融合热惰性先验位置编码的 Transformer 区域供热日耗气量预测模型——以乌鲁木齐市为例

张淑雅 [1,*]，李振飞 [2]

[1] 天津商业大学理学院，天津 300134

[2] 天津商业大学信息工程学院，天津 300134

摘要：近年来，随着生态保护日益受到重视，中国政府确立了“碳中和”的发展目标，北方城市区域供热系统的能源供给正逐步由煤炭转向更清洁的天然气。天然气负荷的准确预测对能源调度规划与需求响应至关重要。然而，天然气负荷不仅受气象因素随机波动的影响，还因供热系统固有的热惰性而呈现出复杂的非线性与显著滞后特性，导致现有方法的预测精度难以满足实际应用需求。为此，本文以某市三家供热企业的实测运行数据为对象，提出一种基于热惰性位置编码序列 Transformer 模型（TI-STM）的区域供热系统天然气负荷预测算法。该方法针对供热负荷“近端影响强、远端影响弱”的物理规律，设计指数衰减形式的位置编码，以逐维可学习的衰减系数刻画热惰性在时间尺度上的差异，替代传统正弦位置编码；同时将未来时刻的气象要素作为已知协变量注入解码器，引导模型利用温度先验。模型借助多头注意力机制挖掘负荷序列中隐含的复杂特征，快速聚焦序列中不同的关键部分，在避免长距离依赖问题的同时减少信息损失。本文与 SVM、LSTM、MLP、RNN、TCN 和 STM 等先进算法进行了详细的对比实验。结果表明，所提模型具有更高的预测精度，其平均绝对百分比误差分别达到 1.77%、2.05%和 2.65%。

关键词：耗气量预测；热惰性; Transformer 模型；深度学习；超参数调节

DOI: 10.57237/j.jest.2026.02.002

District Heating Daily Gas Consumption Prediction Based on Transformer with Thermal Inertia Prior Position Encoding: A Case Study of Urumqi

Zhang Shuya[1,*], Li Zhenfei[2]

[1]School of Science, Tianjin University of Commerce, Tianjin 300134, China

[2]School of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China

Abstract: In recent years, with the increasing attention to ecological protection, the Chinese government has set the carbon neutrality development target. The energy supply of district heating systems in northern cities is gradually shifting from

*通信作者：张淑雅，shuyazhang_123@163.com

收稿日期: 2026-09-30；接受日期: 2026-10-24；在线出版日期: 2026-11-02

coal to cleaner natural gas. Accurate prediction of natural gas load is critical for energy dispatch planning and demand response. Nevertheless, natural gas load is not only affected by random fluctuations of meteorological factors, but also exhibits complex nonlinearity and significant hysteresis characteristics due to the inherent thermal inertia of heating systems, making it difficult for existing methods to meet the prediction accuracy requirements in practical applications. To address this issue, this paper adopts measured operational data from three heating enterprises in a city and proposes a natural gas load prediction algorithm for district heating systems based on a sequential Transformer model (TI-STM) with thermal inertia position encoding. According to the physical law that recent observations exert stronger impacts while distant ones exert weaker impacts on heating load, an exponentially decaying position encoding is designed. Dimension-wise learnable decay coefficients are adopted to characterize the temporal differences of thermal inertia, replacing the traditional sinusoidal position encoding. Meanwhile, meteorological variables of future time steps are fed into the decoder as known covariates to guide the model to utilize temperature priors. With the multi-head attention mechanism, the model mines complex implicit features within the load sequence and quickly focuses on different key segments of the sequence, mitigating long-range dependency problems and reducing information loss. Comprehensive comparative experiments are carried out against advanced algorithms including SVM, LSTM, MLP, RNN, TCN and TI-STM. The results demonstrate that the proposed model achieves superior prediction accuracy, with mean absolute percentage errors (MAPE) of 1.77%, 2.05% and 2.65%, respectively.

Keywords: Gas Consumption Prediction; Thermal Inertia Position Encoding; Transformer Model; Deep Learning; Hyperparameter Tuning

1 引言

近年来，中国北方城市供热规模持续扩张，建筑供暖能耗大幅增长，由此产生的碳排放与大气污染问题日益突出，严重制约了城市绿色低碳发展进程[1]。为推进“碳中和”发展战略，北方区域供热系统正逐步由传统燃煤供热向清洁天然气供热模式转型，天然气已成为区域供热领域的核心替代清洁能源。相较于煤炭，天然气具备热值高、碳排放强度低、污染物排放少等优势，但采购、储运与调峰成本相对较高，负荷受气象波动等因素影响随机性强，因此精准预测区域供热系统天然气消耗量，对保障供热系统稳定运行、优化能源调度配置、提升能源利用效率具有重要的工程价值与现实意义。建筑供热天然气消耗受气象条件、建筑特性、管网结构、系统热惰性等多重因素耦合影响，呈现出显著的非线性、滞后性与动态不确定性，精准建模预测难度较大，国内外学者围绕供热负荷与天然气能耗预测开展了大量研究。

在预测尺度与影响机理研究方面，现有供热能耗预测按时间尺度可分为年度、月度、日度及小时级等多个层级，建模输入特征主要包含历史能耗序列与气象环境数据。从热力学机理角度来看，供热负荷变化受热惰性、温度、压力、流量等物理规律约束，建筑形态与城市管网布局也会对能耗变化趋势产生深远影响。Zhou 等[2]结合管网热力学特性，探究了天然气管网运行优化与能量平衡耦合机制，为能耗机理建模提供了理论支撑；Szega 等[3]围绕热电联产关键设备开展热力学建模与校准研究，有效提升了热力负荷预测精度，其方法可为建筑供热天然气预测提供借鉴。

传统统计与回归模型是早期供热负荷预测的主流方法，凭借结构简单、可解释性强的优势被广泛应用。学者们基于相关性分析、主成分分析、偏最小二乘回归等方法，构建了适配不同场景的供热负荷预测模型。Lim 等[4]考虑建筑热惰性与时滞效应，构建了供暖负荷回归预测模型；Ding 等[5]结合小波分解与主成分分析，实现了区域供热系统短期与超短期负荷的精准预测；Yuan 等[6]通过特征筛选与偏最小二乘回归，建立了严寒地区农村住宅供暖能耗预测模型；Zhang 等[7]结合建筑温度特性，构建了基于模型校准的区域供热住宅供暖负荷预测模型。此类模型易于实现、适用性广，但也存在明显短板：对复杂多维的非线性能耗数据拟合能力不足，且易受异常数据干扰，预测精度难以满足高精度供热调度的需求。

