Data generation method for power system operation considering geographical correlations and actual operation characteristics

Yingming Mao
Xi’an Jiaotong University
Systems Engineering Institute, MOEKLINNS
16-18 December, 2022

Content

1. Background

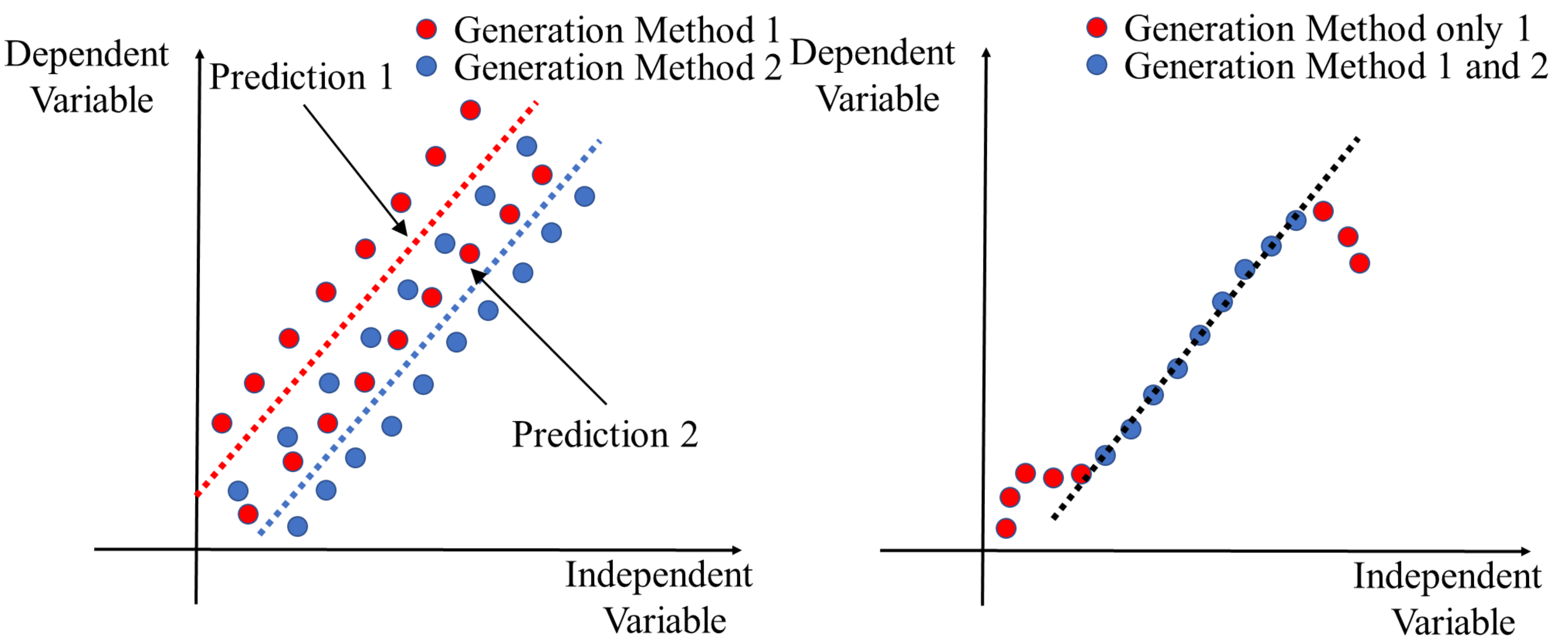
2. Data generation method considering geographical correlations and actual operation characteristics

3. Numerical Tests

4. Summary

1 Background

□ Operation data affects model performance and testing



(a) The impact on performance　　**(b) The impact on model testing**

How to generate the appropriate power system operating data?

These three factors need to be considered.

