$\begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} = \begin{bmatrix} \mathbf{H} & \mathbf{N} \\ \mathbf{M} & \mathbf{L} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{V} \end{bmatrix} + \begin{bmatrix} \mathbf{C}_P \\ \mathbf{C}_Q \end{bmatrix}$$
- Potential application: Data Synchronization
 - The “power flow analysis” is necessary to synchronize the data stream of different types of measurements from PMUs, PV inverters, and smart meters.
 - Forward regression can be used to recover the power injection from the voltage measurements.

J. Yu, Y. Weng and R. Rajagopal, "Mapping Rule Estimation for Power Flow Analysis in Distribution Grids," arXiv preprint arXiv:1702.07948, 2017.



Technology roadmap



Inverse regression:

- Improvement:
It can calculate PF when considering different bus types

- Formulation:

$$\begin{aligned}
 \mathbf{P} &= [\mathbf{P}_L^T \quad \mathbf{P}_S^T \quad \mathbf{P}_R^T]^T \\
 \mathbf{Q} &= [\mathbf{Q}_L^T \quad \mathbf{Q}_S^T \quad \mathbf{Q}_R^T]^T \\
 \mathbf{V} &= [\mathbf{V}_L^T \quad \mathbf{V}_S^T \quad \mathbf{V}_R^T]^T \\
 \boldsymbol{\theta} &= [\boldsymbol{\theta}_L^T \quad \boldsymbol{\theta}_S^T \quad \boldsymbol{\theta}_R^T]^T
 \end{aligned}
 \xrightarrow{\text{Unknown}}
 \begin{bmatrix} \boldsymbol{\theta}_L \\ \boldsymbol{\theta}_S \\ \mathbf{P}_R \\ \mathbf{V}_L \\ \mathbf{V}_S \\ \mathbf{V}_R \end{bmatrix}
 =
 \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} & \mathbf{A}_{14} & \mathbf{A}_{15} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} & \mathbf{A}_{24} & \mathbf{A}_{25} \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} & \mathbf{A}_{34} & \mathbf{A}_{35} \\ \mathbf{A}_{41} & \mathbf{A}_{42} & \mathbf{A}_{43} & \mathbf{A}_{44} & \mathbf{A}_{45} \\ \mathbf{A}_{51} & \mathbf{A}_{52} & \mathbf{A}_{53} & \mathbf{A}_{54} & \mathbf{A}_{55} \\ \mathbf{A}_{61} & \mathbf{A}_{62} & \mathbf{A}_{63} & \mathbf{A}_{64} & \mathbf{A}_{65} \end{bmatrix}
 \begin{bmatrix} \mathbf{P}_L \\ \mathbf{P}_S \\ \mathbf{Q}_L \\ \mathbf{Q}_S \\ \mathbf{Q}_R \end{bmatrix}
 +
 \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \\ \mathbf{C}_4 \\ \mathbf{C}_5 \\ \mathbf{C}_6 \end{bmatrix}$$

Unknown

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}
 =
 \begin{bmatrix} \tilde{\mathbf{A}}_{11} & \tilde{\mathbf{A}}_{12} \\ \tilde{\mathbf{A}}_{21} & \tilde{\mathbf{A}}_{22} \end{bmatrix}
 \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}
 +
 \begin{bmatrix} \tilde{\mathbf{C}}_1 \\ \tilde{\mathbf{C}}_2 \end{bmatrix}
 \xrightarrow{\text{Unknown}}
 \begin{aligned}
 \mathbf{x}_2 &= \tilde{\mathbf{A}}_{22}^{-1} (\mathbf{y}_2 - \tilde{\mathbf{A}}_{21} \mathbf{x}_1 - \tilde{\mathbf{C}}_2) \\
 \mathbf{y}_1 &= \tilde{\mathbf{A}}_{11} \mathbf{x}_1 + \tilde{\mathbf{A}}_{12} \mathbf{x}_2 + \tilde{\mathbf{C}}_1
 \end{aligned}$$

- Potential applications:
Power flow calculations, probabilistic power flow, optimal power flow, voltage control



Technology roadmap



Challenges of regression:

- To address the collinearity of data:
 - Collinearity among the voltage angle and magnitude data is inevitable because of the **similar rise and fall patterns** among the different buses
 - Result in ill-conditioned regression and **larger errors** of PF calculation
- To avoid overfitting:
 - The number of variables in the regression parameter matrices **for large power systems** may be far greater than the amount of historical operation data that represents the current system situation
- A PLS-based regression and a BLR-based regression is proposed



Technology roadmap



Relationship with Physical Parameter Matrices:

Power system matrices

ACPF function

Jacobian matrix

$$\begin{bmatrix} B \cos \theta & -G \cos \theta \\ G \cos \theta & B \cos \theta \end{bmatrix} - \begin{bmatrix} G \sin \theta & B \sin \theta \\ -B \sin \theta & G \sin \theta \end{bmatrix} - \begin{bmatrix} -Q & P \\ P & Q \end{bmatrix}$$

Constant Jacobian matrix

$$\begin{bmatrix} B & -G \\ G & B \end{bmatrix}$$

DCPF function

X matrix

$$B^{-1}$$

Regression matrices

Regression matrix of forward regression

$$\begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} H & N \\ M & L \end{bmatrix} \begin{bmatrix} \theta \\ v \end{bmatrix} + \begin{bmatrix} C_P \\ C_Q \end{bmatrix}$$

Regression matrix of inverse regression

$$\begin{bmatrix} \theta_L \\ \theta_S \\ P_R \\ V_L \\ V_S \\ V_R \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} \\ A_{21} & A_{22} & A_{23} & A_{24} & A_{25} \\ A_{31} & A_{32} & A_{33} & A_{34} & A_{35} \\ A_{41} & A_{42} & A_{43} & A_{44} & A_{45} \\ A_{51} & A_{52} & A_{53} & A_{54} & A_{55} \\ A_{61} & A_{62} & A_{63} & A_{64} & A_{65} \end{bmatrix} \begin{bmatrix} P_L \\ P_S \\ Q_L \\ Q_S \\ Q_R \end{bmatrix} + \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{bmatrix}$$

These relationships can serve as an indicator of overfitting



Numerical results



Data generation

- Monte Carlo simulation:
 - Meshed transmission grids: IEEE 5, 30, 57, and 118-bus systems
 - Radial distribution grids: IEEE 33-bus system, the modified 123-bus system

- Public testing data:
 - The NREL-118 test system (data collinearity)

I. Pena, C. Brancucci and B. M. Hodge, "An Extended IEEE 118-bus Test System with High Renewable Penetration," IEEE Trans. Power Syst., vol. PP, p. 1-1, 2017



Numerical results



Basic results

TABLE I. ERRORS OF FORWARD, INVERSE AND BRANCH CALCULATION ON DIFFERENT CASES

Cases	Size of training data	Size of testing data	Forward calculation					Inverse calculation				Branch PF calculation				
			Errors	DCPF	DLPF	PLS	BLR	Errors	DLPF	PLS	BLR	Errors	DCPF	DLPF	PLS	BLR
IEEE 5	100	300	P	24.11	1.117	0.412	0.615	θ	0.020	8.2e-4	6.8e-3	PF	8.120	8.120	0.126	0.609
			Q	---	66.21	0.940	1.065	V	7.8e-4	2.0e-5	4.1e-3	QF	---	---	8.934	256.1
IEEE 30	100	300	P	12.49	0.578	0.034	0.238	θ	0.154	1.9e-3	0.071	PF	7.734	7.562	0.104	0.825
			Q	---	12.66	0.404	0.471	V	9.9e-4	1.0e-5	1.4e-3	QF	---	---	1.340	226.9
IEEE 33	100	300	P	67.05	1.114	0.012	0.012	θ	0.028	4.3e-4	0.011	PF	1.142	1.142	5.0e-3	8.8e-3
			Q	---	0.759	0.044	0.027	V	2.0e-3	7.3e-6	6.5e-4	QF	---	---	0.013	0.497
IEEE 57	300	300	P	98.11	7.343	0.262	2.132	θ	0.215	0.036	0.218	PF	19.16	13.22	0.395	0.965
			Q	---	26.83	0.300	2.990	V	7.1e-3	2.1e-4	1.1e-3	QF	---	---	5.227	24.71
IEEE 118	300	300	P	16.89	4.546	0.061	1.385	θ	2.593	0.074	0.296	PF	86.96	86.04	2.263	7.078
			Q	---	77.85	1.096	31.73	V	1.9e-3	1.2e-4	8.1e-4	QF	---	---	5.570	68.27
NREL 118	300	300	P	85.90	9.486	0.161	1.207	θ	3.003	0.622	0.271	PF	33.08	29.37	10.59	4.326
			Q	---	107.4	0.486	3.982	V	2.3e-3	6.3e-4	7.6e-4	QF	---	---	28.07	36.53
Modified 123	300	300	P	12.49	0.512	0.007	0.452	θ	0.091	3.2e-4	2.6e-3	PF	0.319	0.319	5.1e-4	6.6e-3
			Q	---	2.071	0.003	0.003	V	2.3e-3	3.2e-6	3.5e-6	QF	---	---	3.6e-3	7.4e-3

✧ The errors of P, Q, PF, and QF are in mean absolute percentage error with the unit of 100%, whereas the errors of θ and V are in mean absolute error.

✧ The errors of Q correspond to DCPF are not shown because DCPF is not able to calculate reactive power. The errors of QF that correspond to DCPF and DLPF are not shown because DCPF and DLPF cannot calculate the reactive power flow.

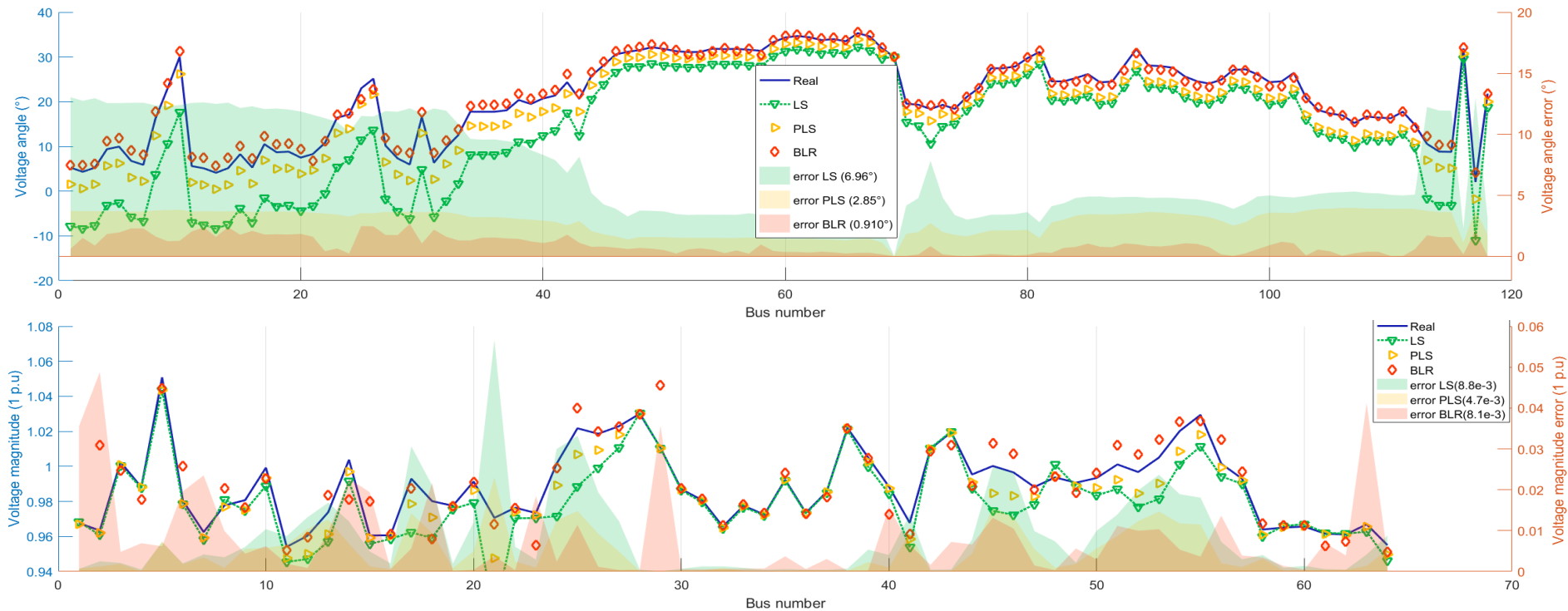


Numerical results



Calculation results under data collinearity

➤ the NREL-118 test system



To show the robustness of the algorithm, the error in the figure is the largest among all groups in the NREL-118 test system.



