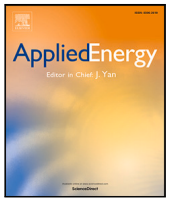
Applied Energy 314 (2022) 118872

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



A dual-driven linear modeling approach for multiple energy flow calculation in electricity–heat system ☆

Hang Tian [a], Haoran Zhao [a,*], Chunyang Liu [a], Jian Chen [a], Qiuwei Wu [b], Vladimir Terzija [c]

[a] Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Shandong University, Jinan 250061, China
[b] Center for Electric Power and Energy, Department of Electrical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[c] Center of Energy Science and Technology, Skolkovo Institute of Science and Technology, Moscow, Russia

ARTICLE INFO

Keywords:
Multiple energy flow
Integrated energy system
Linear model
Data-driven
Electricity–heat system
Partial least squares regression

ABSTRACT

The multiple energy flow (MEF) model is inherently nonlinear, which challenges the fast steady-state analysis and the efficient optimization of the integrated energy system (IES). This paper presents a hybrid mechanism- and data-driven (dual-driven) modeling approach to generate accurate linear models for MEF calculation. A linear regression model for the electricity–heat system is proposed to represent the underlying linear relationship among electrical, thermal, and hydraulic variables. Partial least squares (PLS) regression is used to deal with data collinearity which may lead to over-fitting. Furthermore, a dual-driven modeling approach is proposed to realize the complementarity between mechanism- and data-driven approaches. In this approach, mechanism analysis screens critical features before data processing and provides a fundamental linear model based on physical mechanism. Data-driven regression generates linear models to replace the complicated nonlinear elements and trains an approximation error model to further improve the overall accuracy. Numerical findings on a basic and a complex electricity–heat system confirmed the dual-driven linear model's accuracy advantage over existing linear models. Overall, this study proposes a dual-driven modeling approach that provides high-precision linear models while avoiding the intractable convergence problem in the nonlinear MEF calculation. It also reveals that linear models driven by data and mechanisms can closely fit the energy flows in the nonlinear electricity–heat system.

1. Introduction

Integrated energy system (IES) is a new energy solution that integrates regional energy resources such as coal, oil, natural gas, electricity, heat, etc., achieving synchronized planning, optimized operation, coordinated management, interactive response, and mutual complementarity among various heterogeneous energy subsystems. Multiple energy flow (MEF) analysis is a fundamental issue in the study of IES [1,2]. The MEF problem aims to find the static operating point of IES. The solution to the MEF problem is an essential component of IES planning, operation, control, and safety evaluation [3–5]. It also serves as the initial point for dynamic simulation of IES components. However, the high non-linearity of MEF equations brings heavy computation burdens and poses challenges in the convergence of MEF calculation [6,7]. In terms of practical application, how to deal with MEF calculation accurately and efficiently remains a topic that requires additional investigation.

The core of MEF calculation is to solve nonlinear energy balance equations. Many studies adopt Newton–Raphson (NR) method due to its fast convergent rate [2,6,8–10]. However, the NR method has some defects. One problem is that the global convergence cannot be guaranteed. Therefore, the initial value should be relatively accurate. Several methods have been proposed for selecting initial guesses for the NR method [11–14]. However, with the increasing system scale, a large number of electrical, gaseous, and heating variables are introduced, which reduces the feasible region of MEF calculation. Minor changes in the initial guesses of variables may result in non-convergence [14]. Moreover, the increasing number of variables enlarges the dimension of the Jacobian matrix, which usually leads to slower convergence [15]. Other iterative approaches for root-finding of nonlinear equations, such as Gauss–Seidel (GS), Fast Decoupled Load Flow (FDLF), and their various variations, are widely employed in the power industry. However, these methods exhibit similar limitations, like being initial estimate dependent and possibly locally convergent. A globally accepted approach to resolve the convergence problems of these methods has not been invented yet. Therefore, to perform the MEF calculation effectively and efficiently, a non-iterative method is needed.

☆ This document is the results of the research project funded by National Key R&D Program of China under grant 2018YFA0702200.
* Corresponding author.
E-mail address: hzhao@sdu.edu.cn (H. Zhao).

https://doi.org/10.1016/j.apenergy.2022.118872
Received 15 December 2021; Received in revised form 28 January 2022; Accepted 26 February 2022

There have been studies in power flow (PF) analysis where a linear model such as the DC power flow (DC-PF) model is used instead of the nonlinear model. However, the DC-PF model suffers from low precision. Inaccurate approximations of quantities can, in the worst case, result in substantial economic losses or system failures. Several studies have been carried out to improve its accuracy, generally divided into three categories, i.e., the mechanism-driven approach, the data-driven approach, and the dual-driven approach (hybrid mechanism- and data-driven approach). For the mechanism-driven approach, researchers focus on linearizing precise mathematical models. The formulation of reactive power injection and voltage magnitude is considered to improve its accuracy beyond the DC-PF model [16–20]. Mechanism-driven models have clear logistical structures and solid derivations. However, the trade-off between computing efficiency and accuracy limits their application in most instances, as boosting computing efficiency by simplifying models usually causes a reduction in accuracy [21]. For the data-driven approach, the linear model is generated by implementing the methods such as statistical analysis and machine learning techniques. In [22], the partial least squares (PLS) regression and the Bayesian linear regression are taken to obtain linear PF models. The results show that PLS regression is superior in most cases. The data-driven techniques generally show better performance in computing efficiency and accuracy. However, such methods rely heavily on expert experience in modeling, and their performance is also influenced by the quantity and quality of data samples [23]. For the dual-driven approach, it combines the advantages of the aforementioned approaches. A combination of linear PF model and data-driven error correction model is proposed in [24], which outperforms the data-driven model proposed in [22]. The modeling approaches above have significantly improved the accuracy of linear PF models. However, so far, little attention has been paid to accurate linear modeling in MEF calculation.

Some nonlinear models can also avoid iteration, thus having no convergence issues. Holomorphic Embedding (HE) is a non-iterative method for addressing the PF problem [25]. Unlike its iterative counterparts, it is guaranteed to identify the operational solution when it exists and to explicitly inform when such a solution does not exist. Furthermore, artificial intelligence techniques like neural networks are employed to provide data-driven nonlinear models for PF calculation in some recent research. In [26], the deep neural network (DNN) is utilized to approximate the PF calculation and trains it using physical PF equations to enhance the learning performance. In [27], a physics-guided DNN is proposed to perform the PF calculation, along with an auxiliary task to reconstruct the PF model. A fast calculation approach for probabilistic PF under topological changes is proposed in [28], which is realized by knowledge transfer of DNN parameters. These non-iterative nonlinear models, however, are primarily suited to power systems; their applicability in multi-energy systems with complicated characteristics requires additional exploration.

The presented paper goes two steps further than Tan et al. [24], a state-of-the-art study presenting a hybrid framework for linearizing branch power flow. First, in terms of the modeling target, this paper broadens the scope from the power system to the electricity–heat system, which is, to the best of our knowledge, the first attempt to adopt a data-driven technique in linear modeling of MEF. Second, in terms of modeling methodology, this study provides a more comprehensive dual-driven linear modeling approach than [24], a relatively straightforward combination of a linear PF model and a linearized error model. The proposed approach is specifically designed to combine the strengths of the mechanism- and data-driven approaches and compensate for each other's weaknesses, containing four steps: (1) the mechanism-driven approach filters incorrect data for data-driven modeling; (2) the mechanism-driven model is used to supplement the interpretability that the data-driven model lacks; (3) the data-driven model is integrated into the mechanism-driven model for complicated element substitution; (4) the data-driven model works as a post-processing correction for the mechanism-driven model.

Given that data-driven techniques, despite their potential, are rarely used in MEF modeling, our main goal is to present a dual-driven modeling approach to generate accurate linear models for MEF and demonstrate their interests and competitiveness. The research gaps motivating this paper are two-fold:

1. At present, the mainstream solution to MEF calculation is the iterative root-finding algorithm such as the NR method, which cannot guarantee global convergence and fast calculation.
2. The mechanism-driven linear model and the data-driven linear model both have benefits and limitations in this field. Comparatively speaking, data-driven models are more accurate but less transparent and adaptive; mechanism-driven models are more interpretable but less accurate. This hinders an extensive application of linear models.

In light of the research gaps, this paper's key contributions are two-fold:

1. The PLS regression technique is used to extract the implicit linear relationship in the data of the electricity–heat system, forming a general linear regression model for MEF. It can avoid the limitations of iterative algorithms while providing a variety of computational advantages.
2. A dual-driven modeling approach is developed to realize the complementarity between mechanism- and data-driven methods, which further improves the accuracy of the generated linear model.

In addition to the aforementioned contributions, the provision of such a dual-driven modeling approach for MEF calculation would assist the researchers by offering both linear and accurate models that can be integrated into practical engineering applications. A non-exhaustive list of four potential implementations is shown below. (1) Probabilistic MEF analysis. With growing uncertainty in IES due to stochastic factors in renewable energy, demand sides and markets, such models can support large numbers of fast, repeated calculations, easing the cumulative computational burden without sacrificing too much accuracy. (2) Optimal MEF analysis. The provided linear MEF formulations can serve as an alternative to the nonlinear MEF constraints and, therefore, ensure optimization convergence. (3) Large-scale MEF calculation. The convergence issue will be aggravated if the root-finding or the interior-point searching is performed on a large-scale nonlinear system. The generated linear models have good scalability by avoiding such convergence issues. (4) Parameter-free modeling. The proposed modeling approach is mainly based on the historical measurement and does not need to know the exact parameters of the network. Furthermore, with the help of the latest data, it can consider the time-varying parameter deviation caused by actual operating status. Therefore the model's accuracy may be superior to that of nonlinear empirical models. The framework of this paper is illustrated in Fig. 1 and Table 1 presents a comparison of the present study to three relevant state-of-the-art papers.

The remainder of the paper is organized as follows. Section 2 introduces the nonlinear models and iterative calculation methods of MEF in an electricity–heat system. The linear regression model for MEF calculation using the PLS technique is given in Section 3. Section 4 elaborates the proposed dual-driven modeling approach, which contains four functional modules. The accuracy of the proposed approaches is verified through case studies of different sizes in Section 5. Finally, the conclusions are drawn in Section 6.

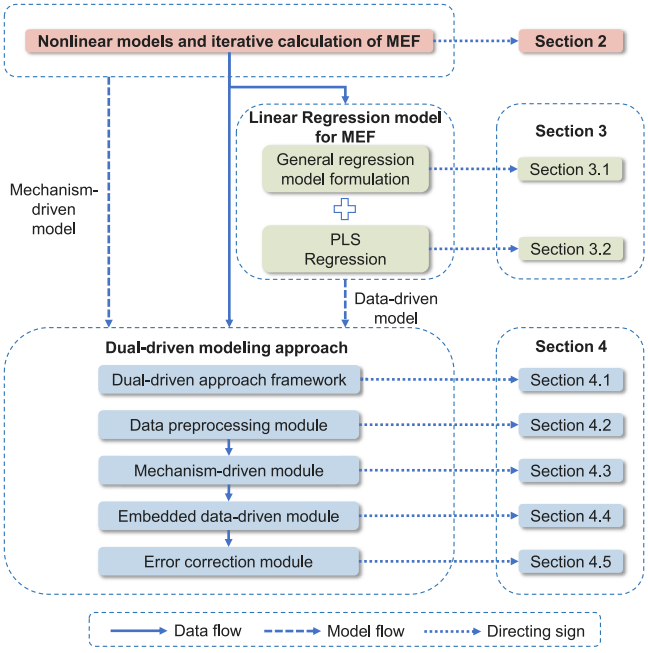


Fig. 1. Framework of this paper.

Table 1
Comparison of the paper's contributions to three relevant state-of-the-art studies.

