Load Data Valuation in Multi-energy Systems: An End-to-end Approach

Yangze Zhou, *Student Member, IEEE*, Qingsong Wen, *Member, IEEE*, Jie Song, *Member, IEEE*, Xueyuan Cui, *Student Member, IEEE*, and Yi Wang, *Member, IEEE*

***Abstract*—Accurate load forecasting serves as the foundation for the flexible operation of multi-energy systems (MES). Multi-energy loads are tightly coupled and exhibit significant uncertainties. Many works focus on enhancing forecasting accuracy by leveraging cross-sector information. However, data owners may not be motivated to share their data unless it leads to substantial benefits. Ensuring a reasonable data valuation can encourage them to share their data willingly. This paper presents an end-to-end framework to quantify multi-energy load data value by integrating forecasting and decision processes. To address optimization problems with integer variables, a two-stage end-to-end model solution is proposed. Moreover, a profit allocation strategy based on contribution to cost savings is investigated to encourage data sharing in MES. The experimental results demonstrate a significant decrease in operation costs, suggesting that the proposed valuation approach more effectively extracts the inherent data value than traditional methods. According to the proposed incentive mechanism, all sectors can benefit from data sharing by improving forecasting accuracy or receiving economic compensation.**

***Index Terms*—Multi-energy systems, data valuation, data sharing, end-to-end modeling, load forecasting.**

I. Introduction

MULTI-energy systems (MES) are considered as a promising approach for enhancing energy efficiency and promoting renewable energy accommodation [1]. MES coordinates different energy sectors, such as power, cooling, heat, and gas, to satisfy multi-energy loads with high overall energy efficiency [2], [3]. Multi-energy load forecasting is the basis for the operations of various energy converters within the MES. For example, in northern China, how much power is generated by combined heat power (CHP) is jointly determined by heat and electricity loads in this area [4].

Since multi-energy loads are deeply coupled, some studies have utilized cross-sector information to improve load forecasting accuracy. In [5], historical load features of different sectors were selected as inputs of the designed parallel architecture. Then, transfer learning is adopted to cope with the issue of data deficiency in MES. [6] introduced multi-task learning with homoscedastic uncertainty to consider the coupling relations, simultaneously achieving better prediction of different sectors. [7] proposed an attention mechanism to enhance the impact of extracted dynamic coupling relationship between different sectors. These results all point to an essential fact: data sharing between different sectors in MES is beneficial for forecasting accuracy enhancement.

However, power, gas, and heat/cooling load data are probably owned by different system operators separately. These data owners tend to prioritize their own economic benefits over social benefits when making decisions [8]. They are generally not motivated to share data, as doing so by sharing cross-sector information does not usually result in substantial advantages for themselves. Only by reasonably valuing the data and fairly allocating the additional profits from data sharing among sectors in MES will they be willing to share their data assets.

Current research about data valuation in energy systems can be primarily divided into two categories but is very limited. The first is to analyze the influence of data on forecasting accuracy. A regression market framework was proposed in [9] to support agents sharing their features with the central agent to train a better regression/forecasting model. The support agents can obtain monetary rewards based on their contribution to improving accuracy. A framework for day-ahead load forecast trading and valuation in an ensemble model was proposed in [10]. Nevertheless, the value of data assets is closely related to the application scenarios and can vary significantly [11]. Thus, the second is to study how forecasting accuracy improvements affect operation costs, leading to the quantification of the economic value of data [12]. [13] utilized Shannon entropy and non-noise ratio as two metrics to measure the quality of photovoltaics-related data and then constructed a reflection from selected metrics to the load forecasting accuracy. On this basis, the economic value of data assets was quantified by simulating a stochastic unit commitment using forecasting results of various accuracy. [14] proposed a data market based on the electricity market with dual price imbalances to measure the value of the renewable energy data from different owners.

Although the economic value of data can be quantified by solving an optimization problem to minimize operation costs, the approach of forecasting-then-optimization (FTO) handles forecasting and decision-making as two separate processes. This separated approach may not accurately capture the actual value of data. In FTO, the forecasting model is trained with traditional loss functions such as mean square errors (MSE). These loss functions treat the positive prediction errors and

The work was supported in part by the Research Grants Council of the Hong Kong SAR under Grant HKU 27203723 and in part by in part by the National Natural Science Foundation of China under Grant 72131001. (*Corresponding authors: Jie Song and Yi Wang*)

Yangze Zhou, Xueyuan Cui, and Yi Wang are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR, China (e-mail: yzzhou@connect.hku.hk, xycui@eee.hku.hk, yiwang@eee.hku.hk).

Qingsong Wen is with the DAMO Academy, Alibaba Group (U.S.) Inc., Bellevue, WA 98004, USA (e-mail: qingsongedu@gmail.com).

Jie Song is with the Department of Industrial Engineering and Management, College of Engineering, Peking University, Beijing, China (email: jie.song@pku.edu.cn).

negative prediction errors equally using a quadratic function, whereas they exert different impacts on costs [15]. Specifically, under-forecasts may result in purchases of expensive peaking power while over-forecasting leads to extra generation capacity to be committed [16]. This fact demonstrates that, in power systems, higher accuracy does not necessarily imply lower costs and there is a discrepancy between minimizing forecasting errors and operation costs [17]. Such discrepancy is more apparent in MES, making it more challenging for the FTO approach to precisely quantify the economic data value. Hence, a more rational approach is to quantify the data value in an end-to-end approach by integrating forecasting and optimization into a single model.

To fill the above research gap, this paper proposes an end-to-end data valuation method within MES. End-to-end refers to combining forecasting and decision-making as a whole. [18] introduced a bilevel programming model that improves both forecasting quality and decision performance by utilizing cost-oriented prediction intervals. [19] proposed an end-to-end approach in probabilistic machine learning for downstream tasks such as electrical grid scheduling and energy storage arbitrage. [20] applied the thought of "smart prediction then optimization" in [21] to construct the electricity price forecasting model focusing on the operation of the energy storage system. End-to-end modeling is an area that has received relatively little attention, let alone its application in value valuation. This paper makes the following contributions:

1) Study multi-energy load data valuation problem in MES which has been rarely touched. Specifically, an end-to-end framework is established to quantify the data value from a cost-oriented perspective so that various sectors in MES can collaborate with each other and achieve cost savings. Compared to existing valuation methods, the proposed framework can fully uncover the data value.
2) Develop a two-stage end-to-end model solution to train the forecasting and decision-making integrated end-to-end model, where the decision-making model is not limited to linear programming or quadratic programming problems but can incorporate integer variables. In this way, the method is applicable to various scenarios with integer variables in MES, such as (dis)charging of energy storage and piecewise linear approximation of nonlinear constraints.
3) Proposes an intensive mechanism to allocate additional profits from data sharing to different sectors in MES according to their contributions. According to the experiment result, all sectors can benefit from data sharing by either improving forecasting accuracy or receiving economic compensation.

The rest of the paper is structured as follows. Section II defines the problem of end-to-end data valuation within MES. Section III given the preliminaries of this work. Section IV elaborates on the end-to-end modeling to integrate forecasting and optimization. Section V illustrates the details of the proposed data valuation framework via the end-to-end approach. Section VI analyzes and visualizes the experiment results. Conclusions and future works can be reached in Section VII.

II. Problem Statement

We are studying the following scenario shown in Fig. 1: In MES with $|\mathcal{N}|$ sectors, an operator is responsible for the scheduling of all equipment within MES based on multi-energy load forecasts $M_n(X_n, w_n)|_{n\in\mathcal{N}}$, submitted by sectors individually, where X_n and w_n denote the input features and parameters of sector n's forecasting model M_n. In our study, we assume that all sectors will report their forecast results honestly, which is commonly satisfied in real-world MES systems. Mathematically, it can be presented as:

$$\min_{z} \; C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}}) \tag{1}$$

where C and z are the cost and decision variables for the scheduling of MES, respectively.

Traditionally, w_n and z in (1) are determined sequentially, i.e., first training forecasting models by sectors and then scheduling MES by the operator. This practice has two potential limitations:

1) Cross-sector data/information has not been shared and fully utilized to enhance forecasting accuracy or reduce operation costs;
2) The forecasting and decision-making processes are treated separately so that data cannot directly serve final decision-making in MES.

To overcome the above limitations, the operator can integrate forecasting model training and MES scheduling to form an end-to-end model:

$$\min_{z;w_n|_{n\in\mathcal{N}}} C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}}) \tag{2}$$

where w_n and z are optimized as a whole; the operation costs is denoted as $C_{\mathcal{N}}$ if all sectors cooperate with the MES operator. The cooperation means the sectors share their data X_n with the operator indirectly.

