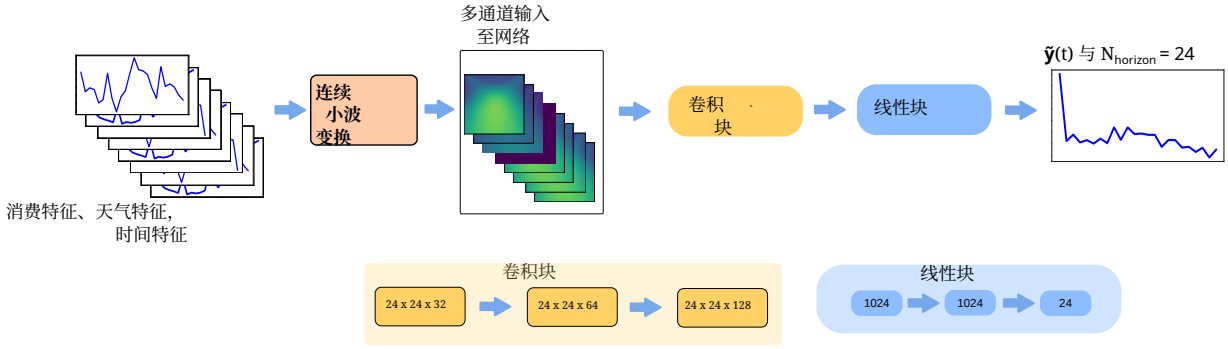
arXiv:2603.01137v1 [cs.LG] 1 Mar 2026

图文摘要

一种利用分解特征的时频表示进行热需求预测的深度学习框架

Adithya Ramachandran、Satyaki Chatterjee、Thorkil Flensmark B. Neergaard、Maximilian Oberndoerfer、Andreas Maier、Siming Bayer



亮点

一种利用分解特征的时频表示进行热需求预测的深度学习框架

Adithya Ramachandran、Satyaki Chatterjee、Thorkil Flensmark B. Neergaard、Maximilian Oberndoerfer、Andreas Maier、Siming Bayer

- 提出了一种利用小波进行热需求预测的新型深度学习方法。
- 特征的时间序列分解显著提高了准确率。
- 系统评估需求、天气和日历特征的影响。
- 提前24步的逐小时热需求预测的最先进基准。
- 展现了稳健性，并通过多年真实世界数据进行了验证。

一种利用分解特征时频表示进行热需求预测的深度学习框架

Adithya Ramachandran[a,*], Satyaki Chatterjee[a], Thorkil Flensmark B. Neergaard[b], Maximilian Oberndoerfer[c], Andreas Maier[a], Siming Bayer[a]

[a]模式识别实验室，弗里德里希·亚历山大大学，*Martenstrasse 3*，埃尔兰根，*91054*，巴伐利亚，德国[b]布伦讷斯莱夫 *Forsyning A/S*，*Virksomhedsvej20*，布伦讷斯莱夫，*9700*，丹麦[c]*SBU Analytics and Services*，*Diehl Metering GmbH*，*Donaustrasse 120*，纽伦堡，*90451*，德国

摘要

区域供热系统是可持续地向某一地理区域内的消费者输送热能的重要基础设施，但要实现高效管理，必须根据波动的需求优化木材、天然气、电力和太阳能等多种能源。使供给与需求相匹配至关重要，不仅关系到确保热能可靠输送，也有助于减少碳排放，并通过降低运行温度延长基础设施的使用寿命。然而，由于用能模式复杂且呈非线性，并受到外部因素影响，要实现支持这些目标的准确多步预测仍面临挑战。在本研究中，我们提出了一种新颖的深度学习框架，用于日前热需求预测，该框架利用历史数据的时频表示。通过对分解后的需求数据和外部气象因素应用连续小波变换，我们的方法使卷积神经网络能够学习标准时域模型通常难以捕捉的层级时间特征。我们使用来自三个不同丹麦区域、一个丹麦城市和一个德国城市的多年数据，系统地将该方法与统计基线、最先进的Transformer模型及新兴基础模型进行评估。结果表明，该方法取得了显著进步：与最强基线方法相比，平均绝对误差降低了36%至43%，在年度测试数据集上的预测准确率最高达到95%。定性和统计分析进一步证实了该方法的准确性与稳健性：在其他方法失效时，它仍能可靠地追踪波动的需求峰值。本研究既提出了一种高性能预测架构，也提供了有关最佳特征组合的重要见解，为现代能源应用提供了经过验证的解决方案。

关键词：需求预测、区域供热、时间频率表示、时间序列预测、分区计量区域、深度学习

1. 引言

在寒冷气候中，空间供暖和热水对人们的福祉至关重要，因此区域供热系统（DHS）是可持续基础设施的重要组成部分，能够提供可扩展性、脱碳途径以及燃料来源多样化 [1]。区域供热系统在能源供应链中占据关键位置，通过地下管网将集中式发电设施产生的热能输送至各类住宅、商业和工业终端用户。为实现高效监测，这些网络通常划分为若干地理区域，称为分区计量区域（DMA） [2]。随着全球积极推动气候中和 [3], ，优化供需平衡已成为当务之急：供给受资源可用性制约，而需求则受到复杂消费者行为的影响。

在此背景下，准确的热需求预测是实现系统优化的前提。智能计量基础设施 [4] 的出现彻底改变了这一领域，使消耗模式的高分辨率监测成为可能，让运营商能够分析外部因素和内部因素如何影响需求 [5]。随着区域供热系统在

*通讯作者地址：德国埃尔兰根，弗里德里希-亚历山大大学模式识别实验室电子邮箱：adithya.ramachandran@fau.de（Adithya Ramachandran）

主要国家 [6]不断扩展，且日益依赖波动性可再生能源，预测建模变得不可或缺。可靠的预测有助于改进生产规划、网络平衡和需求侧管理，最终通过实现能源系统的智能控制来减少能源损耗和峰值负荷，从而打造更具成本效益且可持续的供热解决方案。

1.1. 相关研究

预测方法旨在以较高的准确率和稳定性预测热需求，使运营商能够从被动应对转向有计划、以数据为驱动的能源管理。短期预测使运营商能够优化锅炉、燃气发动机和热泵等热能生产设备的运行，从而减少燃料消耗、运营成本、排放，以及停机和启动带来的惯性效应 [7]。预测还有助于提前调整分配网络中的供水温度和流量 [8], ，既减少热损失，又确保为所有用户提供充足供应。除了提升运营效率，可靠的预测还能支持参与能源市场，并为燃料采购策略提供依据，从而促进更广泛能源网络的稳定性[9][10]。然而，开发有效的预测模型颇具挑战，因为模型必须考虑历史消耗模式、天气变化、消费者的心理特征，以及会引入非线性和噪声的非确定性成分 [11]。此外，扩展至城市级预测需要汇总来自数千个异质来源的数据，这可能导致测量规程不一致 [12]。

在过去二十年中，如何应对区域供热系统热需求预测固有的复杂性一直是研究重点。区域供热系统的预测方法已从经典统计方法发展到先进的深度学习架构。当前文献大体可分为两类：时间序列架构的演进，以及表征学习领域的新兴范式。

1.1.1. 统计学习与机器学习

早期的基础性研究主要集中于捕捉线性依赖关系。Dotzauer [13] 建立了基于线性最小二乘法的模型，主要将热需求与室外温度相关联。鉴于热需求具有非平稳性，经典时间序列方法逐渐得到广泛应用。Grosswindhager 等人 [14] 在状态空间框架中采用了季节性自回归积分滑动平均（SARIMA），而 Fang 和 Lahdelma[15] 则证明，与多元线性回归（MLR）模型相比，在季节性自回归积分滑动平均外生回归模型（SARIMAX）中引入室外温度作为外生变量能够带来益处。

机器学习（ML）的出现为非线性建模开辟了新途径。Idowu 等人[16], Petkovic 等人 [17], 和Protic 等人 [18]研究了一系列浅层算法，包括支持向量机（SVR）、回归树和前馈神经网络（FFNN）。Potocnik 等人 [19]的一项全面比较发现，高斯过程回归（GPR）和经过良好调优的神经网络在多步超前预测方面尤为有效。然而，这些方法通常需要大量人工特征工程，并且往往缺乏有效建模长程时间依赖关系的能力 [19]。此外，它们对现代能源系统中出现的剧烈、非线性转变适应性较差。

1.1.2. 深度学习架构

为捕捉复杂的时间动态，深度学习方法获得了显著关注。Suryanarayana等人 [20]证明了深度前馈神经网络在处理更高维输入方面优于线性模型。为应对序列依赖关系，循环架构成为标准方法。Yao等人 [21] 提出的集成式D-CNN-长短期记忆网络等先进变体，将卷积神经网络（CNN）的特征提取能力与长短期记忆网络（LSTM）的时间记忆相结合。同样，Xue等人 [22] 探索了支持向量回归、深度神经网络和XGBoost在多步超前预测中的应用。尽管长短期记忆网络（LSTM）等早期深度学习模型以及浅层架构相较于线性方法取得了重大进展并确立了优势，但由于梯度消失和表征能力不足，它们往往难以捕捉高频波动性和长序列。