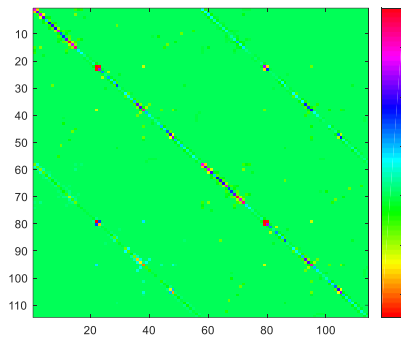
Numerical results



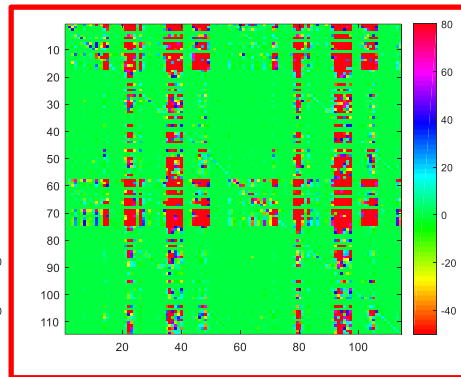
Regression Parameters

➤ IEEE 57-bus systems

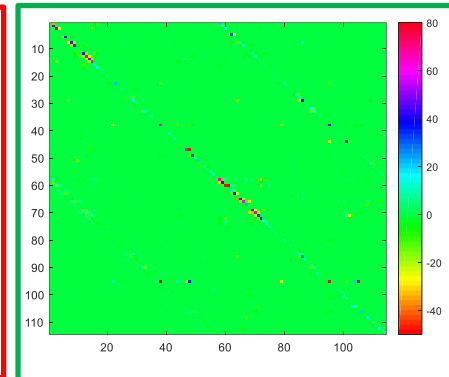
Overfitting



Constant Jacobian matrix

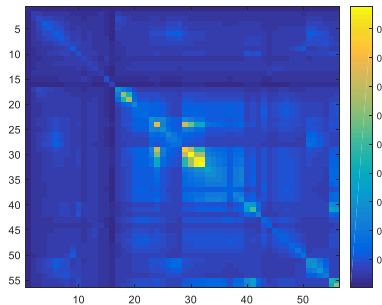


Forward regression parameter matrix under the
PLS-based algorithm

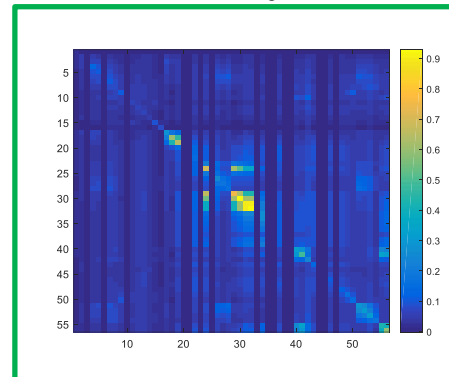


Forward regression parameter matrix under the
BLR-based algorithm

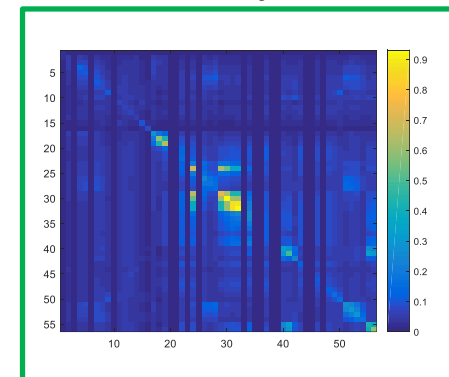
Well-fitted



The inverse matrix of \mathbf{B} in DCPF



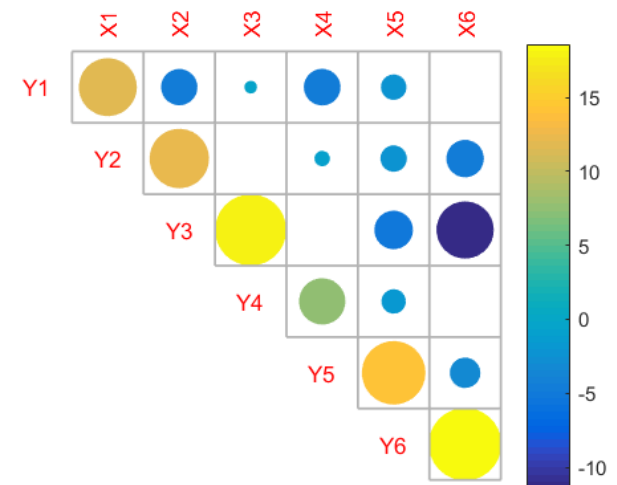
Inverse regression parameter matrix under the
PLS-based algorithm



Inverse regression parameter matrix under the
BLR-based algorithm



TOPOLOGY IDENTIFICATION AND LINE PARAMETER ESTIMATION FOR **NON-PMU** DISTRIBUTION NETWORK



Jiawei Zhang, Yi Wang, Yang Weng, and Ning Zhang, "Topology Identification and Line Parameter Estimation for non-PMU Distribution Network: A Numerical Method." IEEE Transactions on Smart Grid, accepted, in press, doi: 10.1109/TSG.2020.2979368



Background

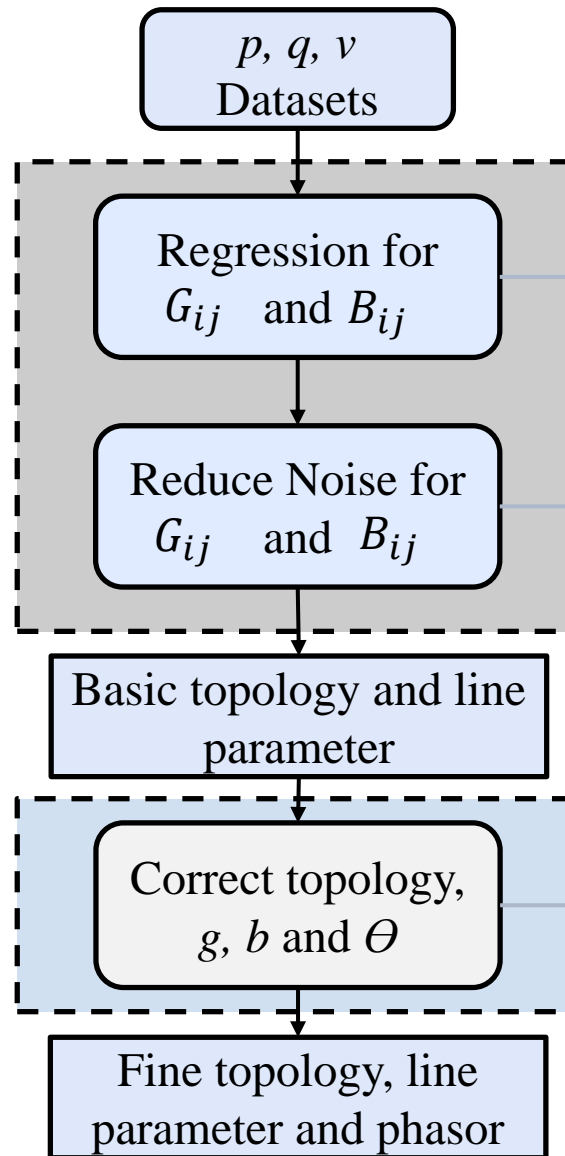
- **Background:** The operation and analysis of distribution network requires the information of system topology and line parameter.
- However, the real-time information of topology and line parameter, especially branch's conductance and susceptance may not be available in distribution networks, since there is fewer monitoring devices for distribution network than those in transmission network.
- **Aim:** identify the topology, estimate line parameter and recover missing voltage angle at the same time **without the measurement of voltage angle.**



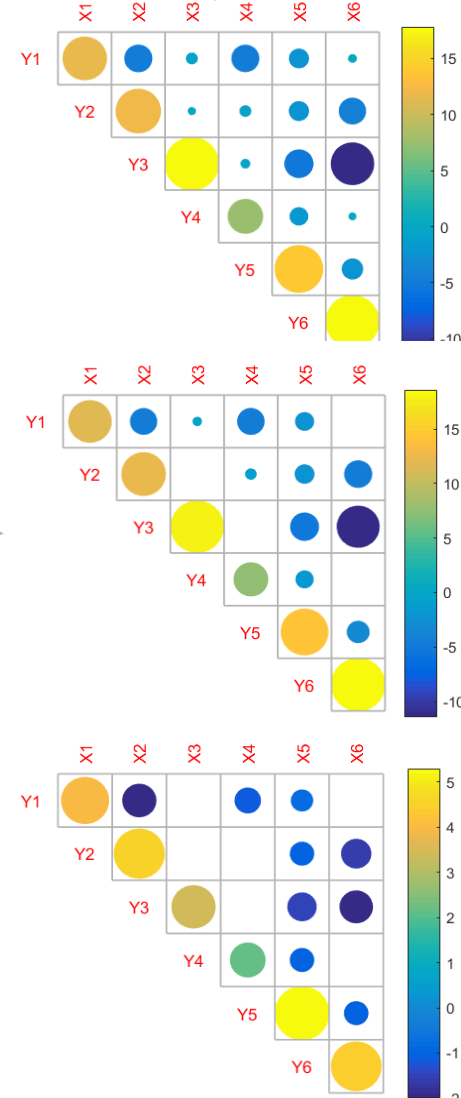
Framework

- Estimation of the distribution network's topology and line parameter is equivalent to estimate matrix G_{ij} , B_{ij} .

Step 1:
Basic
Identification



G_{ij} after different processes
(6-bus test case)



- A two-step model

Step 2:
Fine
Identification



Step 1: Basic Identification

- Basic identification aims to provide approximate topology and line parameter for fine identification.