Criteria	[14]	[22]	[24]	This study
Power flow	✓	✓	✓	✓
Multiple energy flow	✓	×	×	✓
Iterative root-finding algorithm	✓	×	×	×
Initial guess selection	✓	×	×	×
Linear regression	×	✓	✓	✓
Data preprocessing	×	×	×	✓
Linear physical model	×	×	✓	✓
Nonlinear element substitution	×	×	×	✓
Postprocessing correction	×	×	✓	✓

2. Nonlinear models and iterative calculation of MEF

The system of nonlinear equations of the electricity–heat system in the integrated calculation can be expressed as

$$\begin{cases} P_i - V_i \sum\limits_{j\in\Omega_E} V_j \left(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}\right) = 0 \leftarrow \text{Active power balance equation} \\ Q_i - V_i \sum\limits_{j\in\Omega_E} V_j \left(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}\right) = 0 \leftarrow \text{Reactive power balance equation} \\ C_p \boldsymbol{A}_h \dot{\boldsymbol{m}}_q \left(\boldsymbol{T}_s - \boldsymbol{T}_o\right) = \boldsymbol{\Phi} \leftarrow \text{Heat transfer equation} \\ T_n^{end} = \left(T_n^{start} - T_a\right) e^{-\frac{\lambda_n L_n}{C_p \dot{m}_n}} + T_a \leftarrow \text{Heat loss equation} \\ \left(\sum_{n\in\Omega_H} \dot{m}_{mn}^{out}\right) T_m^{out} = \sum_{n\in\Omega_H} \left(\dot{m}_{mn}^{in} T_{mn}^{in}\right) \leftarrow \text{Nodal balance equation of mixture temperature} \\ \boldsymbol{B}_h \boldsymbol{K}\dot{\boldsymbol{m}}|\dot{\boldsymbol{m}}| = 0 \leftarrow \text{Loop head loss balance equation} \end{cases} \tag{1}$$

where P_i and Q_i are the active power injection (MW) and reactive power injection (MVar) at bus i; V_i and V_j are the voltage magnitude (p.u.) at bus i and bus j; θ_{ij} is the voltage angle difference (rad) between bus i and bus j; G_{ij} and B_{ij} are the electrical conductance (S) and susceptance (S) of transmission line ij; $\dot{\boldsymbol{m}}_q$ is the vector of mass flow rates (kg/s) supplied at each heating node; $\boldsymbol{\Phi}$ is the vector of heat power (MW) consumed or supplied at each heating node; $\boldsymbol{T}_s$ is the vector of nodal temperature (°C) at each heating node in the supply network; $\boldsymbol{T}_o$ is the vector of outlet temperature (°C) at each heating load; C_p is the specific heat of water $\left(\mathrm{MJkg^{-1}°C^{-1}}\right)$; $\boldsymbol{A}_h$ is the incidence matrix of heating network; T_n^{start} and T_n^{end} are the temperature (°C) at the start and the end node of pipe n; $\dot{m}_n$ is the mass flow rate (kg/s) of pipe n; T_a is the ambient temperature (°C); λ_n is the overall heat transfer coefficient per unit length (MW/(m °C)) of pipe n; L_n is the length (m) of pipe n; $\dot{m}_{mn}^{out}$ and $\dot{m}_{mn}^{in}$ are the mass flow rates (kg/s) within pipe n leaving or entering node m; T_m^{out} is the mixture temperature (°C) at node m; T_{mn}^{in} is the temperature (°C) of flow at the end of pipe n entering node m; $\boldsymbol{K}$ is the vector of the resistance coefficients of each pipe; $\dot{\boldsymbol{m}}$ is the vector of mass flow rates (kg/s) of each pipe; $\boldsymbol{B}_h$ is the loop incidence matrix relating loops to pipes; Ω_E is the set of buses in the electrical network; Ω_H is the set of pipes in the heating network.

In the current research, two iterative calculation techniques, i.e., the decomposed calculation and the integrated calculation, are mainly used to obtain the operating points of the electricity–heat system, which are illustrated briefly in Fig. 2. In the decomposed calculation, hydraulic equations, thermal equations, and electrical equations are divided into three groups, which are calculated sequentially and connected by coupling equations. This sequential process iterates over each time step until the solution with an acceptable tolerance is converged. In the integrated calculation, hydraulic equations, thermal equations and electrical equations are integrated and solved simultaneously.

The precisions of the two calculation techniques are very close [13]. In this paper, the integrated calculation is taken to obtain the result of the MEF calculation since it is more straightforward and requires fewer iterations.

Table 2
Node types of heating network.

Node types	Known variables	Unknown variables
Slack node	T_s, H	$T_r, \boldsymbol{\Phi}$
Source node	T_s	$T_r, \boldsymbol{\Phi}, H$
Balance node	$\boldsymbol{\Phi}$	T_s, T_r, H
Load node	$T_r, \boldsymbol{\Phi}$	T_s, H

3. Linear regression model for MEF

In this section, a general regression model is formulated for MEF calculation in the electricity–heat system. Besides, the PLS algorithm is performed for regression to better deal with the collinearity so as to avoid the overfitting of the operational data.

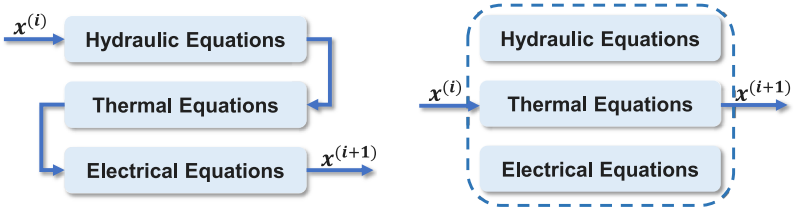
3.1. General regression model formulation

Like the three bus types defined in the electrical network (PV bus, PQ bus, and slack bus), the nodes in the heating network can also be categorized into four types: slack node, source node, balance node, and load node, which are shown in Table 2. In the MEF calculation of an electricity–heat system, the known and unknown variables are listed in Table 3.

Different electrical bus types and heating node types are defined by subscript characters $\mathcal{R}$, $\mathcal{S}$, $\mathcal{L}$ and $\mathcal{B}$, as shown in Table 4. This helps the variables of electrical network and heating network to be labeled by bus/node types for a clear classification of known and unknown variables, which is done in Eq. (2).

$$\begin{cases} \boldsymbol{P} = \left[\boldsymbol{P}_{\mathcal{R}}^T, \boldsymbol{P}_{\mathcal{S}}^T, \boldsymbol{P}_{\mathcal{L}}^T\right]^T \\ \boldsymbol{Q} = \left[\boldsymbol{Q}_{\mathcal{R}}^T, \boldsymbol{Q}_{\mathcal{S}}^T, \boldsymbol{Q}_{\mathcal{L}}^T\right]^T \\ \boldsymbol{V} = \left[\boldsymbol{V}_{\mathcal{R}}^T, \boldsymbol{V}_{\mathcal{S}}^T, \boldsymbol{V}_{\mathcal{L}}^T\right]^T \\ \boldsymbol{\theta} = \left[\boldsymbol{\theta}_{\mathcal{R}}^T, \boldsymbol{\theta}_{\mathcal{S}}^T, \boldsymbol{\theta}_{\mathcal{L}}^T\right]^T \\ \boldsymbol{\Phi} = \left[\boldsymbol{\Phi}_{\mathcal{R}}^T, \boldsymbol{\Phi}_{\mathcal{S}}^T, \boldsymbol{\Phi}_{\mathcal{L}}^T, \boldsymbol{\Phi}_{\mathcal{B}}^T\right]^T \\ \boldsymbol{T}_s = \left[\boldsymbol{T}_{s,\mathcal{R}}^T, \boldsymbol{T}_{s,\mathcal{S}}^T, \boldsymbol{T}_{s,\mathcal{L}}^T, \boldsymbol{T}_{s,\mathcal{B}}^T\right]^T \\ \boldsymbol{T}_r = \left[\boldsymbol{T}_{r,\mathcal{R}}^T, \boldsymbol{T}_{r,\mathcal{S}}^T, \boldsymbol{T}_{r,\mathcal{L}}^T, \boldsymbol{T}_{r,\mathcal{B}}^T\right]^T \\ \boldsymbol{m} = \left[\boldsymbol{m}_{\mathcal{R}}^T, \boldsymbol{m}_{\mathcal{S}}^T, \boldsymbol{m}_{\mathcal{L}}^T, \boldsymbol{m}_{\mathcal{B}}^T\right]^T \end{cases} \tag{2}$$

where $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{V}, \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{T}_s, \boldsymbol{T}_r, \boldsymbol{m}$ represent the vectors of active power injection, reactive power injection, voltage magnitude, voltage angle,



(a) Process of decomposed calculation. (b) Process of integrated calculation.

Fig. 2. Schematic diagrams of iterative methods for MEF calculation in electricity–heat system.

Table 3
Known and unknown variables of electrical network and heating network.

Network	Variables	Known at	Unknown at
Electrical network	Active power P	Slack bus	All buses except slack bus
	Reactive power Q	All buses except slack bus	Slack bus
	Voltage magnitude $\|V\|$	PV bus	PQ bus
	Voltage angle θ	PQ bus	PV bus
District heat network	Heat power Φ	Fixed heat source, balance node, load node	Slack node, source node
	Mass flow rate m	–	Each pipe
	Pressure head H	Slack node	All nodes except slack nodes
	Supply temperature T_s	Slack node, source node	Balance node, Load node
	Return temperature T_r	Load node	Slack node, Source node, Balance node

Table 4
Definition of subscript characters of electrical bus and heating node.

Subscript character	$\mathcal{R}$	$\mathcal{S}$	$\mathcal{L}$	$\mathcal{B}$
Bus of electrical network	$V\theta$	PV	PQ	–
Node of heating network	Slack	Source	Load	Balance

heat power, nodal temperature in supply network, nodal temperature in return network, mass flow rate, respectively. The regression equation is expressed as

$$\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{V}\\ \boldsymbol{m}\\ \boldsymbol{T}_{\mathrm{s}}\\ \boldsymbol{T}_{\mathrm{r}}\end{bmatrix}=\begin{bmatrix}A'_{11} & \cdots & A'_{13}\\ \vdots & \ddots & \vdots\\ A'_{51} & \cdots & A'_{53}\end{bmatrix}\begin{bmatrix}\boldsymbol{P}\\ \boldsymbol{Q}\\ \boldsymbol{\Phi}\end{bmatrix}+\begin{bmatrix}C_1\\ \vdots\\ C_5\end{bmatrix} \tag{3}$$

where the matrix $\boldsymbol{A}'$ contains regression parameters and the matrix $\boldsymbol{C}$ contains constants. It is worth noting that this regression is based on the premise that both $[\boldsymbol{\theta}\quad \boldsymbol{V}\quad \boldsymbol{m}\quad \boldsymbol{T}_{\mathrm{s}}\quad \boldsymbol{T}_{\mathrm{r}}]^{\mathrm{T}}$ and $[\boldsymbol{P}\quad \boldsymbol{Q}\quad \boldsymbol{\Phi}]^{\mathrm{T}}$ are known.

To simplify the expression, the generalized regression equation is presented as

$$\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X} \tag{4}$$

$$\boldsymbol{Y}=\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{V}\\ \boldsymbol{m}\\ \boldsymbol{T}_{\mathrm{s}}\\ \boldsymbol{T}_{\mathrm{r}}\end{bmatrix}=\begin{bmatrix}\theta^1 & \ldots & \theta^z & \ldots & \theta^Z\\ V^1 & \ldots & V^z & \ldots & V^Z\\ m^1 & \ldots & m^z & \ldots & m^Z\\ T_{\mathrm{s}}{}^1 & \ldots & T_{\mathrm{s}}{}^z & \ldots & T_{\mathrm{s}}{}^Z\\ T_{\mathrm{r}}{}^1 & \ldots & T_{\mathrm{r}}{}^z & \ldots & T_{\mathrm{r}}{}^Z\end{bmatrix},$$

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{P}\\ \boldsymbol{Q}\\ \boldsymbol{\Phi}\\ \boldsymbol{1}\end{bmatrix}=\begin{bmatrix}P^1 & \ldots & P^z & \ldots & P^Z\\ Q^1 & \ldots & Q^z & \ldots & Q^Z\\ \Phi^1 & \ldots & \Phi^z & \ldots & \Phi^Z\\ 1 & \ldots & 1 & \ldots & 1\end{bmatrix} \tag{5}$$

where $\boldsymbol{A}$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ represent the matrices of regression parameters, independent variables and dependent variables, respectively, for samples $z=1,\ldots,Z$.

