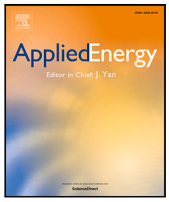
Applied Energy 350 (2023) 121753

Contents lists available at ScienceDirect

Applied Energy

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



Explainable district heat load forecasting with active deep learning

Yaohui Huang [a,d,1], Yuan Zhao [c,1], Zhijin Wang [a,*], Xiufeng Liu [b,*], Hanjing Liu [a], Yonggang Fu [a]

[a] College of Computer Engineering, Jimei University, Yinjiang Road 185, 361021 Xiamen, China
[b] Department of Technology, Management and Economics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[c] School of Business Administration, South China University of Technology, Wushan Road 381, 510630, Guangzhou, China
[d] College of Electronic Information, Guangxi Minzu University, Daxue East Road 188, 530006 Nanning, China

ARTICLE INFO

Keywords:
District heating
Active learning
Explainability
Prediction
Graph neural network

ABSTRACT

District heat load forecasting is a challenging task that involves predicting future heat demand based on historical data and various influencing factors. Accurate forecasting is essential for optimizing energy production and distribution in district heating systems. However, most existing forecasting models lack transparency and interpretability and fail to capture the spatial–temporal dependencies in the data. Moreover, they often require a large amount of annotated data for training, which can be costly and time-consuming to obtain. In this paper, we present a novel approach to district heat load forecasting, which involves predicting future heat demand based on historical data and various influencing factors. The proposed approach is based on an Active Graph Recurrent Network (Ac-GRN), which leverages the strengths of active deep learning and graph neural networks to capture the complex spatial–temporal dependencies in the data. The approach also provides explainability for its predictions by using correlation-based attribution methods. The active deep learning component can effectively select the most informative and representative samples from a large pool of data, reducing the frequency and cost of data collection and human effort. The graph neural network component can model both linear and nonlinear relationships among heat meters using bidirectional recurrent connections, enhancing the accuracy and robustness of the predictions. We conduct extensive experiments and compare our approach with eleven state-of-the-art models on a real-world dataset of district heating consumption data from Danish residential buildings. Our results show that our approach outperforms other models in terms of accuracy, robustness, reliability, and computational efficiency for multi-horizon multi-step district heat load forecasting. Our approach also provides meaningful explanations for its predictions by highlighting the most influential factors and heat meters for each prediction. This paper makes a novel contribution to district heat load forecasting with explainability.

1. Introduction

District heating is a system for distributing heat generated in a centralized location, such as a power plant or a waste incinerator, through a pipeline network to residential, commercial, and industrial buildings for space heating and water heating [1]. District heating can provide energy-efficient, cost-effective, and environmentally friendly heating solutions for urban areas, especially in cold climates [2]. However, to optimize the operation and planning of district heating systems, accurate forecasting of the district heat load is essential. District heat load is the amount of heat demand that needs to be met by the heat supplier at a given time [3]. The forecasting of the district heat load is a challenging task due to the non-linear and non-stationary nature of heat demand, which is influenced by various factors such as weather conditions, building characteristics, human behavior, and calendar effects [4–6] (see Table 1).

District heat load forecasting is a challenging task that involves predicting future heat demand based on historical data and various influencing factors such as weather conditions and building characteristics [3,7]. The non-linear and non-stationary nature of heat demand, coupled with the multitude of influencing factors, makes accurate forecasting difficult [6]. Therefore, researchers have explored advanced machine learning (ML) techniques, particularly deep learning, for their ability to learn complex patterns and dependencies from data [8,9]. Among these techniques, active deep learning is a promising approach that allows the model to actively query the most informative data points

* Corresponding authors.
E-mail addresses: yhhuang5212@gmail.com (Y. Huang), bmyzhao@outlook.com (Y. Zhao), zhijinecnu@gmail.com (Z. Wang), xiuli@dtu.dk (X. Liu), hanjingliu@jmu.edu.cn (H. Liu), yonggangfu@jmu.edu.cn (Y. Fu).
[1] Equal contributor

https://doi.org/10.1016/j.apenergy.2023.121753
Received 4 June 2023; Received in revised form 25 July 2023; Accepted 10 August 2023

Table 1
Abbreviation and description.

Abbreviation	Description
Ac-GRN	Active Graph Recurrent Network
ALM	Active Learning Module
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CVRMSE	Coefficient of Variation of Root Mean Square Error
DHS	District Heating System
GCN	Graph Convolutional Network
GGNN	Gated Graph Neural Network
GRAN	Graph Recurrent Attention Network
GRN	Graph Recurrent Network
GRU	Gated Recurrent Unit
LEU	Linear Representation Unit
LSTM	Long Short-term Memory Network
MAE	Mean Absolute Error
MCFM	Multi-Component Fusion Module
MSE	Mean Squared Error
PCC	Pearson Correlation Coefficient
RMSE	Root Mean Square Error
SCC	Spearman's Rank Correlation Coefficient
SCN	Single-layer Convolution Network
STME	Spatio-Temporal Memory Enhanced Module
TCM	Temporal Convolutional Module
TSFCN	Temporal–spatial Fully Connected Network
XGBoost	Extreme Gradient Boosting

from a large and unlabeled dataset, thus improving the forecasting accuracy while diminishing the costs associated with data collection [10]. This approach is well suited for district heat load forecasting, where data can be highly variable and complex [8].

District heating systems need accurate and transparent forecasting models to optimize their heat production and supply operations. However, most existing models that use active deep learning techniques for district heat load forecasting are not very interpretable. These models are like "black boxes" that do not explain how they make their predictions or what features are important [11]. This can be a major drawback for critical applications like district heating, where explainability is important for trust, reliability, and decision support [12]. Moreover, many models do not capture the spatial–temporal dependencies in the data, which are essential for modeling the interrelationships between different heat meters and their temporal dynamics [13]. Therefore, there is a need for a novel approach that can achieve both high accuracy and high explainability for district heat load forecasting. One possible way to address these challenges is to leverage the strengths of active deep learning and graph neural networks. Active deep learning is an approach where the model actively queries the most informative data points from a large and unlabeled dataset, thus improving the forecasting accuracy and reducing the frequency and cost of data collection [14]. Graph neural networks are a class of deep learning models that can capture the complex spatial–temporal dependencies in the data by modeling the interdependencies between different heat meters using graph structures [15]. Furthermore, these techniques can provide explainability by leveraging the correlation among heat meters, which can inform the model's predictions and feature importance [16]. However, few studies have applied active deep learning and graph neural networks to district heat load forecasting with explainability [17,18].

In this paper, we propose a novel approach to district heat load forecasting using an Active Graph Recurrent Network (Ac-GRN). Ac-GRN actively learns from the most informative data points and models the interdependencies between different heat meters using graph neural networks. It consists of two main components: an Active Learning Module (ALM) and a Graph Recurrent Network (GRN). The ALM selects the most representative heat meters and prepares the inputs for the GRN. The GRN then uses the inputs to make predictions. The GRN has two core parts: a temporal convolution module and a spatio-temporal memory enhanced module. The former extracts temporal dependencies from the smart heat meter data using graph neural networks and single-layer convolutions. The latter further captures dynamic spatial and temporal correlations using spatial attention and recurrent neural networks. Our approach is novel in its ability to actively learn from the most informative data points using an uncertainty-based sampling strategy and its use of graph neural networks to model the interdependencies between different heat meters. Moreover, our approach provides explainability by using correlation-based methods that highlight the spatial features that influence the predictions. Our approach has several advantages over existing methods that use active deep learning or graph neural networks for district heat load forecasting. First, it enhances forecasting accuracy and diminishes data collection costs by implementing a strategy based on uncertainty sampling. Second, it captures the complex spatial–temporal patterns in the data by using graph neural networks. Third, it provides explainability by using correlation-based methods.

The key contributions of this paper are threefold:

- We propose a novel approach to district heat load forecasting using an Active Graph Recurrent Network, which leverages the strengths of active deep learning and graph neural networks to capture the complex spatial–temporal dependencies in the data.
- We provide explainability for our model's predictions by employing relationship-inspired methods that utilize correlations among heat meters. This approach highlights the spatial features in the input that contribute to pattern discovery and prediction.
- We evaluate our approach on real-world district heating data and compare it with 11 state-of-the-art models in terms of accuracy and robustness. We also conduct ablation studies to analyze the impact of different components of our model, as well as computational efficiency compared with other models.

The rest of the study is structured as follows. Section 2 reviews the literature on heat load forecasting and AL. Section 4 describes the proposed Ac-GRN model. Section 5 conducts the evaluation. Section 6 discusses the contributions and limitations of the proposed model. Section 7 concludes the paper and presents the future research directions.

2. Related work

In this section, we review the existing literature on district heat load forecasting, and highlight the main challenges and limitations of current methods. We also review the literature on active deep learning and graph neural networks, and explain how they can address some of the challenges and limitations of existing methods. Finally, we summarize the main contributions of our paper and how it differs from the existing literature.

2.1. District heat load forecasting

District heat load forecasting is a challenging task that involves predicting future heat demand based on historical data and various influencing factors such as weather conditions and building characteristics [3,19]. Accurate forecasting is essential for optimizing energy production and distribution in district heating systems [2]. Existing methods for district heat load forecasting can be broadly categorized into physical models, black-box models [20], and gray-box models [21]. Physical models use equations that describe the physical behavior of a system to predict an output, while black-box models use supervised ML methods where measurements of input and output variables of a system are collected, and then used to mathematically describe the system [22]. Gray-box models provide a balance between high accuracy and good generalization capabilities, by extracting the mathematical model/structure from the system's physics, and estimation of model parameters from measured data [23,24]. Physical models have the advantage of being easily interpretable, but they require a large number of parameters and detailed information about the system as input,

which can be costly and time-consuming to obtain [22]. Moreover, physical models suffer from poor generalization capabilities, as they may not capture the non-linear and non-stationary nature of heat demand [25]. Black-box models have the advantage of being easy to build and capable of learning complex patterns and dependencies from data [12]. However, black-box models require a large amount of data for training purposes, which may not be available or reliable in some cases [26]. Moreover, black-box models lack transparency and interpretability, as they provide little insight into how they make their predictions [27]. Gray-box models have the advantage of combining the strengths of physical models and black-box models, by exploiting the domain knowledge and data-driven learning [28]. However, gray-box models also face some challenges, including determining the optimal model structure and parameters, and dealing with incomplete or noisy data [29].

Among these three categories of methods, black-box models have received increasing attention lately, owing to their efficacy in learning complex patterns and dependencies from data [3]. Black-box models are ML techniques that do not reveal the internal logic or structure of the model, but only provide the input–output relationship [30]. In particular, black-box models such as artificial neural network (ANN) [31], convolutional neural network (CNN) [32], long short-term memory network (LSTM) [8], attention mechanism [8], extreme gradient boosting (XGBoost) [33], and among others, have been applied to district heat load forecasting with promising results. For example, Xue et al. [17] used SVR, DNN, and XGBoost to validate the accuracy and stability of black-box models for multi-step ahead heat load forecasting in district heating system (DHS). Dang et al. [18] employed XAI methods, such as SHAP and PDP, to explain the significance of exogenous features for district heating prediction using XGBoost. Li et al. [34] proposed a heat load forecasting model based on gated recurrent units with time-varying features using LSTM and attention mechanism. However, these methods require large storage space and high costs of data collection, which can be prohibitive in many cases and affect the efficiency in DHS. In addition, collecting and transferring excessive data for DHS may lead to expensive labor and communication overhead. To address these challenges, we propose an AL-based framework for district heat load prediction, which can utilize less heat meter data to achieve accurate prediction for all smart heat meters in a certain time period. With the AL framework, we select representative heat meters for training the prediction model based on an uncertainty-based sampling strategy, while excluding heat meters with similar patterns.