为突破传统统计模型的局限性，数据驱动的机器学习算法被逐步应用于供热负荷预测领域，随机森林、支持向量机、多层感知机、极限梯度提升等算法成为研

究热点。Geysen 等[8]系统对比了线性回归、决策树、支持向量机、前馈神经网络等模型的热负荷预测性能；Ahmad 等[9]对比了多种数据挖掘算法在冷热负荷预测中的应用效果；Xue 等[10]构建了多步热负荷预测框架，验证了极限梯度提升算法在时序能耗预测中的优越性。诸多研究表明，机器学习模型可自动挖掘能耗数据中的隐含特征，有效拟合能耗与影响因素之间的非线性关联。Cai 等[11]、Moayedi 等[12]通过引入智能优化算法优化模型参数，进一步提升了供热负荷预测精度。但传统机器学习方法高度依赖人工特征工程，特征筛选质量受专家经验制约，对长时序、强滞后特性的供热天然气负荷拟合能力有限，模型泛化能力有待提升。

深度学习技术凭借强大的非线性特征提取能力，成为当前时序能耗预测的主流方法，可有效弥补传统机器学习模型的缺陷。现有研究多基于循环神经网络、长短期记忆网络、卷积神经网络及其混合模型开展供热负荷预测。Wang 等[13]利用长短期记忆网络捕捉时序依赖特征，显著提升了热负荷预测精度；Zhu 等[14]融合卷积神经网络与长短期记忆网络，结合 dropout 技术抑制过拟合，优化了短期负荷预测效果。针对单一模型特征提取不足的问题，诸多混合预测模型被相继提出，通过特征分解、注意力机制、聚类增强等方式优化模型性能。Xue 等[15]构建了基于注意力机制的 LSTM 预测模型，实现了负荷与多影响因素之间更精准的非线性拟合；Du 等[16]构建了卷积自编码器与时间注意力结合的混合深度学习框架，可精准捕捉热负荷时序特征；Liu 等[17]结合聚类算法与长短期记忆网络，通过特征增强的方式有效提升了短期天然气消耗预测精度。

尽管现有深度学习模型取得了较好的预测效果，但主流的循环神经网络及其变体仍存在固有缺陷。此类模型采用串行计算模式，无法实现并行训练，计算效率低下，且面对长时序能耗数据易出现梯度消失或梯度爆炸问题，难以有效捕捉长距离时序依赖关系[18]。长短期记忆网络通过门控机制一定程度上缓解了梯度消失问题，但也使得模型结构更复杂，增加了训练成本，且其串行计算模式效率仍然较低，全局特征提取能力有限，难以同时满足预测精度与实时性要求。同时，现有绝大多数供热负荷预测 Transformer 模型均采用固定正弦位置编码，仅能简单标记时序先后顺序，无法贴合供热系统热惰性滞后衰减的物理特性，不能表征远近时刻气象、负荷数据对当前能耗的差异化影响，存在物理先验融合不足、时序建模贴合度较差等问题[19]。现有研究已证实，建筑与管网热惰性是引发供热负荷时序滞后、波动偏移的核心因素，通过热惰性辨识、补偿与特征优化的方式，可有效提升供热负荷预测精度与调控稳定性[20, 21]。此外，相关研究表明，相较于传统时序模型，Transformer 凭借优异的长序列建模与并行计算能力，在供热能耗时序预测任务中具备显著的性能优势[22]。

基于此，本文引入序列 Transformer 模型（STM）开展区域供热系统天然气日耗气量预测研究，提出一种融合热惰性先验位置编码的改进预测算法（TI-STM）。该算法紧扣供热负荷“近端时序影响强、远端时序影响弱”的热力学规律，设计指数衰减可学习位置编码，通过逐维可学习衰减系数精准量化热惰性在时间维度的差异化作用，替代传统固定三角函数位置编码；同时将未来时刻气象协变量引入模型解码器，充分挖掘气象先验信息，结合多头注意力机制高效捕捉天然气负荷序列的多尺度长短期隐含特征，精准适配供热系统热惰性带来的负荷滞后特性。所提模型可实现时序数据并行运算，在提升长时序预测精度、强化模型物理合理性的同时，显著降低训练成本、提升运算效率。通过与支持向量机、循环神经网络、长短期记忆网络、多层感知机、时间卷积网络等多种主流预测模型开展对比验证，充分证明了所提算法的优越性，可为区域供热系统天然气精准调度、节能优化提供可靠的技术支撑。

2 数据及目的

为验证所提 TI-STM 算法的预测性能和泛化能力，我们收集了 2018 至 2020 年两个供暖季内三个区域供热系统（DHS）的日天然气需求历史数据。气象信息是影响 DHS 天然气需求变化的主导因素。目前，我们可以获取较为准确的 15 天天气预报，但天气预报的精度会随时间延长而下降。供暖负荷还与建筑热惰性密切相关，当前的天然气需求与前几日的消耗量关联更强。因此，我们将预测时域固定设为 15 天，且不随样本量的增加而延长。三个天然气日负荷数据集的统计信息汇总于表 1。

表 1 三家供热企业天然气数据统计信息

Metric	CompanyA	CompanyB	CompanyC
count	553	552	550
mean($\times 10^3 m^3$)	1078.0	125.2	800.3
std($\times 10^3 m^3$)	463.2	50.0	353.9
min($\times 10^3 m^3$)	0.0	0.0	0.0
25%($\times 10^3 m^3$)	710.9	89.8	553.7
50%($\times 10^3 m^3$)	1151.2	135.5	877.1
75%($\times 10^3 m^3$)	1459.6	165.0	1060.7
max($\times 10^3 m^3$)	2656.6	226.6	1694.7

3 算法原理

3.1 Transformer 模型的数学原理

3.1.1 总体结构

本模型采用标准 Transformer 的编码器–解码器（Encoder–Decoder）架构。编码器负责把历史窗口内的耗气量与气象序列压缩为一组上下文表示（memory）；解码器以该表示为条件，通过交叉注意力读取信息并一次性输出预测序列。与原始 Transformer 用于机器翻译不同，本模型的解码器不采用自回归逐词生成，而是使用一组可学习的起始向量并行输出全部预测步，由回归头映射为耗气量。

需要强调的是，本模型相对标准 Transformer 的全部改动集中在位置编码环节：以热惰性指数衰减位置编码取代正弦位置编码，其余结构（多头注意力、残差连接、层归一化、位置前馈网络）保持与标准实现一致，从而保证改进的可解释性与可对照性。