During MEF analysis, the types of electrical buses and heating nodes may change, resulting in recalculation of the regression parameter matrix. To avoid the recalculations, known and unknown variables in Eq. (2) are classified in Eq. (6), where the subscript $\mathcal{N}$ stands for 'unknown' and $\mathcal{K}$ stands for 'known'. According to the definition in Eq. (6), Eq. (3) can be rewritten as Eq. (7) and simplified as Eq. (8). Eq. (8) is further reformulated as Eq. (9).

$$\begin{cases}\boldsymbol{P}_{\mathcal{N}} = [\boldsymbol{P}_{\mathcal{R}}^{\mathrm{T}}]^{\mathrm{T}} & ,\boldsymbol{P}_{\mathcal{K}} = [\boldsymbol{P}_{\mathcal{S}}^{\mathrm{T}}, \boldsymbol{P}_{\mathcal{L}}^{\mathrm{T}}]^{\mathrm{T}}\\ \boldsymbol{Q}_{\mathcal{N}} = [\boldsymbol{Q}_{\mathcal{R}}^{\mathrm{T}}, \boldsymbol{Q}_{\mathcal{S}}^{\mathrm{T}}]^{\mathrm{T}} & ,\boldsymbol{Q}_{\mathcal{K}} = [\boldsymbol{Q}_{\mathcal{L}}^{\mathrm{T}}]^{\mathrm{T}}\\ \boldsymbol{V}_{\mathcal{N}} = [\boldsymbol{V}_{\mathcal{L}}^{\mathrm{T}}]^{\mathrm{T}} & ,\boldsymbol{V}_{\mathcal{K}} = [\boldsymbol{V}_{\mathcal{R}}^{\mathrm{T}}, \boldsymbol{V}_{\mathcal{S}}^{\mathrm{T}}]^{\mathrm{T}}\\ \boldsymbol{\theta}_{\mathcal{N}} = [\boldsymbol{\theta}_{\mathcal{S}}^{\mathrm{T}}, \boldsymbol{\theta}_{\mathcal{L}}^{\mathrm{T}}]^{\mathrm{T}} & ,\boldsymbol{\theta}_{\mathcal{K}} = [\boldsymbol{\theta}_{\mathcal{R}}^{\mathrm{T}}]^{\mathrm{T}}\\ \boldsymbol{\Phi}_{\mathcal{N}} = [\boldsymbol{\Phi}_{\mathcal{R}}^{\mathrm{T}}, \boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}}]^{\mathrm{T}} & ,\boldsymbol{\Phi}_{\mathcal{K}} = [\boldsymbol{\Phi}_{\mathcal{L}}^{\mathrm{T}}, \boldsymbol{\Phi}_{\mathcal{B}}^{\mathrm{T}}]^{\mathrm{T}}\\ \boldsymbol{T}_{\mathrm{s}.\mathcal{N}} = [\boldsymbol{T}_{\mathrm{s}.\mathrm{L}}^{\mathrm{T}}, \boldsymbol{T}_{\mathrm{s}.\mathcal{B}}^{\mathrm{T}}]^{\mathrm{T}} & ,\boldsymbol{T}_{\mathrm{s}.\mathcal{K}} = [\boldsymbol{T}_{\mathrm{s}.\mathcal{R}}^{\mathrm{T}}, \boldsymbol{T}_{\mathrm{s}.\mathcal{S}}^{\mathrm{T}}]^{\mathrm{T}}\\ \boldsymbol{T}_{\mathrm{r}.\mathcal{N}} = [\boldsymbol{T}_{\mathrm{r}.\mathcal{R}}^{\mathrm{T}}, \boldsymbol{T}_{\mathrm{r}.\mathcal{S}}^{\mathrm{T}}, \boldsymbol{T}_{\mathrm{r}.\mathcal{B}}^{\mathrm{T}}]^{\mathrm{T}} & ,\boldsymbol{T}_{\mathrm{r}.\mathcal{K}} = [\boldsymbol{T}_{\mathrm{r}.\mathcal{L}}^{\mathrm{T}}]^{\mathrm{T}}\\ \boldsymbol{m}_{\mathcal{N}} = [\boldsymbol{m}_{\mathcal{R}}^{\mathrm{T}}, \boldsymbol{m}_{\mathcal{S}}^{\mathrm{T}}, \boldsymbol{m}_{\mathcal{L}}^{\mathrm{T}}, \boldsymbol{m}_{\mathcal{B}}^{\mathrm{T}}]^{\mathrm{T}}\end{cases} \tag{6}$$

$$\begin{bmatrix}\boldsymbol{\theta}_{\mathcal{N}}\\ \boldsymbol{V}_{\mathcal{N}}\\ \boldsymbol{m}_{\mathcal{N}}\\ \boldsymbol{T}_{\mathrm{s}.\mathcal{N}}\\ \boldsymbol{T}_{\mathrm{r}.\mathcal{N}}\\ \boldsymbol{\theta}_{\mathcal{K}}\\ \boldsymbol{V}_{\mathcal{K}}\\ \boldsymbol{T}_{\mathrm{s}.\mathcal{K}}\\ \boldsymbol{T}_{\mathrm{r}.\mathcal{K}}\end{bmatrix}=\begin{bmatrix}B_{11} & \cdots & B_{13} & B_{14} & \cdots & B_{16}\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ B_{51} & \cdots & B_{53} & B_{54} & \cdots & B_{56}\\ B_{61} & \cdots & B_{63} & B_{64} & \cdots & B_{66}\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ B_{91} & \cdots & B_{93} & B_{94} & \cdots & B_{96}\end{bmatrix}\begin{bmatrix}\boldsymbol{P}_{\mathcal{K}}\\ \boldsymbol{Q}_{\mathcal{K}}\\ \boldsymbol{\Phi}_{\mathcal{K}}\\ \boldsymbol{P}_{\mathcal{N}}\\ \boldsymbol{Q}_{\mathcal{N}}\\ \boldsymbol{\Phi}_{\mathcal{N}}\end{bmatrix}+\begin{bmatrix}D_1\\ \vdots\\ D_5\\ D_6\\ \vdots\\ D_9\end{bmatrix} \tag{7}$$

$$\begin{bmatrix}\tilde{\boldsymbol{Y}}_1\\ \tilde{\boldsymbol{Y}}_2\end{bmatrix}=\begin{bmatrix}\tilde{\boldsymbol{B}}_{11} & \tilde{\boldsymbol{B}}_{12}\\ \tilde{\boldsymbol{B}}_{21} & \tilde{\boldsymbol{B}}_{22}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{X}}_1\\ \tilde{\boldsymbol{X}}_2\end{bmatrix}+\begin{bmatrix}\tilde{\boldsymbol{D}}_1\\ \tilde{\boldsymbol{D}}_2\end{bmatrix} \tag{8}$$

$$\begin{aligned}\tilde{\boldsymbol{X}}_2 &= \tilde{\boldsymbol{B}}_{22}^{-1}\left(\tilde{\boldsymbol{Y}}_2-\tilde{\boldsymbol{B}}_{21}\tilde{\boldsymbol{X}}_1-\tilde{\boldsymbol{D}}_2\right)\\ \tilde{\boldsymbol{Y}}_1 &= \tilde{\boldsymbol{B}}_{11}\tilde{\boldsymbol{X}}_1+\tilde{\boldsymbol{B}}_{12}\tilde{\boldsymbol{X}}_2+\tilde{\boldsymbol{D}}_1\end{aligned} \tag{9}$$

where $\tilde{\boldsymbol{X}}_2$ and $\tilde{\boldsymbol{Y}}_1$ are unknown variables, $\tilde{\boldsymbol{Y}}_2$ and $\tilde{\boldsymbol{X}}_1$ are known variables. The matrix $\boldsymbol{B}$ contains regression parameters and the matrix $\boldsymbol{D}$ contains constants. $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{D}}$ indicate matrices containing multiple parameters from $\boldsymbol{B}$ and $\boldsymbol{D}$, respectively. In the training stage when all variables are known, $\boldsymbol{B}$ and $\boldsymbol{D}$ can be derived by Eq. (7). In the testing stage, when any type of bus/node changes, the newly altered unknown variables can be determined by Eq. (9). In this way, the regression parameter matrices $\boldsymbol{B}$ and $\boldsymbol{D}$ can be reordered to $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{D}}$ rather than recalculated, saving computational resources.

3.2. PLS regression

In this paper, the PLS regression is used to obtain the matrices of regression coefficients. The PLS regression is an effective regression

approach that combines the advantages of principal component analysis (PCA) with multivariate regression for situations with numerous variables [29].

Suppose p continuous independent variables $X_1, \ldots, X_p$. are used to predict q continuous dependent variables $Y_1, \ldots, Y_q$.

$\left(\dot{\boldsymbol{x}}_e, \dot{\boldsymbol{y}}_e\right)_{e=1,\ldots,Z}$ indicate the available data samples of Z observations, where the eth observation of the independent and dependent variables are denoted by $\dot{\boldsymbol{x}}_e \in \mathbb{R}^p$ and $\dot{\boldsymbol{y}}_e \in \mathbb{R}^q$, respectively. $\mathbb{R}$ represents the set of real numbers. Removing the dotted label means subtracting the sample mean, i.e.

$$\begin{aligned} \boldsymbol{x}_e &= \dot{\boldsymbol{x}}_e - \frac{1}{Z}\sum_{f=1}^{Z}\dot{\boldsymbol{x}}_f \\ \boldsymbol{y}_e &= \dot{\boldsymbol{y}}_e - \frac{1}{Z}\sum_{f=1}^{Z}\dot{\boldsymbol{y}}_f \end{aligned} \tag{10}$$

$\tilde{\boldsymbol{X}}$ is the $Z \times p$ matrix containing $\boldsymbol{x}_e = \left(x_{e1}, \ldots, x_{ep}\right)^{\mathrm{T}}$, $\tilde{\boldsymbol{Y}}$ is the $Z \times q$ matrix containing $\boldsymbol{y}_e = \left(y_{e1}, \ldots, y_{eq}\right)^{\mathrm{T}}$:

$$\tilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_Z^{\mathrm{T}} \end{pmatrix}, \quad \tilde{\boldsymbol{Y}} = \begin{pmatrix} \boldsymbol{y}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{y}_Z^{\mathrm{T}} \end{pmatrix} \tag{11}$$

Because the $p \times p$ covariance matrix $\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{X}}$ (which has rank at most Z-1) is singular when $Z < p$, classical linear regression technique such as ordinary least squares (OLS) cannot be used.

PLS, on the other hand, can be used in situations when $Z < p$. PLS regression is based on the component decomposition

$$\tilde{\boldsymbol{X}} = \boldsymbol{T}\boldsymbol{L}_X^{\mathrm{T}} + \boldsymbol{E} \tag{12}$$

$$\tilde{\boldsymbol{Y}} = \boldsymbol{T}\boldsymbol{L}_Y^{\mathrm{T}} + \boldsymbol{F} \tag{13}$$

where $\boldsymbol{T} \in \mathbb{R}^{n \times c}$ is a matrix for Z observations of orthogonal scores, $\boldsymbol{L}_X \in \mathbb{R}^{p \times c}$ and $\boldsymbol{L}_Y \in \mathbb{R}^{q \times c}$ are matrices of X-loading and Y-loading while $\boldsymbol{E} \in \mathbb{R}^{n \times p}$ and $\boldsymbol{F} \in \mathbb{R}^{n \times q}$ are matrices of residuals. PLS can be seen as a method to construct a matrix of orthogonal scores $\boldsymbol{T}$ as a linear transformation of $\tilde{\boldsymbol{X}}$:

$$\boldsymbol{T} = \tilde{\boldsymbol{X}}\boldsymbol{W} \tag{14}$$

where $W \in \mathbb{R}^{p \times c}$ is the matrix of weights. The columns of $\boldsymbol{W}$ and $\boldsymbol{T}$ are denoted as $\boldsymbol{w}_e = \left(w_{1e}, \ldots, w_{pe}\right)^{\mathrm{T}}$ and $\boldsymbol{t}_e = \left(t_{1e}, \ldots, t_{Ze}\right)^{\mathrm{T}}$, respectively, for $e = 1, \ldots, c$.

PLS is an iterative process, and different types of algorithms define several variants of PLS. SIMPLS is used in this paper since it is fast and straightforward owing to its direct extraction of the correlation between the PLS factors and input variables [30]. An algorithm is developed accordingly to solve the optimality problem

$$\begin{aligned} &\text{Objective function:} \quad \boldsymbol{w}_e = \arg\max_{\boldsymbol{w}} \left(\boldsymbol{w}^{\mathrm{T}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^{\mathrm{T}}\tilde{\boldsymbol{X}}\boldsymbol{w}\right) \\ &\text{s.t.} \quad \boldsymbol{w}_e^{\mathrm{T}}\boldsymbol{w}_e = 1, \boldsymbol{t}_e^{\mathrm{T}}\boldsymbol{t}_f = \boldsymbol{w}_e^{\mathrm{T}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\tilde{\boldsymbol{X}}\boldsymbol{w}_f = 0, \text{ for } f = 1, \ldots, e-1 \end{aligned} \tag{15}$$

The matrix of orthogonal scores is then utilized in place of the original variables to make predictions: $\boldsymbol{O}$ is produced as the least-squares solution of Eq. (13) once $\boldsymbol{T}$ has been constructed:

$$\boldsymbol{O}^{\mathrm{T}} = \left(\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^{\mathrm{T}}\tilde{\boldsymbol{Y}} \tag{16}$$

Finally, the general PLS regression model is written as

$$\boldsymbol{Y}' = \tilde{\boldsymbol{A}}\boldsymbol{X}' + \boldsymbol{\mathcal{E}} \text{ where } \tilde{\boldsymbol{A}}^{\mathrm{T}} = \boldsymbol{W}\boldsymbol{O}^{\mathrm{T}} = \boldsymbol{W}\left(\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}\right)^{-1}\boldsymbol{T}^{\mathrm{T}}\tilde{\boldsymbol{Y}}. \tag{17}$$

where $\boldsymbol{Y}'$ and $\boldsymbol{X}'$ represents the dependent and independent variables in the PLS regression; $\tilde{\boldsymbol{A}}$ is the matrix of regression parameters; $\boldsymbol{\mathcal{E}}$ is the matrix of residuals. This equation can be further reformulated as the form in Eq. (4).