- Mathematical relationships among the variables (AC-Power flow equations) – **Newtown method**
- Geographic correlation of operational Data (A region's loads have similar characteristics)
- Inherent characteristics of power Systems (Thermal unit output is not completely random)

1 Background

Existing methods for geographic correlation

- Sample variables independently — **Ignoring geographic Correlation**
- Use historical data directly
- Use computational platforms

Limiting the scope of the application

- Sample only one variable — **Simplifying geographic Correlation**

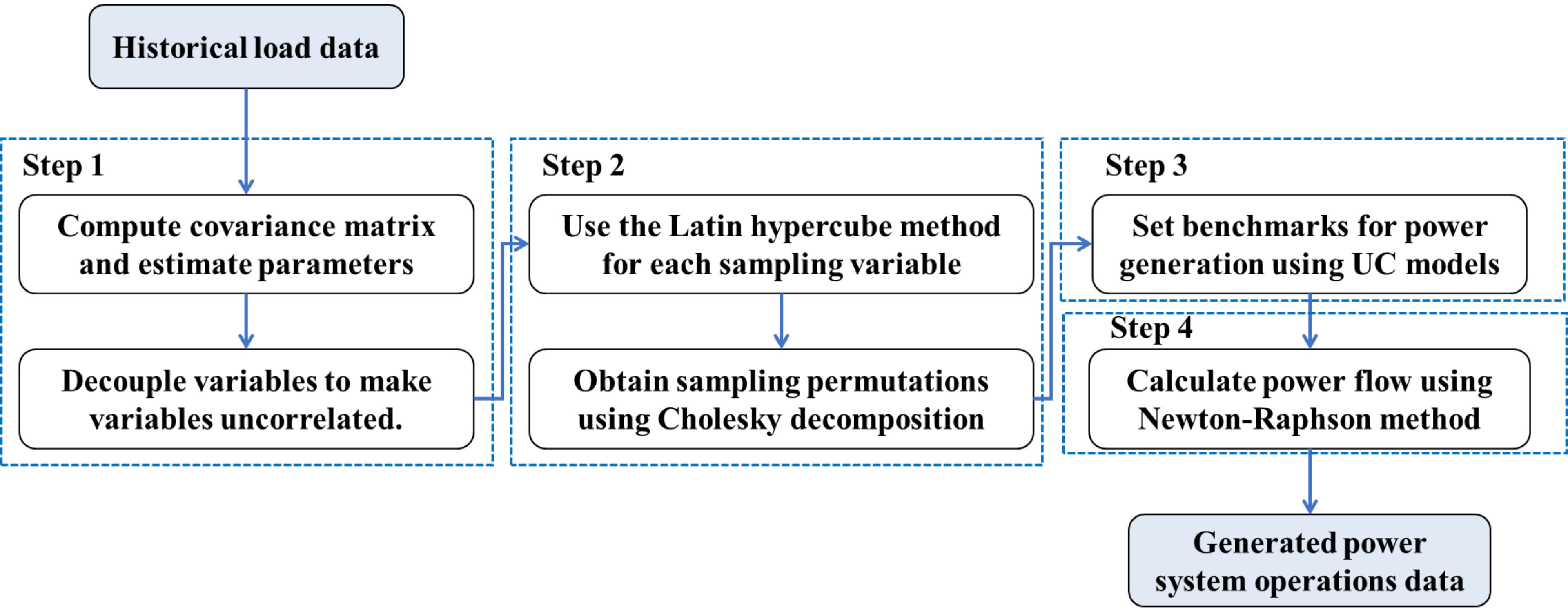
Existing methods considering operational characteristics

- Evenly distribute the load to thermal power units
- Randomly distribute loads to thermal power units

No data generation method that considers the geographic correlations and actual operational characteristics properly