To encourage sectors to participate in the end-to-end model, the value of the data owned by various sectors should be quantified. In this work, the data value is defined as the additional profits derived from data sharing in the cooperation between sectors and MES operators. Hence, two essential problems need to be solved:

1) How many additional profits $V(\mathcal{N})$ can be derived from data sharing of various sectors in MES ?
2) How to make a fair plan $\{v_1, v_2, \cdots, v_{|\mathcal{N}|}\}$ to allocate the profits $V(\mathcal{N})$ to each sector ?

III. Preliminaries

This section provides preliminaries for the two basic components of data valuations: the MES load forecasting model M_n and the operation model C. We choose LSTM and the model proposed in [22] for M_n and C, respectively. Note that in our proposed framework, M_n can be any forecast model that can be trained by gradient descents, and C is not limited to the modeling approach in [22], it can be any QP or LP problem mixed with integers.

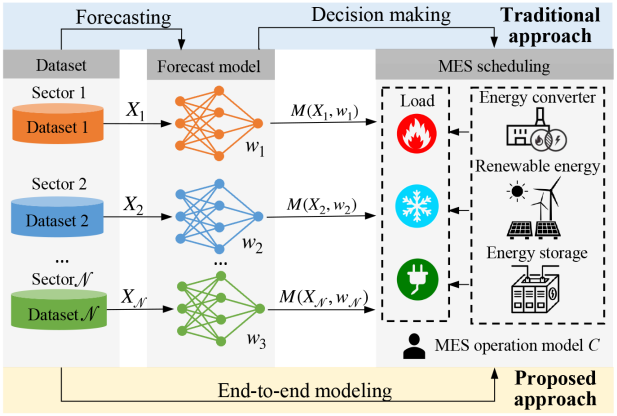

Fig. 1. Sketch of the studied scenario. The blue part shows the traditional approach, where the forecasting process and decision-making process are dealt with sequentially. The yellow part is the proposed end-to-end approach.

A. Forecasting Module M_n

LSTM is a recurrent neural network with several kinds of special gates to capture long-term and temporal dependence [23]. The satisfactory performance of LSTM in load forecasting has been sufficiently proven by existing research [24]–[26]. Therefore, it is introduced as the load forecasting model M_n in this work. An LSTM unit is shown in Fig. 2, There are three special gates in LSTM, which are forgotten gate $\mathcal{F}$, input gate $\mathcal{I}$, and output gate $\mathcal{O}$, respectively. The forget gate determines what information should be discarded from the memory cell (stored in $c_{(t-1)}$). The input gate is designed to determine how much information inputted at time t should be stored in the current memory cell $c_{(t)}$. The output gate controls how much information is worthy of being treated as model output. Its working principles can be formulated as follows:

$$\mathcal{F}(t) = \sigma\left(W_{x,f}^T x(t) + W_{h,f}^T h(t-1) + b_f\right) \tag{3a}$$

$$\mathcal{I}(t) = \sigma\left(W_{x,i}^T x(t) + W_{h,i}^T h(t-1) + b_i\right) \tag{3b}$$

$$\mathcal{O}(t) = \sigma\left(W_{x,o}^T x(t) + W_{h,o}^T h(t-1) + b_o\right) \tag{3c}$$

$$g(t) = tanh\left(W_{x,g}^T x(t) + W_{h,g}^T h(t-1) + b_g\right) \tag{3d}$$

$$c(t) = \mathcal{F}(t) \otimes c(t-1) + \mathcal{I}(t) \otimes g(t) \tag{3e}$$

$$y(t) = h(t) = \mathcal{O}(t) \otimes tanh\left(c(t)\right) \tag{3f}$$

where $x(t)$, $g(t)$, $y(t)$ are inputs features, intermediate state, and output at t, $h(t-1)$ and $c(t-1)$ are hidden state and memory cell propagated from the $t-1$, $W_{x,i}$, $W_{x,f}$, $W_{x,o}$, $W_{x,g}$ are weights related to x, $W_{h,i}$, $W_{h,f}$, $W_{h,o}$, $W_{h,g}$ are weights related to hidden state h, and b_i, b_f, b_o, b_g are corresponding bias terms, σ and $tanh$ are sigmoid and tanh activation function, $\otimes$ denotes the Hadamard product.

In this work, the input features X_n of the forecasting model of different sectors consist of historical load, calendar features (month, weekday), and weather features (temperature and humidity). The output of the forecasting model is the load at 24 timestamps the next day.

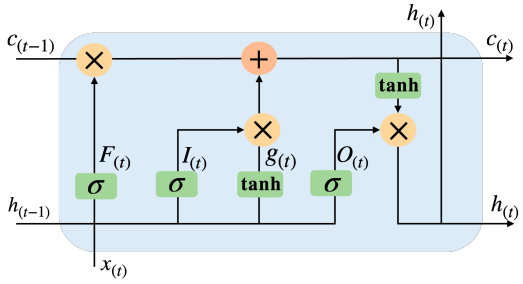

Fig. 2. Sketch of an LSTM unit

B. Decision Making Module C

This work develops the model of MES with the concept of Energy Hub (EH) [27]. EH is defined as a unit that provides the functions of input, output, conversion, and storage of multiple sectors [28]. In this work, variable converter efficiency is considered, which introduces non-linear constraints into EH modeling due to the variable multiplication. The piecewise linear approximation has been applied to depict non-linear input-output relationships linearly. Additionally, this work includes the consideration of energy storage equipment. The comprehensive energy flow equations of the EH are formulated as (4) [22].

$$[I_{\text{EH}}^{\text{T}}, O_{\text{EH}}^{\text{T}}, N_{\text{EH}}^{\text{T}}]V = [V_{\text{in}}^{\text{T}}, V_{\text{out}}^{\text{T}}, \mathbf{0}^{\text{T}}]^{\text{T}} \tag{4}$$

where V_{in} is the input vector, V_{out} is the output vector, V represents energy flows in all branches, I_{EH} is input incidence matrix, O_{EH} is output incidence matrix, matrix N_{EH} combines the nodal energy conversion matrix of all nodes in the EH, $\mathbf{0}$ means the vector with all elements are zero.

In the day-ahead schedule strategy stage, the MES operator schedules the equipment in different sectors according to the day-ahead forecasts. The optimization problem of the day-ahead strategy can be formulated as:

$$\begin{aligned} \min \quad & \sum_{t=1}^{24} F_t \cdot V_{\text{in},t}^{\text{T}} \\ \text{s.t.} \quad & [I_{\text{EH}}^{\text{T}}, O_{\text{EH}}^{\text{T}}, N_{\text{EH}}^{\text{T}}]V_t = [V_{\text{in},t}^{\text{T}}, V_{\text{out},t}^{\text{T}}, \mathbf{0}^{\text{T}}]^{\text{T}} \\ & V_{\text{out},t} = M_n(X_{n,t}, w_n)|_{n\in\mathcal{N}} \end{aligned} \tag{5}$$

where $V_{\text{in},t}$ and F_t denote the vector of input and corresponding prices at time t in the day ahead.

In the intra-day process, the real-time energy loads will likely differ from the forecasting load generally. The goal of the intra-day schedule strategy is to maintain the energy balance of different sectors via temporary scheduling or energy storage charging/discharging with the lowest cost:

$$\begin{aligned} \min \quad & \sum_{t=1}^{24} (\widetilde{F}_t \cdot \widetilde{V}_{\text{in},t}^{\text{T}} + F_{\text{ch}} \cdot Q_{ch,t} + F_{\text{dis}} \cdot Q_{\text{dis},t}) \\ \text{s.t.} \quad & [I_{\text{EH}}^{\text{T}}, O_{\text{EH}}^{\text{T}}, N_{\text{EH}}^{\text{T}}]\widetilde{V}_t = [\widetilde{V}_{\text{in},t}^{\text{T}}, \widetilde{V}_{\text{out},t}^{\text{T}}, \mathbf{0}^{\text{T}}]^{\text{T}} \\ & \hat{M}_t = \widetilde{V}_{\text{out},t} + Q_{\text{dis},t} - Q_{\text{ch},t} \end{aligned} \tag{6}$$

where RD and RU are down/up reserve capacity, $\widetilde{V}_{\text{in},t}$, $\widetilde{V}_{\text{out},t}$ and $\widetilde{V}_t$ denote the input vector, output vector, and corresponding prices at time t in the intra-day, $\hat{M}_t$ are the real-time energy