4. Dual-driven modeling approach

4.1. Dual-driven approach framework

In this section, an integration of mechanism- and data-driven (dual-driven) approach is proposed to generate more accurate linear models for MEF calculation. The framework of the integration process contains the following four modules, illustrated in Fig. 3,

1. Data preprocessing module (module 1). The key features are found through mechanism analysis, and the erroneous data generated by simulation is removed.
2. Mechanism-driven module (module 2). Nonlinear electrical and thermal equations are linearized based on a reasonable, mechanism-driven simplification.
3. Embedded data-driven module (module 3). PLS regression is adopted to generate linear hydraulic models, which are used to replace the original nonlinear equations.
4. Error correction module (module 4): A linear model is obtained from the linearization error regression and further integrated with the previous linear models to improve the overall accuracy.

4.2. Data preprocessing module

4.2.1. Data generation

We intended to use real-world data in our study, but it was not available at this moment. One reason is that the collected measurements of an electricity–heat system are owned by two networks. Due to the consideration of independence, security, and privacy of the systems from the operators, it is not easy to obtain a large number of snapshots from the two networks simultaneously. Another reason lies with the insufficient measurement equipment in IES. However, in recent years, an increasing number of measurement equipment have been placed to increase the observability of IES. Therefore, the absence of measurement will gradually become less of a barrier in more systems [31].

In power system, studies on data-driven PF modeling have commonly taken simulation results as data samples to train models, which is considered as a reasonable alternative for insufficient datasets [22,24,27,28,32,33]. The approach has the benefit of not requiring complete or real-time measurements in all operating points. Thus, data lost in temporary failures of the measurement device or communication infrastructure could be supplemented by simulation data. Therefore, this paper uses the Monte Carlo method to generate random data samples of independent variables, including electrical and heating loads, while the iterative calculation of MEF is adopted to obtain data samples of other dependent variables.

4.2.2. Mechanism analysis

Errors not only exist in the data measurement but also may exist in the simulation results, thus reducing the accuracy of the trained model. In the 50 samples of MEF calculation, which is performed in an electricity–heat system with 32 heating nodes [13], it is found that the limit violation exists in the nodal temperature calculations, as shown in Fig. 4. In these samples, the hot water is supplied by combined heat and power (CHP) units at 70 °C; thus, the nodal temperature should not be above 70 °C in the transmission process (there is no other heat injection during the process). The adopted heating network model is widely tested in numerous literature, including but not limited to [1,13,34–38]. However, these works mainly focus on the iterative process of single MEF calculation. They do not pay attention to the erroneous data that is more likely to be noticed in a large number of simulations. According to Pan et al. [39], Liu's heating network model [13,34] is dependent on the mass flow directions of the pipes, and flow reversals can significantly affect node temperatures, resulting in non-convergence. However, the article does not provide any remedies. To deal with this

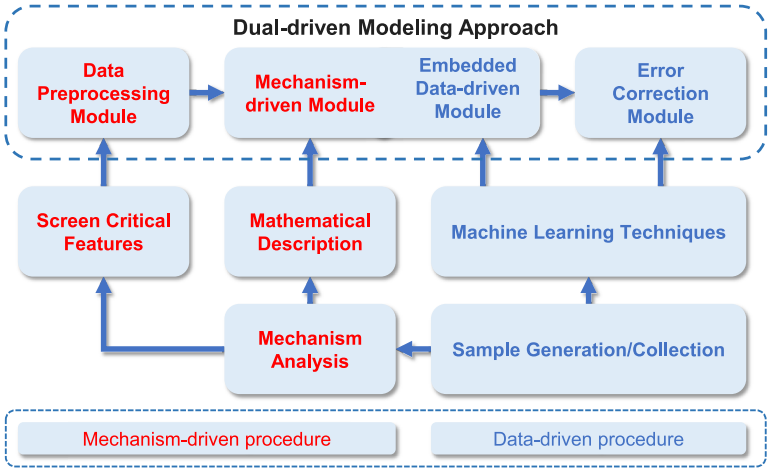


Fig. 3. Framework of the dual-driven modeling approach.

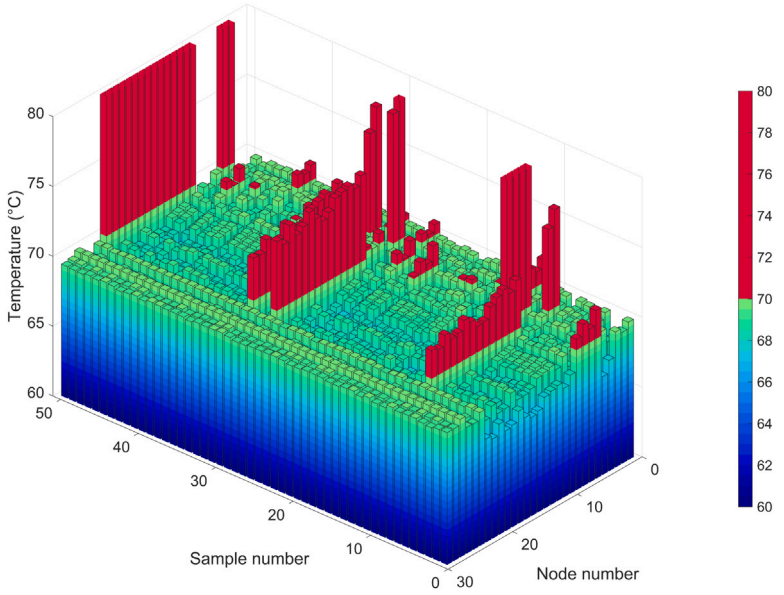
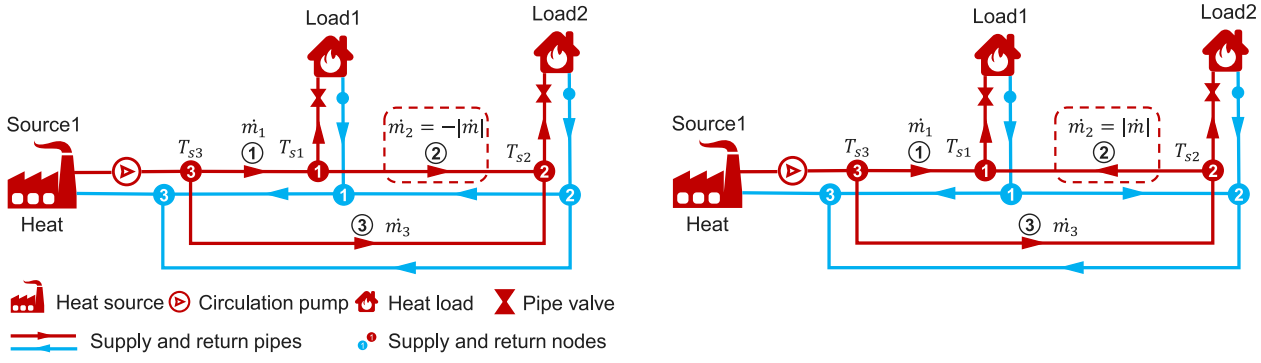


Fig. 4. Upper limit violation of nodal temperatures in MEF calculation of an electricity–heat system.



(a) $m_2 = -|m|$, the direction of water flow in pipe 2 is positive.

(b) $m_2 = |m|$, the direction of water flow in pipe 2 is negative.

Fig. 5. A simple heating network with a loop.

problem, mechanism analysis is carried out in this paper to find the critical cause.

In order to support the mechanism analysis, a looped heating network with 3 nodes is used as a demonstration, illustrated in Fig. 5. The nodal balance equation of mixture temperature [40–42] is shown as

$$\left(\sum_{n\in\Omega_H} \dot{m}_{mn}^{\text{out}}\right) T_m^{\text{out}} = \sum_{n\in\Omega_H} \left(\dot{m}_{mn}^{\text{in}} T_{mn}^{\text{in}}\right). \tag{18}$$

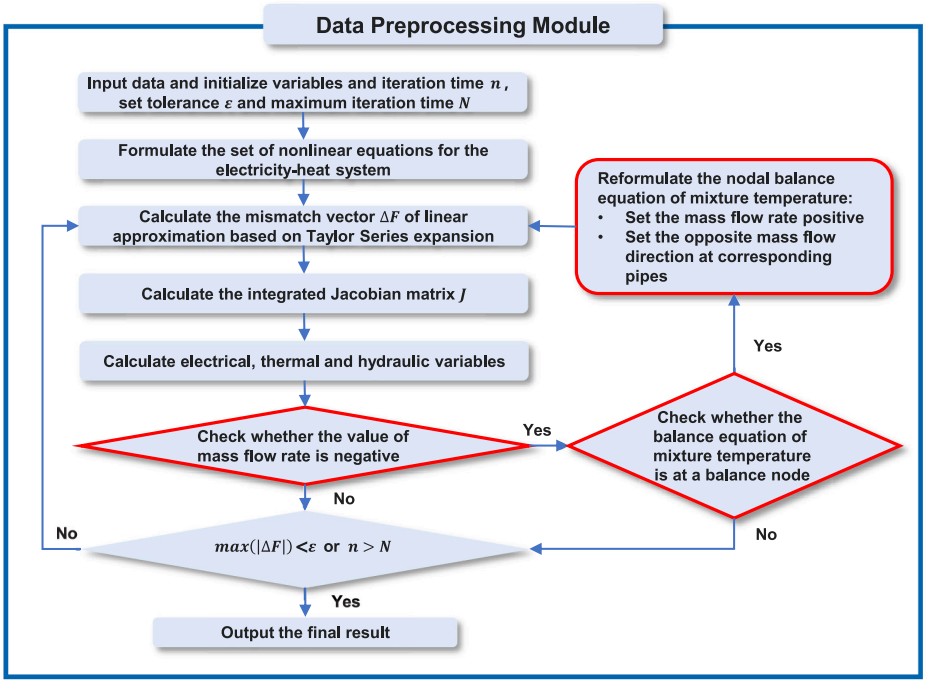


Fig. 6. Modified algorithm of data preprocessing module.

According to Eq. (18), the nodal balance equations of mixture temperature at node 1 and node 2 in Fig. 5(a) can be written as

$$\dot{m}_1(T'_{s3}\Psi_1) = (\dot{m}_1 - \dot{m}_2)T'_{s1} + \dot{m}_2 T'_{s1} \tag{19}$$

$$\dot{m}_2(T'_{s1}\Psi_2) + \dot{m}_3(T'_{s3}\Psi_3) = (\dot{m}_2 + \dot{m}_3)T'_{s2} \tag{20}$$

where T'_{sm} is the relative temperature of heating node m in the supply network, denoting $T'_{sm} = T_{sm} - T_a$; Ψ_n is the temperature drop coefficient of pipe n, denoting $\Psi_n = e^{-\frac{\lambda_n L_n}{C_p \dot{m}_n}}$. Similarly, the nodal balance equations of mixture temperature at node 1 and node 2 in Fig. 5(b) can be written as

$$\dot{m}_1(T'_{s3}\Psi_1) + \dot{m}_2(T'_{s2}\Psi_2) = (\dot{m}_1 + \dot{m}_2)T'_{s1} \tag{21}$$

$$\dot{m}_3(T'_{s3}\Psi_3) = \dot{m}_2 T'_{s2} + (\dot{m}_3 - \dot{m}_2)T'_{s2} \tag{22}$$

Fig. 5(a) and (b) are actually illustrating same mass flow rate conditions. Therefore, Eq. (19) should be equal to Eqs. (21) and (20) should be equal to Eq. (22). Substituting Eq. (19) in Eqs. (21) and (20) in Eq. (22), it can be derived that

$$T'_{s1} = T'_{s2}\Psi_2 \tag{23}$$

$$T'_{s2} = T'_{s1}\Psi_2 \tag{24}$$

It is worth noticing that both Eqs. (23) and (24) must be true at the same time to prove that the nodal balance is reached. However, the direction of mass flow is determined before calculation, resulting in either Eq. (23) or Eq. (24) being true. Therefore, when the actual mass flow direction is different from the predetermined direction, the nodal balance of mixture temperature will be violated, leading to the occasional calculation errors in the generated data samples.

4.2.3. Modified algorithm

In the above mechanism analysis, it is concluded that the upper limit violation of nodal temperatures is caused by the mutually exclusive equations of Eqs. (23) and (24), which can be traced back to the reversible mass flow (RMF) problem in the heating network, as shown in Fig. 5(a) and (b). The RMF problem can be defined as: when the mass flow value of a heating pipe is negative (the actual mass flow direction is different from the predetermined direction), the nodal balance of the mixture temperature at a connected balance node will be violated.

The related literature has not given a solution to the RMF problem since they do not need a large data set. To avoid this conflict which brings inaccuracies into the generated data samples, a data preprocessing module is designed, which includes a modified algorithm for the iterative MEF calculation, as shown in Fig. 6. In this module, two decision procedures are added to the iterative process of the NR method. One is to check whether the mass flow rate is negative, and the other is to check whether the interacting node is a balance node. If both are true, the nodal balance equation of mixture temperature will be reformulated with positive mass flow rates and opposite mass flow direction, like transforming Fig. 5(a) to (b).

4.3. Mechanism-driven module

In order to provide basic linear models for the electricity–heat system, nonlinear electrical and thermal equations are linearized based on a reasonable, mechanism-driven simplification, as shown in detail in Fig. 7. It provides a fast, physics-guided calculation but with unavoidable inaccuracy.