2.2. Active deep learning and graph neural networks

High accuracy, cost-effectiveness, and timeliness are important goals in the field of energy consumption monitoring [27]. Active deep learning offers a potential approach to enhancing cost-efficiency in data acquisition and storage by selecting the most informative or uncertain samples for training, as per specific criteria or strategies [14]. This method optimizes the learning process by focusing on samples that the model finds most challenging to classify or predict [35]. Specifically, the most uncertain samples are those that the model has the most difficulty classifying or predicting. In the context of classification tasks, these typically are the samples where the predicted class probabilities closely resemble a uniform distribution, implying that the model struggles to assign a clear class to these samples. For regression tasks, these are often the samples that yield the widest confidence interval for the model's prediction. This strategic selection of high-uncertainty samples empowers the model to maximize information extraction from a minimal number of samples, thereby fostering accelerated learning and enhanced accuracy. For example, Settles et al. [36] proposed several query strategies, e.g., uncertainty sampling, query-by-committee, expected model change, and expected error reduction. Zhang et al. [10] utilized active learning to counter data bias and achieve high performance in building energy forecasting. Aryandoust et al. [37] proposed a spatio-temporal, multi-modal method for electric load forecasting, which exhibits improved accuracy. Wang et al. [38] designed a selector–predictor framework for electric load forecasting. Active deep learning can be enhanced through the combination with various techniques, such as transfer learning [39], semi-supervised learning [40], and reinforcement learning [41]. These approaches can further improve its performance and expand its applicability.

Graph neural networks have been widely used in various domains such as social networks [42], knowledge graphs [43], computer vision [44], natural language processing [45], etc., where data has rich structural or relational information. Graph neural networks can capture the complex spatial–temporal dependencies in the data by modeling the interdependencies between different nodes using graph structures [15]. For example, Kipf et al. [46] proposed a graph convolutional network (GCN) that can learn node representations by aggregating information from their neighbors using spectral graph theory. Li et al. [47] proposed a gated graph neural network (GGNN) that can learn node representations by propagating information through recurrent neural networks using spatial graph theory.

In this paper, we propose to use active deep learning and graph neural networks for district heat load forecasting. Our approach can address some of the challenges and limitations of existing methods by actively learning from the most informative data samples and modeling the inter-dependencies between different heat meters using graph structures. To the best of our knowledge, this is the first attempt to predict district heat load utilizing a graph neural network and AL framework.

3. Materials

This study uses the public dataset of district heating consumption from residential buildings in Aalborg, Denmark [48]. The dataset covers 3,127 smart heat meters with hourly readings for three years (2018–2020). It also includes contextual information such as dwelling type, construction year, and energy efficiency level. This study primarily targets the prediction of heat load for single-family houses, terraced houses, and apartments, which account for 3,021 smart meters in the dataset. Smart heat meters lacking dwelling type and those that are unoccupied (a total of 105 samples) are excluded from the focus of this study. Additionally, meteorological data from a nearby weather station was collected. This data has nine variables, such as outdoor temperature, solar radiation intensity, wind speed, and others, that may affect heat consumption patterns.

Due to the unavoidable interference for the transmitting process, there are about 0.3% smart heat meter data missing. To ensure compatibility with the model and adhere to the original data processing method described in [48], missing data from the smart meters is filled using a weighted moving average imputation. This method utilizes a symmetric window size of 48 (including 48 preceding and following values) and applies a linear weighting scheme. The meteorological data is complete and devoid of missing values. Then, we conducted a correlation analysis to evaluate the relationship between various meteorological variables and the mean value of heat load. We visualized the results of the correlation analysis using a heatmap (see Fig. 1), which shows the Spearman's rank correlation coefficient (SCC) between each pair of variables. The SCC measures the monotonic relationship between two variables, regardless of whether the relationship is linear or nonlinear. Additionally, the SCC is robust to outliers and apt for handling data that may not follow the same distribution. The coefficient values of the SCC range from −1 to 1. A value of −1 denotes a perfect negative correlation, a value of 0 signifies no correlation, and a value of 1 corresponds to a perfect positive correlation. The heatmap displays the correlation between variables, with darker colors indicating stronger correlations and lighter colors indicating weaker correlations. The diagonal cells display the correlation of each variable with itself, invariably equal to 1.

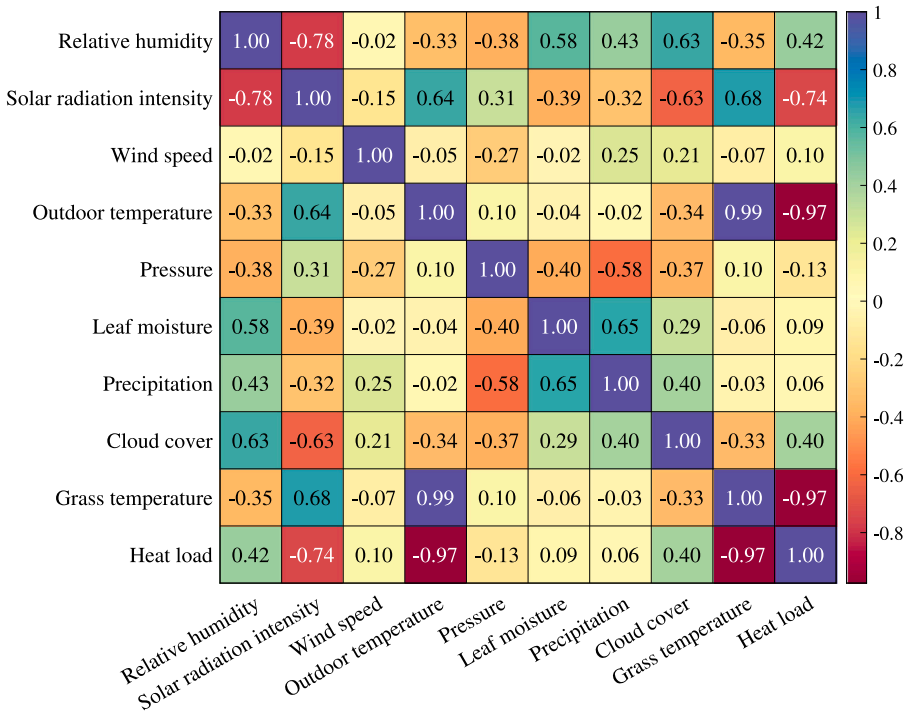


Fig. 1. The Spearman's rank correlation coefficient between heat load observations and meteorological factors.

As illustrated in Fig. 1, outdoor temperature and grass temperature emerge as the most significant meteorological variables for heat load forecasting. The strong correlation between these two factors highlights their closely aligned trend. To reduce the redundancy of temperature information, we choose the outdoor temperature as an exogenous feature in our study. Outdoor temperature directly influences the energy required for heating and cooling in residential and commercial buildings [49]. Solar radiation intensity also significantly correlated with heat load. The solar radiation might contribute to reducing the heating load by naturally warming the building. This effect is particularly significant in structures designed for passive solar heating, where the building's design and orientation synergize to maximize the capture and retention of solar heat [50]. Hence, solar radiation intensity is included as an exogenous meteorological factor in our study. Relative humidity and cloud cover both exhibit a relatively minor impact on heat load. Given the established correlation between relative humidity and building heat transfer, as confirmed by prior research [51], we have included relative humidity as the third meteorological factor in our analysis. The presence of moisture in the air can significantly impact the thermal characteristics of a building, consequently affecting the overall heat load. The variables of wind speed, air pressure, leaf moisture, and precipitation exhibit weak correlations with heat load data. To enhance the diversity of meteorological features without excessive inclusion, we have additionally included wind speed as a factor, as it demonstrates relatively lower correlation with the other three selected variables. Moreover, as wind speed increases, the rate of convective heat transfer from the building surfaces to the outside environment tends to increase. This prompts a need for more heating to keep the inside temperature steady, influencing the overall district heat load.

Above all, the factors of outdoor temperature, solar radiation intensity, relative humidity, and wind speed are selected as additional feature combination for our experiment. We summarized the statistics of heat load and these meteorological factors in Table 2, which shows the minimum, maximum, median, mean, standard deviation, Pearson correlation coefficient (PCC), and SCC. The PCC is another measure that quantifies the strength and direction of a linear relationship between two continuous variables. It also ranges from −1 to 1 and has a similar interpretation as the SCC. Finally, we normalized the data using min–max scaling, a technique that adjusts the range of variable values to a common scale, promoting model stability and learning efficiency.

4. Methodology

4.1. Overview

In this section, we provide a comprehensive description of the proposed Active Graph Recurrent Network (Ac-GRN) model. The Ac-GRN model combines active learning, graph convolutional components, and recurrent learning units to effectively capture both spatial and temporal dependencies in district heating data. An overview of our proposed model is illustrated in Fig. 2. The proposed Ac-GRN is composed of several key modules. First, the training and validation sets of heat load data are pre-processed and randomly segregated into annotated and unannotated groups for subsequent analysis. Second, the active learning module (ALM) is initiated using the annotated meters, while the unannotated meters are stored in the data center. The ALM consists of two components: the *knowledge extractor* and the *feature reconstructor*. The *knowledge extractor* is responsible for selecting the most informative samples using an uncertainty-based sampling strategy. The observations from the selected representative meters are then passed to the *feature reconstructor* component to reconstruct the unknown trends of the unselected heat load meters for prediction. Third, a graph recurrent network (GRN) is designed to capture the latent heat load patterns across multiple heat meters and make predictions. This network comprises three elements: (1) The *temporal convolutional module* (TCM) employs multi-layer graph convolutional networks in the temporal dimension, extracting temporal dependencies from the heat load data. (2) The *spatio-temporal memory enhanced module* (STME) augments the model's capacity to capture dynamic spatial and temporal correlations by implementing a spatial attention mechanism and a recurrent component. (3) The *multi-component fusion module* (MCFM) fuses the linear representation of the input data with the output from the recurrent component to generate the final prediction. The linear representation is a simplified projection of the input data that can alleviate the vanishing gradient problem and enhances the expressiveness

Table 2
The descriptive statistics and correlation coefficients of heat load and meteorological factors. The standard deviation (STD) measures the variability of each variable. The p-value indicates the significance level of the correlation coefficients. A p-value less than 0.001 (***) means that the correlation is very unlikely to occur by chance. The Pearson correlation coefficient (PCC) and Spearman's rank correlation coefficient (SCC) measure the strength and direction of the linear relationship between each meteorological factor and the median heat load.

Symbol	Number of time series	Min	Max	Medium	Mean	STD	PCC	SCC
Outdoor temperature	1	−7.883	23.712	8.450	9.159	6.138	−0.953***	−0.966***
Solar radiation intensity	1	1.804	391.483	94.046	126.140	106.669	−0.709***	−0.738***
Relative humidity	1	45.758	99.892	84.252	82.422	10.930	0.370***	0.424***
Wind speed	1	1.150	12.829	4.685	4.996	2.059	0.102***	0.101***
Heat load	3021	0	462.328	37.110	45.610	36.526	–	–

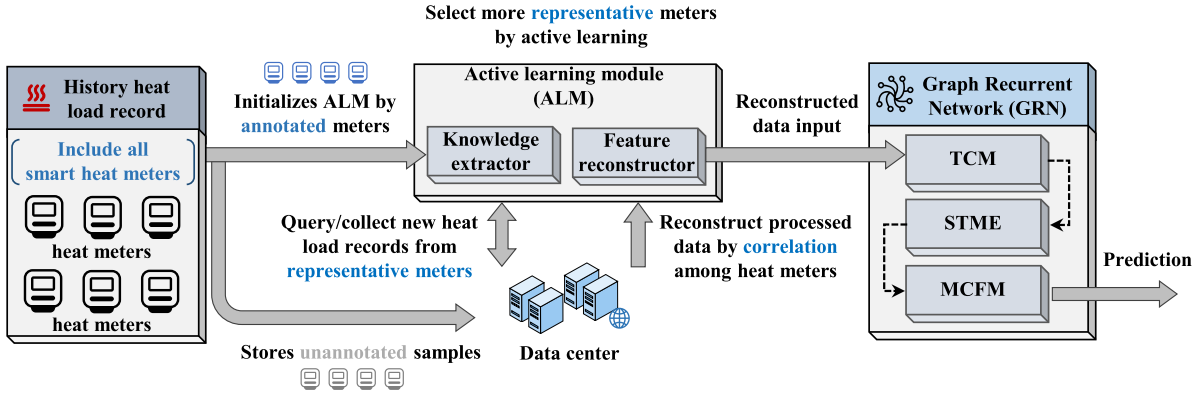


Fig. 2. The overview of the proposed Ac-GRN method.

of the model. With this network structure, the proposed AC-GRN model will be able to capture the complex spatial–temporal dependencies in district heating data, and actively learn from the most informative data samples, thereby improving the forecasting accuracy and reducing the costs of data acquisition.