这些局限推动了架构向更强大的方向演进。近年来，时间序列预测领域的深度学习（DL）取得了显著的架构创新，这在很大程度上得益于 Transformer模型 [23]的成功。Informer [24], 、Autoformer [25], 和 TimesNet [26]等架构已展现出

先进的性能，能够有效捕捉长程依赖关系和复杂的时间模式。它们的效能已在多个不同领域得到持续基准测试，这往往为区域供热系统（DHS）提供了重要启发，例如采用 Informer [27], 和时间融合变换器（TFT）[28]。此外，专门用于小时热需求预测的模型也逐渐崭露头角，例如融合 Bi-Mamba2 与柯尔莫哥洛夫-阿诺德网络（KANs）的 TBMK 网络 [29],，以及将科尔莫戈罗夫-阿诺德网络（KAN）集成到变换器架构中以捕捉长期依赖关系的 Cakformer [30]。

尽管存在这种跨领域影响，但将这些前沿深度学习架构系统地应用于区域供热系统未来24步小时热需求预测，并对其进行严格评估，仍处于起步阶段，且暴露出若干关键不足。例如，Informer 和时序融合变换器等模型的评估所用数据集覆盖时间极短，仅有 33 天 [27] ，或仅来自单一机构，因此其季节稳健性以及对不同用户特征的泛化能力 [28]令人存疑。近期研究中，基于短期数据集开展评估的趋势仍在延续。例如，TBMK 模型仅在为期 68 天的冬季数据集 [29], 上进行了验证，而 Cakformer 模型使用的数据集总共仅涵盖三至四个月 [30] 。如此短的时间窗口无法覆盖需求模式高度非线性的关键过渡季，也无法捕捉年际变化。数据需求量大的基于变换器的模型进一步加剧了这一挑战，而长期、广泛的公共区域供热系统数据集却普遍匮乏。此外，许多研究依赖来自特定工业场所 [31] 或大学建筑 [32], 的私有数据 [33], ，这限制了对实际情况的真实呈现及研究的可复现性，并且侧重于同质化而非多样化的用户区域。

近期对公开数据集的使用，尤其是来自丹麦奥尔堡的数据集 [34], ，标志着朝着提升研究可复现性迈出的积极一步。然而，该数据集在近期重要研究中的应用也凸显了研究的另一面。采用该数据集的研究通过主动深度学习 [35] 多尺度时间表征 [36], 和稀疏动态图学习 [37], ，侧重于对单一、同质用户群体进行日分辨率预测。这些研究的作者也指出，未来需要在用户类型和气候条件更加多样化的数据集上进行验证 [37]。预测目标上的这一差异凸显了热需求预测领域长期存在的一个缺口：针对由异质用户构成、符合现实世界情况的开放式区域供热系统进行需求预测。这对区域供热系统的智能控制至关重要 [28]。

随着架构不断演进，数据表示和特征工程方面的进展也展现出一系列有前景的预测途径，包括时间序列分解 [38], 、相空间表示 [39], 以及小波等时频表示方法 [40]。近期研究表明，分解是工业热负荷预测中一种强大的预处理步骤 [31]。采用连续小波变换（CWT）的技术已证明有助于捕捉标准时域中无法获取的时间周期特征[41], ，但这类技术在提升小时热需求预测方面尚未得到系统性研究 [42]。与此同时，随着基础模型的出现，一个全新的研究前沿也正在形成 [43]。这些模型在海量时间数据语料上进行预训练，为工业应用带来了重大机遇，尤其适用于特定系统的历史数据可能较为稀缺的场景。然而，基础模型在热需求预测中的应用仍完全未被探索，这一关键缺口亟待弥补。

1.2. 贡献

我们的工作旨在通过多方面的方法，解决已发现的区域供热系统预测领域的不足。我们的主要贡献是提出一种新颖的深度学习框架，通过将一维时间序列转换为二维时频表示，重新定义预测问题。具体贡献如下:

- 为解决评估体系碎片化且结果往往缺乏可比性的问题，我们对现代人工智能架构进行了全面基准测试。该评估基于来自多个异质区域、涵盖多年且按小时采样的数据集，范围从经典的长短期记忆网络到最先进的 Transformer 模型和混合模型，其中包括经过微调的 TTM [44] 和 Chronos-2 [45], 基础模型，从而为这一特定预测任务确立了评估标准。此外，为展示泛化能力，我们还展示了现有公开数据集上的预测性能，这些数据集分别涉及德国弗伦斯堡的区域供热系统 [46], 以及丹麦奥尔堡的住宅需求 [34]。

- 为解决特征工程的临时性问题，我们系统评估了特征表示的影响，其中包括对公共假日等复杂、非标准事件的编码方案进行系统评估

并在报告已知相关性的基础上，严谨分析需求和天气序列各自分解出的趋势、季节性和残差分量之间如何相互作用，从而更清晰、以数据为依据地理解热需求的基本驱动因素。

- 为克服标准一维模型的局限性，并挖掘替代数据表示尚未充分探索的潜力，我们提出了一种新颖的时频深度学习框架。通过连续小波变换（CWT）将特定分解后的时间序列转换为二维尺度图，我们的模型使卷积神经网络能够学习层次化时序周期模式。该方法明确确立了新的性能基准测试，显著优于所有基线模型，将平均绝对误差降低约31%至33%，并提供稳定的95%预测准确率。

随着区域供热系统基础设施日益复杂并与波动性可再生能源深度融合，对具备如此前所未有的准确率与可靠性的预测模型的需求已成为运营必需。建模结果在多年数据集上得到验证，展示了一种稳健且可投入实际部署的解决方案。通过提供更准确的日前预测，我们的工作直接贡献于应用价值，例如优化锅炉与热泵调度以降低燃料消耗、更有效地参与能源市场，并服务于区域供热系统模型预测控制等新兴解决方案。

2. 方法

2.1. 数据概述

本文提出的热需求预测框架主要利用一家丹麦公用事业公司的数据进行开发和评估。该公司通过其区域供热系统为布伦讷斯莱夫市供应热水和空间供暖。该数据集包含2015年6月至2020年11月期间，来自三个 DMA（*DMA A*、*DMA B* 和*DMA C*）中4910个智能电表的历史小时热能消耗数据。由于智能电表的原始读数为累积值，因此采用一阶差分来计算每小时的消耗量。随后，将各智能电表的每小时消耗值按区域汇总，以表示任意时刻t的总需求。图1展示了*DMAA*的聚合时间序列数据，其中包含周期性和非周期性异常值。