4.3.1. Linear electrical models

For the electrical model, the AC power flow (AC-PF) equations in Eq. (1) are linearized based on these assumptions

- Active power losses are low and the R/X ratios for transmission lines are small.
- Voltage angle differences are considered to be negligible.
- Voltage magnitudes are close to 1.0 per unit.
- Tap settings of transformers are not considered.

The linearized classical DC-PF equations are expressed as

$$P_i = \sum_{j \in \Omega_E} B_{ij}\left(\theta_i - \theta_j\right) \tag{25}$$

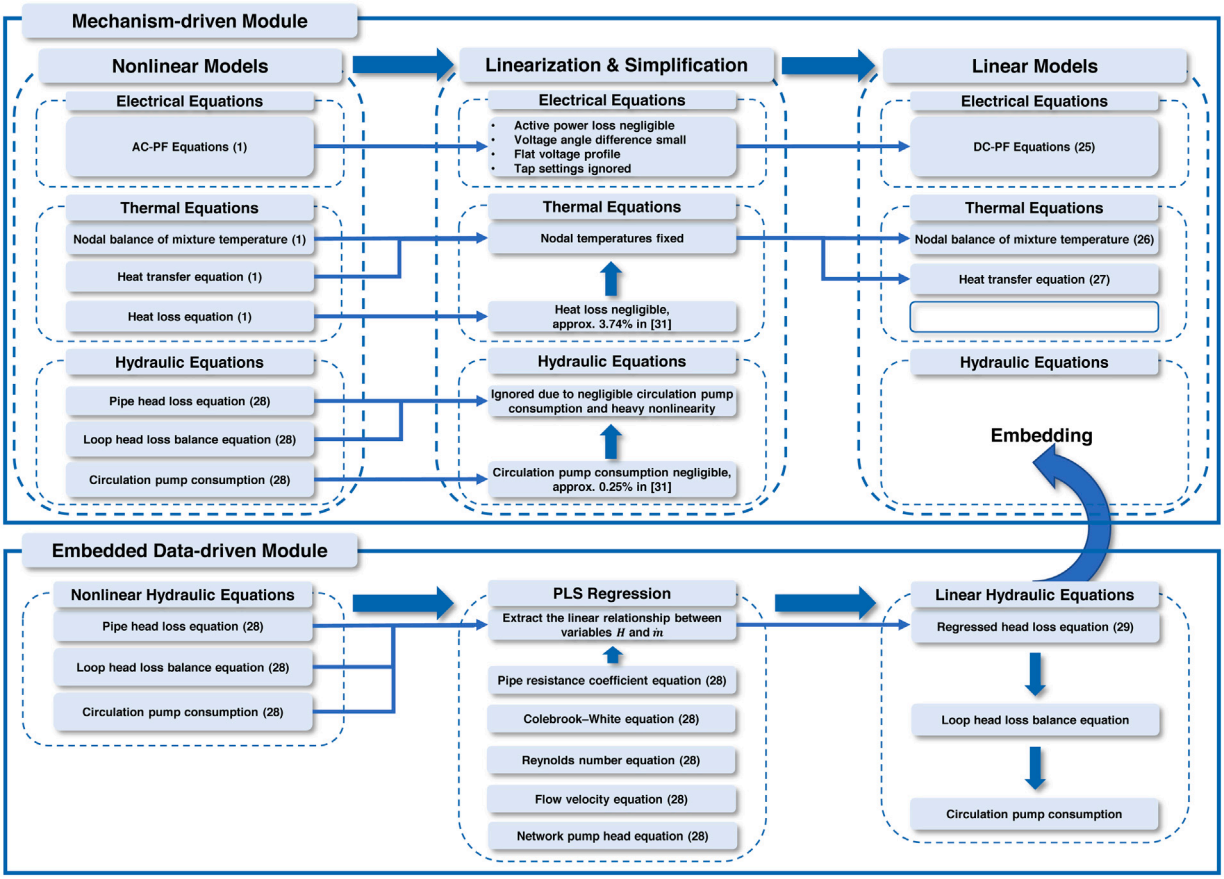


Fig. 7. Framework of mechanism-driven module and embedded data-driven module.

4.3.2. Linear thermal models

Thermal equations in Eq. (1) are linearized based on reasonable assumptions

- As the loss of heat network is usually not significant (approximately 3.74% in a 32-node heating network [34]), the heat loss in the transmission process is considered negligible.
- Since the heat loss in the transmission process is ignored, the nodal temperatures in both supply and return networks are set as constant values.

The linearized nodal balance equation of mixture temperature and heat transfer equation are written as

$$\left(\sum \dot{m}_{mn}^{\text{out}}\right) T_{\text{const}}^{\text{out}} = \sum \left(\dot{m}_{mn}^{\text{in}} T_{\text{const}}^{\text{in}}\right) \tag{26}$$

$$C_p \boldsymbol{A}_h \dot{\boldsymbol{m}}_q \left(\boldsymbol{T}_{\text{s}}^{\text{const}} - \boldsymbol{T}_{\text{o}}\right) = \boldsymbol{\Phi} \tag{27}$$

where $T_{\text{const}}^{\text{out}}$ is the constant mixture temperature (°C) at a heating node; $T_{\text{const}}^{\text{in}}$ is the constant temperature (°C) of flow at the end of a pipe; $\boldsymbol{T}_{\text{s}}^{\text{const}}$ is the vector of constant nodal temperature (°C) at each heating node in the supply network.

4.3.3. Linear hydraulic models

Hydraulic equations of the heating network are totally ignored in this module for two reasons

- According to the MEF calculation results of an islanded electricity–heat system in [34], the power consumption of the circulating pump is relatively small compared to the entire heating consumption of the heat network (approximately 0.24%). Thus, the power consumption of the circulating pump is ignored here, and there is no need to calculate head losses of pipes.
- Due to significant nonlinear characteristics, hydraulic equations are challenging to linearize via mechanism-driven simplifications.

However, ignoring the hydraulic equations may lead to inaccurate mass flow rate calculation, especially for meshed heating networks, since the loop pressure balance is not considered. Therefore, an embedded data-driven module is designed in the following subsection to solve this problem.

4.4. Embedded data-driven module

In this module, PLS regression is used to generate linear hydraulic models. These linear models are then embedded in the mechanism-driven module to realize the functions of nonlinear hydraulic equations. This process is illustrated in Fig. 7. A combination of hydraulic equations is written as

$$\begin{cases} \boldsymbol{B}_{\text{h}} \boldsymbol{K} \dot{\boldsymbol{m}} |\dot{\boldsymbol{m}}| = 0 \leftarrow \text{Loop head loss balance equation} \\ \boldsymbol{H} = \boldsymbol{K} \dot{\boldsymbol{m}} |\dot{\boldsymbol{m}}| \leftarrow \text{Pipe head loss equation} \\ K_n = \frac{8 L_n f_n}{D_n^5 \rho^2 \pi^2 g} \leftarrow \text{Pipe resistance coefficient equation} \\ \frac{1}{\sqrt{f_n}} = -2 \log_{10} \left(\frac{\epsilon_n / D_n}{3.7} + \frac{2.51}{\text{Re}_n \sqrt{f_n}} \right) \leftarrow \text{Colebrook–White equation} \\ \text{Re}_n = \frac{v_n D_n}{\mu_n} \leftarrow \text{Reynolds number equation} \\ v_n = \frac{\dot{m}_n}{\rho \pi D_n^2 / 4} \leftarrow \text{Flow velocity equation} \\ P_{\text{P}} = \frac{\dot{m}_{\text{P}} g H_{\text{P}}}{10^6 \eta_{\text{P}}} \leftarrow \text{Circulation pump consumption equation} \\ H_{\text{P}} = 2 \sum_{n \in \Omega_{\text{P}}} H_n + H_c \leftarrow \text{Network pump head equation} \end{cases} \tag{28}$$

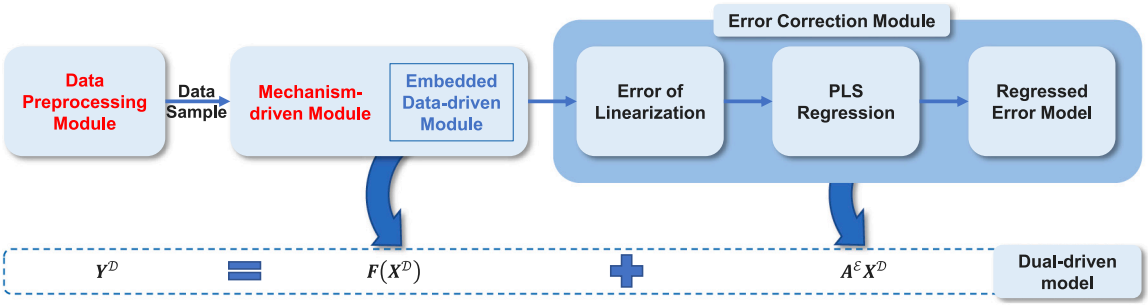


Fig. 8. Framework of error correction module and formulation of dual-driven model.

where $\boldsymbol{H}$ is the vector of head loss of each pipe in the heating network; K_n is the resistance coefficient of pipe n; f_n is the friction factor of pipe n; D_n is the diameter (m) of pipe n; ρ is water density $(\mathrm{kg/m^3})$; g is gravitational acceleration $(\mathrm{kg\ m/s^2})$; ϵ_n is the roughness (m) of pipe n; D_n is the diameter (m) of pipe n; Re_n is the Reynolds number of flow in pipe n; v_n is the flow velocity (m/s) of pipe n; μ_n is the kinematic viscosity of water $(\mathrm{m^2/s})$ of pipe n; P_{P} is the electrical power consumed (MW) by a circulation pump; $\dot{m}_{\mathrm{P}}$ is the mass flow rate (kg/s) through the circulation pump; H_{P} is the pump head (m) of the network; η_{P} is the efficiency of the circulation pump; H_n is the head loss (m) of pipe n; H_c is the minimum allowable head differential (m); Ω_{P} is the set of pipes in the heating network.

Among all hydraulic equations listed above, it is found that pipe head loss variable H_n and mass flow rate variable $\dot{m}_n$ are the key dependent and independent variables, respectively. PLS regression is then used to extract the inexplicit linear relationship between these two variable sets

$$\begin{bmatrix} H_1 \\ \vdots \\ H_n \end{bmatrix} = \begin{bmatrix} A^{\mathcal{H}}_{11} & \cdots & A^{\mathcal{H}}_{1\,n+1} \\ \vdots & \ddots & \vdots \\ A^{\mathcal{H}}_{n1} & \cdots & A^{\mathcal{H}}_{n\,n+1} \end{bmatrix} \begin{bmatrix} \dot{m}_1 \\ \vdots \\ \dot{m}_n \\ 1 \end{bmatrix} \tag{29}$$

where $\boldsymbol{A}^{\mathcal{H}}$ is the matrix of parameters of the hydraulic regression.

4.5. Error correction module

With the embedding of module 3 in module 2, an integration model, i.e., a system of linear equations, can be formed and denoted as $\boldsymbol{F}(\boldsymbol{x})$. With the training data sample $\boldsymbol{X}^{\mathcal{T}}$ and $\boldsymbol{Y}^{\mathcal{T}}$ containing independent variables and dependent variables, the output of the integration model can be expressed as

$$\boldsymbol{Y}^{\mathcal{I}} = \boldsymbol{F}(\boldsymbol{X}^{\mathcal{T}}) \tag{30}$$

Meanwhile, the linearization error of the integrated model in Eq. (30) can be obtained and regressed by PLS

$$\boldsymbol{\Delta E} = \boldsymbol{Y}^{\mathcal{T}} - \boldsymbol{Y}^{\mathcal{I}} \tag{31}$$

$$\boldsymbol{\Delta E} = \begin{bmatrix} \boldsymbol{\Delta E}_{\theta} \\ \boldsymbol{\Delta E}_{V} \\ \boldsymbol{\Delta E}_{m} \\ \boldsymbol{\Delta E}_{T_{\mathrm{s}}} \\ \boldsymbol{\Delta E}_{T_{\mathrm{r}}} \end{bmatrix} = \begin{bmatrix} A^{\mathcal{E}}_{11} & \cdots & A^{\mathcal{E}}_{14} \\ \vdots & \ddots & \vdots \\ A^{\mathcal{E}}_{51} & \cdots & A^{\mathcal{E}}_{54} \end{bmatrix} \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \\ \boldsymbol{\Phi} \\ \boldsymbol{1} \end{bmatrix} \tag{32}$$

where $\boldsymbol{A}^{\mathcal{E}}$ is the matrix of parameters of the linearization error regression. Finally, with the newly-generated matrix of independent variables $\boldsymbol{X}^{D}$, the output of the proposed dual-driven model can be formulated as

$$\begin{aligned} \boldsymbol{Y}^{D} &= \boldsymbol{Y}^{\mathcal{I}} + \boldsymbol{\Delta E} \\ &= \boldsymbol{F}(\boldsymbol{X}^{D}) + \boldsymbol{A}^{\mathcal{E}}\boldsymbol{X}^{D} \end{aligned} \tag{33}$$

The framework of the error correction module and the formulation of the dual-driven model are shown in Fig. 8. In the dual-driven framework, the mechanism-driven approach can perform data filtering and transparency enhancing to cover the shortage of the data-driven model, while the data-driven approach can implement complicated element substitution and error correction to reinforce the mechanism-driven model. In this way, it can achieve higher accuracy than existing mechanism- and data-driven linear models and still maintain the linear property that allows rapid calculation. The performance of the dual-driven model will be validated in the following section.