4.2. Knowledge extractor by active learning

The *knowledge extractor* leverages active learning to efficiently identify the most informative samples among numerous heat meters. This method significantly curtails costs related to data collection and storage, while enhancing the efficiency of the learning process. As illustrated in Fig. 3, the module consists of two phases: a query phase and an update phase. Within the query phase, we develop a novel classifier known as the Temporal–spatial Fully Connected Network (TSFCN). The TSFCN is combined with a temporal convolutional network (Temp-FCN), a spatial convolutional network (Spat-FCN), a fully connected network for output (Out-FCN), and two activation layers. The Temp-FCN and Spat-FCN are employed to capture the temporal and spatial patterns among annotated heat load observations, respectively. The Out-FCN and the softmax function are employed to compute the probabilities corresponding to each class. Specifically, this classifier learns patterns from annotated samples and subsequently classifies the category of unannotated meters. The detail processing of TSFCN classifier $f_\theta(\cdot)$ can be formulated as:

$$f_\theta(x) = \text{softmax}((\boldsymbol{m}_\alpha \odot (\sigma((W_{\alpha,1}^\top x)W_{\alpha,2})))W_{\alpha,3} + b_\alpha), \tag{1}$$

where x is an unannotated sample from data center U, which consists of unannotated heat meters. $W_{\alpha,1}$, $W_{\alpha,2}$, and $W_{\alpha,3}$ are the learnable parameters of Temp-FCN, Spat-FCN, and Out-FCN, respectively. $\odot$ denotes the element-wise multiplication. $\boldsymbol{m}_\alpha$ is the dropout vector to alleviate overfitting. b_α is the bias term. softmax is the activation function that normalizes the output to a probability distribution over the possible classes. $\top$ is the transpose operation that changes the orientation of a matrix or a vector.

After obtaining the probability map, we incorporate an uncertainty sampling method to select the most uncertain samples, designated as representative meters for annotation. The uncertainty of the sample can be measured by different criteria, including entropy, least confidence, or margin. In this study, we employ the least confidence criterion, which selects the sample where the model exhibits the lowest confidence. The least confidence criterion can be formulated as follows:

$$x^* = \arg\min_{x \in U} f_\theta(x), \tag{2}$$

where x^* are samples from the unannotated data pool U. $f_\theta(x)$ is the predicted label of sample x by the model with parameters θ, which is trained on a subset of annotated meters.

In the update phase, the identified representative meters are incorporated into the annotated data subset and concurrently removed from the unannotated data center U. The classifier is subsequently retrained to update the model parameters. This iterative process, involving both the query strategy and update process, continues until the proportion of annotated to unannotated meters achieves a specified target ratio or a predetermined number of iterations is reached. The targeted annotated data ratio is a hyperparameter that controls the trade-off between the data collection cost and the model performance.

4.3. Feature reconstructor for incomplete data

As illustrated in Fig. 4, traditional district heat load prediction utilizes historical records (including recent trends) to forecast the target period for each heating meter. However, when using the active learning framework, the unrepresentative meters inevitably lack information regarding recent short-term trends. This absence can significantly impact the accuracy and efficiency of predictions. Incorporating recent short-term trends can effectively help capture rapid changes, thus allowing for more agile and accurate responses to dynamic circumstances in real-time [49]. To address this issue, our feature reconstructor module leverages observations from representative heat meters to supplement the incomplete data acquired through the Active Learning framework, as shown in Fig. 5. This module consists of two components: a relation discovery module and a representation generation module.

The relation discovery module establishes associations between unannotated heat meters and selected representative heat meters based on their correlation. Specifically, we first calculate the PCC for each pair of unannotated and annotated heat meter data. The trend among

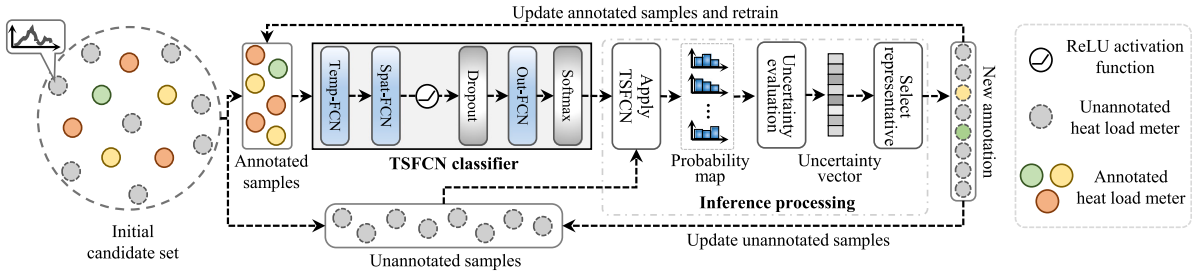


Fig. 3. Illustration of the uncertainty-based query processing in ALM.

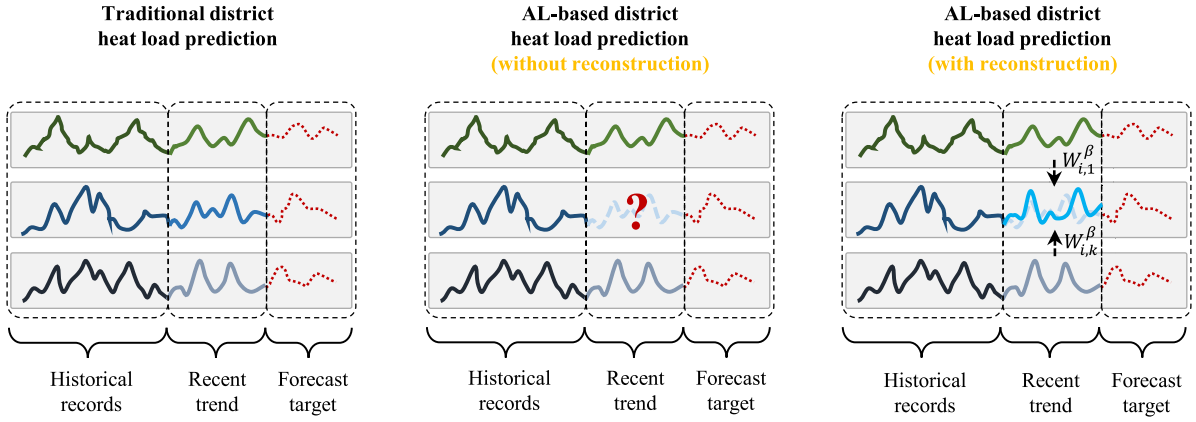


Fig. 4. The illustration of motivation for feature reconstructor.

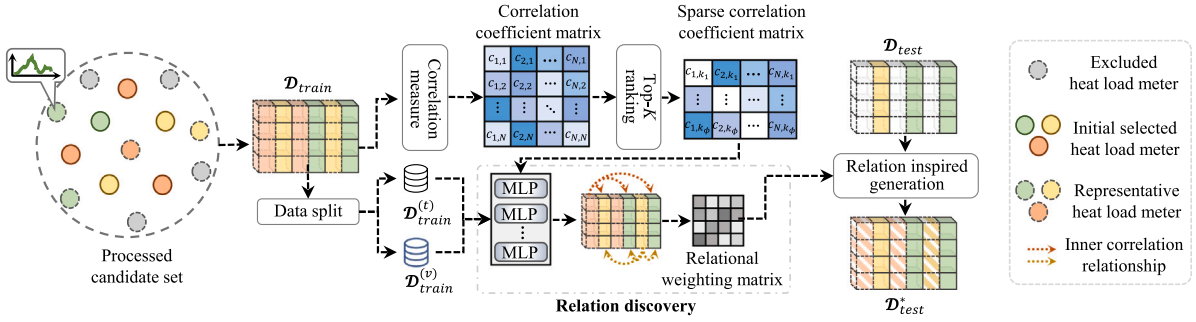


Fig. 5. The illustration of relation discovery and representation generation module.

heat load meters may exhibit some common characteristics or usage patterns [52], PCC is more suitable to measure these linear relationships. Then, for each unselected heat load meter, we choose the top-K most correlated meters to construct a sparse correlation coefficient matrix $\boldsymbol{C}$, where $c_{i,j}$ denotes the correlation value between heat meter i and j. This method can reduce the information redundancy inherent in these relationships, and reconstruct the sequence by exploiting the low-rank relationship structure [53]. The learning process of the sparse relationship can be formulated as follows:

$$\arg\min_{W^{\beta}\in\boldsymbol{W}} f_{\beta}(W_{i,1}^{\beta}\cdot x_{i,1}^{*},\ldots,W_{i,K}^{\beta}\cdot x_{i,K}^{*}), \tag{3}$$

where x_i^* represents the correlated representative samples for ith unannotated meter, and f_β is the function for relation learning. W_i^β denotes the learnable weighting parameters for the ith unannotated meter.

The representation generation module operates as a linear weighting process based on the learned weighting matrix $\boldsymbol{W}^\beta$. This process adaptively emphasizes certain relevant heat meters in the sparse matrix by multiplying them with their corresponding weights. The aim of this is to improve the prediction of the heat load for the target, specifically for the unannotated meters. The generation process can be described as follows:

$$\mathcal{M}_i = W_i^{\beta}\cdot[x_{i,1}^{*},\ldots,x_{i,K}^{*}], \tag{4}$$

$\mathcal{M}_i$ represents the generation for the ith unannotated meter, and W^β is the obtained relational weighting matrix. $\mathcal{M}_i$ encompasses trend information that can provide linear characteristics for the prediction model, thereby helping to reduce the effects of missing data in the heat load predictions for unannotated heat meters.

4.4. Graph recurrent network

The overview of graph recurrent network (GRN) is shown in Fig. 6, which consists of three main parts: temporal convolutional module, spatio-temporal memory enhanced module, and multi-component fusion module. They are described in the following.

4.4.1. Temporal convolutional module

The temporal convolution component comprises two temporal graph convolution networks (TGCNs) along the temporal dimension and a single-layer convolution network (SCN) in the spatial dimension. The

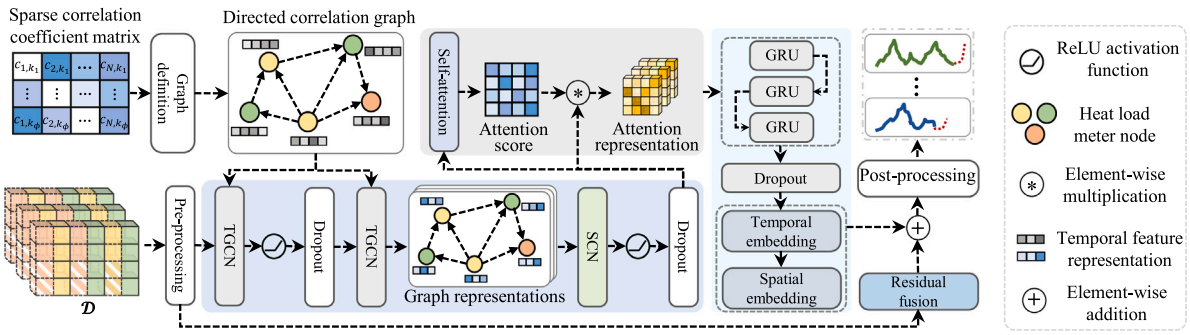


Fig. 6. The overview of the proposed graph recurrent network.