3.1.2 编码器（Encoder）

编码器由 N_{enc} 个结构相同的层堆叠而成。每一层包含两个子层：多头自注意力子层与位置前馈网络子层。设第 l 层输入为 $H^{(l-1)}$，则该层的运算为：

$$Attention(Q,K,V)=softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

式（1）即缩放点积注意力。其中 Q、K、V 分别为查询（Query）、键（Key）与值（Value）矩阵，均由同一输入经不同线性投影得到（自注意力）。分母中的 $\sqrt{d_k}$ 为缩放因子：当 d_k 较大时点积的数值方差随维度增长，若不缩放会使 $softmax$ 进入饱和区、梯度趋近于零。

单一注意力头只能刻画一种关联模式，为此引入多头机制，把表示空间划分为 h 个子空间并行计算，再拼接后线性融合：

$$head_i=Attention(QW_i^Q,KW_i^K,VW_i^V) \tag{2}$$

$$MultiHead(Q,K,V)=Concat(head_1,...,head_h)W^O \tag{3}$$

在编码器中 $Q=K=V=H^{(l-1)}$，即自注意力。其物理含义是：对每一个历史时刻，模型自适应地衡量它与窗口内其他时刻的关联强度，并据此加权聚合信息。这一机制使模型能够在一次运算中捕捉跨越多天的负荷变化模式，而无需像循环网络那样逐步传递状态，从根本上避免了长距离依赖的梯度衰减问题。

第二个子层为位置前馈网络（FFN），对每个时间位置独立施加同一个两层全连接变换，用于引入非线性映射能力：

$$FFN(x)=\max(0,xW_1b_1)W_2b_2 \tag{4}$$

每个子层外围都包裹残差连接与层归一化（Add&Norm），用于稳定深层网络的训练：

$$y=LayerNorm(x+Sublayer(x)) \tag{5}$$

$$LayerNorm(x)=\frac{x-\mu}{\sqrt{\sigma^2+\varepsilon}}\gamma+\beta \tag{6}$$

式（6）中 μ、σ^2为该样本在特征维度上的均值与方差，γ、β为可学习的缩放与平移参数，ε 为防止除零的小常数。残差连接使梯度可以绕过子层直接回传，层归一化则把每层输入重新标准化，二者共同保证编码器堆叠时训练不发散。

3.1.3 解码器（Decoder）

解码器由 N_{dec} 个层堆叠。每个解码器层包含三个子层：掩码多头自注意力、多头交叉注意力与位置前馈

网络。与标准翻译任务不同，本模型采用非自回归方式，因此不施加因果掩码（causalmask），自注意力可在输出位置之间自由交互。

解码器的查询来自一组可学习的起始向量 D_{start}，目标日的温度协变量经线性投影后与其相加，再施加热惰性位置编码：

$$S^{(0)} = PE(D_{start} + W_m M^{fut} b_m) \tag{7}$$

解码器层的运算为：

$$S^{(l)} = DecoderLayer(S^{(l-1)}, memory) \tag{8}$$

其中交叉注意力的 Q 取自解码器状态 $S^{(l-1)}$，K 与 V 取自编码器输出 $memory$，即模型在每个预测步上主动从历史窗口中检索相关信息。最后由线性回归头输出标量预测：

$$\hat{y} = W_o S^{N_{dec}} + b_o \tag{9}$$

3.1.4 注意力机制

本模型涉及三类注意力，其 $Q/K/V$ 来源不同，作用也各不相同：

表 2 三类注意力

位置	类型	Q 来源	K、V 来源	作用
编码器	多头自注意力	历史序列 H	历史序列 H	挖掘历史负荷内部的时序关联
解码器	多头自注意力	解码状态 S	解码状态 S	协调各输出步之间的关系
解码器	多头交叉注意力	解码状态 S	编码器 memory	从历史窗口中检索预测所需信息

三类注意力的计算均复用式（1）的缩放点积形式，仅 $Q/K/V$ 的取法不同，这正是 Transformer 结构高度统一的体现。

3.2 嵌入方式

模型中共有四类嵌入操作：输入特征的线性嵌入、正弦位置编码、热惰性指数衰减位置编码以及未来气象协变量嵌入。

3.2.1 输入特征的线性嵌入

编码器输入为历史日耗气量与同期气象协变量的拼接。耗气量为单变量标量序列，气象部分包含最低气温、最高气温、平均气温三个通道，故拼接后特征维度为 1+3=4。通过一个线性层把 4 维特征映射到 d_{model} 维隐空间：

$$X^{in} = (x; m), x, m \in R^{(L\times 1)}$$

式中 x 为历史耗气量序列，m 为气象协变量序列。线性嵌入的作用是把原始物理量投影到统一的隐空间中，使后续注意力运算可以在同一尺度上进行加权。

3.2.2 正弦位置编码

Transformer 本身不含任何循环或卷积结构，若不加位置信息，模型对输入序列的排列将完全无感。原始 Transformer 使用固定正弦位置编码：

$$PE(t,2i) = sin(\frac{t}{10000^{\frac{2i}{d_{model}}}}), PE(t,2i+1) = cos(\frac{t}{10000^{\frac{2i}{d_{model}}}}) \tag{10}$$

式中 t 为时间步位置，i 为维度序号。该编码由不同频率的正弦、余弦函数组成，使不同位置具有唯一的编码向量，并隐含相对位置可线性表示的性质。但它是通用的、与任务无关的，不包含供热系统的任何领域先验。

3.2.3 热惰性指数衰减位置编码

供热系统存在显著的热惰性：管网与建筑围护结构具有蓄热能力，室外温度的变化不能立即反映到耗气量上，而由历史累积效应共同决定。因此，在负荷预测中“越靠近预测日的历史耗气量，对当前时刻的影响越大”，而远期历史的影响随时间指数式衰减。标准正弦位置编码无法表达这一先验。

为此，本文提出热惰性指数衰减位置编码，把该物理规律直接编码进位置表示中：

$$PE(t,i) = \exp(-\lambda_i \cdot (L-1-t)) \tag{11}$$

式中 $t \in [0, L-1]$ 为窗口内的时间步（$t = L-1$ 表示距离预测日最近的一天），$i = 0, \ldots, d_{model-1}$ 为维度序号，$\lambda_i > 0$ 为第 i 维的衰减系数。该式的性质为：

（1）当 $t=L-1$ 时指数项为 $exp(0)=1$，取最大值；随着 t 减小，指数项按 e 的负指数单调衰减并趋近于 0。这直接实现了“近端影响强、远端影响弱”的热惰性先验。

（2）不同维度 i 拥有不同的衰减率 λ_i。小 λ_i 的维度衰减缓慢，用于刻画持续多天的缓慢热惰性；大 λ_i 的维度衰减迅速，用于刻画仅影响近一两天的快速响应。模型由此获得多时间尺度（multi-scale）的衰减表示能力。