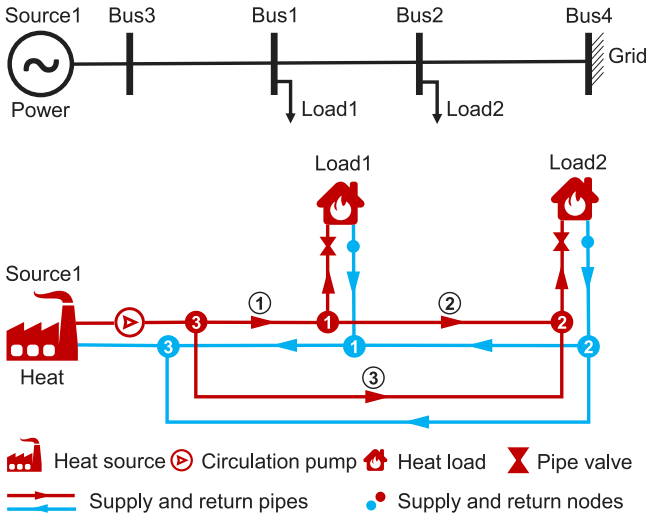


Fig. 9. Schematic diagram of electrical and heating networks in Case A.

5. Case study

5.1. Simulation setups

The Monte Carlo method produces random data samples for independent variables, while the iterative calculation of MEF is used to calculate data samples for dependent variables. The load consumption of active power is set as the reference value multiplied by a stochastic factor drawn from a random distribution for 0.75 to 1.25. The load consumption of heat is obtained the same as above. The load consumption of reactive power is set as the active power multiplied by a stochastic factor drawn from a random distribution for 0.15 to 0.25.

Two electricity–heat systems cases are considered in this section: a small grid-connected system and a large islanding system [34], denoting as Case A and Case B, respectively. The network structure of Case A is illustrated in Fig. 9. The schematic diagrams of Case B are demonstrated in Figs. 10 and 11, for heating network and electrical network. All related parameters concerning Figs. 9, 10, and 11 can be found in [34].

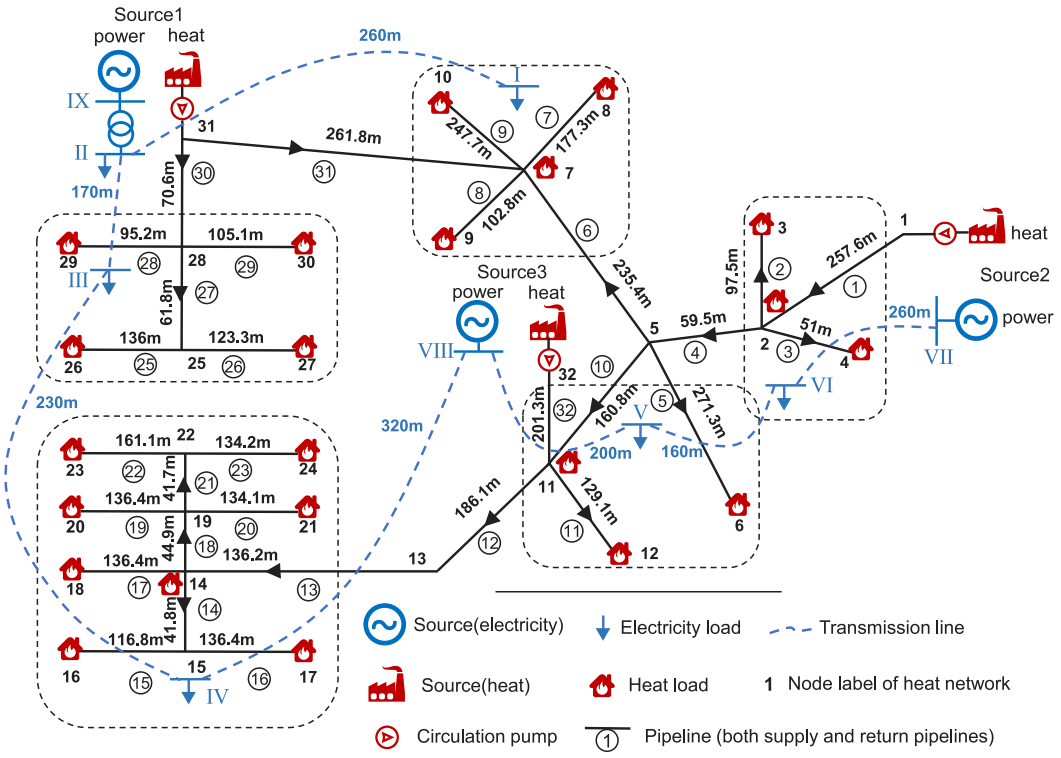


Fig. 10. Schematic diagram of electrical and heating networks in Case B.

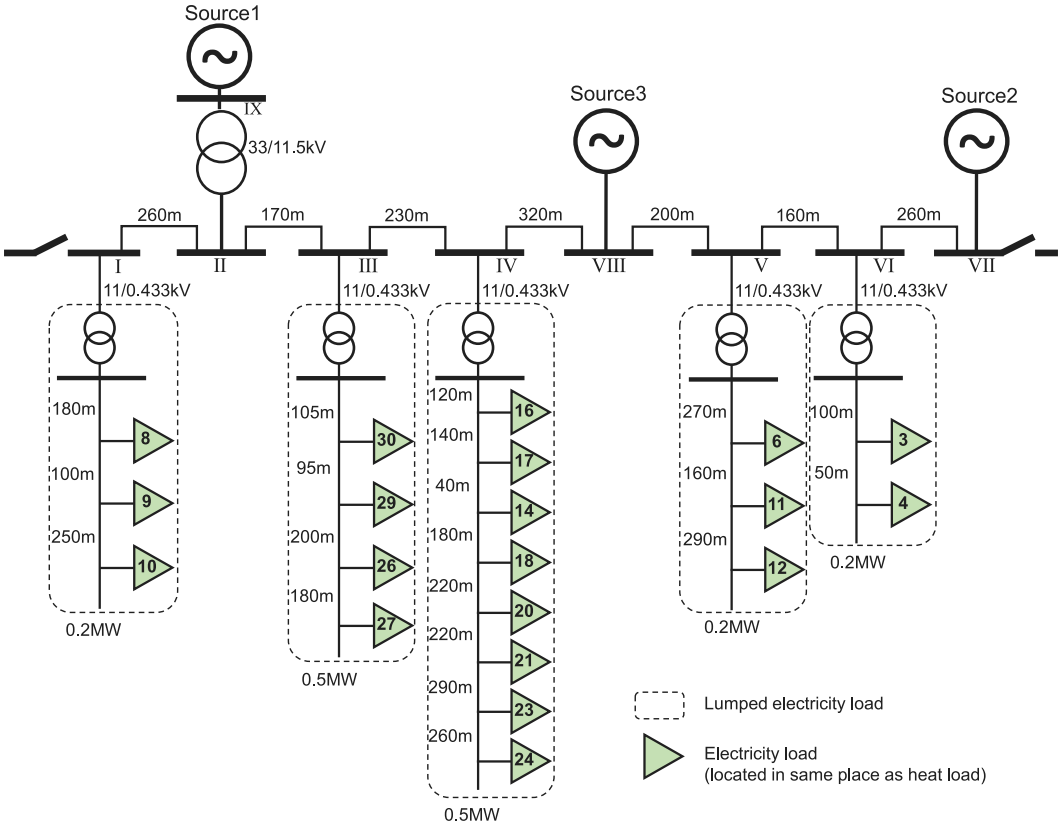


Fig. 11. Schematic diagram of power distribution network in Case B.

Three different modeling approaches are compared by analyzing the calculation results of generated linear models. The modeling approaches are defined as:

1. The mechanism-driven approach: the linear model of the electricity–heat system is derived by mechanism-driven linearization.
2. The data-driven approach: based on the data samples of MEF calculation, PLS regression is conducted to generate linear models.
3. The dual-driven approach: linear models are generated by a combination of four functional modules based on mechanism- and data-driven modeling methods.

All of the linear models generated by the aforementioned approaches are tested by the MEF calculation results on 500 newly generated samples.

5.2. Simulation results of Case A

A small grid-connected electricity–heat system is used to illustrate the preliminary comparison of three modeling approaches. The small system is helpful to demonstrate the accuracy of each variable for detailed analysis. Mean absolute error (MAE) and mean absolute percentage error (MAPE) of unknown variables calculated by different approaches are shown in Table 5. From this table, several observations can be obtained:

(1) Clearly, the dual-driven approach obtains the least MAEs and MAPEs of variables among all approaches in MEF calculation. Meanwhile, the data-driven approach provides the second-best solution. The mechanism-driven linear model performs much worse than the previous two, mainly because much accuracy is sacrificed due to rough linearization. The data-driven and dual-driven approaches show remarkable performance in the accuracies of generated models. However, it remains to be tested in a larger case.

(2) It is worth noticing that the MAPEs of m_2 are significantly larger than that of other variables. The reason is that in Case A, the base values of two heat loads are set the same, resulting in the small value of m_2 because most of the hot water flows through pipe 1 and pipe 3; Besides, the MAEs of m_1, m_2 and m_3 are on the same level, proving that the high MAPEs of m_2 are mainly due to its small base values. A simple test is done to validate the above judgment: when the base value of Load 2 in the heating network is increased by 20%, the MAPEs of m_2 by three approaches are decreased to 38.04%, 3.01% and 1.04% respectively.

(3) The calculation of electrical variables is significantly more accurate than that of the thermal and hydraulic variables, which implies a higher degree of linearity between variables exists in the electrical model than in the thermal and hydraulic model.

5.3. Simulation results of Case B

The three modeling approaches are further compared in Case B, an island-level, self-sufficient electricity–heat system. Electrical and heating networks are interconnected by CHP units. A gas turbine, an extraction steam turbine, and a reciprocating engine are used in CHP units corresponding to Source 1,2, and 3. The equations and parameters of the three types of CHP units are detailed in [34]. Several findings can be drawn from the simulation results in Table 6.

(1) For different variables, The MAPEs of the data-driven approach are less than 1.2%, while the MAPEs of the dual-driven approach are less than 0.2%, which are far less than the mechanism-driven approach. Thus, it indicates that the participation of data-driven methods can significantly improve the accuracy of generated linear models for MEF calculation. Besides, the dual-driven approach can improve the accuracy to a higher level.

(2) From Case A to Case B, the number of variables increases from 13 to 93, and the results of the data-driven approach and the dual-driven approach still show considerable accuracy, which validates the effectiveness of the proposed approach in a large-scale system.

5.4. Error distribution analysis

The error distribution of the variables calculated by the data-driven approach and dual-driven approach is illustrated by the histograms in Fig. 12, where the occurrence frequency of the absolute errors in voltage angle, voltage magnitude, mass flow rate, and nodal temperature (supply) are shown in details. The test is done in Case B with 500 newly generated samples. It is apparent that in all graphs, the error distribution of the dual-driven approach is significantly narrower and more to the left than that of the data-driven approach, which proves that the dual-driven approach has improved the accuracy through combining with the mechanism-driven approach. In addition, the accuracy of thermal and hydraulic variables is improved more significantly than that of electrical variables.

5.5. Impact of training data size

The relationship between model accuracy and training sample size by different approaches in Case B is illustrated in Fig. 13. Obviously, the MAEs of variables of Fig. 13(a) is lower than that of Fig. 13(b), indicating the accuracy advantage of the dual-driven approach over the data-driven approach. It is found that in both Fig. 13(a) and (b), the lines of MAE decline significantly until the sample size of 50 is reached, which is nearly 1.2 times of the sum of electrical buses and heating nodes. However, for the data-driven approach, MAEs of nodal temperatures (both supply and return) have significant fluctuations even after 50 samples, which shows that the dual-driven approach has more stable regression performance than the data-driven approach under different data volumes.

5.6. Regression index analysis

In regression, the R^2 coefficient of determination is a statistical measure of how well the regression predictions approximate the real data, based on the proportion of total variation in the results explained by the model [43,44]. It ranges from −1 to 1, and an R^2 of 1 indicates that the regression predictions perfectly fit the data. The most general definition of the coefficient of determination can be expressed as

$$R^2 = 1 - \frac{S_{\text{res}}}{S_{\text{tot}}} = 1 - \frac{\sum_z \left(y_z - f_z\right)^2}{\sum_z \left(y_z - \bar{y}\right)^2} \tag{34}$$

where S_{res} is the sum of squares of residuals; S_{tot} is the total sum of squares; y_z is the data set value; f_z is the fitted value; $\bar{y}$ is the mean of the data set value.