2 Data generation method considering geographical correlations and actual operation characteristics

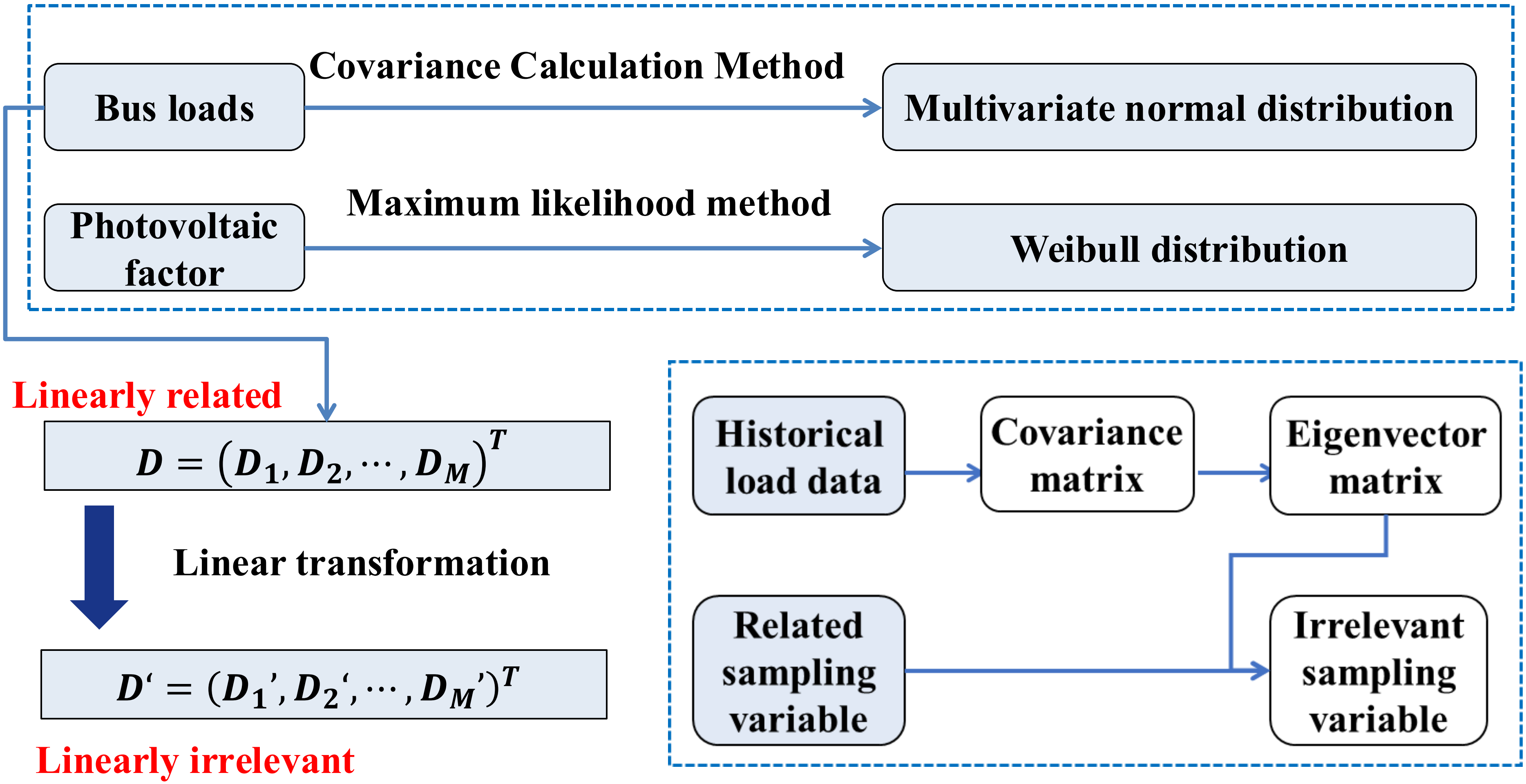
2.1 Overall process



Data Augmentation & Data Generation

2 Data generation method considering geographical correlations and actual operation characteristics

2.2 Parameter Estimation and Sampling Variable Acquisition



2 Data generation method considering geographical correlations and actual operation characteristics

2.2 Parameter Estimation and Sampling Variable Acquisition

Historical load data → **Covariance matrix** → **Eigenvector matrix** → **Related sampling variable** → **Irrelevant sampling variable**

$$\boldsymbol{D} = \left(\boldsymbol{D}_1, \boldsymbol{D}_2, \cdots, \boldsymbol{D}_M\right)^T$$

$$\Downarrow$$

$$Cov(D)$$

$$\Downarrow \text{ Eigendecomposition}$$

$$Q\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}Q^T$$

$$Cov(Q'D) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$\Downarrow$$

$$\boldsymbol{D}' = \boldsymbol{Q}'\boldsymbol{D} = (\boldsymbol{D}_1', \boldsymbol{D}_2', \cdots, \boldsymbol{D}_M')^T$$

$$\Downarrow$$

$$\boldsymbol{S} = \left(\boldsymbol{D}_1', \boldsymbol{D}_2', \cdots, \boldsymbol{D}_M', \eta_{s,1}, \eta_{s,2}, \cdots, \eta_{s,R}\right)^T$$