衰减系数不设为固定常数，而是在对数空间参数化后随训练更新，使模型能够从数据中微调各时间尺度的重要性：

$$\lambda_i=\lambda_0\cdot 10^{\frac{i}{d-1}}(i=0,...,d-1) \tag{12}$$

即 λ_i 在 $[\lambda_0,\ 10\lambda_0]$ 之间按几何分布取值，实现从“慢衰减”到“快衰减”的多尺度覆盖。实现上以 $log\lambda_i$ 作为可学习参数，保证训练过程中 λ_i 恒为正。最终模型取 $\lambda_0=0.20$、$d_{model}=128$。

此外，模型还提供一种阻尼正弦变体用于消融对照，在指数衰减基础上叠加逐维周期项以增强位置辨识度：

$$PE(t,i)=\exp(-\lambda_i\cdot(L-1-t))\cdot\sin(\frac{t}{\tau_i}) \tag{13}$$

式中逐维周期 τ_i 同样可学习。最终模型采用式（11）的纯指数衰减形式；式（13）的阻尼正弦形式与式（10）的标准正弦编码共同作为对照设置，用于验证性能提升确实来自热惰性先验而非位置编码形式的任意改变。

3.2.4 未来气象协变量嵌入

日耗气量对气温高度敏感，且预测日的气温在预测时点是已知的，属于“已知协变量”（knowncovariate）。若忽略这一信息，模型将被迫仅凭历史外推，精度受限。为此，解码器侧额外引入一个线性层，把目标日的温度协变量投影到隐空间后与解码器起始向量相加（见式（7））。

这一设计的含义是：模型在做预测之前就被明确告知“明天有多冷”，从而把“温度到负荷”的映射与“历史负荷到未来负荷”的动态外推解耦，二者在解码器中融合。需要注意的是，该设计成立的前提是目标日气温可获取；若用于提前多日的预测，需替换为相应的气温预报值。

3.3 防过拟合方法

本研究的训练样本规模有限（约 474 天），而 Transformer 参数较多，过拟合是主要风险。代码中采取了以下六类措施：

（1）Dropout 正则化——在编码器层与解码器层的每个子层输出上以概率 p=0.1 随机将神经元置零，推理阶段关闭并按 1/(1−p)等比例缩放，通过破坏神经元之间的共适应迫使网络学习冗余而鲁棒的特征表示；

（2）早停——每一轮结束后在验证集上计算真实尺度的 MAPE，若连续 12 轮未出现超过 1e-4 的改善即终止训练，由于监控量直接取最终评价指标，可避免“训练损失仍在下降但泛化性能已经变差”的典型过拟合现象，同时显著缩短训练时间；

（3）最优权重回滚——早停只负责终止训练，而验证性能的最低点通常出现在终止之前若干轮，因此每轮验证后均与历史最优比较并保存参数快照，训练结束后回滚至该快照，确保最终用于预测的是泛化性能最好的一组参数，而非训练结束时的最后一组参数；

（4）学习率衰减——采用 ReduceLROnPlateau 调度器监控验证 MAPE，连续 5 轮不再下降时把学习率减半，使模型在最优解附近精细搜索，避免在最优点附近震荡；

（5）梯度裁剪——把梯度的全局 L2 范数限制在 1.0 以内，防止节假日、寒潮等突变点产生异常大的梯度而将参数带偏；

（6）结构与归一化层面的约束——位置编码的逐维衰减率 λ_i 虽可学习，但以对数参数化、由单一 λ_0 初始化并受单调几何分布的结构约束，仅引 d_{model} 个自由度，属于强正则化的参数化方式，不会在小样本上自由拟合噪声；同时 MinMax 归一化器严格只在训练集上拟合，再用于验证集与测试集，避免验证与测试的统计信息经归一化参数泄漏进训练过程、造成评估结果虚高。

4 改进算法的预测原理

4.1 模型原理

改进算法的预测过程为纯前馈运算。对长度为 L=7 的历史窗口，先把历史耗气量与同期气象协变量按特征维拼接，经线性层映射为 d_{model} 维并叠加热惰性位置编码，得到编码器输入 $H_{(0)}$；$H_{(0)}$送入编码器，通过多头自注意力在 7 个历史时刻之间加权聚合，输出上下文表示 memory；解码器以可学习起始向量为查询，加上目标日温度的嵌入并叠加位置编码，再依次通过 3 层解码器，每层内部先做自注意力以协调各输出步之间的关系，再做交叉注意力从 memory 中检索历史信息，最后经前馈网络做非线性变换；解码器最终状态经线性回归头映射为标量，即次日耗气量预测值。整个流程不含循环结构，可完全并行计算。由于预测步长 T=1，解码器采用非自回归设计，用 out_{len} 个可学习起始向量一次性给出全部输出，避免了自回归的串行推理与误差累积。训练阶段，数据以滑动窗口方式组织为（输入，目标）样本对，在训练集上更新参数、在验证集上监控泛化性能；推理阶段参数冻结，把待预测日之前 7 天的真实耗气量与其气象数据构成窗口输入，并附上目标日温度，前向一次即得预测值。本研究采用逐日滚动的单步预测模式，每日用最新观测重新构造窗口，不使用模型自身的预测值作为后续输入，因此不存在多步外推的误差累积。

4.2 评价指标

本文采用以下评价指标对算法性能进行评估。平均绝对误差（MAE）、平均绝对百分比误差（MAPE）与均方根误差（RMSE）的数学表达式如下：

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{14}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|(y_i - \hat{y}_i)/y_i| \tag{15}$$

$$RMSE = (\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2)^{\frac{1}{2}} \tag{16}$$

式中：y_i 表示热负荷实测值，$\hat{y}_i$ 表示热负荷预测值，n 表示样本数。

5 实验和讨论

5.1 超参数优化结果

本文所提 TI-STM 算法的超参数通过树结构帕尔森估计器（TPE）进行优化。表 2 汇总了本文核心超参数的取值范围及最优值。不同超参数下的性能对比绘制于图 1。

表 3 TEP 对 TI-STM 的超参数寻优结果

参数名称	参数取值范围	参数值
输入维度	32/64/96/128/160/192/256/320/384/512	256
多头注意力头数	1/2/4/8/16/32/64/128	64
编码器层数	1/2/3/4/5/6/7/8/9/10	7
解码器层数	1/2/3/4/5/6/7/8/9/10	6
前馈网络维度	64/128/192/256/384/512/768/1024/1536/2048	128
Dropout 概率	0.01/0.05/0.1/0.15/0.2/0.25/0.3/0.35/0.4/0.5	0.2
热惰性衰减系数	0.05/0.1/0.2/0.3/0.4	0.4
学习率	0.0001/0.0002/0.0005	0.0001
批大小	16/32/64	16