In this paper, to compare the goodness of fit between the dual-driven approach and the data-driven approach, R^2 of dependent variables at each bus/node in Case B are compared in Fig. 14, where the comparison models are trained from 500 samples. In the first and second subfigure, the R^2 indices are all above 0.9998, which means the independent variables can predict a very large percentage of the variation in the dependent variables. In addition, although the data-driven approach already has high regression performance, the dual-driven approach further improves it to a higher level. In the last three subfigures, it is clear that the R^2 indices for mass flow rate, nodal temperature (supply), and nodal temperature (return) are significantly improved by the dual-driven approach, from mean values of 0.9229, 0.2104, and 0.5529 to 0.9998, 0.8741 and 0.9555, respectively. The key observation from these subfigures are twofold:

Table 5
Errors of MEF calculation results by different linear modeling approaches on Case A.

Modeling approach	Relative error												
	Electrical variables					Thermal and hydraulic variables							
	Voltage angle (°)			Voltage magnitude (1 p.u)		Mass flow rate (kg/s)			Nodal temperature (Supply) (°C)		Nodal temperature (Return) (°C)		
	Bus 1	Bus 2	Bus 3	Bus 1	Bus 2	Node 1	Node 2	Node 3	Node 1	Node 2	Node 1	Node 2	Node 3
Mechanism-driven (MAE)	**2.81E−03**	**4.41E−03**	**2.24E−02**	**1.81E−02**	**8.75E−03**	**1.31E−01**	**1.59E−01**	**2.45E−01**	**1.06E+00**	**2.83E+00**	**4.15E−01**	**3.05E−04**	**8.71E−01**
(MAPE)	2.80286%	13.91926%	13.07532%	1.77488%	0.86671%	8.20851%	432.23947%	18.68422%	1.07417%	2.91416%	0.83656%	0.00061%	1.77308%
Data-driven (MAE)	**3.41E−04**	**1.67E−04**	**5.23E−04**	**9.74E−05**	**1.01E−04**	**4.12E−03**	**3.32E−03**	**3.20E−03**	**1.00E−02**	**1.73E−01**	**6.37E−02**	**5.80E−04**	**3.68E−02**
(MAPE)	0.36674%	0.47447%	0.31071%	0.00957%	0.00998%	0.25555%	10.70073%	0.24213%	0.01014%	0.17794%	0.12844%	0.00116%	0.07493%
Dual-driven (MAE)	**1.55E−04**	**7.81E−05**	**2.32E−04**	**8.09E−05**	**8.72E−05**	**2.15E−03**	**2.01E−03**	**1.17E−03**	**9.49E−03**	**6.18E−02**	**2.35E−02**	**5.30E−05**	**1.55E−02**
(MAPE)	0.17172%	0.23825%	0.14040%	0.00795%	0.00864%	0.13400%	2.99984%	0.08631%	0.00960%	0.06370%	0.04733%	0.00011%	0.03147%

*The relative error is expressed in terms of mean absolute error (MAE) and mean absolute percentage error (MAPE).

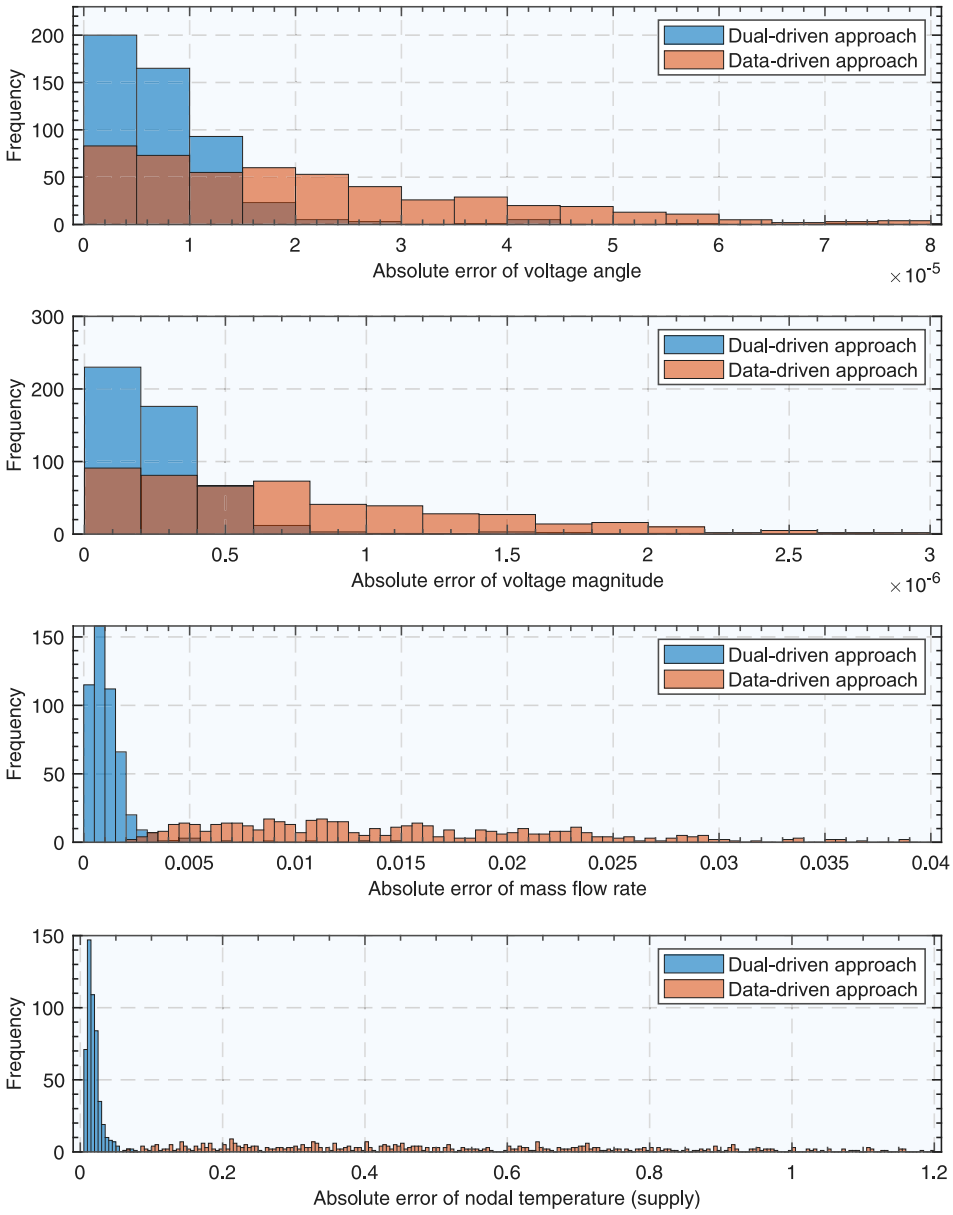
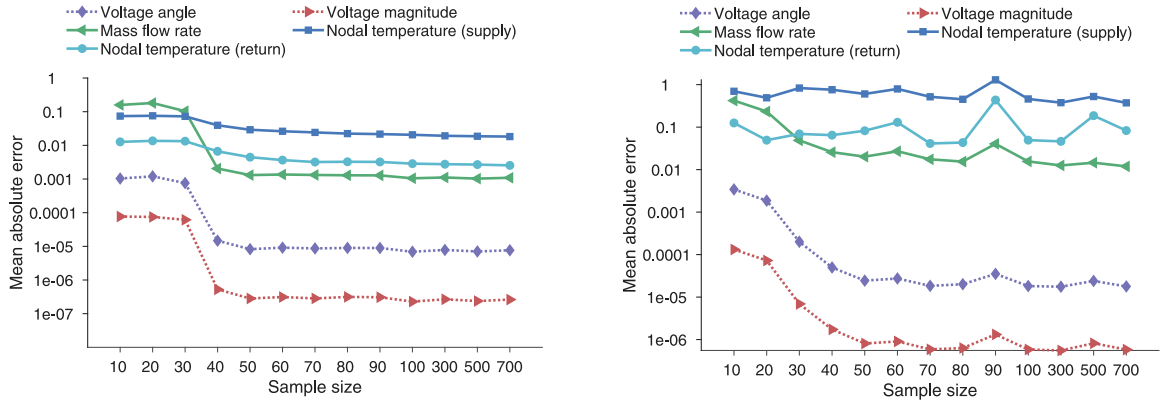


Fig. 12. Error distribution of variables calculated by data-driven approach and dual-driven approach in Case B.



(a) Relationship between model accuracy and training sample size by the dual-driven approach.

(b) Relationship between model accuracy and training sample size by the data-driven approach.

Fig. 13. Relationship between model accuracy and training sample size by different approaches in Case B.

Table 6
Errors of MEF calculation results by different linear modeling approaches on Case B.

Modeling approach	Relative error				
	Electrical variables		Thermal and hydraulic variables		
	Voltage angle (°)	Voltage magnitude (1 p.u)	Mass flow rate (kg/s)	Temperature (°C) (Supply nodes)	Temperature (°C) (Return nodes)
Mechanism-driven (MAE)	**2.3768E−03**	**4.9296E−02**	**3.4727E−01**	**1.0218E+00**	**2.7540E−01**
(MAPE)	17.4451419%	4.6979668%	36.8719178%	1.4869425%	0.9270953%
Data-driven (MAE)	**1.6892E−05**	**5.3795E−07**	**1.1173E−02**	**3.2388E−01**	**4.1315E−02**
(MAPE)	0.1437002%	0.0000513%	1.1361858%	0.4701168%	0.1362907%
Dual-driven (MAE)	**8.0861E−06**	**2.7641E−07**	**1.1545E−03**	**1.9365E−02**	**2.5727E−03**
(MAPE)	0.0603031%	0.0000264%	0.1834494%	0.0280889%	0.0086582%

* The relative error is expressed in terms of mean absolute error (MAE) and mean absolute percentage error (MAPE).

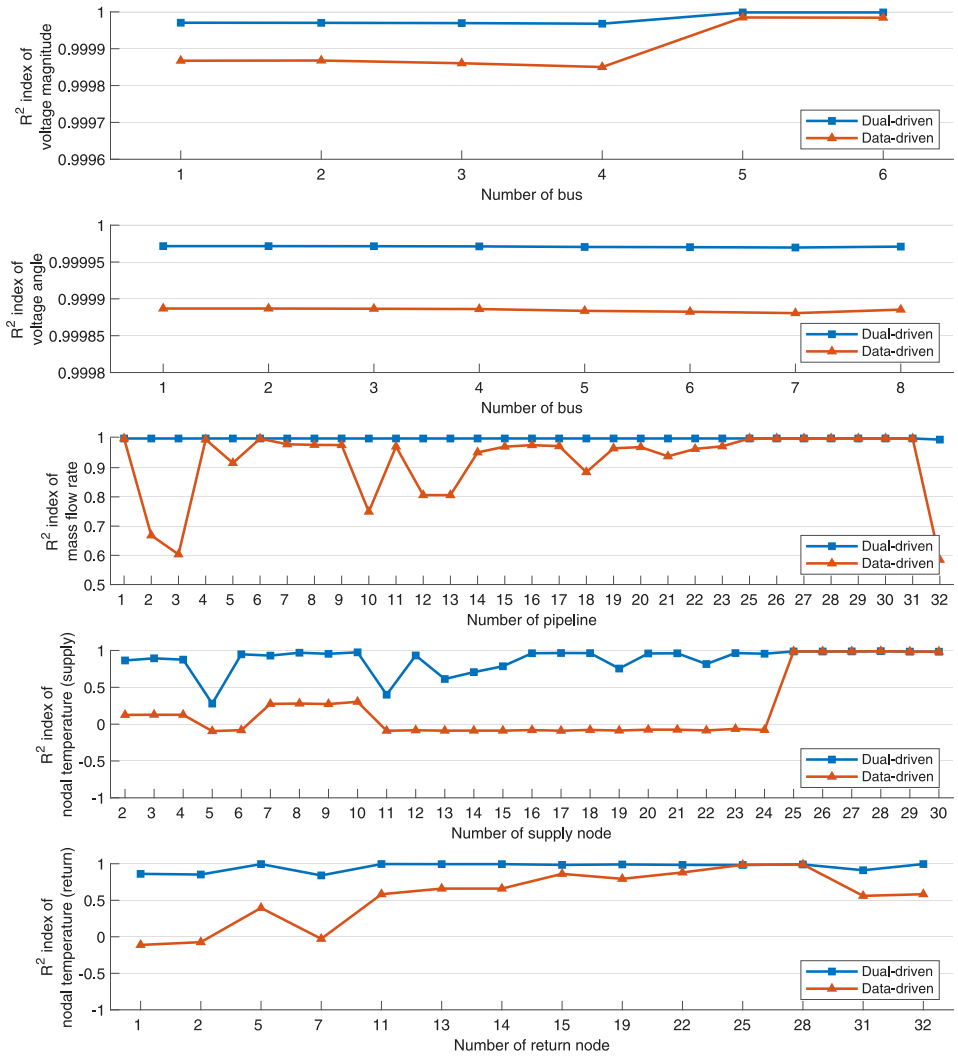


Fig. 14. Comparison between the dual-driven approach and data-driven approach on coefficient of determination for dependent variables in Case B.