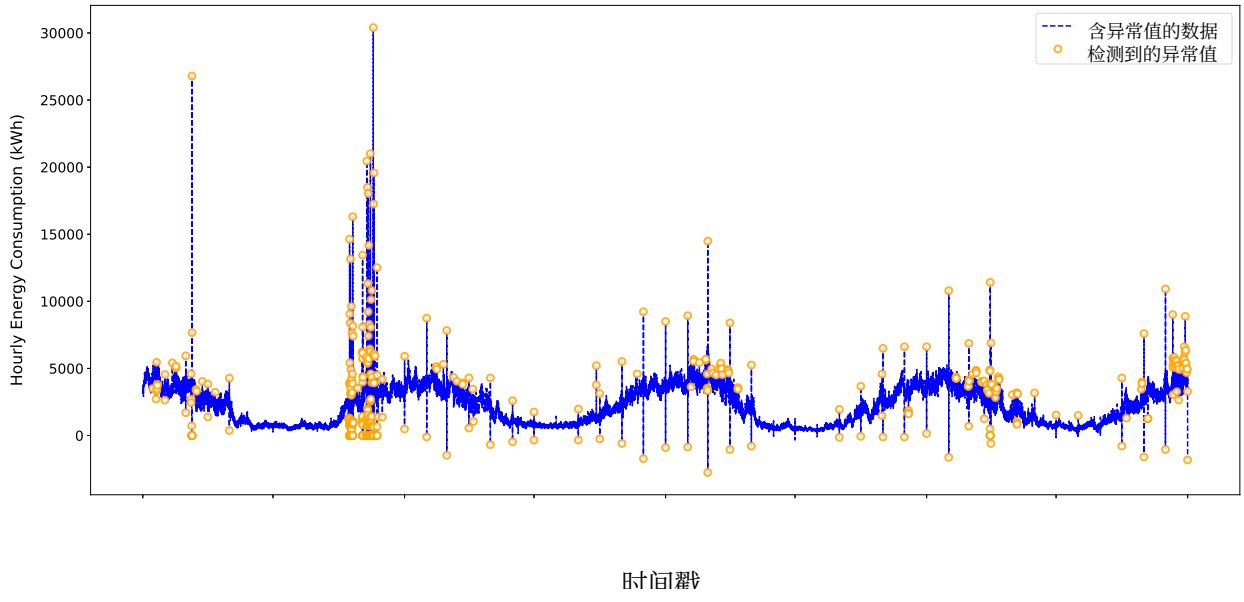


图1：DMA A在2016年1月至2019年12月底期间的每小时热消耗数据。通过 Savitzky-Golay 平滑识别出的数据异常值以橙色圆圈标出。

热需求数据作为内生变量，而天气特征和日历特征等外部独立变量则构成影响热需求的外生变量集合。相较之下

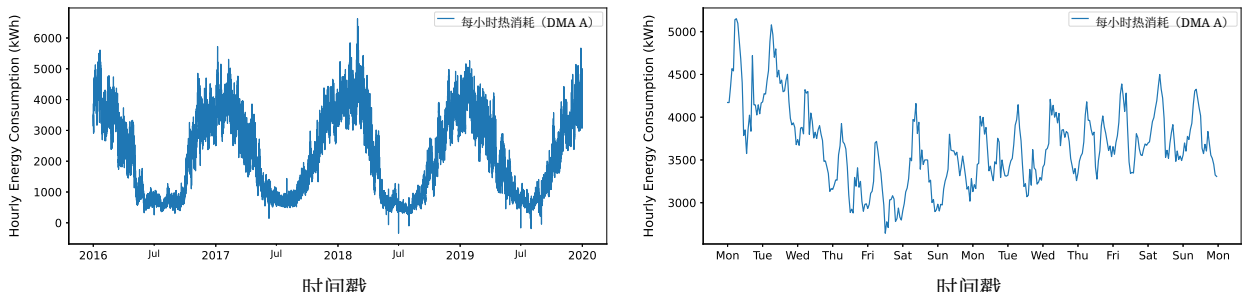
本文获取并使用回顾期的实际天气数据，而非预测天气数据。由于显式的社会经济或人口统计特征已作为固有模式体现在记录的热需求数据中，因此未对其进行建模。数字化成熟阶段内，即2015年6月至2016年期间记录的数据已被排除。在5年的时间跨度内，由于DMA中用于采集各个客户需求的智能电表数量并非恒定，因此将时间t 的聚合需求除以计入总需求的智能电表数量，计算其平均值。这样可以应对数据漂移。由于智能电表数量已知，预测过程结束后，需求时间序列可以重新缩放至原始尺度。此外，为保持行为一致性并消除 COVID-19 相关行为的影响，2020年的数据也已剔除。

为确保框架的稳健性并便于在公共基准数据集上复现评估，我们将评估范围扩展至来自不同地理位置的两个开源数据集。首先，我们使用德国弗伦斯堡的区域供热生产小时数据 [46], ，数据覆盖2017年1月至2024年12月。由于记录精度不一致，未纳入2017年之前的历史数据。对于该数据集，我们通过哥白尼气候数据存储库 [47]，从ERA5再分析数据集中获取历史天气变量，具体包括环境温度和“体感”温度。根据石勒苏益格-荷尔斯泰因州的公共假日生成相应的日历特征。其次，我们使用丹麦奥尔堡的住宅数据集 [34],，该数据集汇总了3021个住宅智能电表的小时消费读数。该地区的气象数据，包括环境温度、最低气温和风速，取自丹麦气象研究所（DMI）开放数据 API [48]。体感温度则明确采用 DMI 使用的标准风寒指数公式计算 [49]。

2.2. 数据预处理

如图1所示，所记录的每小时热量消耗数据呈现出一种行为，即异常值的幅度具有季节性，冬季观测到的偏差较大，夏季偏差较小。此外，数据中包含物理上不合理的负消耗值。这一时间的非线性、非平稳性质

时间序列，加上缺乏明显全局特征的异常值，使传统的异常值处理方法（例如，Z 分数阈值、基于中位数的方法）失效。因此，我们采用一种替代方法，针对时间序列的残差分量，而非其趋势分量和季节分量进行处理。具体而言，对学生化残差应用 Savitzky-Golay 平滑，通过 Bonferroni 校正识别异常值，随后用时间序列的趋势分量与季节分量之和替换这些异常值。图2a展示了 *DMA A* 在整个数据时段内经过异常值处理后的热需求数据，呈现出年度周期性；图2b展示了2017年1月两周内的热消耗数据，显示出非线性、非平稳性，以及日周期性和周周期性。预处理后的热需求数据作为因变量。



（a）2016年1月至2019年12月。　　（b）2017年1月的两周。

图 2：DMA A 的预处理每小时热量消耗数据。(a) 2016年1月至2019年12月剔除异常值后的数据。(b) 2017年1月两周内记录的热消耗详细视图。

2.3. 数据分析

未来热需求与其历史行为有着内在联系，历史行为体现了消耗的基本性质和规模。历史数据还反映了潜在的非确定性影响因素，例如人口统计和心理特征因素，以及明显的长期和短期周期性。相关示例

分别展示了这些长期趋势和短期周期性模式，见图2a和图2b。为了定量评估这些时间依赖关系的强度，我们使用斯皮尔曼相关系数（ρ）进行自相关分析。选择这一非参数度量，是因为它能够捕捉热需求时间序列与其滞后版本之间可能存在的非线性单调关系或分段线性关系，而无需假设二者呈线性关联。为考察即时的时间持续性，我们针对每个 DMA，计算四年间所有连续日对（即当前日与次日）热需求之间的斯皮尔曼相关性 ρ。如图3所示，分析结果表明相关性始终较强，DMA的均值通常在0.8至0.9之间。这说明当日消耗对次日需求有显著影响。此外，观测到连续工作日之间的相关性略高于工作日到周末转换期间的相关性，这种差异可能源于各 DMA 内部社会与消费者行为模式的变化。还应注意，尽管大多数连续日对都呈现强正相关，但也存在正相关较弱甚至负相关的情况。从2016年至2020年，DMA A、DMA B和DMA C分别有25、32和17对负相关。

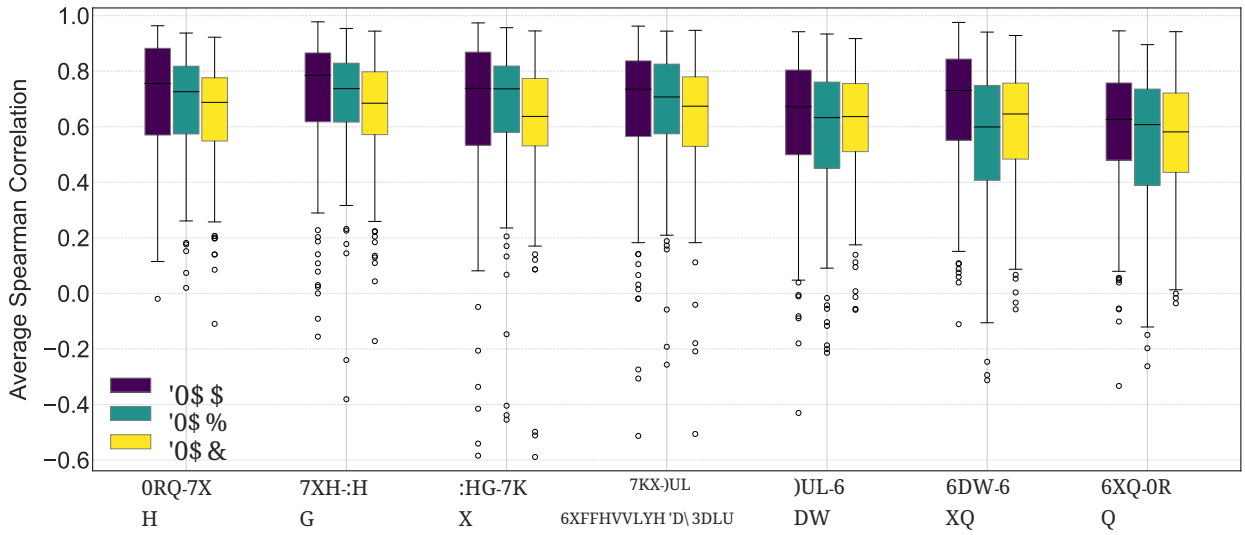
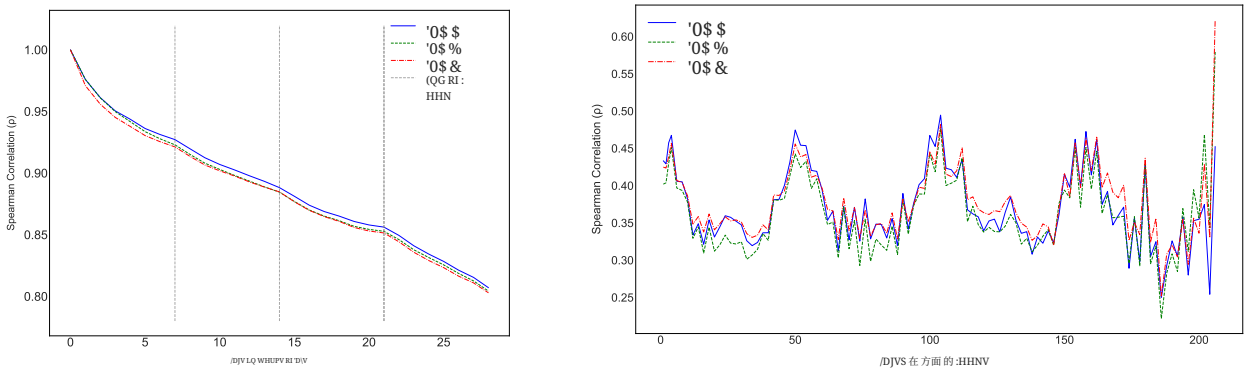


图3：不同 DM 中一周内连续日之间的斯皮尔曼相关系数分布 A.