TGCN module extracts temporal dependencies from adjacent time steps of each smart heat meter, while the SCN module captures spatial dependencies from all smart meters.

To apply the TGCN module, we need to transform the graph structure into an algebraic form that can capture its connectivity and other patterns. One way to do this is to use spectral graph theory and the Laplacian matrix, which are tools for analyzing the properties of graphs in the frequency domain. We define the multi-meters network as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{A})$, where $\mathcal{V}$ represents a finite set of N nodes, corresponding to the smart heat meters; $\mathcal{E}$ is the set of edges, representing the relationships between each heat meter and other meters; and $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is the adjacency matrix of graph $\mathcal{G}$. The normalized Laplacian matrix can be formulated as $\boldsymbol{L}_p = \boldsymbol{I}_{unit} - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}$, where $\boldsymbol{L}_p \in \mathbb{R}^{T\times T}$ is the normalized Laplacian matrix; the diagonal matrix $\boldsymbol{D} \in \mathbb{R}^{T\times T}$ is a degree matrix, $\boldsymbol{D}^{(i,i)} = \sum_j \boldsymbol{A}^{(i,j)}$; $\boldsymbol{I}_{unit}$ denotes a unit matrix.

The embedding process of TGCN can be described as follows: First, we multiply the normalized Laplacian matrix with the input time series and a learnable weight matrix, and add a bias term. This operation performs a linear transformation on each node's features based on its adjacency matrix. Second, we apply a ReLU activation function and a dropout mask to introduce non-linearity and regularization. This operation enhances the model's ability to learn complex and robust features from the data. Third, we multiply the output of the previous operation with another learnable weight matrix. This operation produces the final output of TGCN, which is a node representation that captures the temporal dependencies from adjacent time steps. The mathematical equations for these operations are:

$$\boldsymbol{W}_{g,\theta} *_{\mathcal{G}} \boldsymbol{\mathcal{M}} = \boldsymbol{D}^{-\frac{1}{2}}\hat{\boldsymbol{A}}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{\mathcal{M}}\boldsymbol{W}_{g,\theta} + \boldsymbol{b}_g, \tag{5}$$

$$\boldsymbol{\mathcal{R}}_{g,o} = \boldsymbol{W}_{g,o} *_{\mathcal{G}} (\boldsymbol{m}_g \odot \sigma(\boldsymbol{W}_{g,\theta} *_{\mathcal{G}} \boldsymbol{\mathcal{M}})), \tag{6}$$

where $*_{\mathcal{G}}$ denotes the graph convolution operation. $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_{unit} \in \mathbb{R}^{N\times N}$ is an adjacent matrix with inserted self-loops. $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{B\times L\times N}$ indicates the input time series. $\boldsymbol{\mathcal{R}}_{g,o}$ is the final output of TGCN component. $\odot$ indicates element-wise multiplication, $\sigma(\cdot)$ is rectified linear unit (ReLU) activation, and $\boldsymbol{m}_g$ is a binary mask denotes dropout processing. $\boldsymbol{W}_{g,\theta}$ and $\boldsymbol{W}_{g,o}$ are both the learnable weighting matrix, $\boldsymbol{b}_g$ denotes the biases.

After capturing information from adjacent time steps for each node in the graph's temporal dimension, we proceed to integrate an SCN layer in the spatial dimension to emphasize important pattern information among all smart meters. The SCN module is a single-layer convolutional network that applies a one-dimensional convolution operation to the output of the TGCN along the spatial dimension. This operation involves sliding a kernel or filter over each node's features and computing a dot product. The size of the kernel dictates the number of heat meters considered in each convolution step. Capable of extracting spatial features, the convolutional module effectively discerns similarities or differences between nodes based on their heat meter observations. The spatial convolution can be formulated as follow:

$$\boldsymbol{\mathcal{R}}_{c,o} = \boldsymbol{m}_c \odot \sigma(\boldsymbol{W}_{c,o} * \boldsymbol{\mathcal{R}}_{g,o}), \tag{7}$$

where $\boldsymbol{\mathcal{R}}_{c,o}$ denotes the generated representation of convolution component, $\boldsymbol{W}_{c,o}$ is the learnable parameters of convolution kernel. $\boldsymbol{m}_g$ is a binary mask matrix that prevents the networks from overfitting.

4.4.2. Spatio-temporal memory enhanced module

The spatio-temporal memory enhanced module is designed to further enhance the model's ability to capture dynamic spatial and temporal correlations by applying a spatial attention mechanism and a recurrent component. The spatial attention mechanism assigns different weights to different parts of the input based on their relevance, while the recurrent component stores and updates information over time. This spatial attention is defined as follows:

$$\boldsymbol{S}_o = \text{softmax}(\boldsymbol{W}_s^{(1)} \cdot \sigma(\boldsymbol{W}_s^{(2)}\boldsymbol{\mathcal{R}}_{c,o}) + \boldsymbol{b}_g), \tag{8}$$

$$\boldsymbol{\mathcal{R}}_{s,o} = \boldsymbol{\mathcal{R}}_{c,o} \odot \boldsymbol{S}_o, \tag{9}$$

where $\boldsymbol{S}_o$ denotes the attention score that dynamically adjusts the impact weights in the spatial dimension. $\boldsymbol{\mathcal{R}}_{s,o}$ is the produce of spatial attention mechanism. $\boldsymbol{W}_s^{(1)}$ and $\boldsymbol{W}_s^{(2)}$ are the learnable weighting matrix, $\boldsymbol{b}_g$ is the learnable bias.

The recurrent component is built upon a gated recurrent unit (GRU) capable of learning long-term dependencies in sequential data. The GRU possesses two gates: a reset gate and an update gate. The reset gate determines the degree to which the previous hidden state is forgotten, while the update gate governs how much of the previous hidden state is updated with the new input. The equations governing these gates are:

$$\boldsymbol{e}_t = \varphi(\boldsymbol{W}_{r,e}[\boldsymbol{h}_{t-1}; \boldsymbol{\mathcal{R}}_{s,o}] + \boldsymbol{b}_{r,e}), \tag{10}$$

$$\boldsymbol{u}_t = \varphi(\boldsymbol{W}_{r,u}[\boldsymbol{h}_{t-1}; \boldsymbol{\mathcal{R}}_{s,o}] + \boldsymbol{b}_{r,u}), \tag{11}$$

$$\boldsymbol{z}_t = \tanh(\boldsymbol{W}_{r,z}[\boldsymbol{e}_t \odot \boldsymbol{h}_{t-1}; \boldsymbol{\mathcal{R}}_{s,o}] + \boldsymbol{b}_{r,z}), \tag{12}$$

$$\boldsymbol{h}_t = (1 - \boldsymbol{u}_t) \odot \boldsymbol{h}_{t-1} + \boldsymbol{u}_t \odot \boldsymbol{z}_t, \tag{13}$$

where $\boldsymbol{e}_t$, $\boldsymbol{u}_t$, and $\boldsymbol{z}_t$ are the internal state of the update gate, reset gate, and candidate hidden state at time step t, respectively. $\boldsymbol{W}_{r,e}$, $\boldsymbol{W}_{r,u}$, and $\boldsymbol{W}_{r,z}$ are the learnable weighting matrices. $\boldsymbol{b}_{r,e}$, $\boldsymbol{b}_{r,u}$, and $\boldsymbol{b}_{r,z}$ are the correspondent biases. $\varphi(\cdot)$ denotes sigmoid activation function, and $[;]$ indicates the concatenation operation. In these equations, $\boldsymbol{h}_{t-1}$ is the previous hidden state, and $\boldsymbol{\mathcal{R}}_{s,o}$ is the output of the attention component. The recurrent component takes $\boldsymbol{\mathcal{R}}_{s,o}$ as the input and updates $\boldsymbol{h}_{t-1}$ using the three gates to produce $\boldsymbol{h}_t$, which is the current hidden state. The integration of exogenous meteorological features necessitates an update to the GRU's input, as detailed in the following equation:

$$\boldsymbol{\mathcal{V}}_{s,o} = [\boldsymbol{\mathcal{R}}_{s,o}; \boldsymbol{W}_e \cdot \boldsymbol{\mathcal{E}}], \tag{14}$$

where $\mathcal{E}$ represents the exogenous data. $\mathcal{V}_{s,o}$ signifies the input, combining observations from both the heat meters and meteorological features.

The embedding process maps the output of the recurrent component into an embedding space that matches the desired output shape. The embedding process consists of two steps: first, it applies a temporal embedding function $\mathrm{Emb}_{tmp}(\cdot)$ on $\boldsymbol{h}_t$ to reduce its temporal dimension; second, it applies a spatial embedding function $\mathrm{Emb}_{spt}(\cdot)$ on $\mathrm{Emb}_{tmp}(\boldsymbol{h}_t)$ to reduce its spatial dimension. The equation for this process is:

$$\boldsymbol{\mathcal{R}}_{r,o} = \mathrm{Emb}_{tmp}(\mathrm{Emb}_{spt}(\boldsymbol{m}_r \odot \boldsymbol{h}_t)), \tag{15}$$

where $\boldsymbol{\mathcal{R}}_{r,o} \in \mathbb{R}^{B\times L\times N}$ is the final output of STME module. $\boldsymbol{m}_r$ denotes a binary mask matrix for the dropout layer.

4.4.3. Multi-component fusion module

The nonlinear modules can effectively capture potential dependencies in either the temporal or spatial dimension. However, excessive nonlinearity may result in the vanishing or exploding gradients problem, which in turn can adversely affect the prediction accuracy. Therefore, to balance the nonlinearity and robustness of the proposed Ac-GRN, we introduce a linear representation unit (LEU) that adds a linear combination of the input features $\boldsymbol{\mathcal{M}}$ to the recurrent component output $\boldsymbol{\mathcal{R}}_{r,o}$ as the final prediction:

$$\hat{\boldsymbol{Y}}_{t+h:t+h+L} = \boldsymbol{\mathcal{R}}_{r,o} \oplus (\boldsymbol{W}_{f,o}\boldsymbol{\mathcal{M}} + \boldsymbol{b}_f), \tag{16}$$

where $\hat{\boldsymbol{Y}}_{t+h:t+h+L} \in \mathbb{R}^{B\times L\times N}$ is the predictive values, $\oplus$ indicates the addition operation, $\boldsymbol{W}_{f,o}$ is learnable weighting parameters, and $\boldsymbol{b}_f$ is correspondent bias.

We use the mean square error (MSE) as the loss function to measure the prediction accuracy and add an L_2 regularization term to prevent overfitting. The loss function can be defined as:

$$\mathcal{L}oss(\boldsymbol{Y}, \hat{\boldsymbol{Y}}) = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{N}\sum_{v=1}^{L}(Y_{i,j,v} - \hat{Y}_{i,j,v})^2 + \frac{\lambda}{2}\sum_{i=1}^{\Theta} w_i^2, \tag{17}$$

where K is the length of training samples, N denotes the dimension of target data, and L is the prediction steps. Specifically, λ is a constant that denotes the regularization coefficient, Θ is the number of weights in the model. w indicates the weight parameter of the model.