Equation
$$p_i = \sum_{j=1}^n v_i v_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij})$$

$$q_i = \sum_{j=1}^n v_i v_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij})$$

Approximation
$$G_{ij}^{\#} = G_{ij} + \theta_{ij} B_{ij}$$

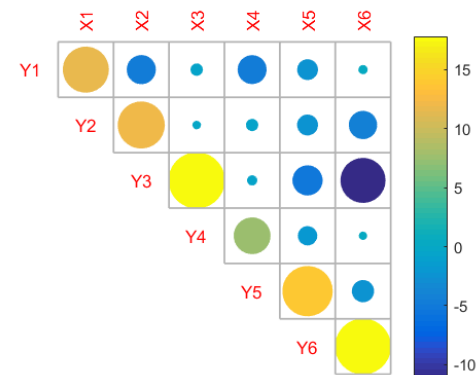
$$B_{ij}^{\#} = B_{ij} - \theta_{ij} G_{ij}$$

Regression
$$[P_i/V_i] = [G_{ij}^{\#}] \cdot [V_j]$$

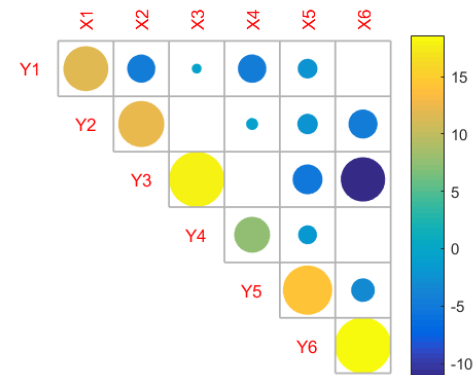
$$[Q_i/V_i] = -[B_{ij}^{\#}] \cdot [V_j]$$

Symmetrization
$$[G_{ij_S}^{\#}] = ([G_{ij}^{\#}] + [G_{ij}^{\#T}])/2$$

$$[B_{ij_S}^{\#}] = ([B_{ij}^{\#}] + [B_{ij}^{\#T}])/2$$



G_{ij} matrix after regression



G_{ij} matrix after reducing noise



Step 2: Fine Identification

- After the basic identification obtains the rough topology and parameters, the fine identification further calculates the more accurate topology and parameters.
- Using known p, q to solve g, b, θ , we can apply Newton-Raphson Method.

$$\begin{array}{l} \text{Newton-} \\ \text{Raphson} \\ \text{Method} \end{array} \quad \begin{bmatrix} \Delta p \\ \Delta q \end{bmatrix}_{[1 \times 2n]} = \begin{bmatrix} \frac{\partial p}{\partial g} & \frac{\partial p}{\partial b} & \frac{\partial p}{\partial \theta} \\ \frac{\partial q}{\partial g} & \frac{\partial q}{\partial b} & \frac{\partial q}{\partial \theta} \end{bmatrix} \cdot \begin{bmatrix} \Delta g \\ \Delta b \\ \Delta \theta \end{bmatrix}_{[1 \times (2m+n-1)]}$$

m: number of branches
n: number of buses

- The amount of variables $(2m+n-1)$ is larger than constraints $(2n)$. Multiple samples would reduce the error from wrong measurements. Also, the number of constraints will be larger than that of variables.

M-Sample Newton-Raphson Method

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix}_{[1 \times 2M \cdot n]} = \begin{bmatrix} \frac{\partial P}{\partial g} & \frac{\partial P}{\partial b} & \frac{\partial P}{\partial \theta} \\ \frac{\partial Q}{\partial g} & \frac{\partial Q}{\partial b} & \frac{\partial Q}{\partial \theta} \end{bmatrix} \cdot \begin{bmatrix} \Delta g \\ \Delta b \\ \Delta \theta \end{bmatrix}_{[1 \times (2m+M \cdot (n-1))]}$$

M: Number of samples



Step 2: Fine Identification

- Yet, Jacobian matrix built from M samples would usually not square. To solve that problem, we can get the unique optimal approximation solution using Penrose-Moore generalized inverse.

generalized inverse

$$\begin{bmatrix} \Delta \mathbf{g} \\ \Delta \mathbf{b} \\ \Delta \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{P}}{\partial \mathbf{g}} & \frac{\partial \mathbf{P}}{\partial \mathbf{b}} & \frac{\partial \mathbf{P}}{\partial \boldsymbol{\theta}} \\ \frac{\partial \mathbf{Q}}{\partial \mathbf{g}} & \frac{\partial \mathbf{Q}}{\partial \mathbf{b}} & \frac{\partial \mathbf{Q}}{\partial \boldsymbol{\theta}} \end{bmatrix}^{\dagger} \cdot \begin{bmatrix} \Delta \mathbf{P} \\ \Delta \mathbf{Q} \end{bmatrix}$$

- The line parameter and voltage angle are renewed in each iteration.

$$\begin{bmatrix} \mathbf{g} \\ \mathbf{b} \\ \boldsymbol{\theta} \end{bmatrix}^{(k+1)} = \begin{bmatrix} \mathbf{g} \\ \mathbf{b} \\ \boldsymbol{\theta} \end{bmatrix}^{(k)} + \begin{bmatrix} \Delta \mathbf{g} \\ \Delta \mathbf{b} \\ \Delta \boldsymbol{\theta} \end{bmatrix}$$

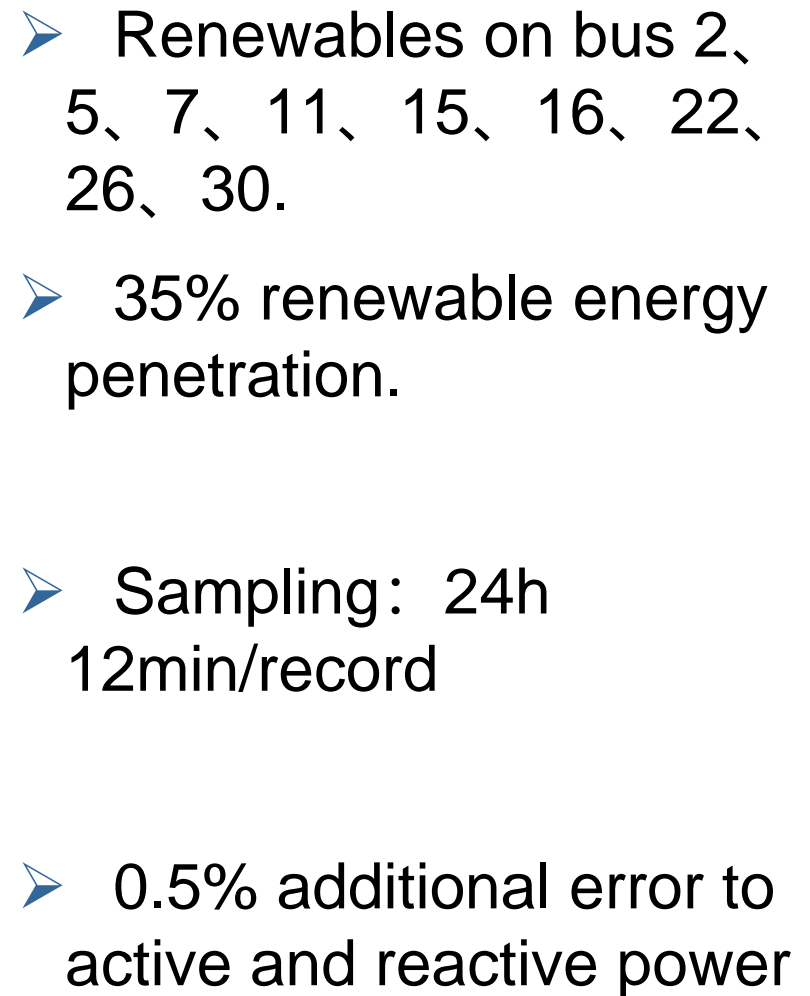
- Another problem: The selection of initial value for $\mathbf{g}, \mathbf{b}, \boldsymbol{\theta}$.



Test cases

- Organization of data
- IEEE 33-bus test case (High renewables penetration)
- IEEE 123-bus test case (Large distribution network)

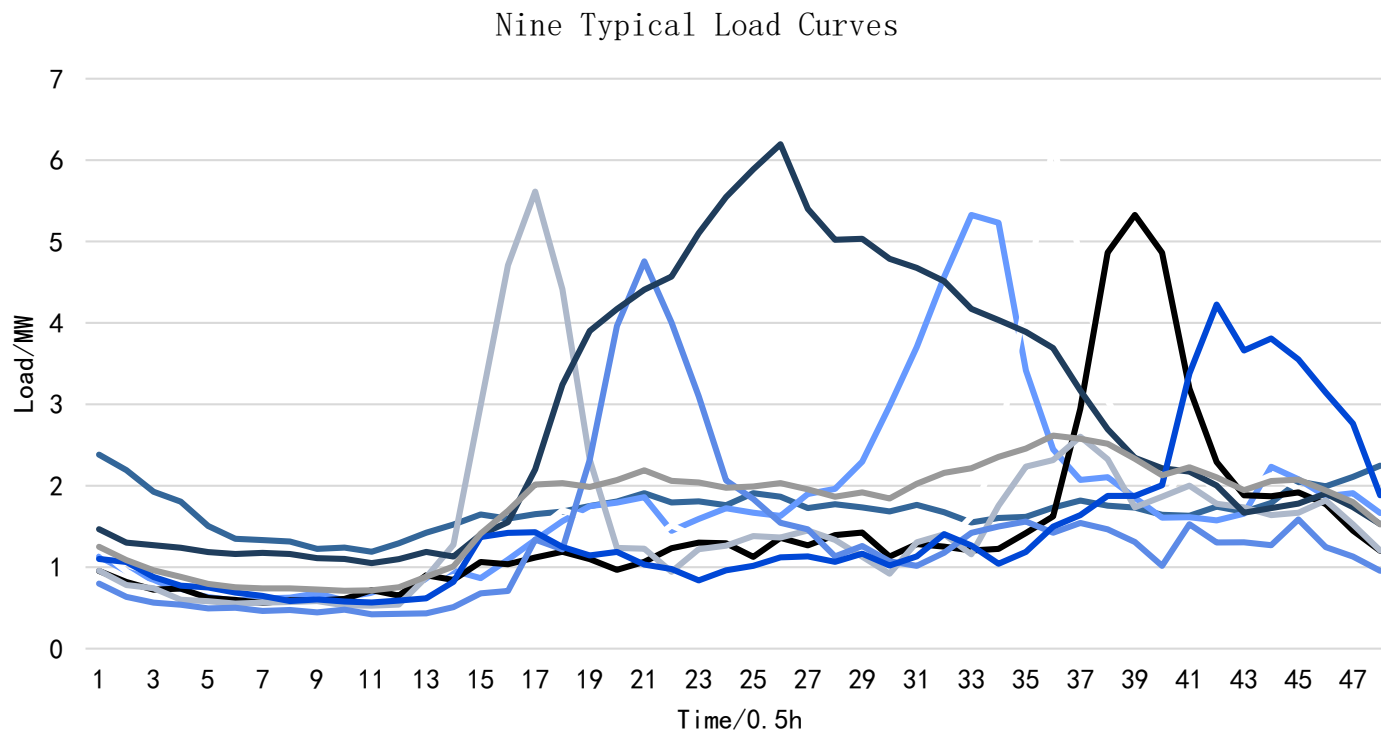
A. IEEE 33-feeder case





Organization of data: Data Source

- Load:
- Smart Meter Electricity Trial data from The Research Perspective Ltd. It includes power load curves from 1000 residents and small companies.