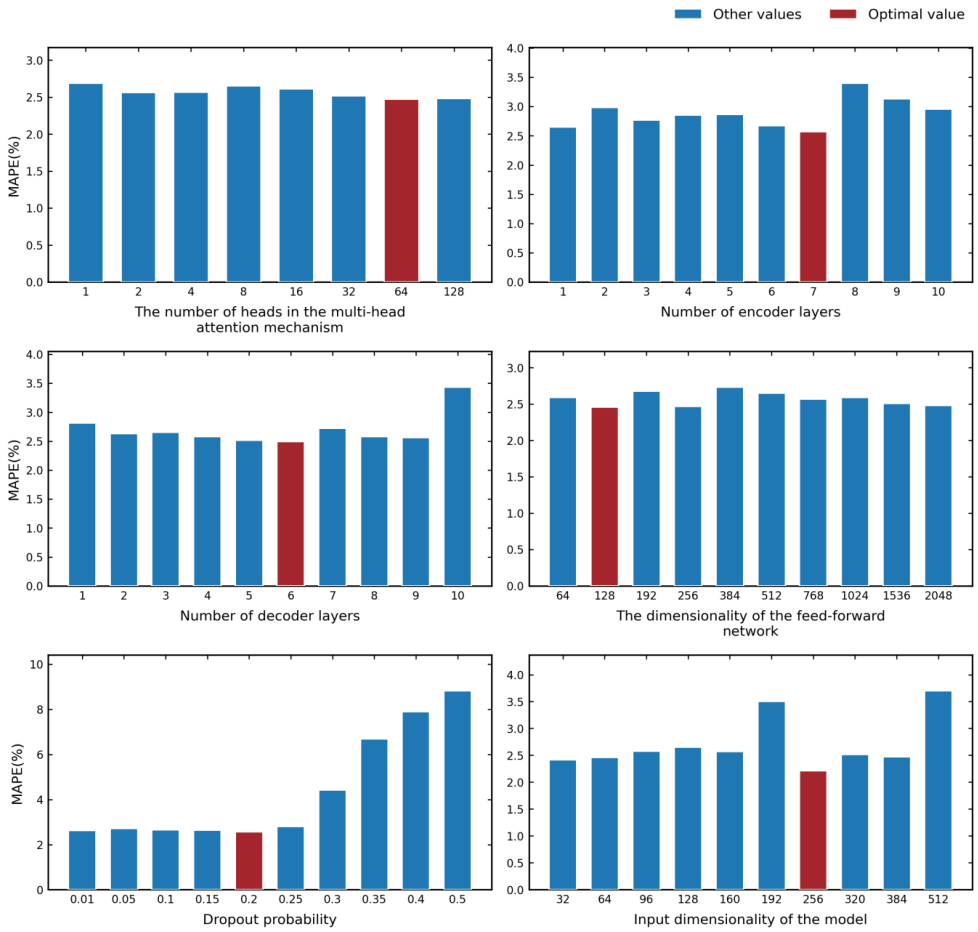


图 1 不同超参数下的性能对比

5.2 性能对比分析

为验证本文所提基于热惰性编码 STM（TI-STM）的天然气需求预测模型的性能，本文采用了三个独立 DHS 的历史天然气需求数据集。结合当前最大可用的 15 天天气预报信息，我们将 TI-STM 模型的预测时域设为 15 天。该预测时域有助于供热企业提前制定天然气采购、调度和储气计划的决策依据，同时通过灵活调整采购时间窗口，帮助供热企业降低运营成本，提高天然气能源的安全性和稳定性。

分别针对三个 DHS 构建了基于 TI-STM 的天然气需求预测模型。未来 15 天的日天然气需求预测结果及误差如图 2 至图 4 所示。以图 2 为例，上子图为预测结果与真实值的对比曲线，下子图为预测误差的直方图。TI-STM 输出连续 15 天的多步预测结果。前五天的预测曲线与实际曲线吻合良好，后十天的预测结果在个别点上略有偏差。这也表明，在天然气需求时间序列的多步预测问题中，新获取的预测样本的输入误差会随着预测步数的增加而累积，导致预测精度下降。本文提出的基于 TI-STM 的日天然气消耗预测方法能够有效预测供热系统历史数据中热惰性的影响。与其他预测算法相比，本文提出的预测模型能够更好地对热惰性时延进行建模，从而获得更高的预测精度和泛化能力。

如图 3 和图 4 所示，TI-STM 预测模型能够显著预测另外两个 DHS 未来 15 天的天然气需求，初步验证了 TI-STM 算法的有效性和泛化性。然而，另一个可观察到的现象是，当天然气需求数据出现较小的突变时，TI-STM 模型的预测误差也会略有增加。TI-STM 模型通过对天然气需求数据和气象数据进行嵌入和热惰性指数衰减位置编码来实现时序特征建模，编码方案在解码突变时可能存在一些微小误差。基于 TI-STM 算法对三个 DHS 的天然气需求预测结果的详细统计指标见表 3，其中 A 公司和 C 公司的天然气需求预测 MAPE 指标均低于 3%，而 A 公司的 MAPE 指标不超过 2%

表 4 三家供热企业所提 Transformer 模型的预测性能指标

SeparateDHS	MAPE(%)	MAE($\times 10^3 m^3$)	RMSE($\times 10^3 m^3$)
CompanyA	1.77	20.99	26.89
CompanyB	2.05	3.10	3.73
CompanyC	2.65	27.10	33.90