5. Experiment

5.1. Experimental setup

To ensure a fair comparison, the same constant parameters related to training are applied to all methods. The Adam optimizer [54] is used to obtain the training models, and the mean squared error (MSE) is selected as the loss function. The batch size is set as 32. The number of training epochs and the learning rate are adjusted to ensure each method's optimal performance. To increase the dependability and stability of the results, every experiment is performed five times using different random seeds, with the reported results being the mean values. The grid search approach is employed to search for relatively optimal hyperparameters. The detailed settings for the hyperparameters of each method are provided in Table 9. The model was implemented using the deep learning framework, Pytorch v1.12.1 [55]. All experiments were carried out on a server equipped with Intel(R) Xeon(R) Gold 6226R CPU (2.90 GHz) with 128G memory and were accelerated by two NVIDIA RTX A6000 GPUs. In this study, we use the first 70% data for training (766 data points), the following 10% data (109 data points) as the validation set, and the remaining 20% data (219 data points) for the test. For the experiment with varying sampling radio, the input to the models was manipulated in the test set. Specifically, the input from unselected heat meters was set to zero for the baseline methods, while these observations were reconstructed using a feature reconstructor for the Ac-GRN model. This processing does not affect the data in the training or validation sets, ensuring that these changes do not influence the models during their training phase. Despite these adjustments, the prediction target remained the true value.

5.2. Baselines

We compare our proposed model with several state-of-the-art methods for multivariate time-series forecasting, including:

- **Autoregressive (AR):** a linear model that predicts the current value of a time series based on its past values [56].
- **Dlinear:** a linear regression model that uses the previous values of all time series as input features [57].
- **MTNet:** a multi-task learning model that combines convolutional neural networks and recurrent neural networks to capture both local and global temporal dependencies [58].
- **TPA:** a temporal pattern attention model that uses an attention mechanism to learn the importance of different temporal patterns [59].
- **LSTNet:** a long- and short-term time-series network that combines convolutional neural networks and recurrent neural networks to capture both short-term and long-term dependencies [60].
- **LSTM:** a type of recurrent neural network that can learn long-term dependencies in sequential data using memory cells and gates [61].
- **ED (GRU):** an encoder–decoder model that uses gated recurrent units to encode the input sequence and decode the output sequence [62].
- **CRNN:** a convolutional recurrent neural network that uses convolutional layers to extract local features and recurrent layers to capture temporal dependencies [63].
- **CRNN (Res):** a variant of CRNN that uses residual connections to improve the information flow and gradient propagation [64].
- **MSL:** a prediction method that uses shapelet learning from multiple variables [65].
- **StemGNN:** a spectral temporal graph neural network that can capture both inter-series correlations and temporal dependencies in the spectral domain. It combines graph Fourier transform and discrete Fourier transform to model the complex patterns and dynamics in multivariate time series data [66].

5.3. Evaluation metrics

In this subsection, we present the evaluation metrics that we use to assess the accuracy and robustness of our proposed model and compare it with several state-of-the-art models for district heat load forecasting. We use three metrics to evaluate the performance of different models: root mean square error (RMSE), mean absolute error (MAE), and coefficient of variation of root mean square error (CVRMSE). RMSE measures the average magnitude of the errors between the predicted and actual values, MAE measures the average absolute errors, and CVRMSE measures the normalized RMSE by dividing it by the mean of the actual values. Lower values of these metrics indicate higher accuracy of the model. These metrics are commonly used for district heat load forecasting because they can capture the overall error distribution and account for different scales of heat demand.

- Root Mean Square Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}, \tag{18}$$

- Mean Absolute Error (MAE):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{Y}_i - Y_i|, \tag{19}$$

Table 3
The multi-horizon one-step performance comparison in three metrics and four horizons on district heat load observations. The best result in terms of each metric is highlighted in gray, and the second best is underlined.

Horizon (h)	Metrics	AR	Dlinear	MTNet	TPA	LSTNet	LSTM	ED (GRU)	CRNN	CRNN (Res)	MSL	StemGNN	Ac-GRN (80%)	Ac-GRN (85%)	Ac-GRN (90%)
15	RMSE	11.983	12.031	11.062	10.998	11.383	23.330	22.294	17.564	13.584	15.690	11.297	12.374	11.035	10.852
	MAE	8.570	8.565	7.772	7.966	8.134	15.329	14.377	11.806	9.720	11.504	7.837	8.239	7.628	7.385
	CVRMSE	0.361	0.362	0.333	0.331	0.343	0.702	0.671	0.529	0.409	0.472	0.340	0.372	0.332	0.327
30	RMSE	14.592	14.459	12.994	11.852	12.930	23.994	23.237	18.265	15.928	16.990	12.947	12.341	11.554	11.074
	MAE	10.826	10.848	9.500	8.737	9.325	17.157	16.154	12.634	11.648	12.668	9.047	8.639	8.214	7.765
	CVRMSE	0.439	0.435	0.391	0.357	0.389	0.722	0.699	0.550	0.479	0.511	0.390	0.371	0.348	0.333
45	RMSE	16.970	16.983	13.835	13.386	15.254	23.504	23.687	18.729	18.450	17.878	16.320	12.537	12.175	11.864
	MAE	13.015	13.048	9.965	9.913	10.979	17.218	16.730	13.166	13.905	13.299	11.909	8.925	8.596	8.397
	CVRMSE	0.511	0.511	0.416	0.403	0.459	0.707	0.713	0.564	0.555	0.538	0.491	0.377	0.366	0.357
60	RMSE	20.454	20.413	17.085	14.811	17.002	24.199	24.823	19.358	20.103	19.850	17.855	14.120	12.706	12.526
	MAE	15.627	15.705	12.545	11.033	12.325	17.790	17.955	14.092	15.772	14.838	13.251	10.374	8.998	9.021
	CVRMSE	0.616	0.614	0.514	0.446	0.512	0.728	0.747	0.583	0.605	0.597	0.537	0.425	0.382	0.377

- Coefficient of Variation of RMSE (CVRMSE):

$$\text{CVRMSE} = \frac{\text{RMSE}}{\bar{Y}}, \tag{20}$$

where $\hat{Y}_i$ is the predicted value, Y_i is the actual value, and $\bar{Y}$ is the mean of the actual values. n is the number of observations.

5.4. Comparison on multi-horizon prediction

We evaluate the performance of our proposed model Ac-GRN and compare it with 11 state-of-the-art baseline models for multi-horizon one-step district heat load forecasting using three metrics: RMSE, MAE, and CVRMSE. All comparable methods are trained with the whole samples (i.e., 100% sampling ratio). Table 3 shows the results of the comparison for four horizons (15, 30, 45, and 60 days). The results demonstrate that our model Ac-GRN with 90% sampling ratio consistently outperforms other models across all metrics and horizons, especially when the prediction horizon is longer. This indicates that our model can effectively capture the complex spatial–temporal dependencies in the data and actively learn from the most informative samples. Moreover, our model shows a remarkable improvement over other models when the prediction horizon is longer, which reflects its robustness and generalization ability. Among the baseline models, TPA and LSTNet perform better than the others, but they still lag behind the Ac-GRN models. This implies that using graph neural networks and active deep learning can improve the performance of district heat load forecasting.

We also evaluate the impact of different sampling ratios on the performance of our model. The sampling ratio is the percentage of samples that are selected by the active learning module. We use three sampling ratios: 80%, 85%, and 90%. The results show that our model can achieve high accuracy with a relatively low sampling ratio (80%), which demonstrates the efficiency and effectiveness of our active learning strategy. However, increasing the sampling ratio to 90% can further improve the accuracy of our model, especially for longer prediction horizons. This suggests that there is a trade-off between the accuracy and efficiency of our model, and that choosing an appropriate sampling ratio can balance the costs associated with data collection and the model performance. We also compare the results of different sampling ratios with each other and with the baseline models. We find that our model with an 80% sampling ratio can outperform most of the baseline models except for TPA and LSTNet, while our model with 90% sampling ratio can outperform all of them. This shows that our model can achieve competitive or superior results with fewer heat meter samples than other models.

Table 4 presents the prediction performance comparison considering the meteorological factors. The Dlinear, MTNet, TPA, MSL, and StemGNN models are excluded from the comparison as they do not take exogenous factors into account in their prior studies. Across all prediction horizons and metrics, Ac-GRN consistently achieves the best performance, with the lowest RMSE, MAE, and CVRMSE at 90% sample ratio. These results underscore the strong predictive performance of Ac-GRN in handling district heat load observations combined with meteorological factors, reflecting its ability to effectively incorporate and utilize exogenous data. For AR, LSTM, ED (GRU), and CRNN (Res), the introduction of weather data generally improved the accuracy, substantiating the importance of external data integration in optimizing district heat load predictions. The meteorological factors are likely to capture more complexity and variability of the district heat load, which are further leveraged by these models to yield more accurate predictions.

It is worth noting that the CRNN model's performance enhancement appears to be limited to a horizon of 15. This limitation could potentially be attributed to the nature of convolutional layers, which primarily extract local features in the data. While beneficial for certain tasks, they may fail to capture long-term dependencies or complex patterns that emerge over larger time horizons. As a result, the CRNN model demonstrates increased performance at a shorter horizon ($h = 15$) and performs relatively less effectively as the horizon increases. Moreover, the results suggest that the addition of residual connections in CRNN (Res) contributes to its superior stability and robustness compared to the standard CRNN. These connections help to alleviate the vanishing gradient problem and allow for more complex feature extraction, ultimately enhancing the model's performance across all horizons.

5.5. Comparison on multi-horizon multi-step prediction

Table 5 presents the results of our proposed model Ac-GRN and four state-of-the-art methods for multi-horizon multi-step district heat load forecasting. We selected these four methods based on their relatively strong performance in the multi-horizon one-step prediction task. We use three metrics: RMSE, MAE, and CVRMSE, to evaluate the accuracy and robustness of the models. We also analyze the impact of different sampling ratios on our model's performance. The results show that our model Ac-GRN consistently outperforms other models across all metrics and horizons, especially when the prediction horizon and step are longer. This indicates that our model can effectively capture the complex spatial–temporal dependencies in the data and actively learn from the most informative samples. Moreover, our model shows a remarkable improvement over other models when the prediction horizon and step are longer, which reflects its robustness and generalization ability. Our model also achieves high accuracy with a relatively low

Table 4
The multi-horizon one-step performance comparison in three metrics and four horizons on district heat load observations with meteorological factors. The best result in terms of each metric is highlighted in gray.

Horizon (h)	Metrics	AR	LSTM	ED (GRU)	CRNN	CRNN (Res)	Ac-GRN (90%)
15	RMSE	11.461	12.266	11.255	17.493	13.096	10.303
	MAE	8.393	9.102	8.050	11.725	9.344	7.130
	CVRMSE	0.345	0.369	0.339	0.527	0.394	0.312
30	RMSE	12.210	14.531	14.825	18.338	15.496	10.728
	MAE	9.033	10.958	11.170	12.798	11.312	7.630
	CVRMSE	0.368	0.437	0.446	0.552	0.466	0.325
45	RMSE	12.791	14.684	16.300	19.088	17.744	10.458
	MAE	9.415	11.202	12.405	13.677	13.357	7.435
	CVRMSE	0.385	0.442	0.491	0.575	0.534	0.317
60	RMSE	13.099	17.140	18.669	19.135	19.820	11.457
	MAE	9.687	13.514	14.788	13.696	15.388	8.277
	CVRMSE	0.394	0.516	0.562	0.576	0.597	0.347

Table 5
The multi-horizon multi-step performance comparison in three metrics and four horizons on district heat load observations.