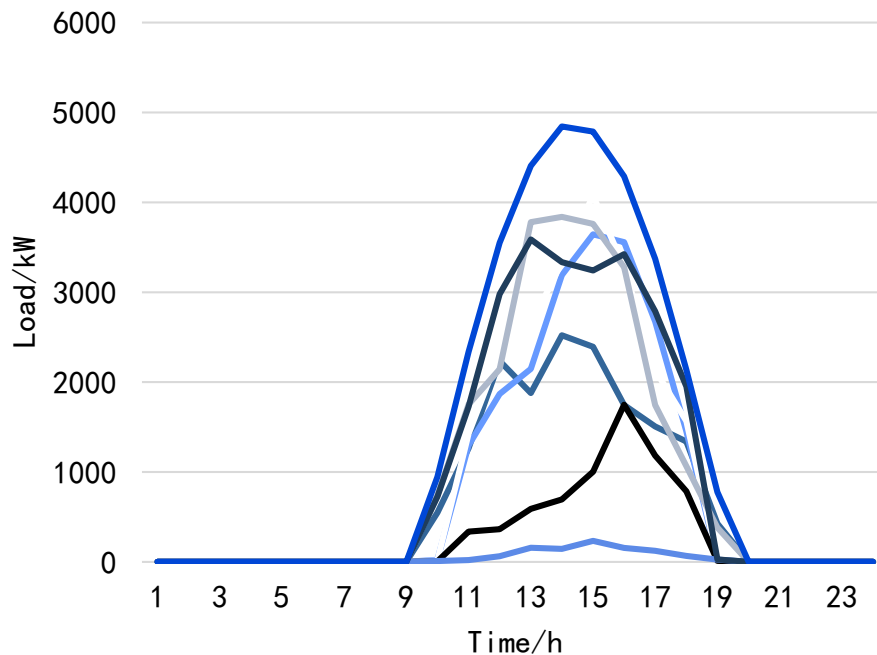


Organization of data: Data Source

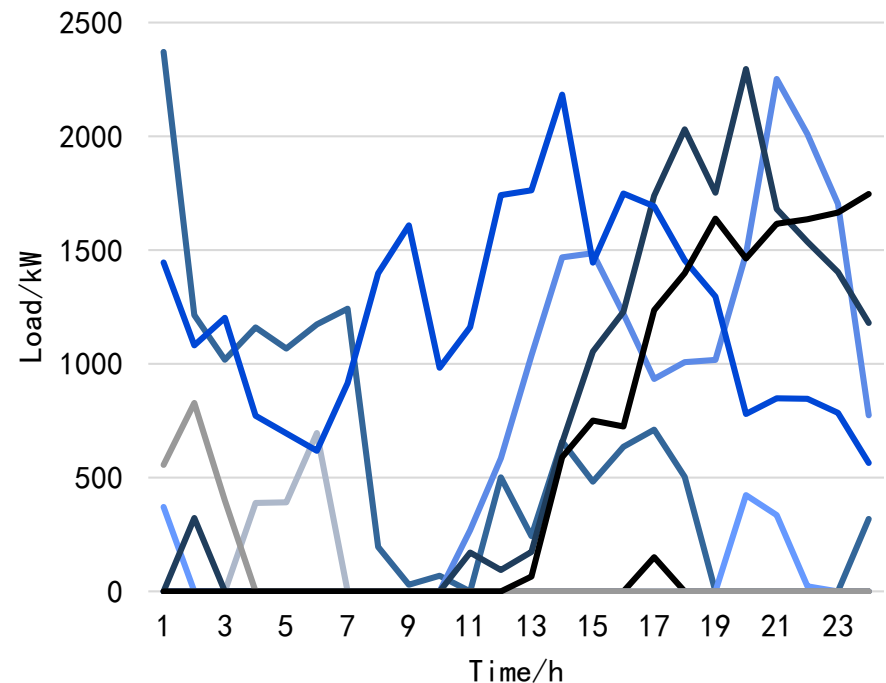
➤ Renewables

- 51 PV stations and 34 wind farms in China

Typical Solar Curves



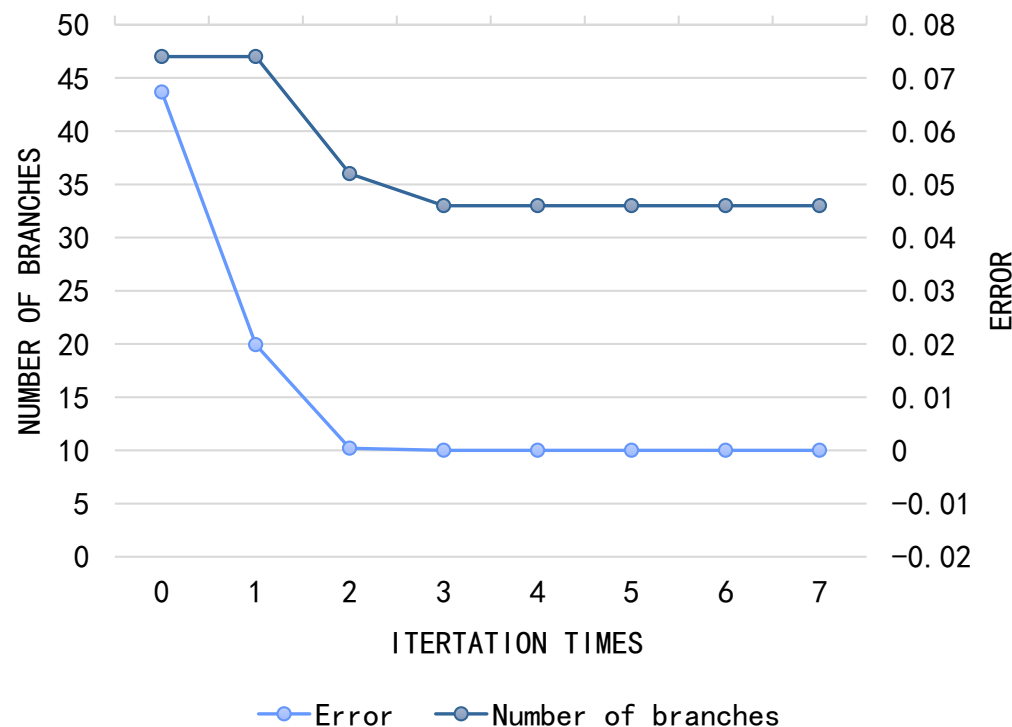
Typical Wind Curves





IEEE 33-bus test case (high renewables penetration)

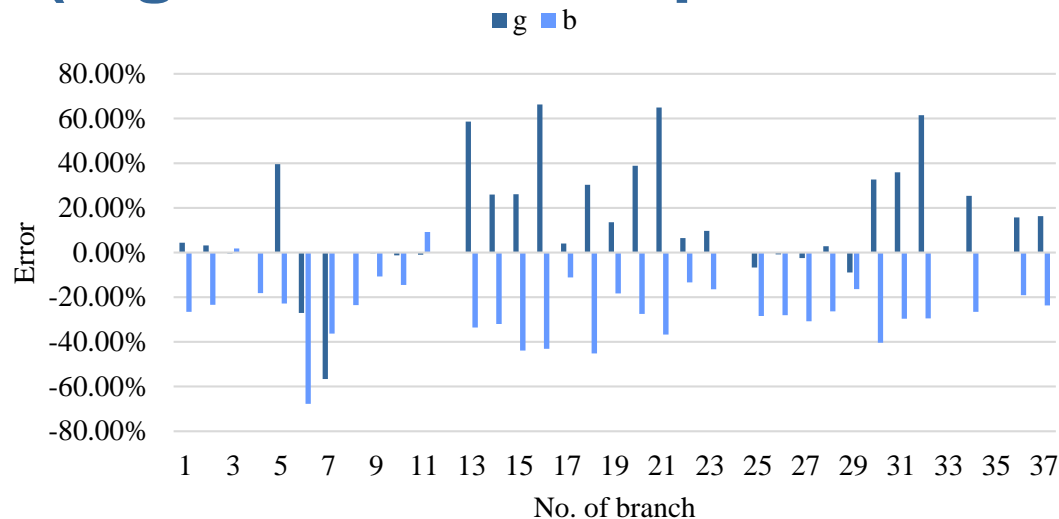
- There are 14 wrong branches after basic identification;
- Topology is corrected in 3 iterations;
- Convergence after 7 iterations.



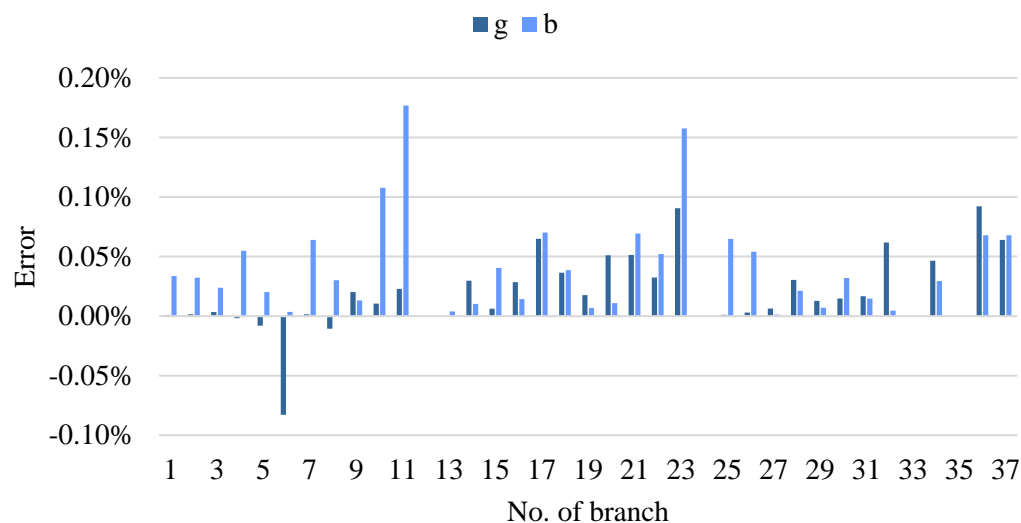


IEEE 33-bus test case (high renewables penetration)

➤ *g , b estimation errors are 20.9% and 26.5% after basic identification.*



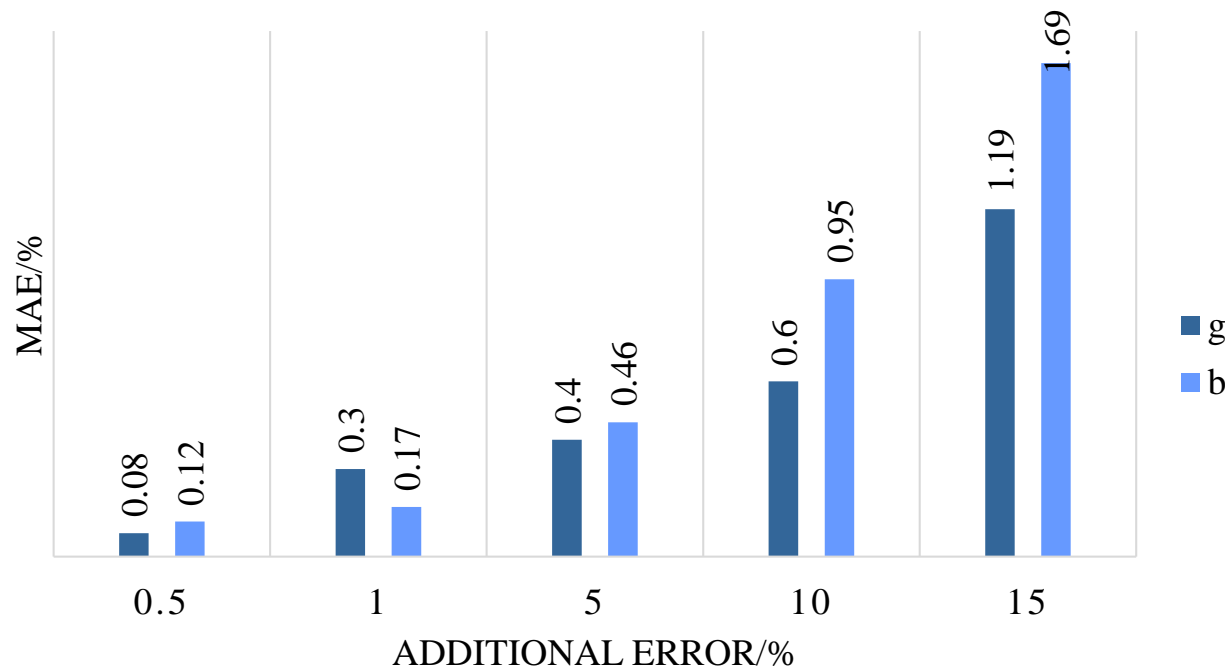
➤ While 0.03%,0.04% after fine identification.