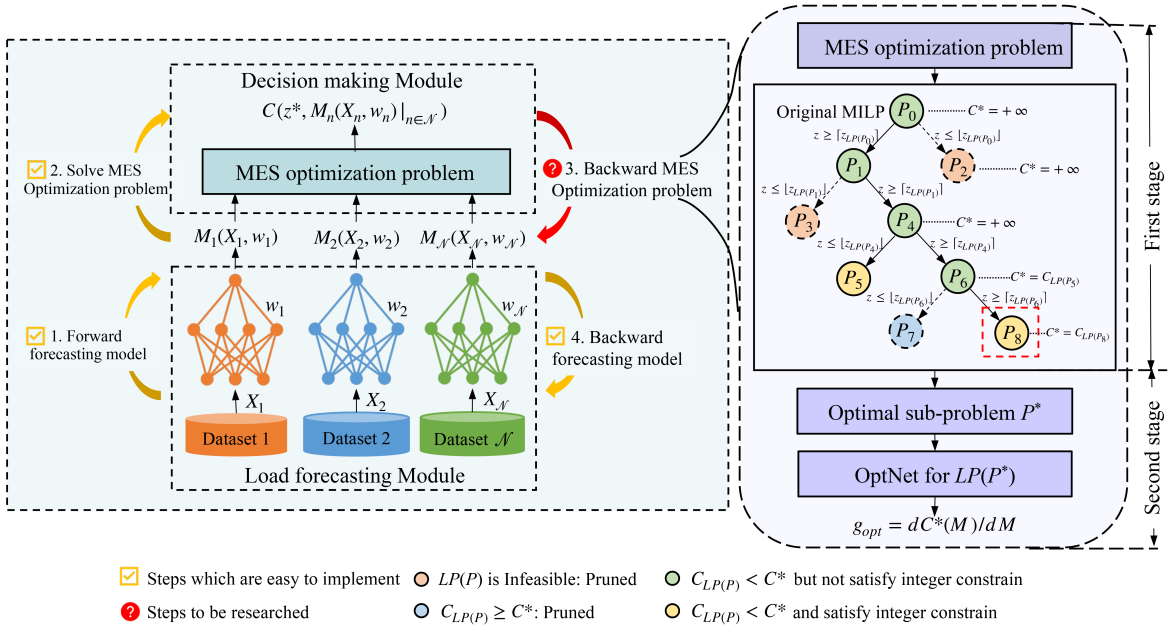

Fig. 3. The sketch of end-to-end modeling. The left-hand part is the four main procedures of end-to-end modeling. The right-hand part is the proposed two-stage end-to-end modeling solution algorithm to integrate forecasting models with an MILP problem.

loads at time t, $Q_{\text{dis},t}$ and $Q_{\text{ch},t}$ are the discharging power and charging power of the energy storage at time t, F_{dis} and F_{ch} are the cost of discharging and charging energy storage.

IV. End-to-end Optimization

This section details the proposed end-to-end model, including modeling and solution algorithm.

A. End-to-end Modeling

End-to-end modeling refers to integrating the training of the local forecasting module M_n and the MES optimization problem C as a whole, as shown in the left-hand part of Fig. 3. An intuitive idea to train the end-to-end model is forward and backward propagation, as used for traditional neural network training. Thus, we have four steps for the model training:

1) Forward load forecasting models: Each sector inputs the feature X_n to the forecasting model and then outputs load forecasts $M_n(X_n, w_n)$.
2) Solve MES optimization problem: The operator collects all sectors' forecasts $M = M_n(X_n, w_n)|_{n\in\mathcal{N}}$ as the inputs of the MES optimization problem to obtain the optimal schedule z^* and operation costs $C(z^*, M_n(X_n, w_n)|_{n\in\mathcal{N}})$.
3) Backward MES optimization problem: The operator gives back information about the relationship (i.e., gradients) between operation costs and forecasts to the load forecasting models.
4) Backward load forecasting models: Each sector receives the gradients from the operator and updates its model parameter w_n by backwarding the forecasting model.

Steps 1, 2, and 4, marked by yellow, can be easily conducted as usual, while the main difficulty is step 3, marked by red: how to obtain the gradient of cost C over load forecasts M.

The MES optimization problem (6) can be abstracted as follows:

$$\begin{aligned} \min \quad & C(z, M), \\ \text{s.t.} \quad & f(z, M) \leq 0, h(z, M) = 0. \end{aligned} \tag{7}$$

where C, z, and M denote the objective function, decision variables, and multi-energy load forecasts, respectively; f and h denote inequality and equality constraints, respectively.

For given M, the optimal operation costs is:

$$C^*(M) = \inf\{C(z, M) \mid f(z, M) \leq 0, h(z, M) = 0\} \tag{8}$$

The optimal decision z^* is dependent on the forecasts M:

$$z^*(M) = \{z \mid C(z, M) = C^*(M), f(z, M) \leq 0, h(z, M) = 0\} \tag{9}$$

To train the end-to-end model that minimizes the operation costs $C^*(M)$, it is necessary to calculate the gradient of $C^*(M)$ over M. For a quadratic programming problem, [29] introduced an optimization differentiable neural network (OptNet) to compute the gradient indirectly by solving equations from the KKT condition of a Lagrangian function.

The Lagrangian function of the optimization problem (7) is:

$$\mathcal{L}(z, \lambda, \mu, M) = C(z, M) + \lambda^T f(z, M) + \mu^T h(z, M), \tag{10}$$

where λ and μ are dual variables. The KKT condition of $\mathcal{L}(z, \lambda, \mu, M)$ is:

$$\begin{cases} f(z, M) \leq 0, \\ h(z, M) = 0, \\ \lambda_i \geq 0, \quad i \in \{1, 2, \ldots, q\} \\ \lambda_i f_i(z, M) = 0, \quad i \in \{1, 2, \ldots, q\} \\ \nabla_z \mathcal{L}(z, \lambda, \mu, M) = 0. \end{cases} \tag{11}$$

where q is the number of inequality constraints.

An implicit function $G(\tilde{z}, M)$ can be derived from (11) if we denoted $[z, \lambda, \mu]$ as $\tilde{z}$.

$$G(\tilde{z}, M) = G(z, \lambda, \mu, M) = \begin{bmatrix} \nabla_z \mathcal{L}(z, \lambda, \mu, M) \\ \lambda f(z, M) \\ h(z, M) \end{bmatrix} = 0. \tag{12}$$

The gradient of $\tilde{z}$ over M can be obtained by the differential principle of implicit function:

$$\frac{d\tilde{z}}{dM} = G_{\tilde{z}}^{-1}(\tilde{z}, M) G_M(\tilde{z}, M), \tag{13}$$

where

$$\frac{d\tilde{z}}{dM} = \begin{bmatrix} \frac{\partial z}{\partial M} & \frac{\partial \lambda}{\partial M} & \frac{\partial \mu}{\partial M} \end{bmatrix}^T$$

$$G_{\tilde{z}}(\tilde{z}, M) = \begin{bmatrix} \frac{\partial \nabla_z \mathcal{L}(z,\lambda,\mu,M)}{\partial z} & \frac{\partial \nabla_z \mathcal{L}(z,\lambda,\mu,M)}{\partial \lambda} & \frac{\partial \nabla_z \mathcal{L}(z,\lambda,\mu,M)}{\partial \mu} \\ \frac{\partial \lambda f(z,M)}{\partial z} & \frac{\partial \lambda f(z,M)}{\partial \lambda} & \frac{\partial \lambda f(z,M)}{\partial \mu} \\ \frac{\partial h(z,M)}{\partial z} & \frac{\partial h(z,M)}{\partial \lambda} & \frac{\partial h(z,M)}{\partial \mu} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial \nabla_z \mathcal{L}(z,\lambda,\mu,M)}{\partial z} & (\frac{\partial f(z,M)}{\partial z})^T & (\frac{\partial h(z,M)}{\partial z})^T \\ \mathrm{diag}(\lambda)\frac{\partial f(z,M)}{\partial z} & \mathrm{diag}(f(z,M)) & 0 \\ \frac{\partial h(z,M)}{\partial z} & 0 & 0 \end{bmatrix}$$

$$G_M(\tilde{z}, M) = \begin{bmatrix} \frac{\partial \nabla_z \mathcal{L}(z,\lambda,\mu,M)}{\partial z} \\ \mathrm{diag}(\lambda)\frac{\partial f(z,M)}{\partial z} \\ \frac{\partial h(z,M)}{\partial z} \end{bmatrix} \tag{14}$$

The gradient of $C^*(M)$ over M can be divided as follows according to the chain principle:

$$g_{\text{opt}} = \frac{dC^*(M)}{dM} = \frac{dC^*(M)}{dz} \frac{dz}{dM}. \tag{15}$$

where $dC^*(M)/dz$ can be easily calculated based on the explicit function $C(z, M)$; dz/dM can be computed using (13). Therefore, the whole forward and backward propagation process for the end-to-end model training in Fig. 3 can be successfully implemented.