Step	Horizon (h)	Metrics	Dlinear	LSTNet	TPA	MTNet	Ac-GRN (80%)	Ac-GRN (85%)	Ac-GRN (90%)
7	15	RMSE	12.556	11.995	11.306	11.789	10.998	10.966	10.821
		MAE	9.064	8.586	8.197	8.207	7.852	7.872	7.791
		CVRMSE	0.388	0.371	0.350	0.365	0.340	0.339	0.335
	30	RMSE	15.294	13.460	12.354	14.648	11.075	11.220	11.143
		MAE	11.292	9.813	9.067	10.382	8.004	8.089	8.005
		CVRMSE	0.473	0.416	0.382	0.453	0.343	0.347	0.345
	45	RMSE	18.101	16.783	14.681	15.323	12.786	12.133	12.039
		MAE	13.289	11.912	10.743	10.875	9.220	8.770	8.751
		CVRMSE	0.560	0.519	0.454	0.474	0.396	0.375	0.372
	60	RMSE	21.974	17.336	17.010	16.678	15.789	15.842	15.683
		MAE	16.184	12.314	12.273	12.010	12.078	12.113	11.994
		CVRMSE	0.680	0.536	0.526	0.516	0.488	0.490	0.485
14	15	RMSE	13.271	13.432	11.391	12.770	11.795	11.769	10.995
		MAE	9.528	9.738	8.276	8.951	8.494	8.474	7.815
		CVRMSE	0.419	0.424	0.359	0.403	0.372	0.371	0.347
	30	RMSE	16.072	14.491	12.104	14.719	12.159	11.829	11.686
		MAE	11.741	10.511	8.924	10.608	8.919	8.588	8.492
		CVRMSE	0.507	0.457	0.382	0.464	0.384	0.373	0.369
	45	RMSE	19.292	17.055	14.714	16.918	14.448	12.611	12.453
		MAE	13.970	12.182	10.799	12.155	10.917	9.296	8.765
		CVRMSE	0.609	0.538	0.464	0.534	0.456	0.398	0.393
	60	RMSE	23.498	17.499	16.303	18.209	16.048	16.000	15.868
		MAE	17.269	12.665	12.024	13.808	12.272	12.243	12.100
		CVRMSE	0.741	0.552	0.514	0.574	0.506	0.505	0.501

Table 6
Ablation analysis. The best result is highlighted in gray, while a wavy line indicates the worst result.

Horizon (h)	Metrics	Ac-GRN	Ac-GRN w/o TCM	Ac-GRN w/o GRU	Ac-GRN w/o Attention	Ac-GRN w/o LEU	Ac-GRN w/o Generate
15	RMSE	10.852	12.074	11.670	11.671	12.820	13.455
	MAE	7.385	8.484	7.813	8.185	9.403	9.871
	CVRMSE	0.327	0.363	0.351	0.351	0.386	0.405
30	RMSE	11.074	12.924	11.958	11.855	13.098	13.751
	MAE	7.765	9.435	8.398	8.301	9.655	10.253
	CVRMSE	0.333	0.389	0.360	0.357	0.394	0.414
45	RMSE	11.864	13.446	12.159	12.448	17.042	14.709
	MAE	8.397	10.041	8.671	8.919	12.971	10.910
	CVRMSE	0.357	0.405	0.366	0.375	0.513	0.443
60	RMSE	12.526	14.910	12.638	15.599	16.902	15.018
	MAE	9.021	11.134	8.920	11.789	12.739	11.232
	CVRMSE	0.377	0.449	0.380	0.470	0.509	0.452

sampling ratio (80%), which demonstrates the efficiency and effectiveness of our active learning strategy. This table complements the results in Table 3, which only considers one-step prediction for different horizons. It shows that our model can also handle multi-step prediction for different horizons, which is more challenging and realistic for district heat load forecasting.

5.6. Ablation and sensitivity analysis

To evaluate the contribution of each component of our proposed Ac-GRN model, we conduct an ablation study by removing one component at a time and comparing the performance with the full model. The components considered for this analysis include the spatial–temporal

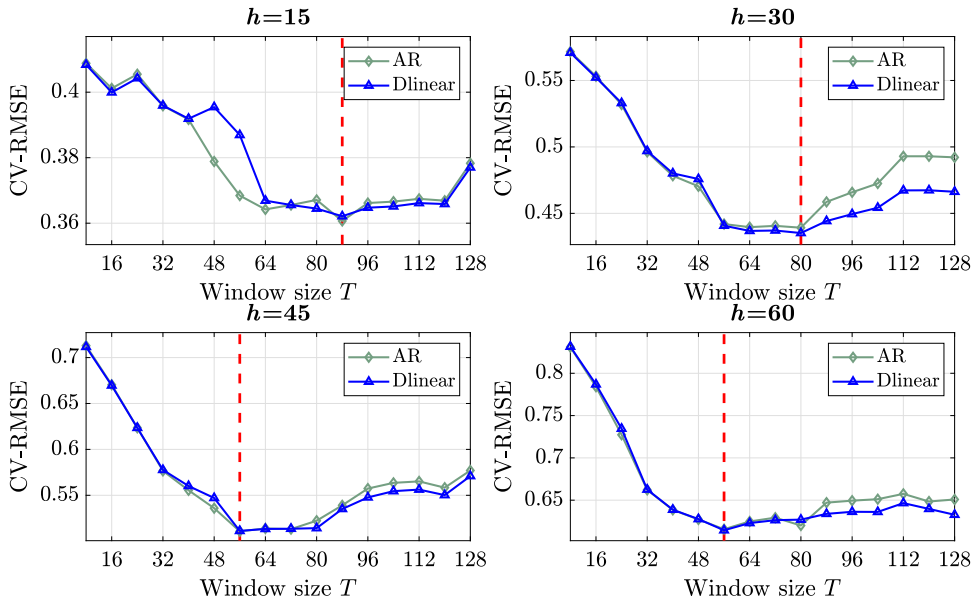


Fig. 7. Sensitivity analysis of window size T in terms of CV-RMSE. The optimal value is found at the red dash line.

convolution (TCM) module, the GRU and Attention in the spatio-temporal memory enhancement module, the linear representation unit (LEU) in the multi-component fusion, and the generation component in the feature-enhanced module. Table 6 shows the results of the ablation analysis in terms of RMSE, MAE, and CV-RMSE metrics for four horizons (15, 30, 45, and 60 days). The results show that removing any component leads to a decrease in accuracy and an increase in error metrics for all horizons. This indicates that each component is essential for the effectiveness of our model and that they work well together. The most significant drop in performance occurs when we remove the Generate or the LEU components, depending on the horizon. The Generation component in the feature-enhanced module plays a crucial role in complementing the incomplete data, which provide by active learning. This mitigates potential accuracy loss caused by missing data from unselected heat meters. The LEU in the multi-component fusion aids in enhancing the trend features of model inputs via linear embedding, which helps improve the model's ability to capture and predict heat load trends accurately. For short horizons (15 and 30 days), removing the Generate component causes the largest performance degradation. This observation implies that generating informative samples is vital to accurately capture the short-term dynamics of heat load. For long horizons (45 and 60 days), removing the LEU component causes the largest performance degradation, which suggests that enhancing the linear features is crucial for capturing the long-term trends of heat load.

To examine how the performance of our model changes with different values of a parameter or a variable, we conduct a sensitivity analysis by varying one parameter or variable at a time and measuring the impact on the performance. Fig. 7 shows the results of the sensitivity analysis for the window size T, which is the number of historical observations used as input for our model. AR and Dlinear are chosen for the experiment as they do not rely on other hyper-parameters, allowing for a more transparent view of how changes in window size T affect outcomes. The results demonstrate how the CV-RMSE metric varies with different window sizes T for four different forecasting horizons (15, 30, 45, and 60 days). These results reveal that an optimal value of T exists for each horizon that minimizes the CV-RMSE metric and maximizes the accuracy of our model. The optimal values of T are 88, 80, 55, and 55 for horizons 15, 30, 45, and 60 days, respectively. This suggests that our model can capture the temporal dependencies in the data better when using an appropriate window size. If the window size is too small, our model may not have enough information to make accurate predictions. If the window size is too large, our model may suffer from overfitting or noise. All comparisons are made using the optimal window size settings for each method.

The ablation analyses demonstrate the robustness and generalization capabilities of our proposed Ac-GRN model. Meanwhile, the sensitivity analyses offer valuable insights into the significance and influence of various parameters within our model for district heat load forecasting. These analyses collectively affirm the potential of the Ac-GRN model as a reliable tool for district heat load prediction.

5.7. Computational efficiency and sampling ratio analysis

In this subsection, we compare the computational efficiency and sampling ratio analysis of different methods for district heat load prediction. The computational efficiency is measured by the running time per batch for prediction processing, while the sampling ratio indicates the proportion of annotated samples used for training. The results are shown in the two tables, which report the running time and the prediction accuracy for different methods under different horizons and sampling ratios, respectively. The relatively worse result in terms of each metric is highlighted in red.

We compared the computational efficiency of the proposed Ac-GRN with other comparable methods for district heat load prediction without meteorological factors, as shown in Table 7. The prediction horizon h and the prediction step s were fixed at 15 and 1, respectively. The performances are evaluated based on two statistical measures: the mean and the standard deviation, computed over 20 repetitions of the experiment for each model. The mean represents the average processing time per batch, while the standard deviation measures the degree of variation in the processing times. As can be seen from Table 7, the AR model has the fastest running time, with a mean value of 0.00103 s per batch. This is likely attributable to the linear structure of the AR model, which, while beneficial for computational efficiency, could limit its performance in long-horizon predictions. The StemGNN model has the slowest running time, with a mean value of 1.74973 s per batch, which is considerably higher than any of the other models. The performance of StemGNN also has a large variation, as indicated by its high standard deviation of 0.00689. The proposed model (Ac-GRN) has a moderate running time, with a mean value of 0.01058 s per batch. Although not the fastest, our proposed method still demonstrates reasonable efficiency, especially when compared to the StemGNN model. Both methods are based on graph neural network structures, with our method being over 160 times faster than StemGNN. This suggests

Table 7
The running time per batch for prediction processing of comparable methods and Ac-GRN. The best and worst methods are highlighted in green and red, respectively.

Time (s/batch)	AR	Dlinear	MTNet	TPA	LSTNet	LSTM	ED (GRU)	CRNN	CRNN (Res)	MSL	StemGNN	Ac-GRN (90%)
Mean value	**0.00103**	0.00182	0.00537	0.00697	0.01104	0.00728	0.00700	0.00571	0.00314	0.00105	**1.74973**	0.01058
Standard deviation	0.00017	0.00022	0.00055	0.00144	0.00309	0.00446	0.00299	0.00180	0.00303	0.00022	0.00689	0.00256

Table 8
The multi-horizon one-step performance comparison in three metrics and four horizons on district heat load observations with different sampling ratios. The relatively worse result in terms of each metric is highlighted in red.