2.3 Sampling

$$S = \left(D'_1, D'_2, \cdots, D'_M, \eta_{s,1}, \eta_{s,2}, \cdots, \eta_{s,R}\right)^T$$

Sampling matrix X

Latin Hypercube Sampling

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} \end{bmatrix}$$

Origin sampling matrix X

One to one mapping

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

Origin order matrix

Random permutating

$$\begin{bmatrix} 1 & \boxed{2 \quad 3} & 4 \\ 2 & 3 \quad 4 & 1 \\ 4 & \boxed{2 \quad 3} & 1 \end{bmatrix}$$

Random order matrix O

linear transformation

$$\begin{bmatrix} 1 & \boxed{2 \quad 3} & 4 \\ 2 & 3 \quad 4 & 1 \\ 3 & \boxed{1 \quad 4} & 2 \end{bmatrix}$$

Updated order matrix G

One to one mapping

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ x_{2,2} & x_{2,3} & x_{2,4} & x_{2,1} \\ x_{3,3} & x_{3,1} & x_{3,4} & x_{3,2} \end{bmatrix}$$

Updated sampling matrix

2.3 Sampling

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 4 & 2 & 3 & 1 \end{bmatrix} \xrightarrow{\text{Linear transformation}} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 1 & 4 & 2 \end{bmatrix}$$

Origin order matrix

Updated order matrix G

Random order matrix O

Correlation coefficient matrix ρ_O

Cholesky decomposition

$\rho_O = TT^T$

$G = T^{-1}O$

Updated order matrix G

2 Data generation method considering geographical correlations and actual operation characteristics

2.3 Sampling

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} \end{bmatrix}$$

Origin sampling matrix X

Bijection

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

Origin order matrix

Random permutating

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 4 & 2 & 3 & 1 \end{bmatrix}$$

Random order matrix O

Cholesky decomposition

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 1 & 4 & 2 \end{bmatrix}$$

Updated order matrix G

Bijection

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ x_{2,2} & x_{2,3} & x_{2,4} & x_{2,1} \\ x_{3,3} & x_{3,1} & x_{3,4} & x_{3,2} \end{bmatrix}$$

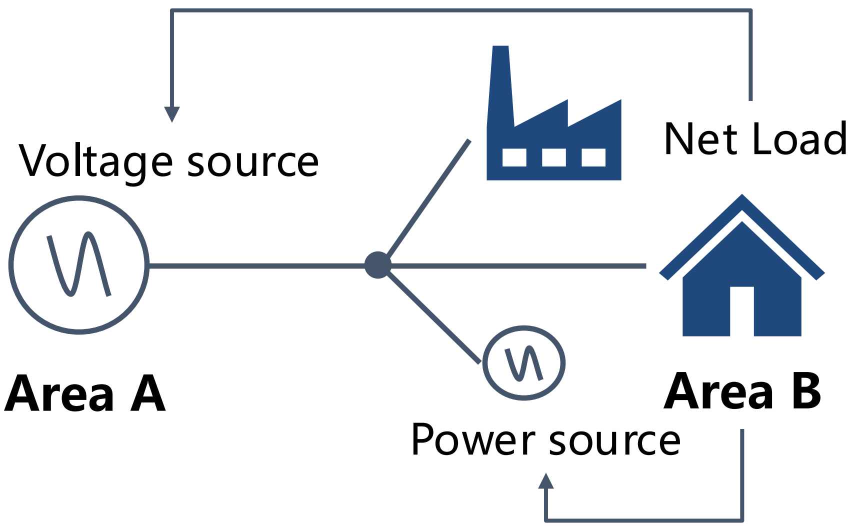
Updated sampling matrix

$$\boldsymbol{D} = (\boldsymbol{Q}^{'})^{-1}\, \boldsymbol{D}'$$

Net load = load - new energy output

Net load samping matrix $\widetilde{D}$

2.4 Generation Benchmark Settings



How to distribute the net load to thermal units?