Fig. 4 illustrates the forward and backward of the end-to-end model training where the optimization consists of day-ahead and intra-day schedules. In forward propagating, the OptNet, which represents the day-ahead scheduling, takes the forecasts $M_n(X_n, w_n)|_{n\in\mathcal{N}}$ as input and gives the decision variable z_{ahead} by minimizing the objective in (5). Then, the decision variables z_{ahead} and the actual load $\hat{M}$ are fed to the second OptNet, and it outputs the intra-day scheduling cost C_{intra} as shown in as (6). It is noted that the actual load $\hat{M}$ guides the model training by involving intra-day scheduling rather than calculating the MSE between forecasts and $\hat{M}$. In the backward propagating, the second OptNet transfers the gradient of C_{intra} over z_{ahead} to the OptNet for day-ahead scheduling. This gradient will be adopted to calculate the g_{opt} for the fine-tuning of forecasting models.

The OptNet was originally designed for linear or quadratic programming problems. However, the MES optimization problem will inevitably introduce integer variables if we want to model battery status, nonlinear efficiencies, etc. [30], which is usually modeled as an MILP problem. OptNet is not applicable in this case. The next subsection will explain how to solve the end-to-end model when the forecasting models are integrated with an MILP problem.

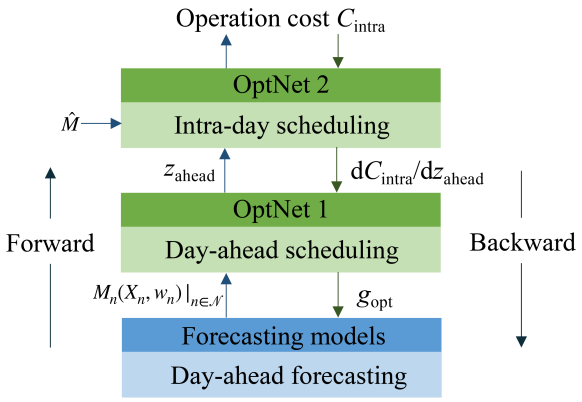

Fig. 4. The forward and backward of the end-to-end modeling where the optimization consists of day-ahead and intra-day schedules.

B. Solution Algorithm

When the operation of the MES is modeled as an MILP problem, the integer variables introduce non-convexity. This makes it challenging to solve the end-to-end modeling by directly utilizing the original OptNet. This makes it challenging to solve the end-to-end modeling by directly utilizing the original OptNet. To address this issue, this study proposes a two-stage end-to-end model solution method, as shown in the right-hand part of Fig. 3.

To clearly illustrate the proposed algorithm, some variables need to be defined. For an MILP problem P, its linear relaxation is represented as $LP(P)$. The optimal cost and decision variables of its linear relaxation, $LP(P)$, are denoted as $C_{LP(P)}$ and $z_{LP(P)}$, respectively. Temporary variables C^*, z^*, and P^*, are defined to store the temporary search results in the solution process. The notation $\lceil\cdot\rceil$ represents the smallest integer greater than or equal to the specified number, while $\lfloor\cdot\rfloor$ represents the largest integer less than or equal to the specified number. The expression $P \wedge f(z)$ refers to a new optimization problem based on the constraints of P and an additional constraint $f(z)$.

Algorithm 1 details the proposed method, which includes two main stages: optimal sub-problem selection and optimal sub-problem backward propagation. In the first stage, the branch and bound method is utilized to select the optimal sub-problem of the original MILP problem. Firstly, a search tree T is initialized with the original MILP problem $P_0(z, M)$, and each node of T is its sub-problem. Then, we move to the branch and bound phases. In the bound phase, the child nodes are solved to obtain the upper and lower bounds of operation costs [18]. Specifically, a sub-problem P is popped from the search tree T unless it is empty, and $LP(P)$ is solved. If $LP(P)$ is unfeasible or $C_{LP(P)}$ is larger than C^*, it ought to be pruned; otherwise, the integer constraints should

be checked. If the $z_{LP(P)}$ satisfies the integer constraints, temporary variables should be updated. If a certain variable z_j violates the integer constraint, it turns to the branch phase. In the branch phase, P will be divided into child nodes by branching on an integer variable [31]. The feasible space of z_j will be divided into two parts and generate two constraints. These two new constraints will be added to the constrain of P to create two new sub-problems, P_l and P_r, which will be added to the search tree T. The branch and bound phases will continue until no child nodes have been fathomed yet. The first stage will return the optimal operation costs $C^*(M)$, decision variable z^* and sub-problem P^* of $P_0(z, M)$.

In the second stage, the backward propagation is implemented for the optimal sub-problems $LP(P^*)$. An OptNet is first developed with the $LP(P^*)$. On this basis, the gradient of $C^*(M)$ over the M, i.e., g_{opt} can be computed by back-propagating through the OptNet.

Discussion: Another intuitive idea to address the issue of non-differentiability caused by integer variables is to incorporate OptNet into the branch and bound search process, i.e., the OptNet-embedded branch and bound method. This solution requires a new temporary variable g^* to store the temporary gradient calculation result and OptNet is embedded in the bound phase. Specifically, for each child node that satisfies $C_{LP(P)} < C^*$ and integer constrain (yellow node in Fig. 3), we need to construct an OptNet for $LP(P)$ and backward it to update g^* accordingly in line 8 of Algorithm 1.

Remark IV.1. *The gradient of $C^*(M)$ over M derived by the OptNet-embedded branch and bound solution method is the same as the two-stage solution approach.*

Proof. The incorporation and backpropagation of OptNet have no impact on the bound and branch processes. As a result, the OptNet-embedded branch and bound method will select the same optimal sub-problem as the first stage of the proposed method. It is evident that the gradient of the objective function for the same sub-problem over M remains unchanged. □

Given that the proposed two-stage end-to-end model solution enables the gradient calculation with less computational complexity and storage requirements, we have opted for it for our end-to-end model training.

Algorithm 1: Two-stage End-to-end Model Solution

Input : Original MILP problem $P_0(z, M)$
Output: The gradient of $C^*(M)$ over M

1 **Optimal sub-problem selection:**
2 Initialize $T := [P_0(z, M)]$, $C^* := +\infty$, $z^* = \varnothing$, $P^* = \varnothing$
3 **while** $T \neq \varnothing$ **do**
4 Pop P from T;
// Bound Phase
5 $C_{LP(P)}, z_{LP(P)}$ =Solve $LP(P)$;
6 **if** $LP(P)$ *is feasible* ***and*** $C_{LP(P)} < C^*$ **then**
7 **if** $z_{LP(P)}$ *satisfy the integer constraints* **then**
8 $C^* = C_{LP(P)}$, $z^* = z_{LP(P)}$, $P^* = P$;
9 **else**
// Branch Phase
10 $P_l = P \wedge (z_j \leq \lfloor z_{LP(P),j} \rfloor)$;
11 $P_r = P \wedge (z_j \geq \lceil z_{LP(P),j} \rceil)$;
12 $T \cup \{P_l, P_r\}$;
13 **Return** C^*, P^*, **and** z^*
14 **Optimal sub-problem backward propagation:**
15 Construct OptNet for $LP(P^*)$;
16 g_{opt} = Backward(OptNet);
17 **Return** g_{opt} ;

V. Data Valuation Framework

This section elucidates the proposed data valuation framework based on the end-to-end model, including additional profit quantification and allocation.

A. Additional Profit Quantification

The end-to-end model enables various sectors to achieve data indirect data sharing: cross-sector information is first integrated into the optimization problem of the MES schedule; this integrated information is then transmitted back to sectors in MES by the gradient g_{opt}. Compared to traditional FTO methods, the indirect cross-sector information sharing in the cooperation between sectors and the MES operator helps to decrease operation costs $C^*(M)$. The reduced operation costs can be regarded as the additional profits derived from the data sharing. The additional profits $V(\mathcal{N})$ is defined as:

$$V(\mathcal{N}) = C_{\mathcal{N}} - C_{\varnothing} \tag{16}$$

where $C_{\mathcal{N}}$ is the operation costs if all sectors cooperate with the MES operator in the end-to-end model, and $C_{\varnothing}$ is the operation costs of the traditional FTO approach.

The proposed method for quantifying the additional profits is illustrated in Algorithm 2, which consists of four steps.

1) Each sector $n \in \mathcal{N}$ utilizes its own data to develop the basic forecasting model M_n. This step can be used for the next step calculation and also provides good initial parameters for end-to-end model training. To prevent overfitting of the basic model, early stopping is introduced to terminate the training process. The model training will stop when the forecasting error on the validation dataset does not decrease for a certain number of epochs.
2) The operation costs of the traditional FTO approach $C_{\varnothing}$ are computed with the forecasts provided by the basic forecasting model M_n.
3) Operator integrates the forecasting model with the MES optimization problem for end-to-end model training. Firstly, the MES operator receives the forecasts M and selects the optimal sub-problems P^* of the original MILP issue $P(z, M)$. Secondly, the MES operator develops the OptNet for the $LP(P^*)$ and acquires the gradient of C^* over M by backward propagating the OptNet. Thirdly, sectors in MES receive the gradient of the operation

Algorithm 2: Additioanl Profit Quantification

Input : Sectors set $\mathcal{N}$, Traditional loss function L_{MSE}, Maximum basic model training epoch E_1, Maximum end-to-end model training epoch E_2, MILP optimization problem $P(\cdot,\cdot)$

Output: Value of data owned by $\mathcal{N}$.