在此分析基础上，我们考察当前热需求与不同滞后期（具体为1至28天，相当于24、48.....672小时）的历史值之间的关系。图4a显示，尽管相关性很强，但会随着滞后期增加而减弱。当滞后期为7天（168小时）的倍数时，相关性会略有周期性上升，这表明存在周季节性。这凸显了热需求时间序列的自回归特性。这些发现对于有效进行特征选择至关重要，尤其是在采用机器学习和深度学习方法对需求建模时。为避免冗余、增加计算成本以及对某些行为的过度表征，必须优先选取近期的特征（前一天的需求）和反映周周期的特征（前一周同一天的需求），同时有策略地限制纳入历史数据中相同特征的数量。

此外，为探究长期关系模式，我们对覆盖整个四年数据集、以168小时为单位的每周需求进行了相关性分析。图4b展示了这一更宏观的视角，进一步凸显了年度周期性以及周际相关性较强的周期性重复变化，从而全面揭示了热需求数据中的多尺度时间依赖关系。表1通过斯皮尔曼相关系数，从统计角度证明了需求在24小时和168小时滞后期下呈现的这种依赖性。

除了热需求的自回归特性外，外生因素，尤其是气象条件，众所周知会显著影响热能消耗。为了量化这些因素在我们数据集中的关系，



(a) 4周相关性分析

(b) 204周相关性分析

图4：热需求数据的短期和长期相关性分析。(a) 热需求数据与热需求数据滞后版本之间的斯皮尔曼相关 ρ。为评估短期关系，考虑了相当于1、2、.....、28天的小时滞后。(b) 四年热需求数据集包含的204周中，不同DMA之间每周需求的斯皮尔曼相关 ρ。

表1：小时分辨率下可用的内生和外生特征，以及它们与需求的斯皮尔曼相关系数。

特征类型	**特征**	**斯皮尔曼相关系数 (ρ)**		
		DMA A	DMA B	DMA C
消耗	24小时滞后（$\mathbf{c}_{24}$）	0.973	0.975	0.970
	168小时滞后（$\mathbf{c}_{168}$）	0.923	0.920	0.918
天气	温度（$\mathbf{t}_{amb}$）	-0.907	-0.914	-0.911
	最低气温（$\mathbf{t}_{min}$）	-0.902	-0.910	-0.907
	最高温度（$\mathbf{t}_{max}$）	-0.921	-0.927	-0.924
	体感温度（$\mathbf{t}_{feels}$）	-0.915	-0.922	-0.919
	降雨量（1 小时、3 小时）	-0.046/0.021	-0.050/0.017	-0.051/0.016
	降雪量（1 小时，3 小时）	0.130/0.106	0.131/0.106	0.130/0.106
	风速	0.121	0.123	0.116
	风向	-0.114	-0.115	-0.112
	阵风	0.065	0.064	0.064
	湿度	0.263	0.255	0.255
	露点（$\mathbf{t}_{dew}$）	-0.847	-0.857	-0.854
	气压	-0.073	-0.062	-0.062
	云量	0.223	0.213	0.210
Time	一天中的小时	周期性编码为具有相应周期的正弦-余弦对		
	星期			

通过计算小时热需求与一组全面天气变量之间的斯皮尔曼相关系数 ρ，进行初步评估。表1总结了三个DMA 中的这些相关性。我们使用正式符号表示关键特征，其中滞后h 小时的历史消耗量表示为$\mathbf{c}_h$。天气特征表示为$\mathbf{t}_x$，具体天气特征中的环境温度、最低温度、最高温度、体感温度和露点分别表示为$\mathbf{t}_{amb}$、$\mathbf{t}_{min}$、$\mathbf{t}_{max}$、$\mathbf{t}_{feels}$和$\mathbf{t}_{dew}$。

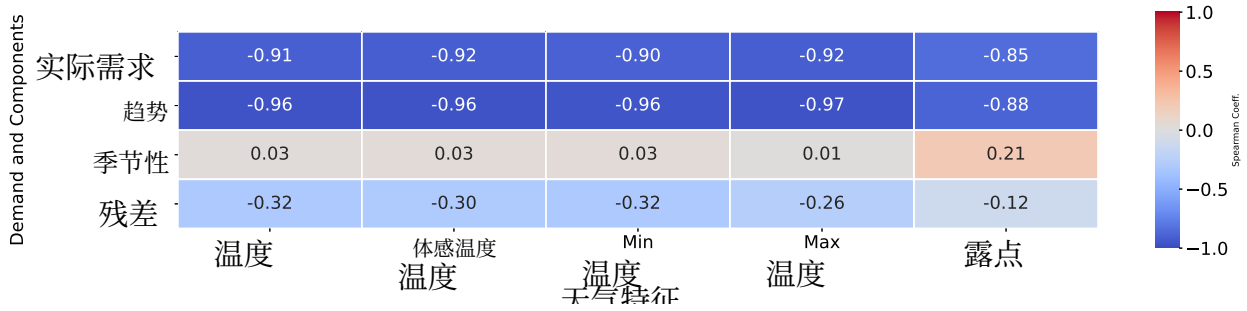
从表1可以看出，温度相关变量（环境温度、最低/最高温度和体感温度）在所有 DMA 中均与热需求呈现强且一致的负相关，ρ 数值通常低于 -0.9。露点$\mathbf{t}_{dew}$ 也呈现较强的负相关（约为 -0.85）。相比之下，降雨量、积雪量、风速、风向、阵风、气压和云量等其他天气变量与小时热需求的相关性则弱得多。

鉴于温度相关变量和露点的显著影响，为了理解这些关键气象特征如何与热需求相关——不仅体现在其原始形式上，也体现在其不同的潜在时间分量上——本文进行了更为细致的分析。为此，对热需求时间序列以及这些选定关键气象特征的时间序列（$\mathbf{t}_{amb}$、$\mathbf{t}_{min}$、$\mathbf{t}_{max}$、$\mathbf{t}_{feels}$、$\mathbf{t}_{dew}$）进行加法季节分解，以获得其趋势、季节和残差分量。随后，计算热需求各对应分量与每个关键气象特征相应分量之间的斯皮尔曼相关系数 ρ。

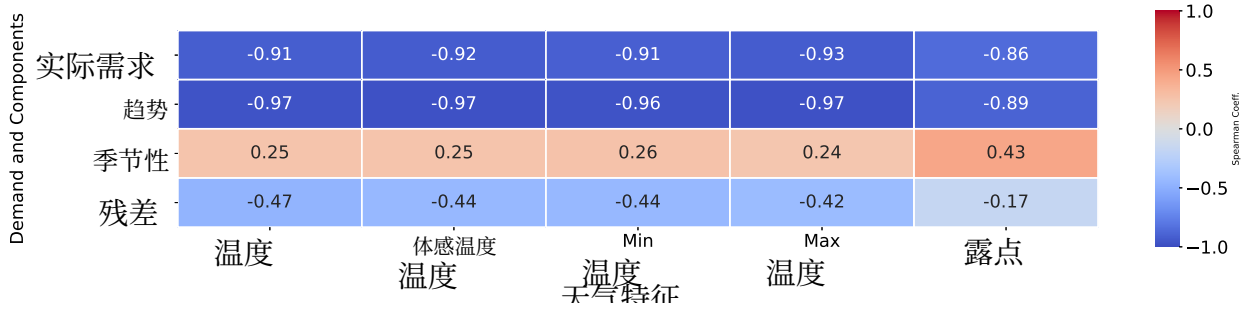
这些多分量相关性以图5的形式进行可视化。图5中的热图分为四行，依次展示实际热需求与实际天气数据、各自的趋势分量、各自的季节分量以及各自的残差分量之间的相关性。对图5的分析揭示了热需求与所选天气变量在不同时间分量上的关系中存在若干关键模式。第一行表示实际热需求与实际天气数据之间的相关性，即表1中呈现的初步结果。对热图趋势分量（第二行）的考察表明，热需求的长期趋势与温度相关天气变量（包括露点）的趋势之间存在极强的负相关。这种稳健的关系表明，热消耗潜在的长期变化与关键气象指标趋势中相应的长期变化密切对应。季节分量（第三行）涉及的相关性呈现出更加多样化的情况。对于DMA A和DMA C，各种温度指标的季节分量与热需求季节分量之间的相关性可忽略不计。在DMA B中，季节分量之间观测到弱相关。然而，露点的季节分量不同于温度，其与各DMA中的季节性热需求呈现弱到中等程度的相关性。最后，与趋势或实际数据的相关性相比，残差分量（第四行）之间的相关性总体有所减弱。对于温度相关天气变量，残差相关性尽管仍为负相关，但已显著减弱，处于弱到中等相关范围内。这意味着，在考虑趋势之后，热需求中剩余的短期、不规则波动与这些天气变量之间的耦合程度降低了，尽管二者之间仍然存在一定关系。