UC Model

- **Conditions to be met**
  - Load balancing
  - Anti-risk ability
  - Power generation cost

2.4 Generation Benchmark Settings

$$\min_{p_i^g, z_i} \sum_{i=1}^{I} C_i p_i^g$$

Fuel costs

Generation benchmark

$$\sum_{i=1}^{I} p_i^g - \sum_{m=1}^{M} \mathring{D}_m = 0$$

Load balancing constraints

$$z_i \underline{P}_i \leq p_i^g \leq z_i \overline{P}_i ; i = 1, 2, \ldots, I$$

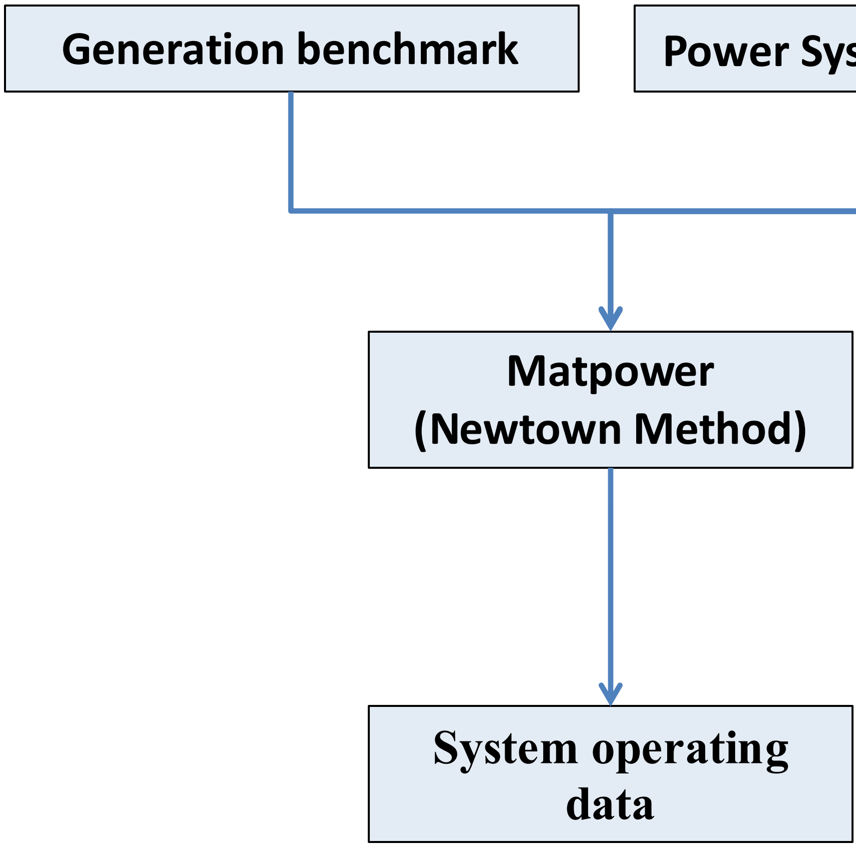
Generator output limit constraints

$$\sum_{i=1}^{I} \left(z_i \overline{P}_i - p_i^g \right) \geq SR^{+}$$

$$\sum_{i=1}^{I} \left(p_i^g - z_i \underline{P}_i \right) \geq SR^{-}$$

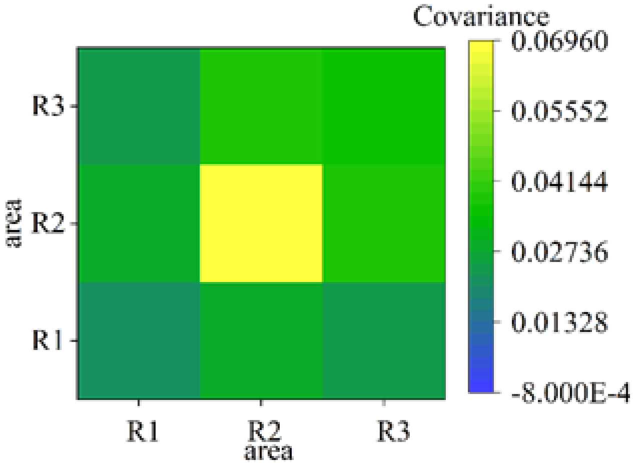