1 **Basic model development:**
2 **for** *each sector* $n \in \mathcal{N}$ **do**
3 Random initialize parameters $w_n|_{n\in\mathcal{N}}$
4 **for** $k \in [0, E_1]$ **do**
5 $M_n = M_n(X_n; w_n^{(k)})$
6 $g_n = \text{Backward}(L_{MSE}(M_n, \hat{M}_n))$
7 $w_n^{(k+1)} = w_n^{(k)} - lr \cdot g_n$
8 **Return** $M_n|_{n\in\mathcal{N}}$
9 **End-to-end data valuation:**
10 $C_\varnothing$ **Calculation:**
11 $C_\varnothing = \min_z C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}})$
12 **End-to-end modeling:**
13 **for** $k \in [0, E_2]$ **do**
14 $M = M_n(X_n, w_n)|_{n\in\mathcal{N}}$
15 $P^* =$ Optimal sub-problem of $P(z, M)$
16 Construct OptNet for $LP(P^*)$
17 **for** *sector* $n \in \mathcal{N}$ **do**
18 $g_{\text{opt},n} = \text{Backward(OptNet)}$
19 $g_n = \text{Backward}(g_{\text{opt},n})$
20 $w_n^{(k+1)} = w_n^{(k)} - lr \cdot g_n$
21 $C_\mathcal{N}$ **Calculation:**
22 $C_\mathcal{N} = \min_z C(z, M_n(X_n, w_n)|_{n\in\mathcal{N}})$
23 **Additioanl profits quantification:**
24 $V(\mathcal{N}) = C_\mathcal{N} - C_\varnothing$
25 **Return** $V(\mathcal{N})$

costs $C^*(M)$ over their forecasts $M_n(X_n, w_n)$, which is denoted as $g_{\text{opt},n}$, and update their forecasting model w_n. These three steps will continue until convergences or reach the maximum training epoch.

4) Operator forward propagates the end-to-end model to calculates the operation costs $C_\mathcal{N}$ and then quantifies the additional profits using (16).

After quantifying the additional profits, it is also essential to allocate them fairly among sectors to encourage their cooperation with the MES operator.

B. Additional Profit Allocation

1) Shapley Value: Energy sectors are playing a cooperative game to reduce the total operational cost. Thus, we design a Shapley value-based incentive mechanism for additional profit allocation. Shapley value has been widely adopted to measure the members' contributions to the collaboration earning [32] because of various game theoretical properties [33]. However, the allocation of Shapley value may be negative, which makes some sectors reluctant to join the end-to-end model [9]. The zero-Shapley value [34] is considered in this work and the profit allocated to section n is:

$$v_n = \sum_{S\subseteq\mathcal{N}\setminus\{n\}} \frac{|S|!(\mathcal{N}-|S|-1)!}{|\mathcal{N}|!}[V(S\cup\{n\}) - V(S)]^+ \tag{17}$$

where $V(S)$ is the value of the cooperation formed by sectors union S, $V(S\cup\{n\})$ is the value of the cooperation formed by sectors union $S\cup\{n\}$, $[\cdot]^+ = \max\{0,\cdot\}$.

However, the zero-shapley value does not satisfy the budget balance property: the sum of revenues is equal to the sum of payments. So we need to adopt a normalized function Γ to allocate the profits according to $v_n, n\in\mathcal{N}$.

$$\Gamma(v_n) = \frac{v_n}{\sum_{i\in\mathcal{N}} v_i}(V(\mathcal{N}) - V(\varnothing)) \tag{18}$$

Fig. 5 illustrates the basic idea of Shapley value-based profit allocation with three sectors in MES. The profits allocated to sector n are measured by calculating the difference in profit before and after the sector participates in cooperation with different sector combinations [35]. Therefore, we need to consider the situation where some sectors within the MES do not participate in the end-to-end modeling, which is referred to as "partially integrated" cooperation.

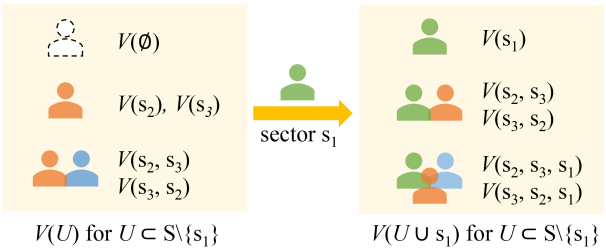

Fig. 5. Profits allocation strategy with three sectors.

2) Value of "Partially Integrated" Cooperation: This subsection first describes the scenario of "partially integrated" cooperation and then quantifies its value.

When only the sectors in U cooperate with the MES operator, the sectors in $\mathcal{N}\setminus U$ will remain their model parameters unchanged (denoted as $\overline{w}_n$) and only submit final forecasts $M_n(X_n, \overline{w}_n)|_{n\in\mathcal{N}\setminus U}$ to the operator. This scenario is described as:

$$\min_{z;w_n|_{n\in U}} C(z, M_n(X_n, w_n)|_{n\in U}, M_n(X_n, \overline{w}_n)|_{n\in\mathcal{N}\setminus U}) \tag{19}$$

where U denotes the set of sectors that cooperate with the operator; C_U denotes the operation costs of the "partially integrated" end-to-end model. When U equals $\varnothing$, this scenario degrades to the traditional FTO approach.

Algorithm 2 can also be applied to scenarios where only parts of sectors are willing to cooperate with the MES operator. After the basic model development procedure, the parameters of models $\overline{w}_n$ belonging to $\mathcal{N}\setminus U$ are fixed. At the end-to-end modeling steps, only the sectors in U update their model parameters $w_n|_{n\in U}$ with the gradient given by the operator (lines 17 to 20). Once the model updating of sectors in U is

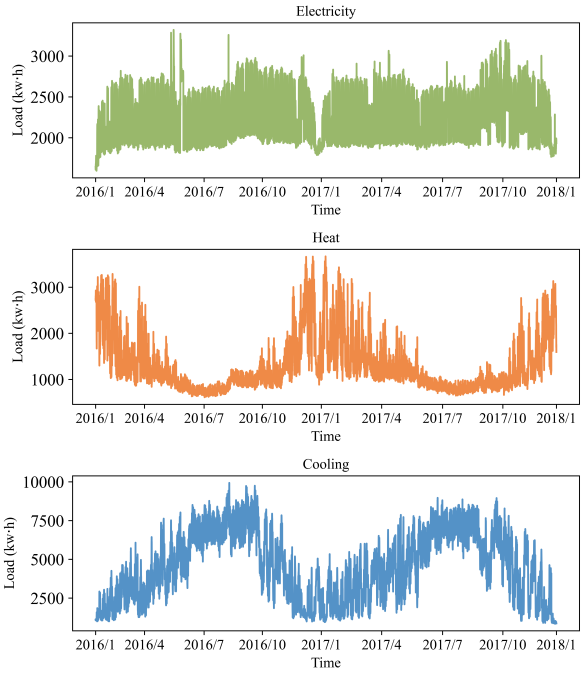

Fig. 6. Multi-energy load profiles in MES in 2016 and 2017.

completed, the operation costs C_U will be calculated with (19). The additional profits derived by the "partially integrated" end-to-end model are regarded as the value of the cooperation formed by sectors in U:

$$V(U) = C_U - C_{\varnothing} \tag{20}$$

where C_U is the operation costs if only sectors in U cooperate with the MES operator.

VI. Case Studies

This section conducts comprehensive experiments to examine the end-to-end model and showcase how to quantify the value of the multi-energy load data.

A. Experimental Setups

The case studies are conducted on a public dataset called Building Data Genome Project 2 [36]. Specifically, the Austin station dataset is employed, which includes hourly load data for electricity, heat, and cooling collected at the University of Texas. The multi-energy loads are shown in Fig. 6. The load data in 2016 and 2017 are used as the training and testing datasets, respectively.

Fig. 7 shows the studied MES, including combined cooling, heat, and power (CCHP), gas boiler, electric boiler, electric refrigerator, and different storages. The turbine used in the studied MES is a backpressure turbine, which means that the generated electricity and heat are in a certain proportion. The inputs of this MES include electricity from the external power grid and gas purchased from the gas station. The outputs comprise electricity, cooling, and heating to satisfy the demands.