Horizon (h)	Metrics	Dlinear (80%)	Dlinear (90%)	Dlinear (100%)	MTNet (80%)	MTNet (90%)	MTNet (100%)	TPA (80%)	TPA (90%)	TPA (100%)	LSTNet (80%)	LSTNet (90%)	LSTNet (100%)	Ac-GRN (80%)	Ac-GRN (90%)
15	RMSE	18.160	16.644	12.031	12.363	12.315	11.062	15.244	13.068	10.998	13.684	12.516	11.383	12.374	10.852
	MAE	12.719	12.121	8.565	8.581	8.529	7.772	10.526	9.158	7.966	9.284	8.633	8.134	8.239	7.385
	CVRMSE	0.547	0.501	0.362	0.372	0.371	0.333	0.459	0.393	0.331	0.412	0.377	0.343	0.372	0.327
30	RMSE	20.320	18.168	14.459	14.450	13.873	12.994	14.798	13.869	11.852	15.741	14.550	12.930	12.341	11.074
	MAE	15.771	13.680	10.848	10.207	9.785	9.500	10.700	9.955	8.737	11.187	10.415	9.325	8.639	7.765
	CVRMSE	0.612	0.547	0.435	0.435	0.418	0.391	0.445	0.417	0.357	0.474	0.438	0.389	0.371	0.333
45	RMSE	21.257	19.214	16.983	15.656	**15.689**	13.835	16.712	15.503	13.386	16.629	15.748	15.254	12.537	11.864
	MAE	16.859	15.087	13.048	11.451	11.312	9.965	12.369	11.392	9.913	11.871	11.218	10.979	8.925	8.397
	CVRMSE	0.640	0.578	0.511	0.471	**0.472**	0.416	0.503	0.467	0.403	0.501	0.474	0.459	0.377	0.357
60	RMSE	23.490	22.351	20.413	17.852	**17.977**	17.085	17.790	16.389	14.811	21.136	19.168	17.002	14.120	12.526
	MAE	18.980	17.868	15.705	13.036	12.838	12.545	13.209	12.164	11.033	15.841	13.968	12.325	10.374	9.021
	CVRMSE	0.707	0.673	0.614	0.537	**0.541**	0.514	0.535	0.493	0.446	0.636	0.577	0.512	0.425	0.377

that Ac-GRN is capable of balancing the computational efficiency and prediction performance more effectively than some of the other models. The standard deviations for each method are relatively small compared to their corresponding means, indicating a relatively tight distribution of running times. The largest standard deviation is for StemGNN, which could suggest variability in its performance.

Table 8 shows the multi-horizon one-step performance comparison of different methods for district heat load prediction with different sampling ratios. The sampling ratios indicate the proportion of annotated samples used for training, ranging from 80% to 100%. The prediction accuracy is measured by three evaluation metrics: MAE, RMSE, and CVRMSE. The relatively worse result in terms of each metric is highlighted in red. As can be seen from Table 8, our proposed model (Ac-GRN) achieves the best performance in most cases, especially in longer horizons (45 and 60), where it significantly outperforms other methods. The only exception is in horizon 15, where Dlinear has a slightly lower MAE and RMSE than our proposed model with 80% and 90% sampling ratios, but our proposed model still has a lower CVRMSE than Dlinear with all sampling ratios. This demonstrates that our proposed model is more accurate and robust than other methods under different sampling ratios. The other methods have varying performances under different horizons and sampling ratios, but generally, they have higher MAE, RMSE, and CVRMSE than our proposed model. The Dlinear model has the worst performance in most cases, which is likely due to its linear structure that cannot capture the complex spatial–temporal dependencies in the data. The MTNet model has a relatively good performance in shorter horizons (15 and 30), but it deteriorates in longer horizons (45 and 60), which could indicate its difficulty in handling long-term dependencies. The TPA model has a moderate performance in most cases, but it has a high CVRMSE in horizon 45 with an 80% sampling ratio, which could suggest its instability under data scarcity. The LSTNet model has poor performance in most cases, which could be attributed to its high complexity and computational cost.

In summary, our proposed model (Ac-GRN) outperforms other methods in terms of prediction accuracy and computational efficiency under different sampling ratios. This shows that our proposed model can effectively leverage the spatial–temporal correlations among heat meters using active deep learning and graph neural networks, while other methods may suffer from data sparsity or complexity issues.

6. Discussions

In this paper, we proposed Ac-GRN, a novel model that combines active deep learning and graph neural networks for district heat load forecasting. Our model consists of three main components: a knowledge extractor that uses active learning to select the most informative heat meters, a feature reconstructor that imputes the missing data of unselected meters using correlation analysis, and a graph recurrent network that models the spatial–temporal dependencies among heat meters. Our model can capture the complex patterns and relationships in the data and provide accurate and robust forecasts for different horizons and steps.

One of the main contributions of our paper is that we use active learning to reduce the frequency and cost of data collection and human effort. By selecting only a small subset of heat meters that are most representative and informative, we can train our model with fewer data and achieve better performance than using all the data. We also use correlation analysis to reconstruct the missing data of unselected meters, which improves the data quality and completeness. We show that our model can achieve high accuracy with only 80% of the data, which demonstrates the efficiency and effectiveness of our active learning strategy.

Another contribution of our paper is that we use graph neural networks to model the spatial–temporal dependencies among heat meters. We represent the heat meters as nodes in a graph, where the edges are defined by the heat load similarity. We use a graph convolutional network to learn the spatial features of the nodes, and a gated recurrent unit to learn the temporal features of the sequences. We also use an attention mechanism to learn how to weigh different nodes and time steps for prediction. We show that our model can capture both linear and nonlinear relationships in the data, and outperform other methods that use different types of neural networks.

A third contribution of our paper is that we provide explanations for our model and its predictions. Our proposed Ac-GRN model leverages active deep learning methods and correlation-based techniques to offer explainability. A novel framework was designed to select representative smart heat meters by active deep learning. This framework identifies both representative and non-representative meter groups using the least confidence measure. The annotated results of these meter groups can provide actionable feedback to corporations or governmental bodies, thereby aiding in heating regulation and management. Further, we introduce a feature reconstructor to interpolate incomplete data based on correlations among heat meters. This reconstruction process is both

transparent and interpretive, effectively linking unselected heat meters with their correlated, annotated counterparts. In addition, we use attention mechanisms to explain how our model attends to different nodes in the graph and different time steps in the time series. These attention weights can be visualized, demonstrating how they adjust depending on the input data and the horizon. To provide a concrete example of our model's explainability, consider a scenario where our model is predicting heat load for a specific time period. The active deep learning component identifies the heat meters that are most influential for the prediction. The feature reconstructor then fills in any missing data in the heat meter readings, using correlations among the meters. This reconstructed data is transparent and interpretable, offering a clear view of how the model arrived at its prediction.

Despite the promising results of our model, we acknowledge that there are some limitations and challenges that need to be addressed in future research. First, our model relies on a fixed graph structure to represent the spatial dependencies among heat meters, which may not capture the dynamic changes in the network topology or node attributes over time. For example, there may be new nodes added or removed, or node features updated over time. Second, our model uses a single prediction horizon for all heat meters, which may not be optimal for each meter. For example, some meters may have more stable or predictable patterns than others, and may require different prediction horizons. Third, our Ac-GRN model requires a large amount of computational resources and time to train and update the model parameters, especially when dealing with large-scale and high-dimensional district heating data. This may limit the applicability and scalability of our model in real-time and resource-constrained settings, where timely and efficient forecasting is essential for optimizing energy production and distribution. Fourth, our model does not provide interpretable explanations for its predictions, which may limit its applicability and trustworthiness in real-world scenarios. For example, it may be useful to know why some meters have higher or lower heat load than others, or what factors contribute most to the prediction.

To address these limitations, we propose some possible directions for future research. First, we can use dynamic graph neural networks to model the spatial dependencies among heat meters, such as temporal graph convolutional networks [42], recurrent graph neural networks [47], or graph attention networks [46]. These networks can handle the changes in the graph structure or node attributes over time and learn more expressive representations of the data. Dynamic graph neural networks have been shown to improve the accuracy and robustness of district heat load forecasting in previous studies [67,68]. Second, we can use multi-task learning or multi-output learning to predict different horizons for different heat meters simultaneously. These learning methods can leverage the shared information and correlations among different horizons and meters and improve the accuracy and efficiency of our model. Multi-task learning or multi-output learning can also reduce the complexity and redundancy of the model by sharing parameters and features across different tasks or outputs [69]. Third, we can use explainable artificial intelligence techniques to provide interpretable explanations for our predictions, such as SHAP values [70], LIME [71], or counterfactuals [72]. These techniques can help us understand how our model works and what factors influence its predictions. Explainable artificial intelligence techniques can also enhance the trust and reliability of our model and support decision-making for optimizing energy production and distribution [73]. Last, to further validate and refine our model's explanations, we plan to conduct user studies or expert evaluations. These studies will involve presenting the model's explanations to users or experts and collecting their feedback on the clarity, utility, and relevance of the explanations. This feedback will provide valuable insights into how well users or experts understand the explanations and how they might use the explanations in real-world decision-making scenarios. We believe that this additional validation will not only strengthen our model's ability to provide "meaningful explanations" but also provide directions for further improving the model's explainability.

7. Conclusions and future work

District heat load forecasting is an important task for improving the efficiency and sustainability of district heating systems. However, existing methods for district heat load forecasting face several challenges, such as capturing the complex spatial–temporal dependencies among heat meters, reducing the frequency and cost of data collection and human effort, and providing interpretable explanations for the predictions. In this paper, we addressed these challenges by proposing Ac-GRN, a novel model that combines active deep learning and graph neural networks for district heat load forecasting. Our model can effectively select the most informative and representative samples from a large pool of data, and use them to train a graph recurrent neural network with bidirectional recurrent connections. Our model can capture both linear and nonlinear relationships in the data and provide accurate and robust forecasts for different horizons and steps. We evaluated our model on a real-world dataset of district heating consumption data from Danish residential buildings and compared it with 11 state-of-the-art methods. We conducted extensive experiments to test the performance of our model for different prediction horizons, steps, and sampling ratios. The results indicate that our model achieved superior performance in terms of accuracy, robustness, and reliability for multi-horizon multi-step district heat load forecasting. We also compared the computational efficiency of our model with other methods and found that our model has a reasonable running time per batch for prediction processing, especially when compared to other graph neural network-based methods. This shows that our model can balance the trade-off between computational efficiency and prediction performance more effectively than some of the other methods.

For future work, we plan to extend our model to handle dynamic graph structures and node attributes, to predict different horizons for different heat meters, and to provide interpretable explanations for our predictions. We also plan to apply our model to other domains that involve spatial–temporal data, such as traffic flow forecasting, air quality forecasting, and social network analysis. We hope that our work can inspire more research on active deep learning and graph neural networks for district heat load forecasting and other related tasks.

CRediT authorship contribution statement

Yaohui Huang: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Conceptualization. **Yuan Zhao:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Zhijin Wang:** Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Xiufeng Liu:** Writing – review & editing, Writing – original draft, Supervision, Data curation, Conceptualization. **Hanjing Liu:** Writing – review & editing, Writing – original draft, Methodology. **Yonggang Fu:** Writing – review & editing, Project administration, Conceptualization.

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.

Data availability

The authors do not have permission to share data.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used the language polishing service provided by ChatGPT in order to the readability and clarity. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

This research was supported in part by the Fujian Province Natural Science Foundation of Fujian Province (CN) (nos. 2021J01857, 2021J01859, and 2022J01335), and the EMB3Rs project (no. 84712) funded by the European Union Horizon 2020 research and innovation programme. We thank the editors and anonymous reviewers for their valuable comments and suggestions, which greatly benefited this paper.

Appendix

A.1. Hyper-parameter setting

The parameters setting of the proposed method and benchmarks are listed in Table 9.

Table 9
Hyper-parameter settings.