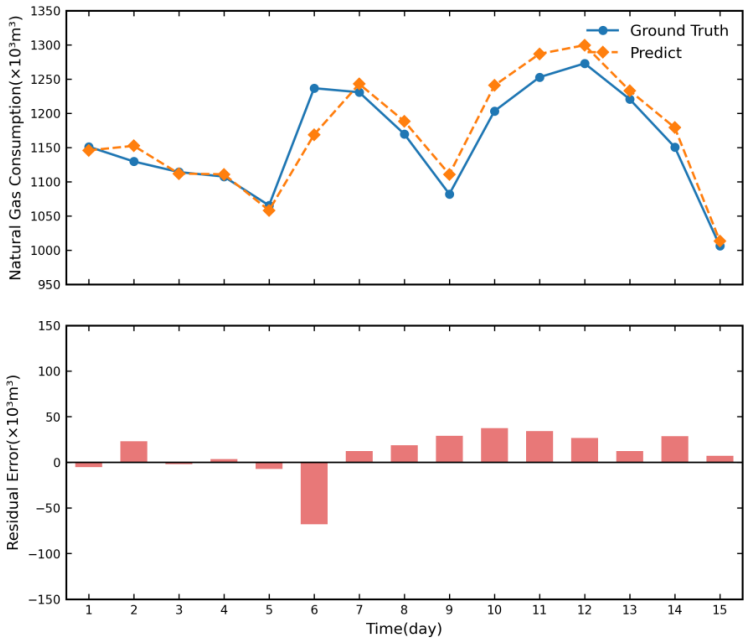


图 2 基于 TI-STM 模型的 A 企业预测结果与误差曲线

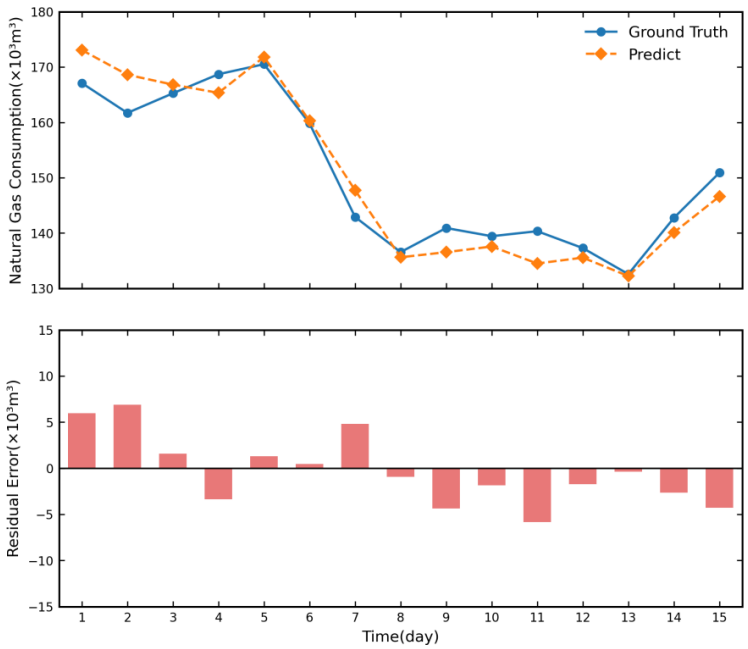


图 3 基于 TI-STM 模型的 B 企业预测结果与误差曲线

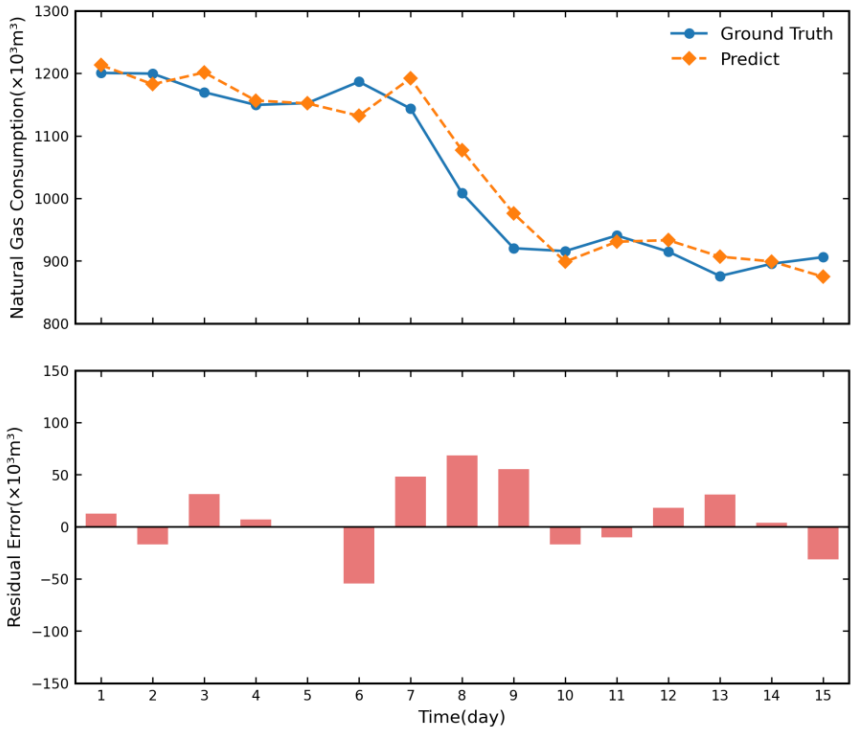


图 4 基于 TI-STM 模型的 B 企业预测结果与误差曲线

6 与其他算法的性能对比

为进一步横向评估所提 TI-STM 算法的预测性能，本文将其与当前先进的时间序列预测模型（如 SVM、LSTM、MLP、RNN、TCN 和 STM 等）进行了比较分析。实验仍采用三家供热公司的历史数据，并结合最大 15 天的天气预报信息，预测未来 15 天的天然气需求。通过对比 MAPE、MAE 和 RMSE 指标，TI-STM 模型与其他算法相比展现出优异的预测性能。三家公司的预测结果曲线分别如图 5 至图 7 所示，预测性能统计指标见表 4 至表 6。

从图 5 至图 7 可以看出，TI-STM 算法在三个不同 DHS 数据集上的实验结果相比其他算法更接近真实天然气消耗数据曲线。表 4 至表 6 的性能统计数据也表明 STM 算法优于其他算法。具体而言，A 公司的 MAPE 为 1.77%，MAE 为 20.99×10³m³，RMSE 为 26.89×10³m³；B 公司的 MAPE 为 2.05%，MAE 为 3.10×10³m³，RMSE 为 3.73×10³m³；C 公司的 MAPE 为 2.65%，MAE 为 27.10×10³m³，RMSE 为 33.90×10³m³。TI-STM 算法能够准确预测三个不同 DHS 的天然气需求，充分展示了其在预测精度和泛化能力方面的卓越性能。TI-STM 算法之所以能够取得如此高的预测性能，主要归功于其基于多头注意力机制的编码器-解码器网络结构，以及所引入的热惰性指数衰减位置编码。该编码将供热系统“越靠近预测日的历史影响越大”的物理先验直接融入位置表示，使模型无需借助额外的循环结构即可刻画负荷的滞后与累积效应，从而实现了对天然气需求时间序列数据中强非线性和强耦合等复杂特征的提取与函数逼近。