TABLE I
The details of the forecasting model M

Parameters	Value	Parameters	Value
Hidden size	32	Learning rate	0.001
Number of layer	1	Maximum epochs	50
Number of direction	1	Early Stop epochs	5

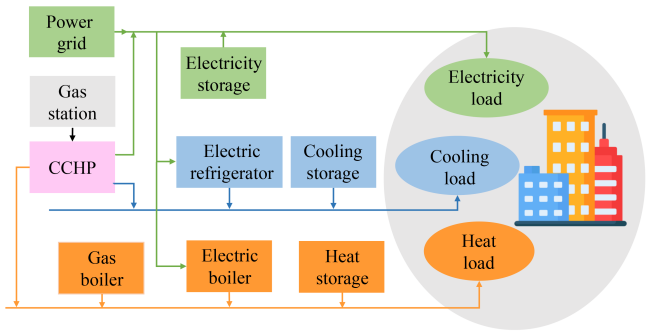

Fig. 7. The structure of the studied MES.

The MES operation includes day-ahead scheduling and intra-day adjustment. The operator first schedules the energy converters based on multi-energy load forecasts in the day ahead. For the electricity sector, the electricity purchased from the power grid and generated from CCHP will be utilized to meet the electricity demands of residential areas as well as power electric boilers and refrigerators (green line in Fig. 7). For the heat sector, the heat produced by gas boilers, electric boilers, and CCHP will be utilized to meet the heat demand (orange line). The cooling sector will utilize both CCHP and electric refrigerators to meet the cooling load (blue line). During daily operations, due to the potential gap between actual load and load forecast, operators have three ways to maintain supply-demand balance: (1) purchase electricity from the grid temporarily, (2) adjust the output of energy converters within backup capacity constraints, and (3) utilize of energy storage for energy charging and discharging.

The experiment in this work is conducted in a virtual environment with Python 3.8.13 and Pytorch 2.1.2. The computing machine adopted in this work is an Intel Xeon W-3335 CPU and an NVIDIA GTX 3080 Ti GPU. The details of the forecasting model M_n are shown in Table I.

B. End-to-end Modeling Results

This subsection depicts the end-to-end modeling results, including forecasting performance and operation costs.

Mean absolute error (MAE), root mean square error (RMSE), and mean average percent error (MAPE) are used to evaluate the forecasting performance:

$$\text{MAE} = \frac{1}{|D|}\sum_{i\in D}|M_i - \hat{M}_i| \tag{21}$$

$$\text{RMSE} = \sqrt{\frac{1}{|D|}\sum_{i\in D}(M_i - \hat{M}_i)^2} \tag{22}$$

TABLE II
LOAD FORECASTING PERFORMANCE OF THREE SECTORS IN MES

	Model	MAE	RMSE	MAPE
Electricity sector	Benchmark	82.592	113.102	3.565
	End-to-end	82.383	112.951	3.562
	Accuracy Variation	0.25%	0.13%	0.00%
Heat sector	Benchmark	120.967	178.530	9.074
	End-to-end	121.331	181.348	9.216
	Accuracy Variation	-0.30%	-1.58%	-0.14 %
Cooling sector	Benchmark	487.406	626.781	12.895
	End-to-end	477.677	615.956	12.799
	Accuracy Variation	2.00%	1.73%	0.10%

$$\text{MAPE} = \frac{100\%}{|D|} \sum_{i \in D} \left| \frac{M_i - \hat{M}_i}{\hat{M}_i} \right| \tag{23}$$

where M_i, $\hat{M}_i$, and D denote the i^{th} load forecasts, the i^{th} actual loads, the index set of the dataset, respectively.

The load forecasting performance of the three sectors is given in Table II, where the benchmark refers to the forecasting models that are separately trained with MSE. The accuracy of the electricity and cooling sector measured by MAE, RMSE, and MAPE are similar to the benchmarks. The accuracy of heat load prediction has experienced a slight decrease.

In the absence of a significant impact on accuracy, we will focus on MES operation costs. Table III shows the average daily operation costs in 2017. FTO refers to the average daily operation costs when M is given by the models trained with MSE separately; Ideal corresponds to the operation costs when the day-ahead forecasts for M are perfectly accurate. The operation costs during the summer are higher compared to other seasons due to the smaller heat load and higher cooling load. In this situation, the output of CCHP is determined by the heat load, which requires a significant amount of electricity to meet the cooling demand, leading to increased operation costs.

The additional operation costs of FTO and end-to-end modeling compared to ideal costs are illustrated in Fig. 8. The black point in the graph represents the monthly operation cost savings compared to FTO achieved by end-to-end modeling. The average daily costs of the end-to-end modeling for each month in 2017 significantly decreased compared to FTO. The additional operation cost of the end-to-end model has been significantly reduced, especially in Sep. and Oct. The annual total operation costs in the ideal case are 31012.06 kCNY. The values for the end-to-end model and FTO are 31294.04 kCNY and 31418.71 kCNY, respectively. Compared with the ideal operation costs, the forecasting errors result in 1.31% additional operation cost when models are trained with MSE. In comparison, the forecasting errors just lead to 0.91% additional operation cost when all sectors are enrolled in end-to-end modeling. This indicates that the operation achieves a 0.40% reduction, resulting in annual cost savings of 124.66 kCNY. These findings highlight the significance of indirect data sharing between sectors.

The variation of the daily average costs of the training and testing dataset in the end-to-end modeling process is shown in Fig. 9. It shows the model can reach convergence in just two epochs after the basic model development procedure, indicating that the proposed method possesses favorable convergence properties.

TABLE III
DAILY OPERATION COSTS FOR 12 MONTHS IN 2017 (KCNY)

	FTO	End-to-end	Ideal	Improvement
Jan.	88.517	88.115	86.808	0.454 %
Feb.	87.500	87.064	85.730	0.498 %
Mar.	86.687	86.294	85.174	0.453 %
Apr.	88.015	87.579	86.170	0.495 %
May	88.282	87.873	86.338	0.463 %
Jun.	92.834	92.442	91.786	0.423 %
Jul.	95.141	94.821	94.402	0.336 %
Aug.	93.729	93.373	92.961	0.379 %
Sep.	92.823	92.477	92.375	0.373 %
Oct.	87.576	87.301	87.197	0.314 %
Nov.	86.505	86.231	85.681	0.316 %
Dec.	85.328	85.140	84.527	0.220 %

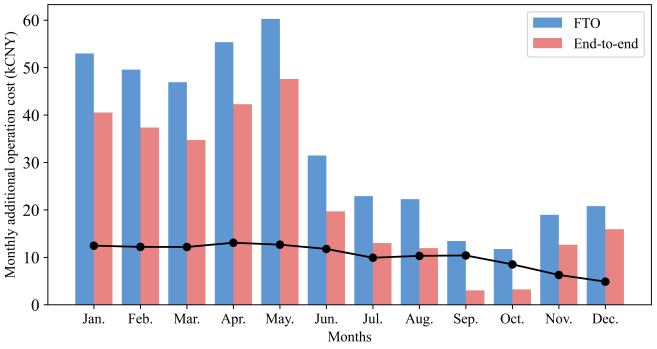

Fig. 8. Monthly additional operation cost of FTO and end-to-end model compared to ideal cost (kCNY)

C. Data Valuation Results

Fig. 10 shows the additional profits from indirect data sharing in cooperation combination. The electricity, heat, and cooling sectors in cooperation are denoted as *e*, *h*, and *c* for the convenience of illustration. The red bar, green bar, and black point denote the $V(S)$, $V(S \cup \{n\})$, and $V(S \cup \{n\}) - V(S)$ in (17), respectively. In (**a**), the enrollment of the electricity sector does not contribute significantly to additional profits. There are two possible reasons: (1) the forecasting accuracy of the electricity sector is relatively high, and the variation of forecasting accuracy within the reserve capacity in the electricity sector has little effect on operation costs. (2) The deviation of the electricity price in intra-day and day-ahead is relatively small, and the additional operation costs resulting from electricity forecasting errors are relatively small. In (**b**), it is evident that the earned additional profits when the heat sector cooperates with the MES operator can be markedly improved. When the heat load is relatively small, the output of the CCHP unit will be determined by the heat load. As shown in Fig. 11, the distribution of forecasting errors moves in the direction of over-forecasting. Over-forecasting the heat load prediction can effectively increase the CCHP output and thus achieve a significant cost reduction. In (**c**), the enrollments of the cooling sector can slightly decrease the operation costs,

TABLE IV
PROFIT ALLOCATION RESULT (KCNY)

Combinations		e, h, c	e, h	e, c	h, c	e	h	c	$\varnothing$
Operation costs $C_{(\cdot)}$		31294.04	31291.83	31311.15	31403.95	31412.30	31314.79	31410.94	31418.71
Data valuation $V(\cdot)$		124.66	126.87	107.56	14.76	6.40	103.92	7.77	0
Contributions	Electricity sector	$\Gamma(\frac{[V(e,h,c)-V(h,c)]^+ + \frac{1}{2}[V(e,h)-V(h)]^+ + \frac{1}{2}[V(e,c)-V(c)]^+ + [V(e)-V(\varnothing)]^+}{3}) = 3.89$							
	Heat sector	$\Gamma(\frac{[V(e,h,c)-V(e,c)]^+ + \frac{1}{2}[V(h,c)-V(c)]^+ + \frac{1}{2}[V(e,h)-V(e)]^+ + [V(h)-V(\varnothing)]^+}{3}) = 107.35$							
	Cooling sector	$\Gamma(\frac{[V(e,h,c)-V(e,h)]^+ + \frac{1}{2}[V(e,c)-V(e)]^+ + \frac{1}{2}[V(h,c)-V(h)]^+ + [V(c)-V(\varnothing)]^+}{3}) = 13.43$							