IEEE 33-bus test case (high renewables penetration)

- With the increasing of additional error, the estimation error of g and b line parameters also increases. The method proposed in this paper has a high degree of redundancy when the measurement error is significant.





IEEE 123-bus Test Case (Large Distribution Network)

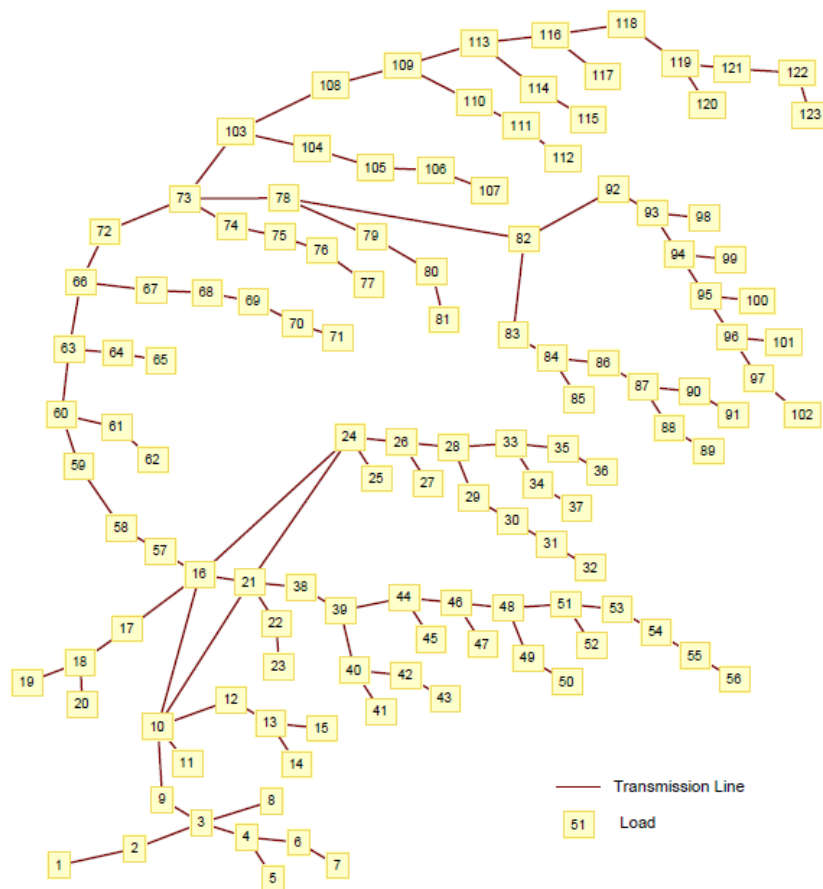


Fig. 7. IEEE 123 Bus Feeder

- In this test case, we investigate the impact of the size of measurement samples in basic identification on the whole algorithm.
- i.e. 24-hour dataset with 10, 15, 20 and 25 samples per hour
- In fine identification, we only select the last 20 samples in basic identification's dataset .



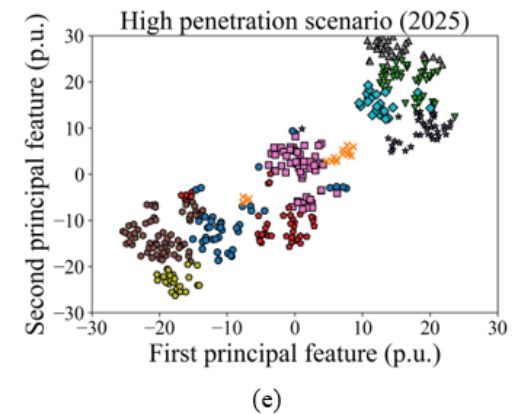
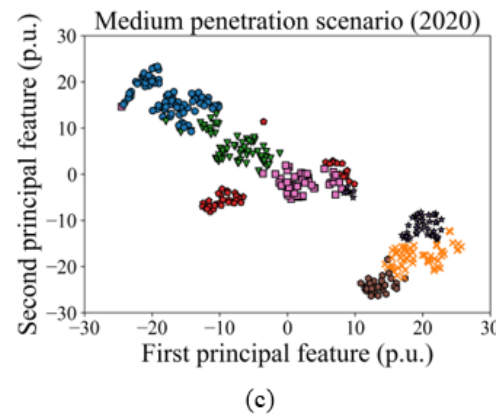
IEEE 123-bus Test Case (Large Distribution Network)

- Though larger data size would increase time consumed in basic identification, with larger data size, the wrong branches after basic identification reduce, and it may save time for the proposed method in fine identification.

Basic identification Data Size	MAPE of g	MAPE of b	Number of Iterations	Unidentified Branches after Basic Identification	Time Consumption/s
10×24	0.158%	0.057%	6	32	141.3
15×24	0.133%	0.055%	9	24	150.4
20×24	0.133%	0.051%	7	5	108.6
25×24	0.144%	0.047%	6	1	107.3



IMPACT OF HIGH RENEWABLE PENETRATION ON THE POWER SYSTEM OPERATION MODE: A DATA-DRIVEN APPROACH

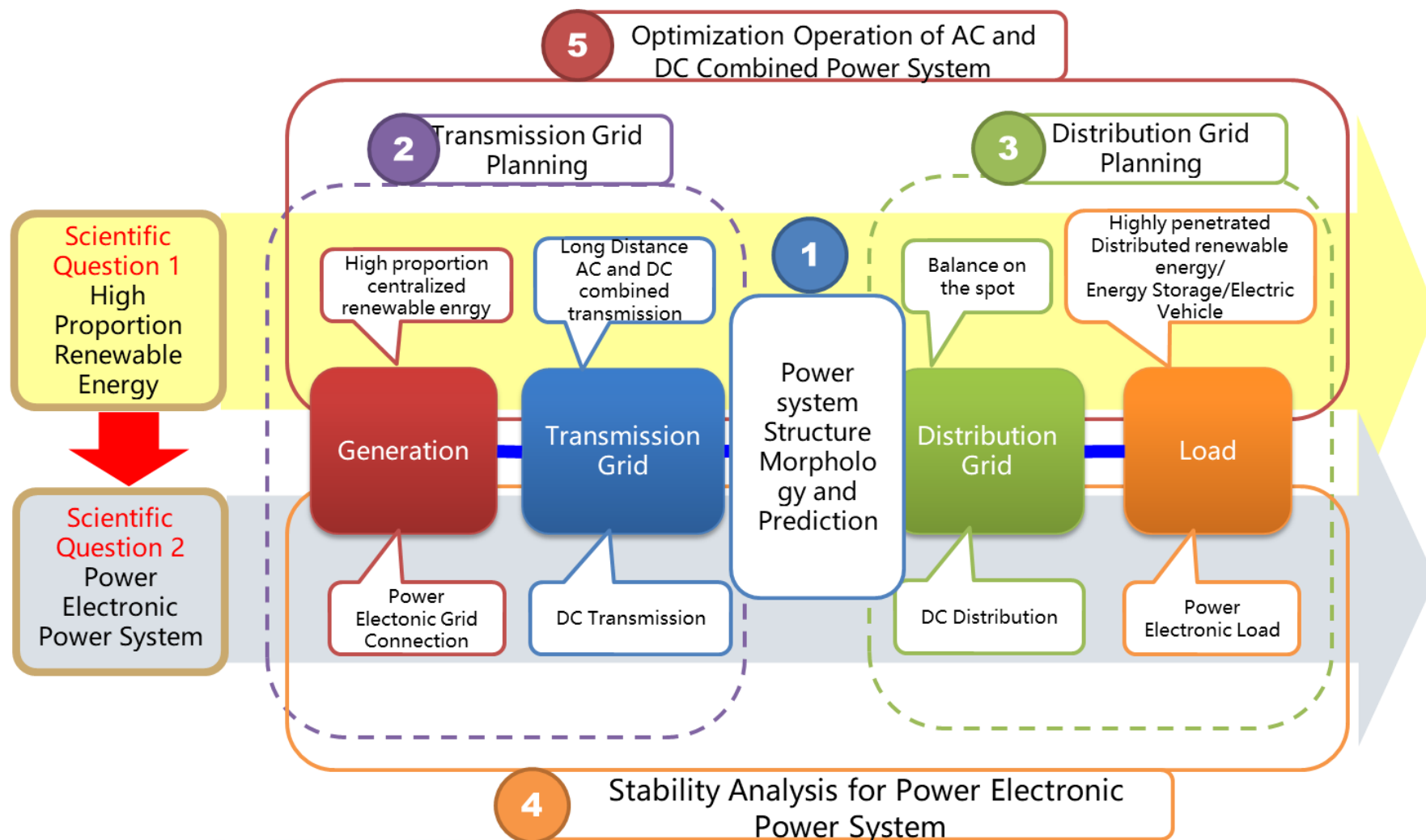


Qingchun Hou, Ershun Du, Ning Zhang, Chongqing Kang, Impact of High Renewable Penetration on the Power System Operation Mode: A Data-Driven Approach. IEEE Transactions on Power Systems, 2020, 35(1): 731-741.