图 6 对这些分解关系进行了定性示意，展示了一个月内热需求与最高温度在实际分量和分解分量之间的相互作用。该可视化突显了热需求的季节分量主要反映其自身固有的历史季节性。同时，热需求与最高温度各自趋势分量之间的反向关系清晰可见。此外，这种分解可视化有助于洞察温度对需求的影响在不同时间尺度和分量上的变化特征及一致性。

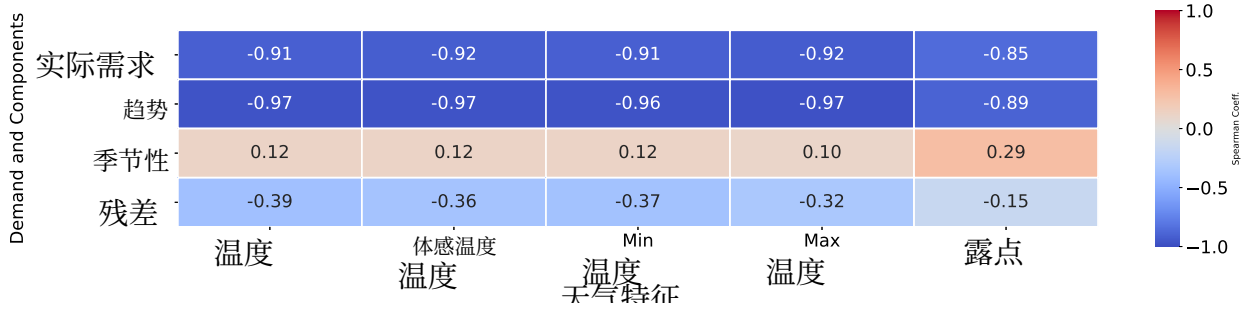
此外，日历特征也会影响热需求。图4b和图4a中的分析表明，消耗具有固有的周期性，因此需要恰当地表示时间指标。因此，一天中的时间和星期等特征会转换为正弦-余弦对，以有效模拟其各自的日、周和年周期性模式。为进一步丰富特征集，还纳入了公共假日和其他特殊场合的特定指标，以捕捉它们独特但往往并不频繁的需求影响。我们将这些明确的节假日特征与常规周末模式区分开来，因为周末已通过“星期”的周期性编码自然地表示。



(a) 计算得到的 DMA A 热需求与天气变量之间的斯皮尔曼相关。



(b) 计算得到的 DMA B 热需求与天气变量之间的斯皮尔曼相关。



(c) 计算得到的 DMA C 热需求与天气变量之间的斯皮尔曼相关。

图5：不同DMA中热需求与关键天气特征之间的斯皮尔曼相关系数。热图分为四行，分别显示以下相关性：(1) 实际热需求与实际天气数据之间的相关性；(2) 各自趋势分量之间的相关性；(3) 各自季节性分量之间的相关性；以及 (4) 各自残差分量之间的相关性。

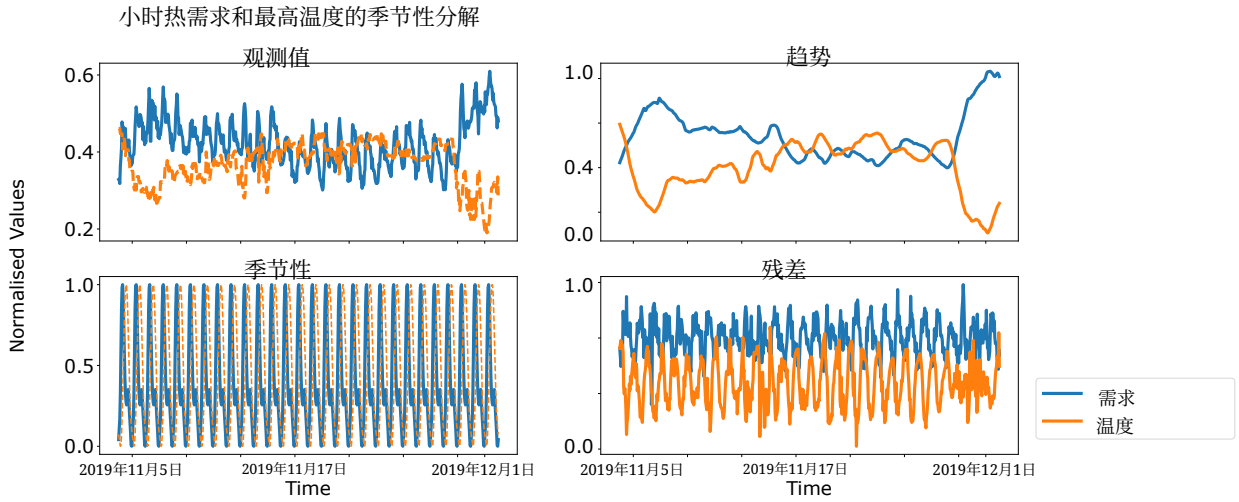


图 6：某一特定 DMA B 在一个月内按小时采样的季节分解需求数据和最高温度，展示了需求各组成部分与外生特征之间的明确关系。

2.4. 特征表示

预测框架的输入特征表示始于对时间数据进行结构化。对于每个输入特征k（如表1所总结），根据历史值，将输入特征向量表示为$\mathbf{s}_k(t) = [s^{t-N_{history}+1}, s^{t-N_{history}+2}, \ldots, s_t]$，其中包含截至当前时间$t$的$N^{history}$个历史观测值。利用这组经过筛选的$N_{features}$个输入变量的历史信息，可预测目标向量$\mathbf{y}(t) = [y_{t+1}, y_{t+2} \ldots, y_{t+N_{horizon}}]$；该向量包含目标变量（即小时热需求）的$N_{horizon}$个未来值。

每个包含 N 个特征的输入特征向量，均通过连续小波变换方法 [50] 从时域转换为二维时频表示，即小波尺度图。尺度图展示信号中固有的频率分量，并能关键性地确定这些分量在时间上的位置。连续小波变换通过将输入信号与母小波 $\psi_{a,\tau}(s_k)$ 进行卷积来运算；母小波是一种原型函数，可按因子 a 系统地缩放，并通过 τ 平移。特征 k 对应的尺度图记为 S_k，其维度为 M $\times N^{\text{history}}$，其中 M 表示小波分析中使用的离散尺度数量。连续小波变换将 s_k 的时间动态有效映射为一个矩阵，其中行对应频率尺度，列对应 N^{history} 窗口内的时间步。最后，将 N_{features} 个独立尺度图 $S_1, S_2, \ldots, S_{N_{\text{features}}}$沿第三个维度拼接，形成单个三维输入张量 X。该输入张量 $X \in \mathbb{R}^{M \times N_{\text{history}} \times N^{\text{features}}}$作为全面的类图像表示，捕捉特定时刻的时频分量。模型的任务是学习从 X 到目标向量 $y(t) = [y_{t+1}, y_{t+2} \ldots, y_{t+N_{horizon}}]$的映射。

选择连续小波变换而非离散小波变换（DWT），因为它有望提供高保真度的时频表示 [51]。小波变换的效果很大程度上取决于母小波的选择 $\psi(\cdot)$，因为不同信号具有独特的内在特征。因此，对于每个N_{features}输入时间序列，都会通过定量方法选择最优母小波。

2.5. 预测模型

将输入时间序列转换为时频域张量 $\mathbf{X} \in \mathbb{R}^{M \times N_{\text{history}} \times N^{\text{features}}}$ 后，我们采用基于卷积神经网络的架构来学习复杂模式，并将其映射到所需的预测时域。图 7 所示的模型架构专为解读图像式尺度图数据、执行多步超前回归任务而设计。该模型整体表示一个映射函数 F_{model}，其可学习参数为 θ，因此：

$$\hat{\mathbf{y}}(t) = F_{\text{model}}(\mathbf{X}; \theta) \tag{1}$$

架构设计直接依据分解热需求数据的特征及其与外生特征之间的关系。由于 CWT 能够明确地在时间维度上定位频率分量，采用二维卷积核可使模型捕捉各特征中的局部时间周期模式。尤为重要的是，通过沿深度方向堆叠尺度图，卷积核能够同时学习跨特征交互，例如时间 t 时温度与需求之间的耦合关系。池化层会进行空间下采样，从而有效降低时间分辨率和频谱分辨率 [52]。在区域供热系统中，准确预测早高峰出现的时间与预测其幅度同样重要。因此，为减少空间不变性，我们不采用池化操作，以保留高保真度的时频定位；这使网络能够维持输入特征与预测向量之间的相位相干，并借助卷积的空间等变性，保留一天中的时间与需求模式之间的关系。因此，我们在整个卷积堆叠中保留完整的空间分辨率，并采用高维展平操作，使后续全连接层能够捕捉最重要的依赖关系，而非薄弱且充满噪声的模式。为降低这种高维表示固有的过拟合风险，我们在线性层之间加入 Dropout 层，确保模型学习到稳健且具有泛化能力的特征，而不是记住噪声。