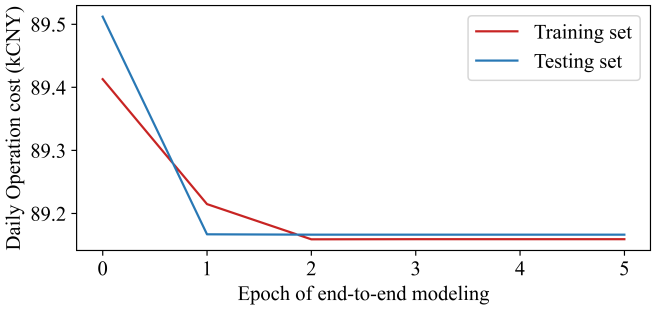

Fig. 9. Daily average cost of the training/testing dataset in the end-to-end modeling process

especially when the heat sector cooperates with the MES operator at the same time.

Table IV lists the detailed profit allocation result and calculation method. The total data value of three sectors $V(e,h,c)$ is 124.66 kCNY ($C_{\{e,h,c\}} - C_{\varnothing}$). According to (17), the data valuations of the electricity, heat, and cooling sectors are 3.89 kCNY, 107.35 kCNY, and 13.43 kCNY, respectively.

The valuation proposed in this paper exhibits significant advantages in terms of both allocation fairness and incentive effect. From the perspective of allocation fairness, the heat sector's enrollment significantly reduces the operation costs, as shown in Fig. 10, justifying the allocation of the largest share of profits to it. As for the electricity sector, it has made the least contribution, so it should share the least share of profits. From the perspective of intensive effect, the electricity and cooling sectors not only achieved improvements in model accuracy but also acquired modest profits through their participation in the end-to-end model. As for the heat sector, the degradation of the forecasting accuracy caused by enrolling the end-to-end model is compensated with a reasonable economic reward.

D. Discussion on Real-world Implmentation

Currently, the research and implementation of data markets, especially within energy systems, are relatively limited. When applying the proposed framework to real-world energy systems, several concerns need to be addressed.

1) The signing of data sharing contract: Before the data owners share their data, data owners must sign a transaction agreement with the MES operator. This agreement serves as a legal contract outlining the data-sharing process's terms, including data usage rights, responsibilities, etc. This kind of contract can be signed over a relatively long time, such as several years. There are two important items in the sharing contract, which are the transmission of price signals and the exit mechanism. As for the transmission of price signals, similar to the electricity market, it can be sent to data owners in real time or daily. It mainly depends on the training patterns (batch training or online learning) and the forecasting step length. Due to the forecasting model giving the day-ahead forecasting in this work, the settlements for each transaction are carried out daily. As for the exit mechanism, when the data owner finds that they cannot obtain profits from data sharing, they can make choices, such as exiting the data market. In our proposed framework, the exit of certain energy sectors won't influence the operation of the MES. However, the exit mechanism is indeed important and should be specified in the contract. It is necessary to determine whether compensation is required if a party decides to exit before the end of the contract.

2) Cost of data sharing: When deciding whether to participate in data sharing, data owners need to consider not only the potential benefits of sharing but also the associated costs. These costs may include expenses related to data processing, storage, transmission, and privacy protection. In this work, our primary focus is on quantifying and allocating the additional profits generated by data sharing in MES. If the cost of sharing becomes significant, the proposed framework can be easily adjusted to address this situation by deducting the costs from the additional profits, either as a fixed amount or a fixed percentage.

3) Platform Support: To support the real-world applications of the proposed data valuation method, well-developed software or platforms are necessary and achievable. In the software or platforms, the contents of the data-sharing agreements should be strictly enforced. Our proposed framework requires a trusted third party to serve as the MES system operator, which is a centralized approach. A centralized market can reduce communication requirements and prevent the participation of dishonest sectors, thereby promoting efficiency and trust within the system. However, a centralized market may raise some concerns, such as potential privacy leaks. Therefore, when designing a platform for the proposed framework, it is essential to consider encryption during the data transmission process to ensure the security and privacy of the shared information.

VII. CONCLUSIONS AND FUTURE WORKS

This paper proposes an end-to-end approach for quantifying the additional profits of the multi-energy load data owned

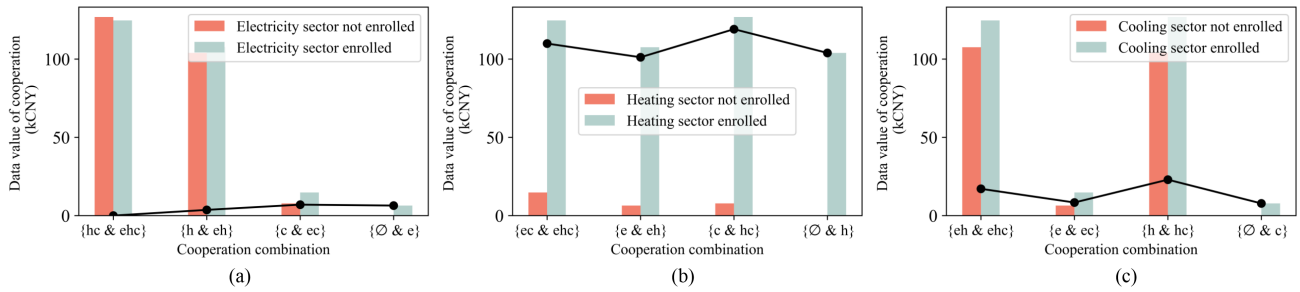

Fig. 10. The data value of cooperation formed by different sectors combination. **(a)** With and without the involvement of the electricity sector. **(b)** With and without the involvement of the heat sector. **(c)** With and without the involvement of the cooling sector.

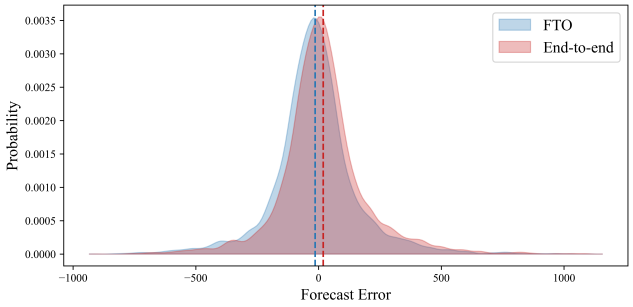

Fig. 11. The distribution of the forecasts error of heat sector

by different sectors in MES and then designs an incentive mechanism to allocate additional profits from data sharing to different sectors in MES according to their contributions.

The experiments indicate that the proposed end-to-end modeling approach can achieve the operation cost reduction under the premise of not affecting the forecasting accuracy of various sectors in MES. The data valuation results illustrate that the proposed valuation approach can thoroughly uncover the value of data through indirect data sharing among sectors in MES. This approach offers a direct economic measurement rather than analyzing data utility. In the profit allocation, the contribution of each sector to the total operation cost-saving has been equitably measured. Our profit allocation strategy can effectively incentivize sectors to cooperate with the MES operator to achieve better operation costs because all sectors can benefit from the cooperation in accuracy improvement or economic compensation.

Although our current valuation framework is effective, it relies on a trusted third party to calculate the gradient in the end-to-end model training process and assumes that all sectors involved are honest. In the future, we will develop a privacy-preserving data valuation framework and explore new design mechanisms that encourage all sectors to be truthful and honest.

REFERENCES

[1] J. Zhu, H. Dong, W. Zheng, S. Li, Y. Huang, and L. Xi, "Review and prospect of data-driven techniques for load forecasting in integrated energy systems," *Applied Energy*, vol. 321, p. 119269, 2022.

[2] Y. Zhou and X. Cui, "Can cross-sector information improve multi-energy demand forecasting accuracy?" *Energy Reports*, vol. 9, pp. 886–893, 2023.

[3] Z. Li and Y. Xu, "Optimal coordinated energy dispatch of a multi-energy microgrid in grid-connected and islanded modes," *Applied Energy*, vol. 210, pp. 974–986, 2018.

[4] C. Kang, X. Chen, Q. Xu, D. Ren, Y. Huang, Q. Xia, W. Wang, C. Jiang, J. Liang, J. Xin *et al.*, "Balance of power: toward a more environmentally friendly, efficient, and effective integration of energy systems in china," *IEEE Power and Energy Magazine*, vol. 11, no. 5, pp. 56–64, 2013.