National Key Research and Development Program of China

Fundamental Theory of Planning and Operation for Power Systems with High Share of Renewable Energy Generations

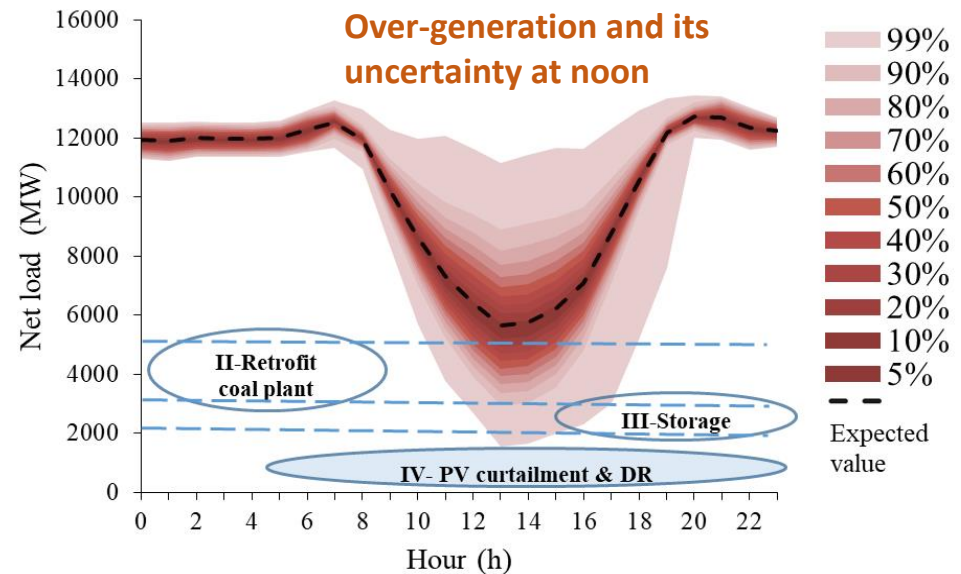




Motivations

- The introduction of highly penetrated renewable energy make the power system operation mode highly diversified and variable.

Probabilistic Duck Curve in Qinghai in 2020



Hou, Q., Zhang, N., Du, E., Miao, M., Peng, F., & Kang, C. (2019). Probabilistic duck curve in high PV penetration power system: Concept, modeling, and empirical analysis in China. *Applied Energy*, 242, 205-215.



Challenges

- How to pick typical/representative days in power system planning/operation analysis?
 - How many should be picked?
 - How to pick?



Power system operation mode from a data-driven prospective

- **The power system operation mode definition**
 - The status of power system operation, which is determined by the generator outputs, load demand, transmission topology, and accordant power flow in a certain period, such as a day, an hour or a snapshot.

- **Identifying the power system operation mode pattern is a typical big-data analytic problem.**
 - These data are inherently high-dimensional and complexly coupled to one another
 - Those operation data will have a significant variation with time, which makes it very hard to find the patterns in large amounts of data.



Challenge of analyzing the operation data

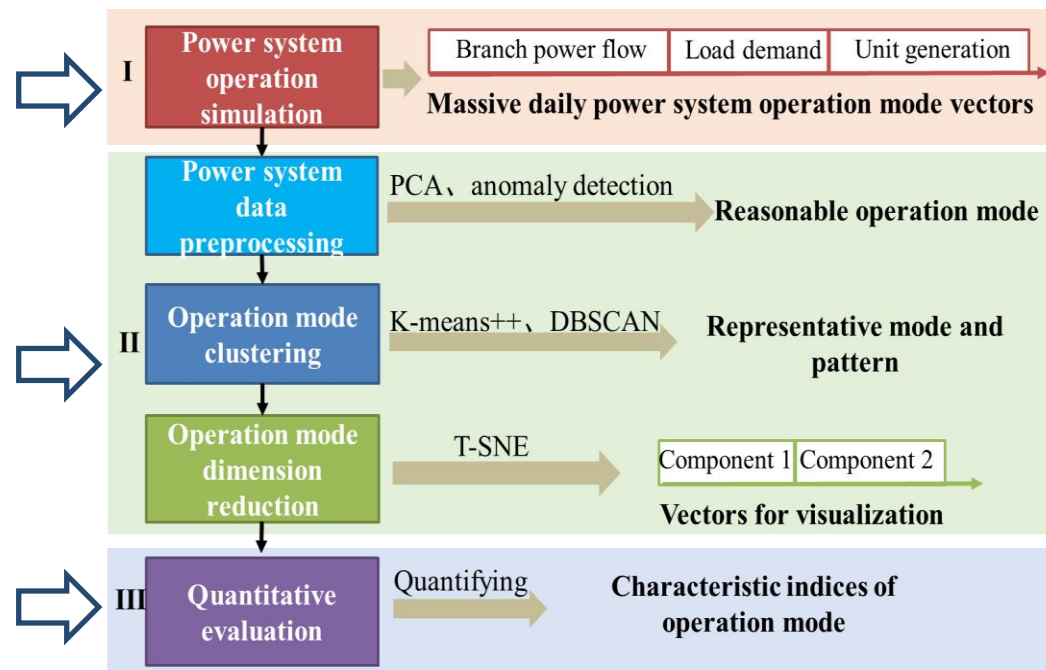
- The operation mode data for real power system is **high-dimensional, correlated, and hard to obtain intuitive understanding**.
 - The selected preprocessing algorithm must be **efficient with large amount of data**;
 - The clustering algorithm must be able to **capture complicate power system patterns under high renewable energy penetration**;
 - The dimension reduction and visualization algorithm should be able to **decouple the correlation among high-dimensional features and map them into 2D/3D space** for visualization and intuitive understanding.



Data-driven Framework

Three issues:

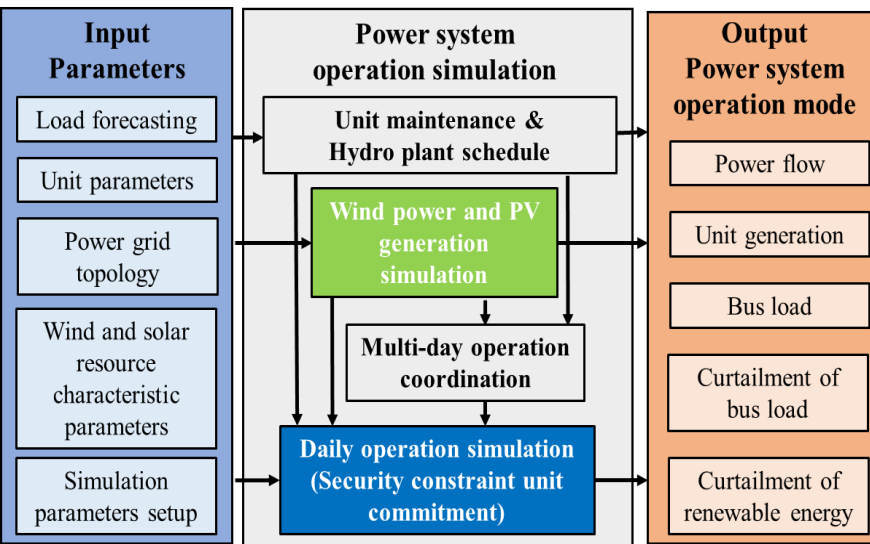
- **Power system operation data acquisition:** Sufficient operation data including power flow, load, and unit output are needed to form a complete year-round operation modes data set.
- **Pattern identification:** How to recognize the key characteristics of massive operation modes and identify the typical patterns and the number of patterns.
- **Visualization and evaluation:** How to visualize the high-dimensional operation mode data to provide intuitive understanding and quantify their characteristics.





Operation data acquisition

➤ Power system chronological operation simulation



An operation mode should include three aspects: energy source side, grid side, and load demand side.

$$\mathbf{p} = (\mathbf{g}_{1 \times (|g| \times T_s)}, \mathbf{r}_{1 \times (|r| \times T_s)}, \mathbf{f}_{1 \times (|f| \times T_s)}, \mathbf{d}_{1 \times (|d| \times T_s)}, \mathbf{d}^c_{1 \times (|d| \times T_s)})^T$$

generation schedule
renewable energy
power flow
load demand
load shedding

➤ Operation data can also be obtained through SCADA system

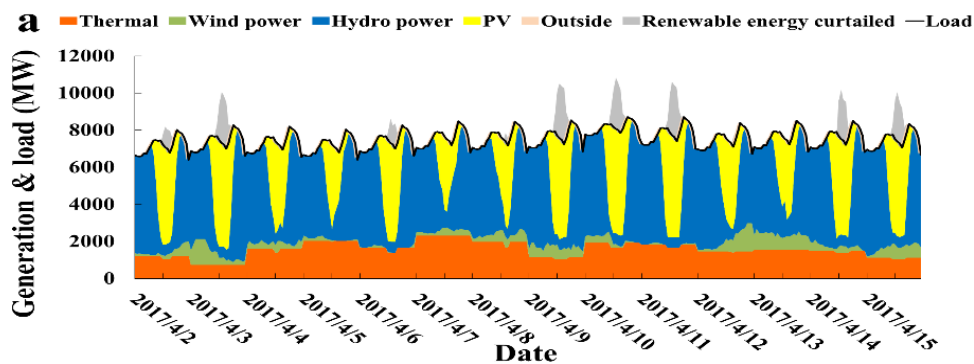




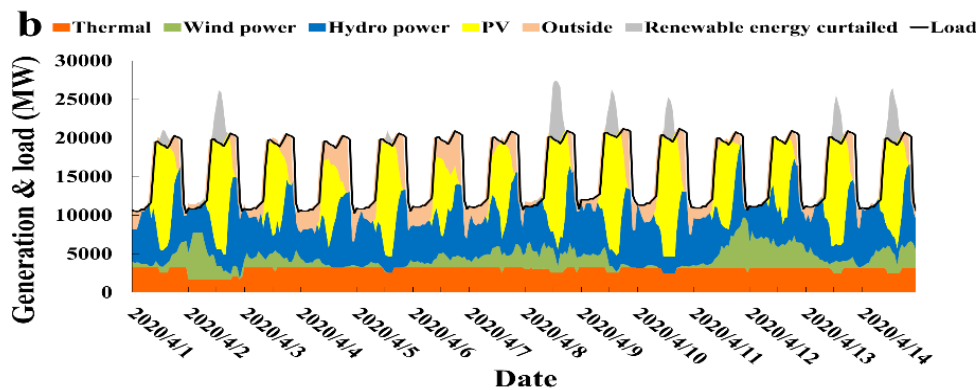
Case Study

- **Three renewable energy penetration scenarios are compared:**
- low penetration (20%) scenarios in 2017, medium penetration (33%) scenarios in 2020, high penetration (40%) scenarios in 2025.