Spinning reserves constraints

2.5 System operation data calculation



$$data_k = \begin{pmatrix} P_1^k, & \dots, & P_M^k, & Q_1^k, & \dots, & Q_M^k, \\ V_1^k, & \dots, & V_M^k, & \theta_1^k, & \dots, & \theta_M^k \\ PF_1^k, & \dots, & PF_L^k, & QF_1^k & \dots, & QF_L^k \end{pmatrix}$$

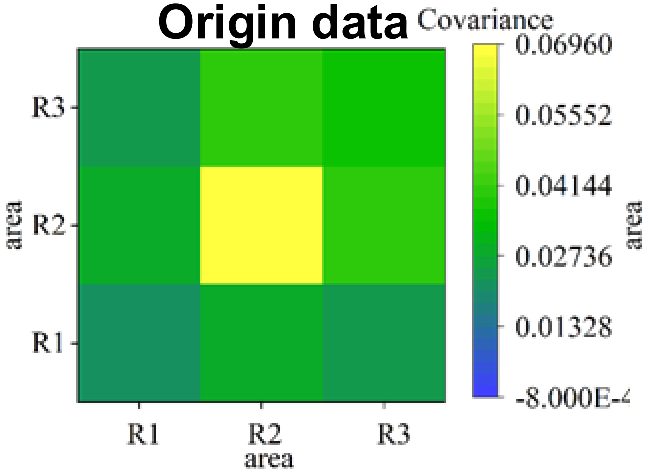
3.1 Comparison of sampling methods



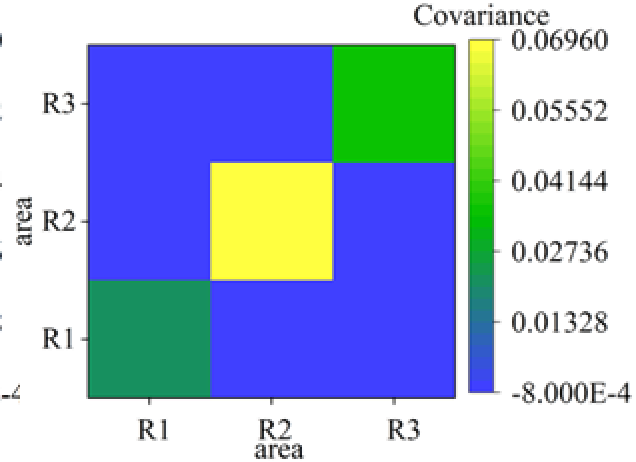
Origin data

Table 1. The overall error of different sampling methods

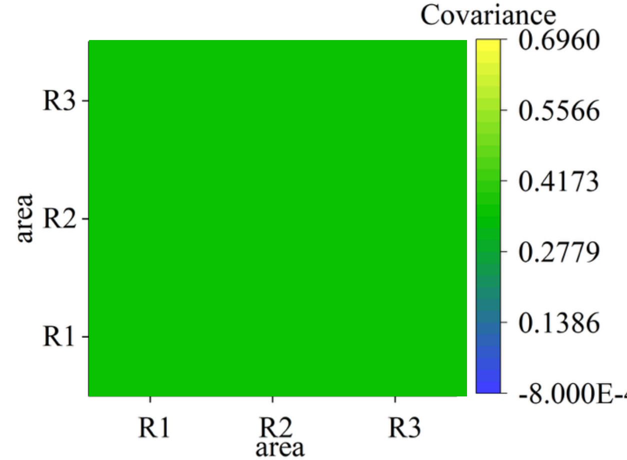
Sampling method	Average error (MW)	Max error (MW)
Method 1	2.81	106
Method 2	128	3176
Method 3	42.6	721