[5] C. Li, G. Li, K. Wang, and B. Han, "A multi-energy load forecasting method based on parallel architecture CNN-GRU and transfer learning for data deficient integrated energy systems," *Energy*, vol. 259, p. 124967, 2022.

[6] W. Xuan, W. Shouxiang, Z. Qianyu, W. Shaomin, and F. Liwei, "A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems," *International Journal of Electrical Power & Energy Systems*, vol. 126, p. 106583, 2021.

[7] D. Niu, M. Yu, L. Sun, T. Gao, and K. Wang, "Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism," *Applied Energy*, vol. 313, p. 118801, 2022.

[8] X. Cao, Y. Chen, and K. R. Liu, "Data trading with multiple owners, collectors, and users: An iterative auction mechanism," *IEEE Trans. Signal and Information Processing over Networks*, vol. 3, no. 2, pp. 268–281, 2017.

[9] P. Pinson, L. Han, and J. Kazempour, "Regression markets and application to energy forecasting," *Top*, vol. 30, no. 3, pp. 533–573, 2022.

[10] Z. Sun, L. Von Krannichfeldt, and Y. Wang, "Trading and valuation of day-ahead load forecasts in an ensemble model," *IEEE Trans. Industry Applications*, vol. 59, no. 3, pp. 2686–2695, 2023.

[11] J. Wang, F. Gao, Y. Zhou, Q. Guo, C.-W. Tan, J. Song, and Y. Wang, "Data sharing in energy systems," *Advances in Applied Energy*, vol. 10, p. 100132, 2023.

[12] B. Wang, Q. Guo, T. Yang, L. Xu, and H. Sun, "Data valuation for decision-making with uncertainty in energy transactions: A case of the two-settlement market system," *Applied Energy*, vol. 288, p. 116643, 2021.

[13] M. Yu, J. Wang, J. Yan, L. Chen, Y. Yu, G. Li, and M. Zhou, "Pricing information in smart grids: A quality-based data valuation paradigm," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3735–3747, 2022.

[14] C. Goncalves, P. Pinson, and R. J. Bessa, "Towards data markets in renewable energy forecasting," *IEEE Trans. Sustainable Energy*, vol. 12, no. 1, pp. 533–542, 2020.

[15] J. Han, L. Yan, and Z. Li, "A task-based day-ahead load forecasting model for stochastic economic dispatch," *IEEE Trans. Power Systems*, vol. 36, no. 6, pp. 5294–5304, 2021.

[16] H. Kebriaei, B. N. Araabi, and A. Rahimi-Kian, "Short-term load forecasting with a new nonsymmetric penalty function," *IEEE Trans. Power Systems*, vol. 26, no. 4, pp. 1817–1825, 2011.

[17] J. Zhang, Y. Wang, and G. Hug, "Cost-oriented load forecasting," *Electric Power Systems Research*, vol. 205, p. 107723, 2022.

[18] C. Zhao, C. Wan, and Y. Song, "Cost-oriented prediction intervals: On bridging the gap between forecasting and decision," *IEEE Trans. Power Systems*, vol. 37, no. 4, pp. 3048–3062, 2021.

[19] P. Donti, B. Amos, and J. Z. Kolter, "Task-based end-to-end model learning in stochastic optimization," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[20] L. Sang, Y. Xu, H. Long, Q. Hu, and H. Sun, "Electricity price prediction

for energy storage system arbitrage: A decision-focused approach," *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 2822–2832, 2022.
[21] A. N. Elmachtoub and P. Grigas, "Smart 'predict, then optimize'," *Management Science*, vol. 68, no. 1, pp. 9–26, 2022.
[22] W. Huang, N. Zhang, Y. Wang, T. Capuder, I. Kuzle, and C. Kang, "Matrix modeling of energy hub with variable energy efficiencies," *International Journal of Electrical Power & Energy Systems*, vol. 119, p. 105876, 2020.
[23] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
[24] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, 2017.
[25] W. H. Chung, Y. H. Gu, and S. J. Yoo, "District heater load forecasting based on machine learning and parallel CNN-LSTM attention," *Energy*, vol. 246, p. 123350, 2022.
[26] C. Zhang, J. Li, Y. Zhao, T. Li, Q. Chen, and X. Zhang, "A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process," *Energy and Buildings*, vol. 225, p. 110301, 2020.
[27] Y. Wang, N. Zhang, Z. Zhuo, C. Kang, and D. Kirschen, "Mixed-integer linear programming-based optimal configuration planning for energy hub: Starting from scratch," *Applied Energy*, vol. 210, pp. 1141–1150, 2018.
[28] M. Mohammadi, Y. Noorollahi, B. Mohammadi-Ivatloo, and H. Yousefi, "Energy hub: From a model to a concept–A review," *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1512–1527, 2017.
[29] B. Amos and J. Z. Kolter, "Optnet: Differentiable optimization as a layer in neural networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 136–145.
[30] L. Urbanucci, "Limits and potentials of Mixed Integer Linear Programming methods for optimization of polygeneration energy systems," *Energy Procedia*, vol. 148, pp. 1199–1205, 2018.
[31] R. Fletcher and S. Leyffer, "Numerical experience with lower bounds for MIQP branch-and-bound," *SIAM Journal on Optimization*, vol. 8, no. 2, pp. 604–616, 1998.
[32] F. Fang, S. Yu, and M. Liu, "An improved Shapley value-based profit allocation method for CHP-VPP," *Energy*, vol. 213, p. 118805, 2020.
[33] S. L. Chau, R. Hu, J. Gonzalez, and D. Sejdinovic, "RKHS-SHAP: Shapley values for kernel methods," *Advances in Neural Information Processing Systems*, vol. 35, pp. 13 050–13 063, 2022.
[34] J. Liu, "Absolute shapley value," *ArXiv Preprint ArXiv:2003.10076*, 2020.
[35] S. R. Dabbagh and M. K. Sheikh-El-Eslami, "Risk-based profit allocation to DERs integrated with a virtual power plant using cooperative Game theory," *Electric Power Systems Research*, vol. 121, pp. 368–378, 2015.
[36] C. Miller, A. Kathirgamanathan, B. Picchetti, P. Arjunan, J. Y. Park, Z. Nagy, P. Raftery, B. W. Hobson, Z. Shi, and F. Meggers, "The building data genome project 2, energy meter data from the ASHRAE great energy predictor III competition," *Scientific Data*, vol. 7, no. 1, p. 368, 2020.

Yangze Zhou received the B.S. degree from the College of Energy Engineering at Zhejiang University in 2023. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Hong Kong. His research includes data analytics and data valuation in energy systems.

Qingsong Wen is currently a Staff Engineer and Manager at DAMO Academy-Decision Intelligence Lab, Alibaba Group, working in the areas of intelligent time series analysis, data-driven intelligence decisions, machine learning, and signal processing. Before that, he worked at Qualcomm and Marvell in the areas of big data and signal processing, and received his M.S. and Ph.D. degrees in Electrical and Computer Engineering from Georgia Institute of Technology, Atlanta, USA. He has published over 50 top-ranked journal and conference papers, received AAAI/IAAI 2023 Deployed Application Award, and won the First Place in 2022 ICASSP Grand Challenge Competition (AIOps in Networks). He is an Associate Editor for Neurocomputing, Guest Editor for Applied Energy, and regularly served as an SPC/PC member of the major AI and signal processing conferences including AAAI, IJCAI, KDD, ICDM, GLOBECOM, EUSIPCO, etc.

Jie Song received her B.S. degree in Applied Mathematics from Peking University, Beijing, China, in 2004 and her Ph.D. degree in Management Science and Engineering from Tsinghua University, Beijing, in 2010. She is currently a full professor (with tenure) in the Department of Industrial Engineering and Management at Peking University and honored as the Chang Jiang Scholar by China's Ministry of Education. Her research interests include stochastic simulation and optimization, and online algorithm design, with applications in resource allocation of complex service systems.

Xueyuan Cui received the B.E. and M.S. degrees from Zhejiang University, Hangzhou, China, in June 2019 and March 2022, respectively, both in electrical engineering. He is currently a Ph.D. student with the Department of Electrical and Electronic Engineering, The University of Hong Kong. His research interests include power system operation and network balancing technologies.

Yi Wang received the B.S. degree from Huazhong University of Science and Technology in June 2014, and the Ph.D. degree from Tsinghua University in January 2019. He was a visiting student with the University of Washington from March 2017 to April 2018. He served as a Postdoctoral Researcher in the Power Systems Laboratory, at ETH Zurich from February 2019 to August 2021. He is currently an Assistant Professor with the Department of Electrical and Electronic Engineering, The University of Hong Kong. His research interests include data analytics in smart grids, energy forecasting, multi-energy systems, Internet-of-things, and cyber-physical-social energy systems.