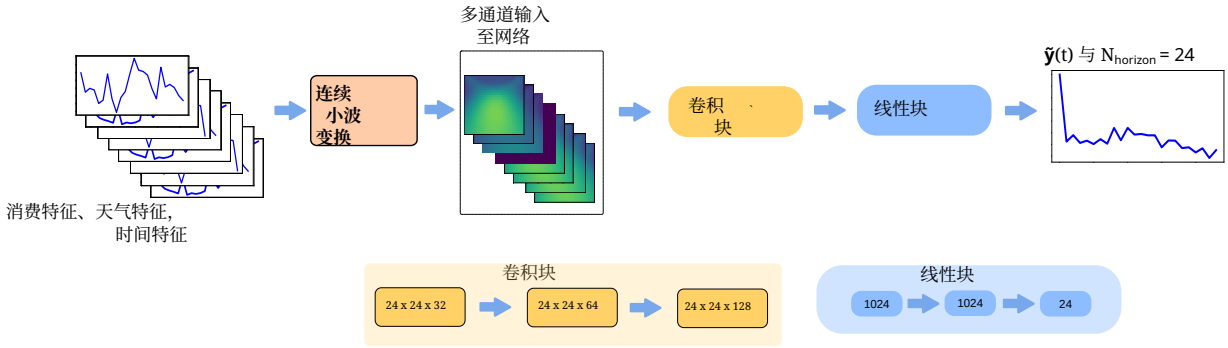


图 7：小波摄氏度 以尺度图的类图像多通道表示为输入的神经网络预测网络，用于预测日前逐小时需求预测。

初始模块由L 个二维卷积层堆叠而成，负责从输入张量$\mathbf{X}$中提取特征。其目的是识别并学习尺度图中的层次化模式，这些模式对应于原始信号中特定的重要时频事件。模型的输入$\mathbf{H}^{(0)} = \mathbf{X}$ 经过l层卷积层，其中一组$C^{(l)}$ 个滤波器作用于前一层的输入特征图$\mathbf{H}^{(l-1)}$。每个滤波器通过卷积运算生成一个输出特征图，并逐元素应用修正线性单元（ReLU）激活函数，以引入非线性。每个卷积层l 上的运算可表示为：

$$\mathbf{H}^{(l)} = \mathrm{ReLU}(\mathrm{Conv2D}^{(l)}(\mathbf{H}^{(l-1)})) \tag{2}$$

其中$\mathbf{H}^{(l)} \in \mathbb{R}^{M \times N_{\text{history}} \times C^{(l)}}$ 是输出特征图张量。

在最后一个卷积层之后，得到的三维特征图张量$\mathbf{H}^{(L)}$ 被展平为一个一维向量，$\mathbf{v}_{\text{flat}}$，用作线性回归头的输入。

$$\mathbf{v}_{\text{flat}} = \mathrm{Flatten}(\mathbf{H}^{(L)}) \tag{3}$$

该向量的维度为 $M \cdot N_{\text{history}} \cdot C^{(L)}$。随后，该向量由多层感知机（MLP）处理，以完成最终回归。MLP 由一系列全连接稠密层组成。每个稠密层都会计算其输入的加权和，并应用激活函数。稠密层 j 的变换表示为：

$$\mathbf{d}^{(j)} = \sigma(\mathbf{W}^{(j)}\mathbf{d}^{(j-1)} + \mathbf{b}^{(j)}) \tag{4}$$

其中，$\mathbf{d}^{(j-1)}$表示前一层的输出（其中 $\mathbf{d}^{(0)} = \mathbf{v}_{\text{flat}}$），$\mathbf{W}^{(j)}$和 $\mathbf{b}^{(j)}$分别为权重矩阵和偏置向量，σ表示 Leaky ReLU 激活函数。回归头将$\mathbf{v}_{\text{flat}}$ 映射到神经元数量分别为$d_{neurons}$ 的全连接层，随后通过最终预测层（含$N_{horizon}$ ）来预测热需求 $\tilde{\mathbf{y}}(t)$。

2.6. 实验设置

2.6.1. 数据集

本研究使用了一个历史数据集，涵盖四年（2016年1月1日至2019年12月31日）的区域供热消耗及相关特征，如第 2.1 节所述。模型配置为根据由 N_{features} 个输入变量构成的、长度为 $N_{\text{history}} = 24$ 小时的历史上下文窗口，预测随后 $N_{\text{horizon}} = 24$ 小时的热需求。研究采用按时间顺序划分数据的方式模拟真实的预测场景，将最后 52 周（364 天）指定为留出测试集。为期一年的测试能够充分涵盖不同季节、假期时段或上学时段等心理特征行为，以及每种假日的至少一次出现，从而确保评估稳健且全面。此前的数据也按时间顺序进一步划分，其中 80% 用于训练，随后 20% 用作验证数据集。训练期间，模型按时间顺序分批迭代。该划分方案产生了 872 个训练样本、219 个验证样本和 364 个测试样本，每个样本均基于输入特征表示一个 24 小时预测窗口，且上下文长度为 24 小时。

为验证泛化能力，我们对公开数据集采用相同的模型配置（$N_{\text{horizon}} = 24$,$N_{\text{history}} = 24$）重复这一评估方案。弗伦斯堡数据集 [46], 覆盖了从 2017 年 1 月 1 日至 2024 年 12 月 31 日的 8 年数据，用于评估德国城市需求预测的多年表现。奥尔堡数据集 [34], 涵盖 ≈ 3年数据，时间范围为 2018 年 1 月 2 日至 2020 年 12 月 31 日，代表 3021 个住宅单元的住宅供热消耗。对于这两个公开数据集，测试集之前剩余的历史数据均按照相同的时间顺序以 80/20 的比例划分为训练集和验证集，确保所有实验配置中均未发生数据泄漏。

2.6.2. 架构

该框架的核心由三个二维卷积层组成，滤波器数量分别为 32、64 和 128。所有卷积层均采用 3×3 卷积核、步幅为 1，并使用相同填充，以保留输入张量的空间维度。线性层包含两个隐藏全连接层，每层有 1024 个神经元，并采用斜率为 0.3 的 Leaky ReLU 函数进行激活，以确保梯度流稳定，随后接一个丢弃率为 0.1 的 Dropout 层。最后一层为含 24 个神经元的全连接输出层，用于生成最终预测。最终模型配置未采用池化或 L2 正则化（权重衰减）。有关超参数搜索过程以及针对池化、Dropout 和正则化的消融研究，详见附录。模型概述见表 2。

表 2：所提出的卷积神经网络模型的详细架构，共有 76,669, 976 个参数。输入张量维度对应于 $(\mathrm{M} \times N_{history} \times N^{features})$。

层	算子	输出维度	卷积核	步幅	填充
输入	-	$24 \times 24 \times N_{\text{features}}$	-	-	-
卷积 1	二维卷积 + ReLU	$24 \times 24 \times 32$	3×3	1	1（相同填充）
卷积 2	二维卷积 + ReLU	$24 \times 24 \times 64$	3×3	1	1（相同填充）
卷积 3	二维卷积 + ReLU	$24 \times 24 \times 128$	3×3	1	1（相同填充）
展平	展平	73728	-	-	-
全连接层 1	全连接层 + 带泄漏的线性整流函数	1024	-	-	-
全连接层 2	全连接层 + 带泄漏的线性整流函数	1024	-	-	-
丢弃率	丢弃率（$p = 0.1$）	1024	-	-	-
输出	全连接层	24	-	-	-

模型的可训练参数使用Adam算法进行优化，学习率为 $\eta = 0.001$。此外，还采用了学习率调度器，在验证损失进入平台期时将学习率按0.9的因子衰减。训练目标是最小化均方误差（MSE）损失函数，该函数用于量化预测需求值与实际需求值之间的平方差。

模型以每批32个样本进行训练，最多训练1000轮。为防止过拟合并提升模型的泛化能力，训练循环中集成了早停机制。该机制会监测验证集上的均方误差损失；如果连续50个训练轮次未观测到改善，则终止训练。终止后，将验证损失最低的训练轮次对应的模型权重恢复，用于在测试集上进行最终评估。有关架构、超参数搜索空间、硬件和计算效率的消融研究记录在附录A中。