1 Proposed in this paper



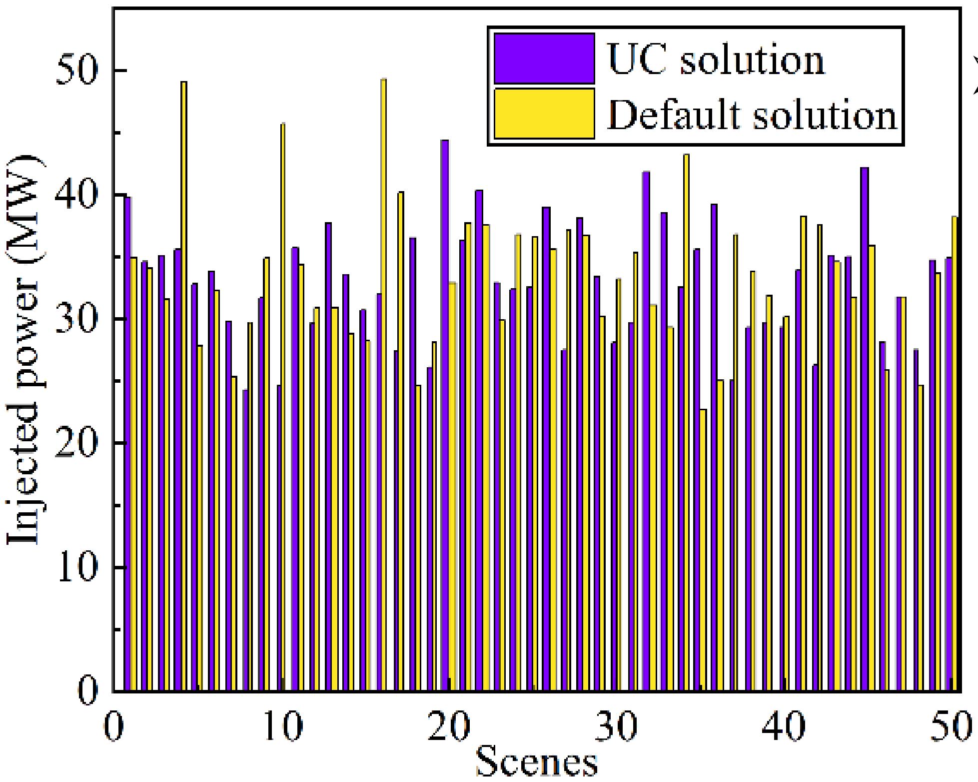
2 Sampling individually



3 Sampling only one factor

3 Numerical Tests

3.2 Comparison of generation benchmark settings



- The injected power of a bus was simulated 50 times
- The power generation benchmark has a very large impact on the data

3 Numerical Tests

3.2 Comparison of generation benchmark settings

Table 2. The overall error of different generation benchmark settings.

Benchmark-setting strategy	Average error (MW)	Max error (MW)
Default power generator benchmark	20.0964	444
Benchmark solved using UC	2.81	106

- Comparing the performance of different generation benchmark setting methods in data-driven models
- The method proposed in this paper significantly improves the performance of data-driven models

3.3 Influence analysis of renewable energy

Table 3. The overall error of different systems

Type of error	Average error (MW)	Max error (MW)
Only test set access to renewable energy	3.42	160
Both sets access to renewable energy	2.81	106
No set access to renewable energy	0.11	6.02

- We designed an experiment to test the impact of renewable energy on the data-driven linear power flow model.
- Access to renewable energy has a very large impact on the data-driven model.
- Considering renewable energy when generating the data allows for better simulations.

4 Summary

- This paper proposes a data generation method with consideration of geographical correlations and actual operational characteristics.
- Eigendecomposition and UC model are introduced into the proposed method to accurately characterize geographic correlations
- UC models are introduced to accurately characterize actual operating characteristics of thermal units.
- In the future, nonparametric estimates can be used, and more categories of renewable energy could be considered in the future to generate datasets with wider applicability.

Thank you for your listening!

mao1234@.xjtu.edu.cn