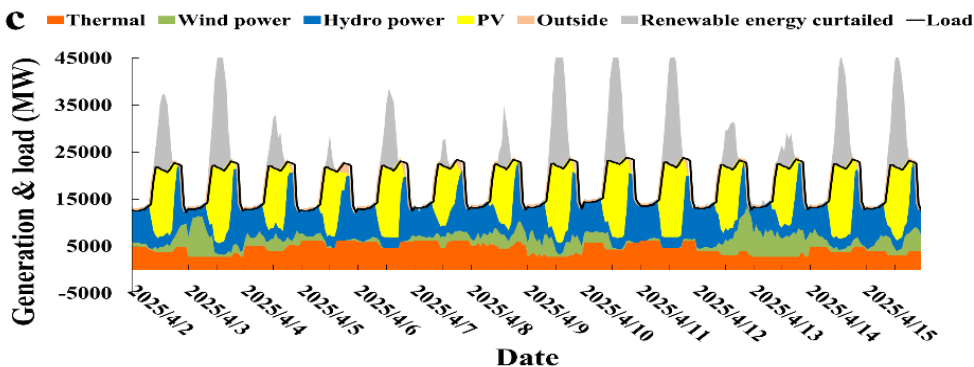
Year	2017	2020	2025
Hydro-power (MW)	1169	1637	1900
Thermal Power(MW)	360	510	850
Wind Power (MW)	162	700	1081
PV (MW)	790	2000	3800
PV and Wind Power Capacity/ Total Capacity (%)	38	56	64
Total Load (GWh)	88000	141300	161300
Maximal Load (MW)	10000	22000	25000
PV and Wind Power Generation/Total Load (%)	20	33	40



20%
Hydro-dominated



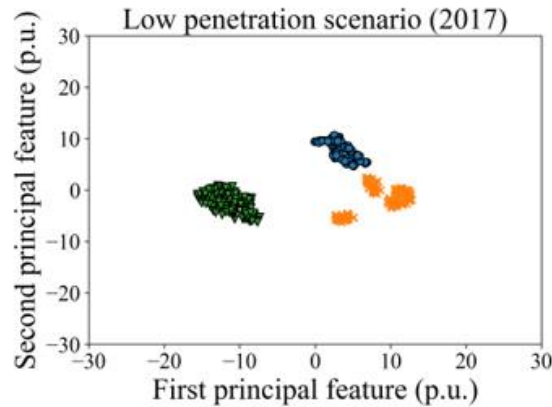
33%
PV and wind play
an important role



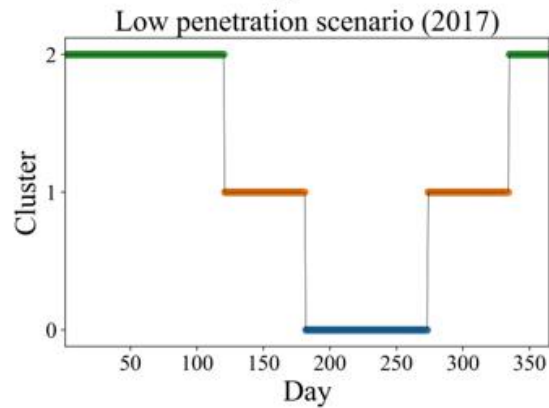
40%
Intermittent renewable
energy-dominated



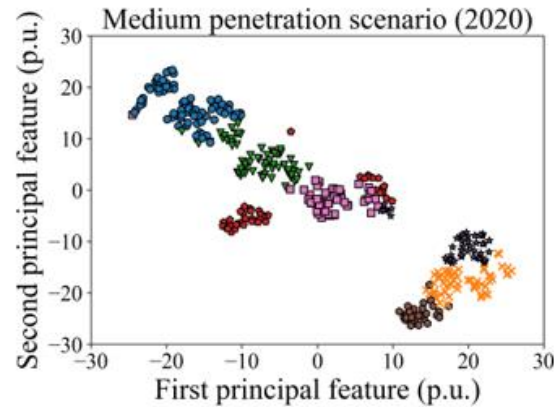
Visualization of data-driven results



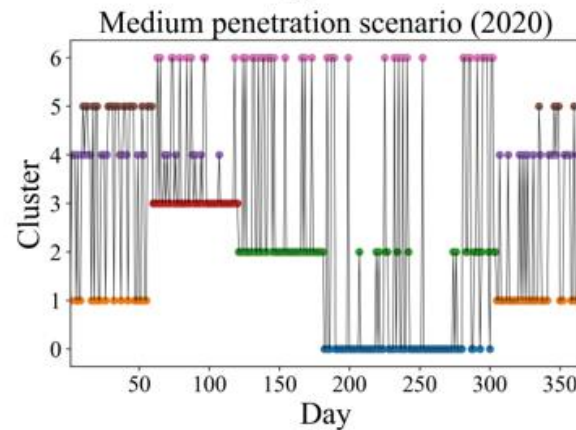
(a)



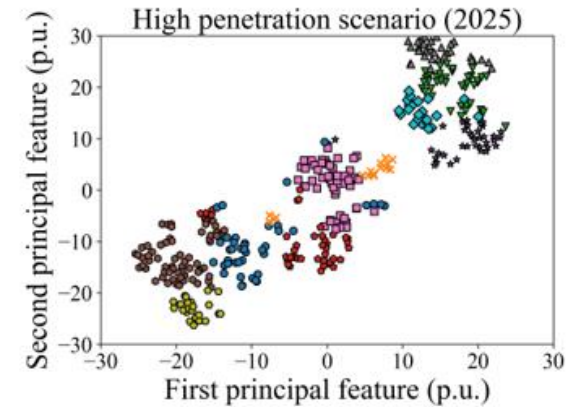
(b)



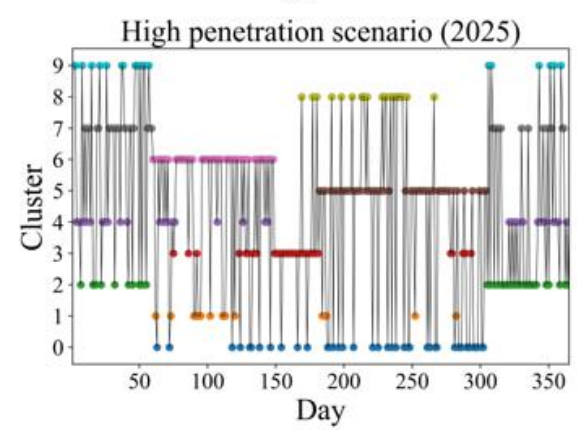
(c)



(d)



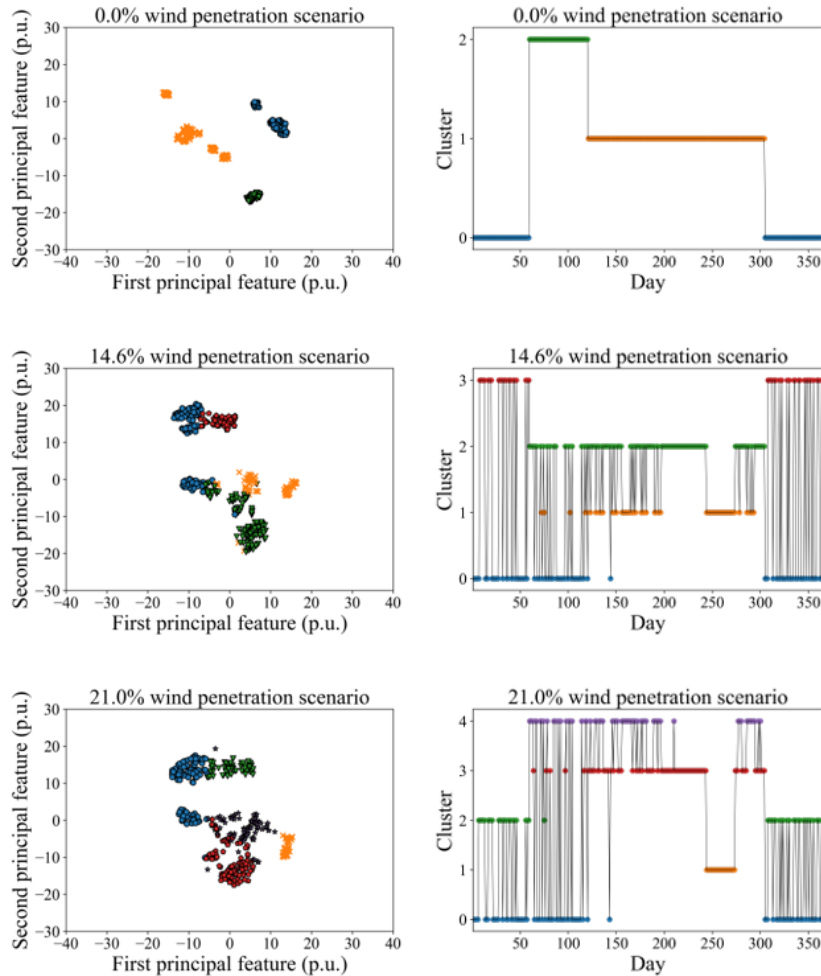
(e)



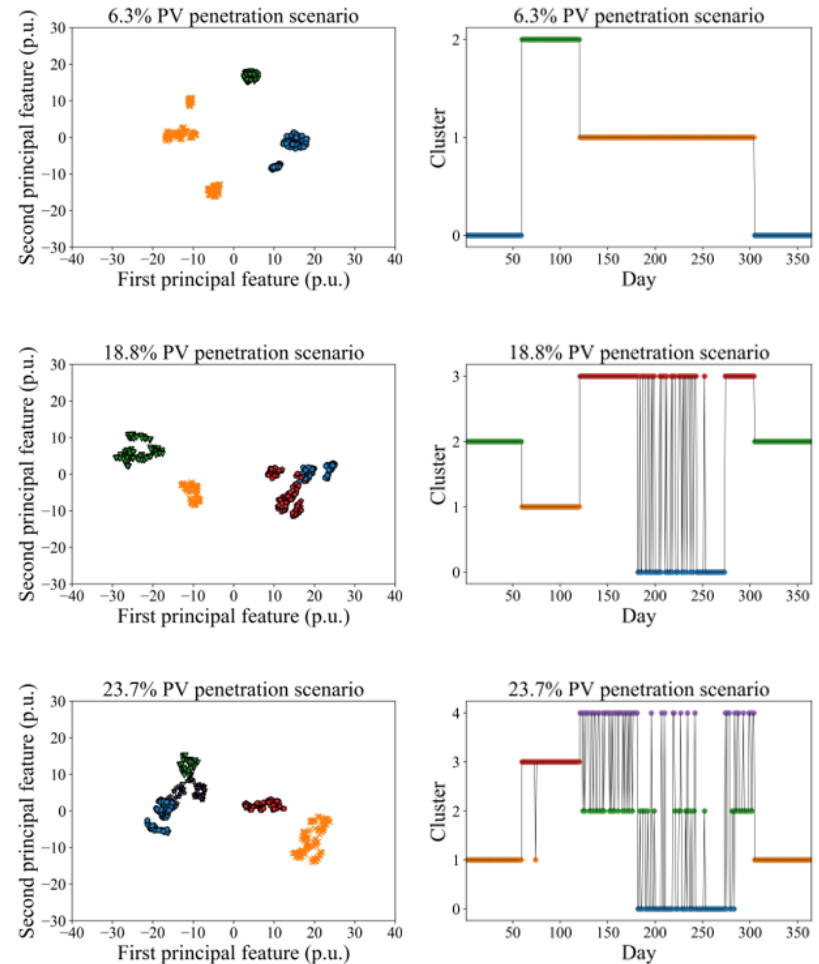
(f)