2.6.3. 评估策略

使用平均绝对误差（MAE）和平均绝对百分比误差（MAPE）这两项标准指标评估模型的预测准确率。MAE 衡量预测值的平均绝对偏差，而 MAPE 则以实际值的百分比表示误差，提供与尺度无关的视角。为了帮助评估新增特征对异常值的影响，也报告均方误差（MSE）。这些指标的形式化定义如下：

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \tilde{y}_i| \tag{5}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \tilde{y}_i}{y_i}\right| \tag{6}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 \tag{7}$$

其中，n 为预测数据点总数，y_i 表示真实值，$\tilde{}y_i$ 则是其对应的模型预测。由于需求在夏季和冬季之间呈周期性变化，作为相对误差指标的 MAPE 可用于比较不同季节的预测结果；而 MAE 具有绝对量纲，因此能提供更直观的比较。同时，我们也考虑到 MAPE 可能因除以零而不稳定。综合来看，这些指标刻画了预测值相对于观测需求的逐小时准确度。

2.6.4. 实验概述

本研究设计了一系列结构化实验，以系统地评估和优化所提出的预测框架。评估方案包含多个方面，旨在首先确定最佳的特征组合和特征表示。为进行严格的基准测试，最终模型将与一系列成熟方法进行全面比较。实验首先建立一个仅依赖内生（自回归）需求特征的基线模型。在这项初步实验（实验一）中，我们比较24小时滞后期（**摄氏度**$_{24}$）与同时包含周滞后期的组合特征集（**摄氏度**$_{24}$,**摄氏度**$_{168}$）的预测效用，并特别关注图3和图4a所示的模式转变时段的表现。该基础特征集记为 $\mathcal{M}_C$。随后，在实验II中，通过系统地纳入外生气象变量来评估该框架的性能。我们评估了将初步数据分析确定的关键天气特征单独加入以及以不同组合加入已确定的内生基线 $\mathcal{M}_C$ 特征集所产生的影响。此过程旨在确定信息量最高且冗余最少的天气特征集。

为进一步增强模型的预测能力，我们在实验 III 中研究了一种基于时间序列分解的先进特征工程策略。该方法通过添加时间序列的构成分量来扩充原始时间序列特征。

如表 3 所示，消耗特征中加入了季节性、趋势和残差分量，而天气相关特征中仅加入了趋势和残差分量。忽略天气特征的季节分量，因为热需求固有的季节性已由其自身历史特征的季节性分量充分体现，图 6 中的分析也支持这一观察。本文比较了扩展各项需求和天气特征表示方式的效果，并选出最佳特征及其表示方式

表 3：原始特征集和分解特征集的定义。d 上标在$\mathbf{c}_{24}^{d}$、$\mathbf{c}_{168}^{d}$和$\mathbf{t}_{x}^{d}$ 中表示这些表示包含分解分量。

特征类型	原始向量	分解分量	分解向量
消耗	$\mathbf{摄氏度}_{24} \in \mathbb{R}^{24\times1}$ $\mathbf{摄氏度}_{168} \in \mathbb{R}^{24\times1}$	季节性、趋势、残差	$\mathbf{c}_{24}^{d} = \left([\mathbf{摄氏度}_{24}] \cup [\mathbf{摄氏度}_{24}^{trend},\ \mathbf{摄氏度}_{24}^{seas},\ \mathbf{摄氏度}_{24}^{resid}]\right) \in \mathbb{R}^{24\times4}$ $\mathbf{c}_{168}^{d} = \left([\mathbf{摄氏度}_{168}] \cup [\mathbf{摄氏度}_{168}^{trend},\ \mathbf{摄氏度}_{168}^{seas},\ \mathbf{摄氏度}_{168}^{resid}]\right) \in \mathbb{R}^{24\times4}$
天气	$\mathbf{t}_x \in \mathbb{R}^{24\times1}$	趋势，残差	$\mathbf{t}_{x}^{d} = \left([\mathbf{t}_x] \cup [\mathbf{t}_x^{trend},\ \mathbf{t}_x^{resid}]\right) \in \mathbb{R}^{24\times3}$

。随后，在实验 IV 中，我们考察加入显式周期性日历特征的影响，并评估多种嵌入公共假日信息的策略。逐步进行特征评估至关重要，因为在现实世界预测中，即使平均准确率略有提升，也能带来更好的决策和运营效益。精心筛选特征有助于减少非典型或难以预测日期上的较大误差，也有助于了解现有信息在建模中的局限，以及因信息不足而可能出现预测错误的情形。尽管较简单的特征集在大多数情况下可能表现尚可，但即使整体性能提升看似有限，降低这些极端偏差的能力也足以证明深入分析的必要性。

为在当前时间序列预测领域全面比较所提出的框架，我们开展了基准测试研究，将其作为实验 V。评估方案旨在将我们的方法与广泛的模型进行比较，涵盖从经典统计方法到最先进的深度学习架构。选定的基准模型包括传统统计模型（SARIMAX） [53], 基于树的集成方法（XGBoost） [54], 经典循环神经网络（长短期记忆网络） [55], 以及若干以长序列预测见长的现代 Transformer 模型，它们采用仅编码器或编码器-解码器架构，具体包括 Informer [24], Autoformer [25], Preformer [56], PatchTST [57], 和 TimesNet [26]。此外，我们还采用了已被证明能有效进行时间序列预测的时间序列混合模型，例如 DLinear [58], TSMixer [59], 和 TimeMixer [60] ，用于基准测试。此外，为评估模型在新兴的大型预训练基础模型范式下的性能，我们将经过微调的 TTM [44]和 Chronos-2 [45] 纳入比较。为确保比较公平，我们对所有基线方法采用统一的训练方案。虽然针对各架构的特定超参数分别通过网格搜索进行优化，以确保达到最佳性能，但计算预算保持一致。所有模型的批量大小均固定为 32，使用相同的 ReduceLROnPlateau 调度器，最多训练 1000 轮，并采用相同的早停耐心值（50 轮）。各模型的详细超参数搜索空间和最终配置见附录 B。

最后，在实验 VI 中，我们采用滚动区块预测策略进行纵向性能分析。本实验旨在评估所提出模型的时间鲁棒性，以及其适应多年累积历史数据的能力。在此设置中，模型首先使用最早一年的可用数据进行训练，以预测下一年。随后，将训练数据扩展至前两年，并在第三年进行评估。这个迭代过程持续进行，直至预测数据集的最后一年。我们在布伦讷斯莱夫、弗伦斯堡和奥尔堡的全部三个数据集上进行了这一严格的滚动评估，从而全面考察模型在长期运行周期中的稳定性。

3. 结果与讨论

3.1. 特征组合

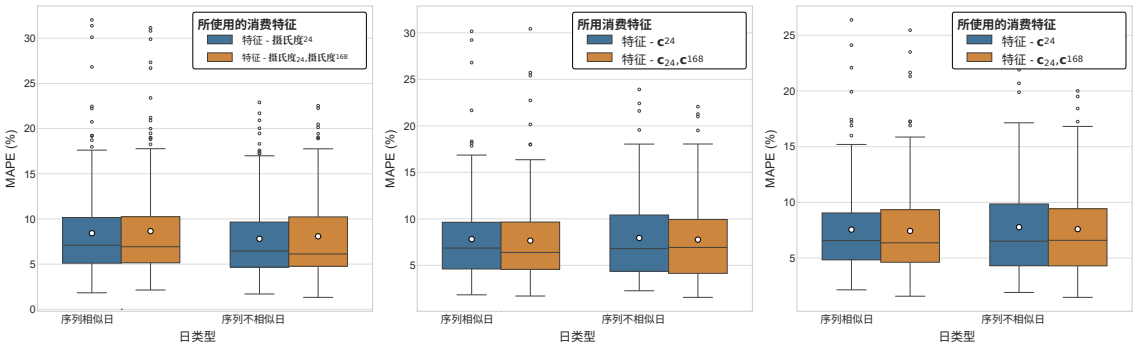
3.1.1. 内生特征 - 实验一

我们最初的特征评估着重考察内生消费特征的影响。我们比较了模型使用仅滞后 24 小时的消耗（$\mathbf{c}_{24}$）作为特征的基线，与在日滞后特征中加入周滞后特征 $[\mathbf{c}_{24}, \mathbf{c}_{168}]$ 后的预测能力。如表4所示的汇总结果，加入$\mathbf{c}_{168}$ 特征仅使 DMA B 和 DMA C 的指标略有改善，对 DMA A 则没有改善。