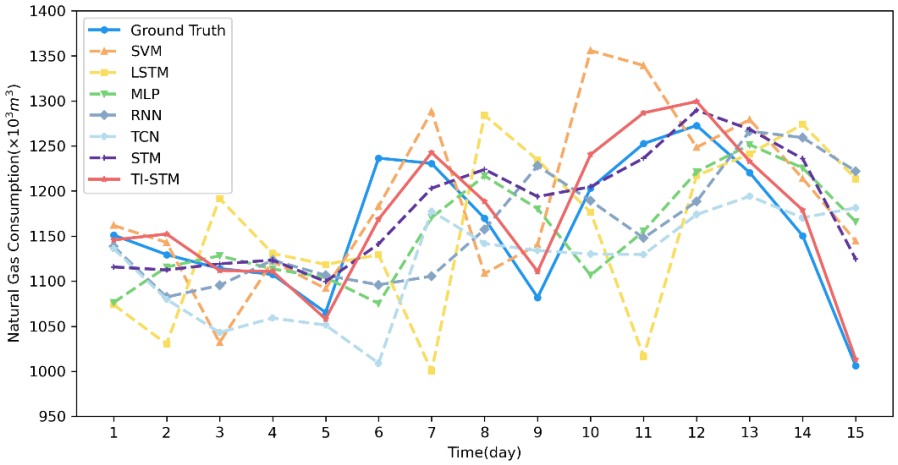


图 5 换热站 A 小时热负荷预测结果

表 5 A 供热企业性能对比统计

Model	MAPE(%)	MAE(×10^3m^3)	RMSE(×10^3m^3)
SVM	5.24	60.23	72.80
LSTM	9.29	107.05	127.69
MLP	5.94	68.44	82.794
RNN	6.61	75.57	96.79
TCN	6.18	71.67	93.26
STM	4.04	45.60	59.19
TI-STM	1.77	20.99	26.89

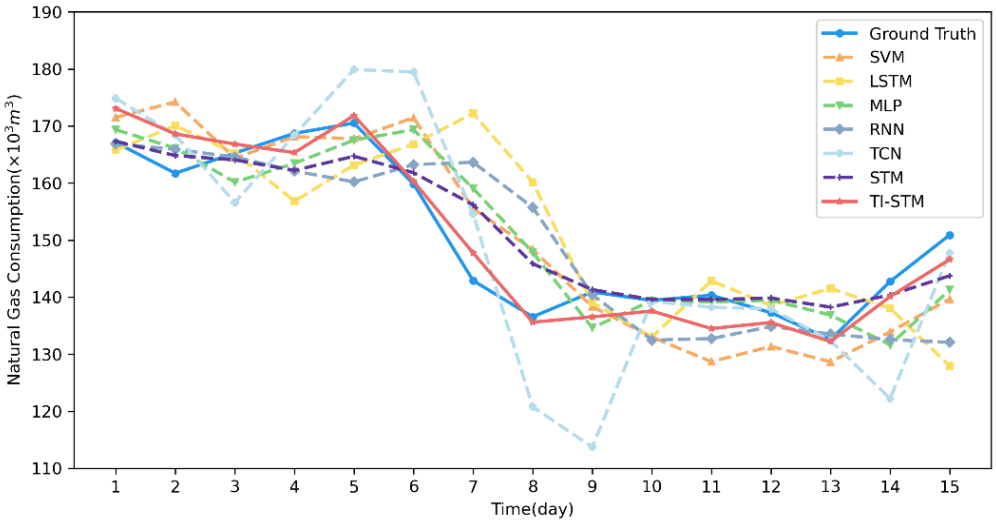


图 6 换热站 B 小时热负荷预测结果

表 6 B 供热企业性能对比统计

Model	MAPE(%)	MAE(×10^3m^3)	RMSE(×10^3m^3)
SVM	4.88	7.18	8.39
LSTM	6.22	9.16	12.68
MLP	4.12	6.11	7.51
RNN	5.11	7.52	10.16
TCN	5.98	8.93	12.15
STM	2.73	4.03	5.48
TI-STM	2.05	3.10	3.73

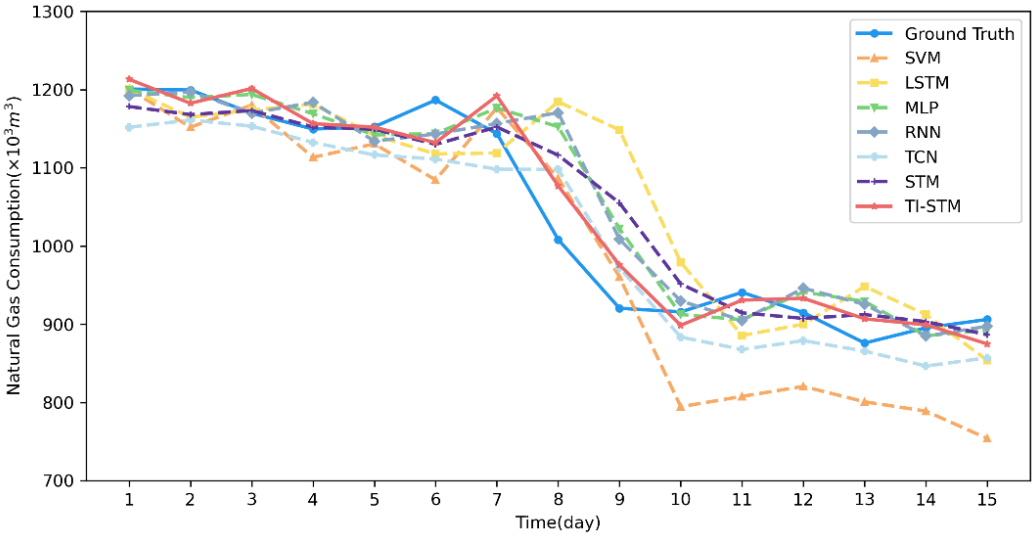


图 7 换热站 C 小时热负荷预测结果

表 7 C 供热企业性能对比统计

Model	MAPE(%)	MAE(×10^3m^3)	RMSE(×10^3m^3)
SVM	7.23	70.13	83.60
LSTM	5.83	56.95	84.16
MLP	3.54	35.24	51.81
RNN	3.50	34.65	53.40
TCN	4.36	44.57	49.50
STM	3.39	33.58	50.55
TI-STM	2.65	27.10	33.90

7 结论

精准的天然气需求预测算法能够提升区域供热系统（DHS）的能源稳定性和安全性。针对由于强耦合、热惰性等复杂特征导致的天然气需求预测精度不足的问题，本文提出了一种基于热惰性位置编码的序列 Transformer 预测算法。该算法以指数衰减形式的位置编码替代传统正弦位置编码，将“越靠近预测日的历史影响越大”这一供热系统热惰性先验直接融入位置表示，使模型能够通过多头注意力机制提取区域供热系统长期历史数据中的抽象特征；同时将目标日温度作为已知协变量注入解码器，强化了气象因素与负荷之间的映射关系。利用实际工程系统中的天然气消耗数据集对所提模型的性能进行了验证，详细实验结果表明，相较于现有先进模型，所提模型能够更稳定、更准确地预测 DHS 系统中的天然气需求，三个数据集的预测精度分别可达 98.23%、97.95%与 97.35%。