Model	Parameter	Option range
LSTM ED (GRU)	Hidden size	$\{2^4, 2^5, 2^6\}$
CRNN	Kernel size	3–9 (2 per step)
	CNN out channels	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
CRNN (Res)	Hidden size	$\{2^4, 2^5, 2^6\}$
CRNN (Res)	Residual window size	1–5 (1 per step)
TPA-LSTM	Kernel size	3–9 (2 per step)
	CNN out channels	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
LSTNet	GRU hidden size	$\{2^4, 2^5, 2^6\}$
MTNet	The number of GRU layers	1–3 (1 per step)
	Highway window size	1–10 (1 per step)
LSTNet	Skip window size	1–3 (1 per step)
	Skip GRU hidden size	$\{2^4, 2^5, 2^6\}$
MTNet	Block size	1–10 (1 per step)
MSL	Shapelet size	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
StemGNN	Encoder layers	1–3 (1 per step)
	The numbers of blocks	1–3 (1 per step)
Ac-GRN	GCN hidden size	$\{2^4, 2^5, 2^6\}$
	GCN out channel	$\{2^3, 2^4, 2^5, 2^6\}$
	GRU hidden size	$\{2^4, 2^5, 2^6\}$
	Kernel size	3–9 (2 per step)
	The number of correlated meters for reconstruction	$\{3, 5, 10, 15, 20\}$
	CNN out channels	$\{2^3, 2^4, 2^5, 2^6\}$

References

[1] Werner S. International review of district heating and cooling. Energy 2017;137:617–31.
[2] Rezaie B, Rosen MA. District heating and cooling: Review of technology and potential enhancements. Appl Energy 2012;93:2–10.
[3] Tardioli G, Kerrigan R, Oates M, James O, Finn D. Data driven approaches for prediction of building energy consumption at urban level. Energy Procedia 2015;78:3378–83.
[4] Ding Y, Timoudas TO, Wang Q, Chen S, Brattebø H, Nord N. A study on data-driven hybrid heating load prediction methods in low-temperature district heating: An example for nursing homes in nordic countries. Energy Convers Manage 2022;269:116163.
[5] Garcia O. A stochastic differential equation model for the height growth of forest stands. Biometrics 1983;1059–72.
[6] Gadd H, Werner S. Daily heat load variations in Swedish district heating systems. Appl Energy 2013;106:47–55.
[7] Dalipi F, Yildirim Yayilgan S, Gebremedhin A. Data-driven machine-learning model in district heating system for heat load prediction: A comparison study. Appl Comput Intell Soft Comput 2016;2016.
[8] Chung WH, Gu YH, Yoo SJ. District heater load forecasting based on machine learning and parallel CNN-LSTM attention. Energy 2022;246:123350.
[9] Thommessen C, Soltysik S, Roes J. Heat load forecasting for district heating systems using neural networks. In: 2022 international conference on electrical, computer, communications and mechatronics engineering. IEEE; 2022, p. 1–6.
[10] Zhang L, Wen J. Active learning strategy for high fidelity short-term data-driven building energy forecasting. Energy Build 2021;244:111026.
[11] Shwartz-Ziv R, Tishby N. Opening the black box of deep neural networks via information. 2017, arXiv preprint arXiv:1703.00810.
[12] Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inform Fusion 2020;58:82–115.
[13] Xie J, Zhong Y, Xiao T, Wang Z, Zhang J, Wang T, et al. A multi-information fusion model for short term load forecasting of an architectural complex considering spatio-temporal characteristics. Energy Build 2022;277:112566.
[14] Liu P, Wang L, Ranjan R, He G, Zhao L. A survey on active deep learning: From model driven to data driven. ACM Comput Surv 2022;54(10s).
[15] Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY. A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst 2020;32(1):4–24.
[16] Ying Z, Bourgeois D, You J, Zitnik M, Leskovec J. Gnnexplainer: Generating explanations for graph neural networks. In: Advances in Neural Information Processing Systems, vol. 32. 2019.
[17] Xue P, Jiang Y, Zhou Z, Chen X, Fang X, Liu J. Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms. Energy 2019;188:116085.
[18] Dang LM, Shin J, Li Y, Tightiz L, Nguyen TN, Song H-K, et al. Toward explainable heat load patterns prediction for district heating. Sci Rep 2023;13(1):7434.
[19] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018;81:1192–205.
[20] Afram A, Janabi-Sharifi F. Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods. Energy Build 2015;94:121–49.
[21] Zhou Q, Wang S, Xu X, Xiao F. A grey-box model of next-day building thermal load prediction for energy-efficient control. Int J Energy Res 2008;32(15):1418–31.
[22] Idowu S, Saguna S, Åhlund C, Schelén O. Applied machine learning: Forecasting heat load in district heating system. Energy Build 2016;133:478–88.
[23] Said Z, Sundar LS, Tiwari AK, Ali HM, Sheikholeslami M, Bellos E, et al. Recent advances on the fundamental physical phenomena behind stability, dynamic motion, thermophysical properties, heat transport, applications, and challenges of nanofluids. Phys Rep 2022;946:1–94.
[24] Lin X, Chen R, Huang L, Liu Z, Niu X, Guo G, et al. ChirpTracker: A precise-location-aware system for acoustic tag using single smartphone. IEEE Internet Things J 2023. Early Access.
[25] Chen S, Zhou X, Zhou G, Fan C, Ding P, Chen Q. An online physical-based multiple linear regression model for building's hourly cooling load prediction. Energy Build 2022;254:111574.
[26] Wang Z, Hong T, Piette MA. Building thermal load prediction through shallow machine learning and deep learning. Appl Energy 2020;263:114683.
[27] Alabi TM, Aghimien EI, Agbajor FD, Yang Z, Lu L, Adeoye AR, et al. A review on the integrated optimization techniques and machine learning approaches for modeling, prediction, and decision making on integrated energy systems. Renew Energy 2022;194:822–49.
[28] Hossain MM, Zhang T, Ardakanian O. Identifying grey-box thermal models with Bayesian neural networks. Energy Build 2021;238:110836.
[29] Thilker CA, Bacher P, Bergsteinsson HG, Junker RG, Cali D, Madsen H. Non-linear grey-box modelling for heat dynamics of buildings. Energy Build 2021;252:111457.
[30] Du M, Liu N, Hu X. Techniques for interpretable machine learning. Commun ACM 2019;63(1):68–77.
[31] Chen S, Ren Y, Friedrich D, Yu Z, Yu J. Sensitivity analysis to reduce duplicated features in ANN training for district heat demand prediction. Energy AI 2020;2:100028.
[32] Zhao A, Mi L, Xue X, Xi J, Jiao Y. Heating load prediction of residential district using hybrid model based on CNN. Energy Build 2022;266:112122.
[33] Guo J, Yun S, Meng Y, He N, Ye D, Zhao Z, et al. Prediction of heating and cooling loads based on light gradient boosting machine algorithms. Build Environ 2023;236:110252.
[34] Li J, Gu C, Wei X, Gil IH, Xiang Y. An IoT-based thermal modelling of dwelling rooms to enable flexible energy management. IEEE Trans Smart Grid 2023;Early Access.
[35] Sener O, Savarese S. Active learning for convolutional neural networks: A core-set approach. 2017, arXiv preprint arXiv:1708.00489.

[36] Settles B, Craven M. An analysis of active learning strategies for sequence labeling tasks. In: Proceedings of the 2008 conference on empirical methods in natural language processing. 2008, p. 1070–9.
[37] Aryandoust A, Patt A, Pfenninger S. Enhanced spatio-temporal electric load forecasts using less data with active deep learning. Nat Mach Intell 2022;4(11):977–91.
[38] Wang Z, Zhao B, Guo H, Tang L, Peng Y. Deep ensemble learning model for short-term load forecasting within active learning framework. Energies 2019;12(20):3809.
[39] Gavves E, Mensink T, Tommasi T, Snoek CG, Tuytelaars T. Active transfer learning with zero-shot priors: Reusing past datasets for future tasks. In: Proceedings of the IEEE international conference on computer vision. 2015, p. 2731–9.
[40] Han W, Coutinho E, Ruan H, Li H, Schuller B, Yu X, et al. Semi-supervised active learning for sound classification in hybrid learning environments. PLoS One 2016;11(9):e0162075.
[41] Fang M, Li Y, Cohn T. Learning how to active learn: A deep reinforcement learning approach. 2017, arXiv preprint arXiv:1708.02383.
[42] Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. 2017, arXiv preprint arXiv:1710.10903.
[43] Wang H, Zhao M, Xie X, Li W, Guo M. Knowledge graph convolutional networks for recommender systems. In: The world wide web conference. 2019, p. 3307–13.
[44] Monti F, Boscaini D, Masci J, Rodola E, Svoboda J, Bronstein MM. Geometric deep learning on graphs and manifolds using mixture model cnns. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 5115–24.
[45] Yu F, Zhu Y, Liu Q, Wu S, Wang L, Tan T. TAGNN: Target attentive graph neural networks for session-based recommendation. In: Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval. 2020, p. 1921–4.
[46] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. 2016, arXiv preprint arXiv:1609.02907.
[47] Li Y, Tarlow D, Brockschmidt M, Zemel R. Gated graph sequence neural networks. 2015, arXiv preprint arXiv:1511.05493.
[48] Schaffer M, Tvedebrink T, Marszal-Pomianowska A. Three years of hourly data from 3021 smart heat meters installed in danish residential buildings. Sci Data 2022;9(1):420.
[49] Wang Z, Liu X, Huang Y, Zhang P, Fu Y. A multivariate time series graph neural network for district heat load forecasting. Energy 2023;278:127911.
[50] Abbas A, Noreen A, Ashraf Ali M, Ashraf M, Alzahrani E, Marzouki R, et al. Solar radiation over a roof in the presence of temperature-dependent thermal conductivity of a casson flow for energy saving in buildings. Sustain Energy Technol Assess 2022;53:102606.
[51] Wang Y, Liu K, Liu Y, Wang D, Liu J. The impact of temperature and relative humidity dependent thermal conductivity of insulation materials on heat transfer through the building envelope. J Build Eng 2022;46:103700.
[52] Bergsteinsson HG, Møller JK, Nystrup P, Pálsson ÓP, Guericke D, Madsen H. Heat load forecasting using adaptive temporal hierarchies. Appl Energy 2021;292:116872.
[53] Candes E, Recht B. Exact matrix completion via convex optimization. Commun ACM 2012;55(6):111–9.
[54] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: Proceedings of the 3rd international conference on learning representations. San Diego, CA, USA: OpenReview.net; 2015, p. Poster.
[55] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. In: Proceddings of the 33rd annual conference on neural information processing systems, vol. 32. Vancouver, BC, Canada; 2019, p. 8024–35.
[56] Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; 2015.
[57] Zeng A, Chen M, Zhang L, Xu Q. Are transformers effective for time series forecasting? 2022, arXiv preprint arXiv:2205.13504.
[58] Huang W, Stokes JW. MtNet: A multi-task neural network for dynamic malware classification. In: Detection of intrusions and malware, and vulnerability assessment: 13th international conference, DIMVA 2016, San SebastiáN, Spain, July 7-8, 2016, proceedings 13. Springer; 2016, p. 399–418.
[59] Shih S-Y, Sun F-K, Lee H-y. Temporal pattern attention for multivariate time series forecasting. Mach Learn 2019;108:1421–41.
[60] Lai G, Chang W-C, Yang Y, Liu H. Modeling long-and short-term temporal patterns with deep neural networks. In: The 41st international ACM SIGIR conference on research & development in information retrieval. 2018, p. 95–104.
[61] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[62] Cho K, Van Merri"enboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of the 2014 conference on empirical methods in natural language processing. 2014, p. 1724–34.
[63] Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K, Woo W-c. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. 2015, p. 802–10.
[64] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 770–8.
[65] Wang Z, Cai B. COVID-19 cases prediction in multiple areas via shapelet learning. Appl Intell 2022;52(1):595–606.
[66] Cao D, Wang Y, Duan J, Zhang C, Zhu X, Huang C, et al. Spectral temporal graph neural network for multivariate time-series forecasting. In: Advances in neural information processing systems, vol. 33. 2020, p. 17766–78.
[67] Benalcazar P, Kamiński J. Short-term heat load forecasting in district heating systems using artificial neural networks. IOP Conf Ser: Earth Environ Sci 2019;214(1):012023.
[68] Golla A, Geis J, Loy T, Staudt P, Weinhardt C. An operational strategy for district heating networks: Application of data-driven heat load forecasts. Energy Inform 2020;3:1–11.
[69] Zhang Y, Yang Q. An overview of multi-task learning. Natl Sci Rev 2018;5(1):30–43.
[70] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Advances in neural information processing systems, vol. 30. 2017.
[71] Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016, p. 1135–44.
[72] Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harv JL & Tech 2017;31:841.
[73] Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. 2017, arXiv preprint arXiv:1702.08608.