Adding more wind power

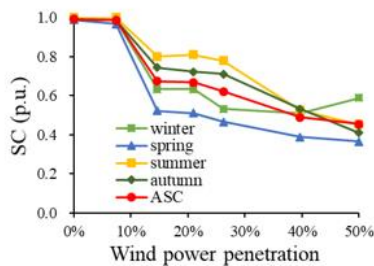
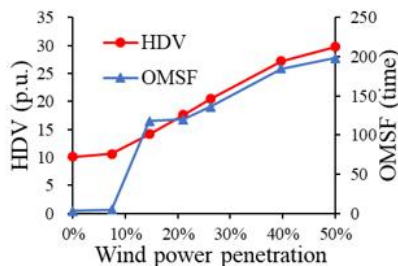
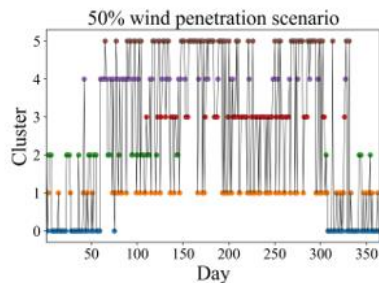
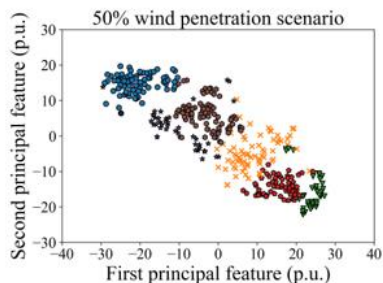
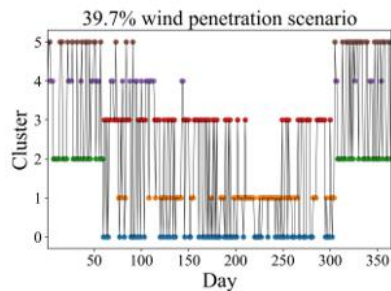
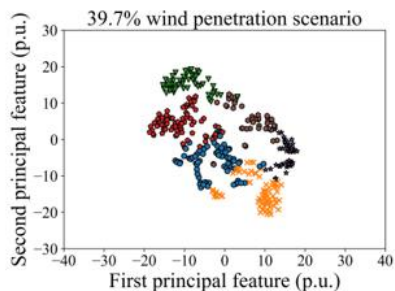


Adding more PV power

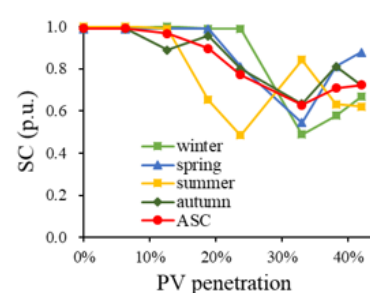
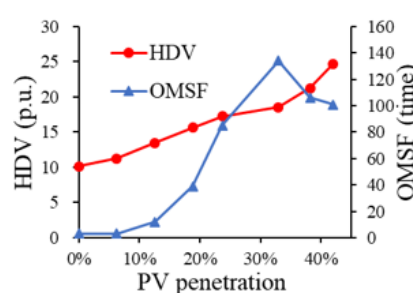
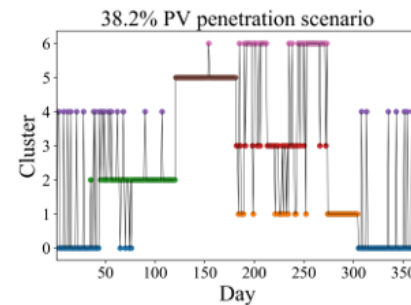
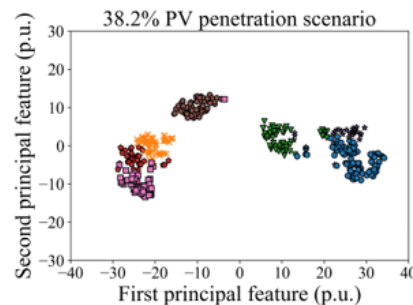
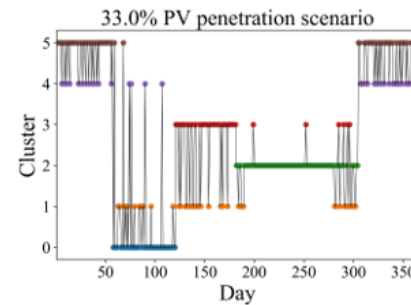
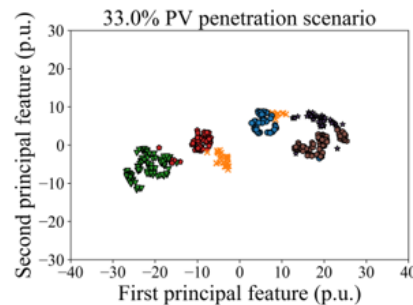




Adding more wind power



Adding more PV power



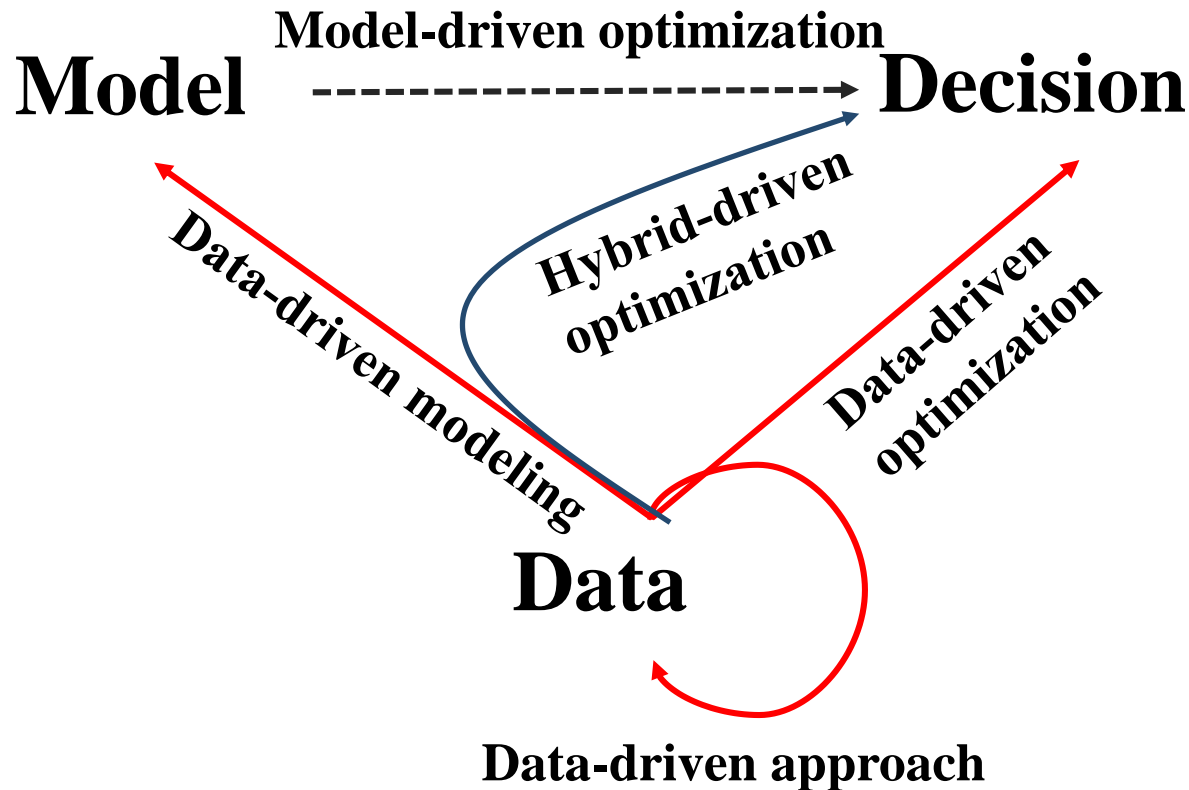


Finding

- Data-driven approach provides more intuitive insight on the diversity of power system operation mode.
- Under low renewable energy penetration, the power system operation mode is dominated by load/hydropower and basically consistent with the season.
- With the growing renewable energy penetration, the power system mode is gradually dominated by intermittent PV and wind power, indicating more representative modes are necessary for power system planning.
- The break point is system-dependent, normally when the VRE penetration is higher than 20%~30%.
- The impact of wind power and PV is distinct. Less daily difference are observed when PV penetration is higher than 30%.

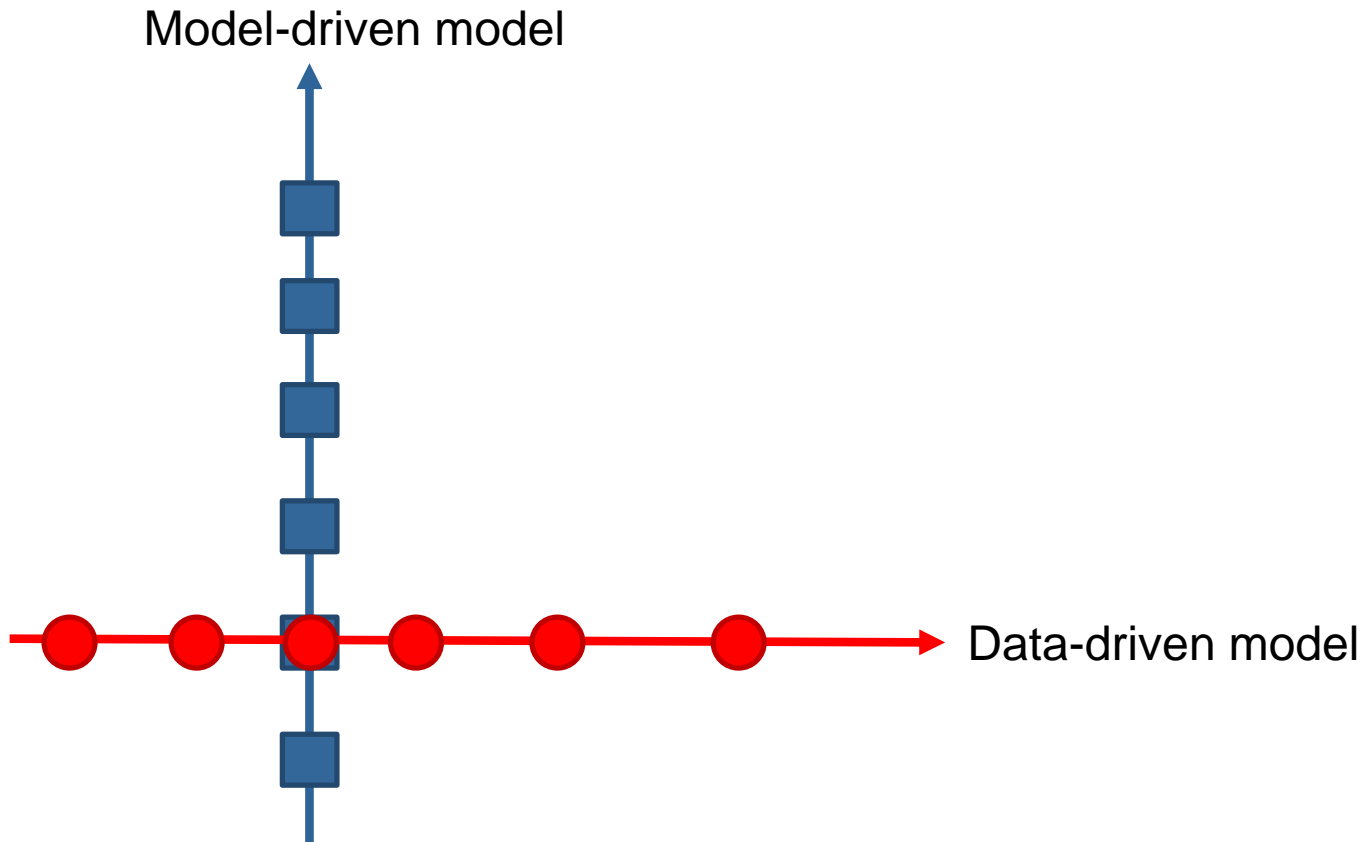


Takeaways: Data-driven or model-driven?





Takeaways: Data-driven or model-driven?





Thank you!

Q&A

Ning Zhang

Department of Electrical Engineering, Tsinghua University,
Beijing, China

ningzhang@tsinghua.edu.cn