在未来的研究中，我们还将考虑供热建筑空间分布对天然气需求的影响，拟利用图神经网络构建供热区域内天然气需求的拓扑结构，分析不同空间区域天然气需求之间的交互作用，以满足 DHS 精准能源需求预测的要求。此外，本文目前是以指数衰减位置编码对供热管网的热惰性影响作隐式刻画，未来计划采用输入与输出序列的互相关方法对热惰性时延进行精确估计，并基于时延估计对热负荷时间序列的预测结果作进一步校正，从而在显式考虑供热管网热惰性时延的条件下实现更准确的热负荷预测。

参考文献

[1] Lin X W, Ma J J, Chen H, Shen F, Ahmad S, Li Z Q. Carbon Emissions Estimation and Spatiotemporal Analysis of China at City Level Based on Multi-Dimensional Data and Machine Learning [J]. Remote Sensing, 2022, 14(13): 3014. https://doi.org/10.3390/rs14133014

[2] Lin XW, Ma JJ, Chen H, Shen F, Ahmad S, Li ZQ. Carbon Emissions Estimation and Spatiotemporal Analysis of China at City Level Based on Multi-Dimensional Data and Machine Learning. Remote Sens-Basel. 2022; 14(13). 3014. https://doi.org/10.3390/rs14133014

[3] Zhou J, Zhao YX, Fu TT, Zhou X, Liang GC. Dimension optimization for underground natural gas storage pipeline network coupling injection and production conditions. Energy. 2022; 256. https://doi.org/10.1016/j.energy.2022.124651

[4] Szega M, Zymelka P, Janda T. Improving the accuracy of electricity and heat production forecasting in a supervision computer system of a selected gas-fired CHP plant operation. Energy. 2022; 239. https://doi.org/10.1016/j.energy.2021.122464

[5] Lim HS, Kim G. Development of Regression Models considering Time-Lag and Aerosols for Predicting Heating Loads in Buildings. Advances in Civil Engineering. 2018; 2018. https://doi.org/10.1155/2018/4878021

[6] Ding Y, Zhang Q, Yuan TH, Yang K. Model input selection for building heating load prediction: A case study for an office building in Tianjin. Energy Build. 2018; 159: 254-70. https://doi.org/10.1016/j.enbuild.2017.11.002

[7] Yuan PL, Lin DM, Wang ZS. Coal consumption prediction model of space heating with feature selection for rural residences in severe cold area in China. Sustainable Cities and Society. 2019; 50. https://doi.org/10.1016/j.scs.2019.101643

[8] Zhang Q, Tian Z, Ma ZJ, Li GY, Lu YK, Niu JD. Development of the heating load prediction model for the residential building of district heating based on model calibration. Energy. 2020; 205. https://doi.org/10.1016/j.energy.2020.117949

[9] Geysen D, De Somer O, Johansson C, Brage J, Vanhoudt D. Operational thermal load forecasting in district heating networks using machine learning and expert advice. Energy Build. 2018; 162: 144-53. https://doi.org/10.1016/j.enbuild.2017.12.042

[10] Xue PN, Jiang Y, Zhou ZG, Chen X, Fang XM, Liu J. Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms. Energy. 2019; 188. https://doi.org/10.1016/j.energy.2019.116085

[11] Cai W, Wen XD, Li CE, Shao JJ, Xu JG. Predicting the energy consumption in buildings using the optimized support vector regression model. Energy. 2023; 273. https://doi.org/10.1016/j.energy.2023.127188

[12] Moayedi H, Bui DT, Dounis A, Lyu Z, Foong LK. Predicting Heating Load in Energy-Efficient Buildings Through Machine Learning Techniques. Applied Sciences-Basel. 2019; 9(20). https://doi.org/10.3390/app9204338

[13] Wang CD, Yuan JJ, Zhang J, Deng N, Zhou ZH, Gao F. Multi-criteria comprehensive study on predictive algorithm of heating energy consumption of district heating station based on timeseries processing. Energy. 2020; 202. https://doi.org/10.1016/j.energy.2020.117714

[14] Zhu RJ, Guo WL, Gong XJ. Short-Term Load Forecasting for CCHP Systems Considering the Correlation between Heating, Gas and Electrical Loads Based on Deep Learning. Energies. 2019; 12(17). https://doi.org/10.3390/en12173308

[15] Xue GX, Qi CY, Li H, Kong XF, Song JC. Heating load prediction based on attention long short term memory: A case study of Xingtai. Energy. 2020; 203. https://doi.org/10.1016/j.energy.2020.117846

[16] Du J, Zheng JQ, Liang YT, Lu XY, Klemes JJ, Varbanov PS, et al. A hybrid deep learning framework for predicting daily natural gas consumption. Energy. 2022; 257. https://doi.org/10.1016/j.energy.2022.124689

[17] Liu JY, Wang SX, Wei N, Qiao WB, Li Z, Zeng FH. A clustering-based feature enhancement method for short-term natural gas consumption forecasting. Energy. 2023; 278. https://doi.org/10.1016/j.energy.2023.128022

[18] Wei N, Yin LH, Li C, Wang W, Qiao WB, Li CJ, et al. Short-term load forecasting using detrend singular spectrum fluctuation analysis. Energy. 2022; 256. https://doi.org/10.1016/j.energy.2022.124722

[19] 任子豪.基于多气象因子耦合与热惰性补偿的集中供热负荷预测与动态调控策略 [C]//中国城镇供热协会供热技术专业委员会,《区域供热》杂志社有限公司.第九届供热学术年会论文集(2026). 2026: 773-790. https://link.cnki.net/doi/10.26914/c.cnkihy.2026.052343

[20] 赵晶巍, 夏学彬. 基于热惰性辨识的采暖系统热负荷修正与室温波幅控制研究 [C]. 中国城镇供热协会供热技术专业委员会,《区域供热》杂志社有限公司. 第九届供热学术年会论文集(2026). 2026: 1373-1380. https://link.cnki.net/doi/10.26914/c.cnkihy.2026.052396

[21] 赵洁, 程鹏月, 陶俊秀.建筑热惰性对供热系统智慧调节的影响分析 [J]. 区域供热, 2024(3): 15-19. https://link.cnki.net/doi/10.16641/j.cnki.cn11-3241/tk.2024.03.003

[22] Wahl M, Bayer D R, Pruckner M. Benchmarking Transformer and xLSTM for time-series forecasting of heat consumption [C]//2026 IEEE Sustainable Power and Energy Conference, 2026: 1-8. https://api.semanticscholar.org/CorpusID:288258460