(1) Although the IES is considered to have more nonlinear characteristics than the power system, the tailored linear regression tactics effectively generate well-fit linear models (with the average R^2 higher than 0.85) to deal with MEF calculation for the electricity–heat system.

(2) Higher nonlinearity exists in thermal and hydraulic calculations than in electrical calculation, which lowers the R^2 indices for thermal and hydraulic variables, especially for the balance node connecting multiple pipes (e.g., node 5 and node 11 in the fourth subfigure). Thus, there is still room to improve the regression performance for thermal and hydraulic parts. In future work, other data-driven tactics will be adopted to better deal with the inherent nonlinearity of the data while preserving good computational efficiency of the generated models.

5.7. Challenges and prospects

Several limitations of this work should be noted. First, data-driven models rely heavily on historical data. Generally, they have poor performance on data samples with unfamiliar characteristics (such as

topology changes). Another limitation is that the samples of variables generated by Monte Carlo simulations have certain numerical relationships described by the nonlinear MEF equations, different from measurement data from the real world.

Future work will focus on improving the current approach to accommodate topology re-configurations and training models with more realistic data.

6. Conclusion

This paper proposes a dual-driven approach to generate linear and accurate models for multiple energy flow calculations in an electricity–heat system. The performance of the dual-driven model is validated in two systems: a small grid-connected system and a large islanding system, denoted as Case A and Case B. The simulation of study cases verified that the dual-driven model has remarkable accuracy, with the mean absolute percentage error less than 0.2% among all dependent variables in Case B, significantly lower than that of the mechanism- and data-driven linear models. Because of its high precision and linear characteristics, the dual-driven model can be used as an alternative to the original nonlinear model, thus preventing the occurrence of slow convergence or non-convergence. Furthermore, the dual-driven model outperforms the data-driven model in terms of error distribution, stability of regression, and goodness of fitting, demonstrating the complementarity between mechanism- and data-driven approaches.

Several benefits can be seen from the proposed dual-driven approach. First, the generated linear models facilitate rapid calculations of the probabilistic multiple energy flow analysis, which is necessary to evaluate the influence of the growing uncertainty on the operating state of the integrated energy system. Second, linear models can be used in optimal multiple energy flow calculations to ensure optimization convergence. Third, even if the exact physical parameters and control logic are challenging to obtain, which is common in the integrated energy system, the proposed dual-driven approach can still well function.

The proposed dual-driven linear modeling approach is intended to serve as the foundation for accurate linearized multiple energy flow calculation and optimization.

CRediT authorship contribution statement

Hang Tian: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization, Resources. **Haoran Zhao:** Methodology, Software, Validation, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Chunyang Liu:** Conceptualization, Methodology, Validation, Writing – review & editing, Project administration. **Jian Chen:** Methodology, Software, Writing – review & editing, Project administration. **Qiuwei Wu:** Writing – review & editing, Supervision, Project administration. **Vladimir Terzija:** Writing – review & editing, Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by National Key R&D Program of China under grant 2018YFA0702200.

References

[1] Qin Xin, Sun Hongbin, Shen Xinwei, Guo Ye, Guo Qinglai, Xia Tian. A generalized quasi-dynamic model for electric-heat coupling integrated energy system with distributed energy resources. Appl Energy 2019;251:113270.

[2] Ayele Getnet Tadesse, Haurant Pierrick, Laumert Björn, Lacarriere Bruno. An extended energy hub approach for load flow analysis of highly coupled district energy networks: Illustration with electricity and heating. Appl Energy 2018;212:850–67.

[3] Li Zhengmao, Xu Yan. Optimal coordinated energy dispatch of a multi-energy microgrid in grid-connected and islanded modes. Appl Energy 2018;210:974–86.

[4] Geidl Martin, Andersson Göran. Optimal power flow of multiple energy carriers. IEEE Trans Power Syst 2007;22(1):145–55.

[5] Liu Xuezhi, Mancarella Pierluigi. Modelling, assessment and sankey diagrams of integrated electricity-heat-gas networks in multi-vector district energy systems. Appl Energy 2016;167:336–52.

[6] Zeng Qing, Fang Jiakun, Li Jinghua, Chen Zhe. Steady-state analysis of the integrated natural gas and electric power system with bi-directional energy conversion. Appl Energy 2016;184:1483–92.

[7] Chen Zhang, Liu Jun, Liu Xinglei. GPU accelerated power flow calculation of integrated electricity and heat system with component-oriented modeling of district heating network. Appl Energy 2022;305:117832.

[8] Lei Yunkai, Hou Kai, Wang Yue, Jia Hongjie, Zhang Pei, Mu Yunfei, Jin Xiaolong, Sui Bingyan. A new reliability assessment approach for integrated energy systems: using hierarchical decoupling optimization framework and impact-increment based state enumeration method. Appl Energy 2018;210:1237–50.

[9] Dancker Jonte, Wolter Martin. Improved quasi-steady-state power flow calculation for district heating systems: A coupled Newton-Raphson approach. Appl Energy 2021;116930.

[10] Zhu Mengting, Xu Chengsi, Dong Shufeng, Tang Kunjie, Gu Chenghong. An integrated multi-energy flow calculation method for electricity-gas-thermal integrated energy systems. Protection and Control of Modern Power Systems 2021;6(1):1–12.

[11] Shabanpour-Haghighi Amin, Seifi Ali Reza. An integrated steady-state operation assessment of electrical, natural gas, and district heating networks. IEEE Trans Power Syst 2015;31(5):3636–47.

[12] Martinez-Mares Alberto, Fuerte-Esquivel Claudio R. A unified gas and power flow analysis in natural gas and electricity coupled networks. IEEE Trans Power Syst 2012;27(4):2156–66.

[13] Liu Xuezhi, Wu Jianzhong, Jenkins Nick, Bagdanavicius Audrius. Combined analysis of electricity and heat networks. Appl Energy 2016;162:1238–50.

[14] Sun Qiuye, Dong Qianyu, You Shi, Li Zhibo, Wang Rui. A unified energy flow analysis considering initial guesses in complex multi-energy carrier systems. Energy 2020;213:118812.

[15] Zhang Gang, Zhang Feng, Meng Ke, Zhang Xin, Dong Zhao Yang. A fixed-point based distributed method for energy flow calculation in multi-energy systems. IEEE Trans Sustain Energy 2020;11(4):2567–80.

[16] Fatemi Seyed Masoud, Abedi Sajjad, Gharehpetian GB, Hosseinian Seyed Hossein, Abedi Mehrdad. Introducing a novel DC power flow method with reactive power considerations. IEEE Trans Power Syst 2014;30(6):3012–23.

[17] Yang Zhifang, Zhong Haiwang, Bose Anjan, Zheng Tongxin, Xia Qing, Kang Chongqing. A linearized OPF model with reactive power and voltage magnitude: A pathway to improve the MW-only DC OPF. IEEE Trans Power Syst 2017;33(2):1734–45.

[18] Yang Jingwei, Zhang Ning, Kang Chongqing, Xia Qing. A state-independent linear power flow model with accurate estimation of voltage magnitude. IEEE Trans Power Syst 2016;32(5):3607–17.

[19] Wang Yi, Zhang Ning, Li Hai, Yang Jingwei, Kang Chongqing. Linear three-phase power flow for unbalanced active distribution networks with PV nodes. CSEE J Power Energy Syst 2017;3(3):321–4.

[20] Li Zhigang, Yu Jinyu, Wu QH. Approximate linear power flow using logarithmic transform of voltage magnitudes with reactive power and transmission loss consideration. IEEE Trans Power Syst 2017;33(4):4593–603.

[21] Wang Qi, Li Feng, Tang Yi, Xu Yan. Integrating model-driven and data-driven methods for power system frequency stability assessment and control. IEEE Trans Power Syst 2019;34(6):4557–68.

[22] Liu Yuxiao, Zhang Ning, Wang Yi, Yang Jingwei, Kang Chongqing. Data-driven power flow linearization: A regression approach. IEEE Trans Smart Grid 2018;10(3):2569–80.

[23] Gurusinghe Dinesh Rangana, Rajapakse Athula D. Post-disturbance transient stability status prediction using synchrophasor measurements. IEEE Trans Power Syst 2015;31(5):3656–64.

[24] Tan Yi, Chen Yuanyang, Li Yong, Cao Yijia. Linearizing power flow model: A hybrid physical model-driven and data-driven approach. IEEE Trans Power Syst 2020;35(3):2475–8.

[25] Trias Antonio. The holomorphic embedding load flow method. In: 2012 IEEE power and energy society general meeting. IEEE; 2012, p. 1–8.

[26] Yang Yan, Yang Zhifang, Yu Juan, Zhang Baosen, Zhang Youqiang, Yu Hongxin. Fast calculation of probabilistic power flow: A model-based deep learning approach. IEEE Trans Smart Grid 2019;11(3):2235–44.
[27] Hu Xinyue, Hu Haoji, Verma Saurabh, Zhang Zhi-Li. Physics-guided deep neural networks for power flow analysis. IEEE Trans Power Syst 2020;36(3):2082–92.
[28] Xiang Mingxu, Yu Juan, Yang Zhifang, Yang Yan, Yu Hongxin, He He. Probabilistic power flow with topology changes based on deep neural network. Int J Electr Power Energy Syst 2020;117:105650.
[29] Rosipal Roman, Krämer Nicole. Overview and recent advances in partial least squares. In: Proc. int. conf. subspace latent struct. feature selection. Springer; 2005, p. 34–51.
[30] De Jong Sijmen. SIMPLS: an alternative approach to partial least squares regression. Chemometr Intell Lab Syst 1993;18(3):251–63.
[31] Bhattarai Bishnu P, Paudyal Sumit, Luo Yusheng, Mohanpurkar Manish, Cheung Kwok, Tonkoski Reinaldo, Hovsapian Rob, Myers Kurt S, Zhang Rui, Zhao Power, et al. Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. IET Smart Grid 2019;2(2):141–54.
[32] Liu Yuxiao, Wang Yi, Zhang Ning, Lu Dan, Kang Chongqing. A data-driven approach to linearize power flow equations considering measurement noise. IEEE Trans Smart Grid 2019;11(3):2576–87.
[33] Chen Yanbo, Wu Chao, Qi Junjian. Data-driven power flow method based on exact linear regression equations. J Mod Power Syst Clean Energy 2021.
[34] Liu Xuezhi. Analysis, modelling and operational optimization of district heating systems (PhD thesis), Cardiff University; 2013.
[35] Wei Zhinong, Sun Juan, Ma Zhoujun, Sun Guoqiang, Zang Haixiang, Chen Sheng, Zhang Side, Cheung Kwok W. Chance-constrained coordinated optimization for urban electricity and heat networks. CSEE J Power Energy Syst 2018;4(4):399–407.
[36] Zhou Yizhou, Shahidehpour Mohammad, Wei Zhinong, Li Zhiyi, Sun Guoqiang, Chen Sheng. Distributionally robust co-optimization of energy and reserve for combined distribution networks of power and district heating. IEEE Trans Power Syst 2019;35(3):2388–98.
[37] Zhou Yizhou, Shahidehpour Mohammad, Wei Zhinong, Sun Guoqiang, Chen Sheng. Multistage robust look-ahead unit commitment with probabilistic forecasting in multi-carrier energy systems. IEEE Trans Sustain Energy 2020;12(1):70–82.
[38] Fu Xueqian, Sun Hongbin, Guo Qinglai, Pan Zhaoguang, Xiong Wen, Wang Li. Uncertainty analysis of an integrated energy system based on information theory. Energy 2017;122:649–62.
[39] Pan Zhaoguang, Guo Qinglai, Sun Hongbin. Interactions of district electricity and heating systems considering time-scale characteristics based on quasi-steady multi-energy flow. Appl Energy 2016;167:230–43.
[40] Zhao Hongping. Analysis, modelling and operational optimization of district heating systems (PhD thesis), Technical University of Denmark; 1995.
[41] Hassine Ilyes Ben, Eicker Ursula. Impact of load structure variation and solar thermal energy integration on an existing district heating network. Appl Therm Eng 2013;50(2):1437–46.
[42] Bøhm Benny, Ha Seung-kyu, Kim Won-tae, Kim Bong-kyun, Koljonen Tiina, Larsen Helge V, Lucht Michael, Park Yong-soon, Sipilä Kari, Wigbels Michael, et al. Simple models for operational optimisation. Contract 2002;524110:0010.
[43] Glantz Stanton, Slinker Bryan. Primer of applied regression & analysis of variance. New York: McGraw-Hill, Inc.; 2001.
[44] Draper Norman R, Smith Harry. Applied regression analysis, Vol. 326. John Wiley & Sons; 1998.