表4：不同内生特征集的模型性能比较（平均绝对误差、平均绝对百分比误差和均方误差）。

DMA	特征	平均绝对误差 (千瓦时)	平均绝对百分比误差 (%)	均方误差($\times 10^4$ 千瓦时^2)
DMA A	$\mathbf{c}_{24}$	160.78 ± 114.56	8.1 ± 4.9	5.65 ± 9.09
	$\mathbf{c}_{24}, \mathbf{c}_{168}$	165.98 ± 119.89	8.4 ± 5.2	5.95 ± 9.35
DMA B	$\mathbf{c}_{24}$	300.87 ± 206.56	7.9 ± 4.6	18.32 ± 25.44
	$\mathbf{c}_{24}, \mathbf{c}_{168}$	293.57 ± 203.57	7.7 ± 4.6	17.44 ± 24.22
DMA C	$\mathbf{c}_{24}$	264.69 ± 172.27	7.7 ± 4.3	13.74 ± 18.52
	$\mathbf{c}_{24}, \mathbf{c}_{168}$	260.04 ± 172.59	7.5 ± 4.2	13.43 ± 18.64

我们仍采用特征集$\mathbf{c}_{24}, \mathbf{c}_{168}$，这是因为预测结果在下游应用中所需的性能存在一个关键细节，尤其是在节假日和罕见事件期间。为进一步探究这一点，我们根据日间模式相似性对预测误差进行了分层，详见图8。分析将日期分为两组：序列相似日，即消费模式预计会延续前一日的模式（例如，星期一之后的星期二）；以及序列不相似日，即出现模式中断的日期，例如周末与工作日转换（星期日到星期一）以及工作日中的节假日。



(a) DMA A (b) DMA B (c) DMA C

图 8：仅使用 24 小时消费滞后（$\mathbf{c}_{24}$）与同时使用 24 小时和 168 小时滞后（$\mathbf{c}_{24}, \mathbf{c}_{168}$）的模型，其平均绝对百分比误差分布。比较了三个 DMA 中序列相似日与出现模式突变日的模型表现。

图 8 的定量结果表明，尽管两种模型在序列相似日上的表现相当，但加入周滞后（$\mathbf{c}_{168}$）后，模型在序列不相似日上具有优势。异常值数量减少与均方误差降低有关。对于这些过渡时期，同时利用$\mathbf{c}_{24}$和$\mathbf{c}_{168}$的模型取得了更低的 MAPE 中位数和更小的四分位距，同时减少了预测中的异常值。这表明，$\mathbf{c}_{168}$ 特征是重要的稳定锚点，尤其在$\mathbf{c}_{24}$ 所捕捉的日模式最不可靠时，能够提高预测稳健性。尽管它对整体平均指标的影响有限，但它能在模式转变期间缓解较大的误差，同时不影响其他样本上的表现，因此值得纳入我们的特征集。我们将 $[\mathbf{c}_{24}, \mathbf{c}_{168}]$ 的内生特征集记为 $\mathcal{M}_C$。纳入$\mathbf{c}_{168}$ 也为我们提供了一种途径，可利用相关历史需求来建模罕见事件的需求特征，而这类事件可能具有可变滞后期。

3.1.2. 外生特征 - 实验*II*

随后，特征分析扩展至纳入外生变量，具体而言，即表1所示相关性分析中筛选出的天气特征。为量化这些特征的影响，我们将其逐一以及以不同组合的方式系统地添加到实验一的基准消费特征集（$\mathcal{M}_C$）中。各项实验的量化结果按 DMA 汇总于表5、表6和表7。

三个区域中均得出一个主要且一致的发现：与仅消费基准相比，加入任一单独的温度相关特征，都能显著提高预测准确率。提升幅度最小的是加入$\mathbf{t}_{\text{dew}}$，DMA C约为7.3%，DMA B约为14.7%。

表5：DMA A 各种天气特征组合的性能指标，在每个部分内按平均绝对误差降序排列。基础模型 $\mathcal{M}_C$ 包含消费特征。表现最佳的模型以粗体标出。

模型特征	**平均绝对误差 (千瓦时)**	**平均绝对百分比误差 (%)**	**均方误差 (千瓦时$^2 \times 10^4$)**
$\mathcal{M}_C$	166.0 ± 119.9	8.40 ± 5.20	5.95 ± 9.35
$\mathcal{M}_C, \mathbf{t}_{\text{dew}}$	149.1 ± 100.8	7.87 ± 4.83	4.73 ± 7.72
$\mathcal{M}_C, \mathbf{t}_{\text{min}}$	130.4 ± 88.5	6.72 ± 3.85	3.65 ± 6.55
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}$	129.1 ± 94.2	6.58 ± 3.90	3.71 ± 6.85
$\mathcal{M}_C, \mathbf{t}_{\text{max}}$	127.0 ± 87.7	6.41 ± 3.44	3.58 ± 6.84
$\mathcal{M}_C, \mathbf{t}_{\text{amb}}$	122.7 ± 80.0	6.45 ± 3.50	3.21 ± 6.16
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{min}}$	126.2 ± 92.2	6.46 ± 3.76	3.60 ± 7.30
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}$	123.4 ± 85.4	6.22 ± 3.29	3.33 ± 6.18
$\mathcal{M}_C, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	121.1 ± 80.5	6.32 ± 3.51	3.17 ± 6.20
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{min}}$	121.1 ± 82.6	6.28 ± 3.59	3.23 ± 6.40
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{amb}}$	120.1 ± 82.2	6.21 ± 3.51	3.17 ± 6.18
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}$	117.6 ± 78.6	6.18 ± 3.46	3.03 ± 6.08
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{min}}$	121.4 ± 82.3	6.47 ± 4.04	3.19 ± 6.20
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	119.3 ± 84.9	6.13 ± 3.49	3.23 ± 6.27
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}$	118.9 ± 81.3	6.24 ± 3.54	3.10 ± 6.20
$\mathcal{M}_C, \mathbf{t}_{\text{feels}}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	**116.0 ± 78.7**	**6.12 ± 3.48**	**2.97 ± 6.02**
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	117.7 ± 81.4	6.15 ± 3.51	3.06 ± 6.22

表6：DMA B 各种天气特征组合的性能指标，按每个部分内平均绝对误差降序排列。基础模型 $\mathcal{M}_C$ 包含消费特征。表现最佳的模型以粗体标出。

模型特征	**平均绝对误差 (千瓦时)**	**平均绝对百分比误差 (%)**	**均方误差 (千瓦时$^2 \times 10^4$)**
$\mathcal{M}_C$	293.6 ± 203.6	7.70 ± 4.60	17.44 ± 24.22
$\mathcal{M}_C, \mathbf{t}_{\text{dew}}$	250.4 ± 151.5	6.85 ± 4.15	12.30 ± 14.71
$\mathcal{M}_C, \mathbf{t}_{\text{min}}$	221.6 ± 119.7	6.11 ± 3.35	9.13 ± 9.74
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}$	216.8 ± 127.1	5.83 ± 3.11	9.10 ± 11.18
$\mathcal{M}_C, \mathbf{t}_{\text{max}}$	212.6 ± 121.0	5.64 ± 2.78	8.71 ± 9.78
$\mathcal{M}_C, \mathbf{t}_{\text{amb}}$	**195.2 ± 104.3**	**5.21 ± 2.57**	**7.22 ± 7.43**
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{min}}$	204.5 ± 120.1	5.45 ± 2.88	8.17 ± 9.88
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}$	201.8 ± 113.8	5.38 ± 2.72	7.95 ± 8.90
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}$	200.3 ± 107.6	5.47 ± 2.86	7.59 ± 8.09
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{min}}$	199.9 ± 111.3	5.37 ± 2.79	7.72 ± 8.62
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{amb}}$	196.2 ± 103.5	5.45 ± 3.00	7.25 ± 7.89
$\mathcal{M}_C, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	195.5 ± 106.6	5.25 ± 2.67	7.34 ± 7.93
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{min}}$	206.4 ± 105.2	5.91 ± 3.81	7.81 ± 7.78
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	201.6 ± 112.7	5.50 ± 3.05	7.83 ± 8.85
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	201.5 ± 114.4	5.52 ± 2.91	7.82 ± 10.18
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}$	199.5 ± 114.1	5.37 ± 2.72	7.74 ± 8.91
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	195.5 ± 106.1	5.26 ± 2.72	7.35 ± 7.76

表7：DMA C 各种天气特征组合的性能指标，按每个部分内平均绝对误差降序排列。基础模型 $\mathcal{M}_C$ 包含消费特征。表现最佳的模型以粗体突出显示。

模型特征	平均绝对误差 (千瓦时)	平均绝对百分比误差 (%)	均方误差 (千瓦时$^2 \times 10^4$)
$\mathcal{M}_C$	260.0 ± 172.6	7.50 ± 4.20	13.43 ± 18.64
$\mathcal{M}_C, \mathbf{t}_{\text{dew}}$	241.0 ± 144.5	7.19 ± 4.31	11.15 ± 13.50
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}$	192.6 ± 116.7	5.63 ± 2.89	7.24 ± 9.59
$\mathcal{M}_C, \mathbf{t}_{\text{max}}$	192.1 ± 103.5	5.74 ± 2.86	6.89 ± 7.84
$\mathcal{M}_C, \mathbf{t}_{\text{min}}$	189.7 ± 106.1	5.64 ± 2.81	6.80 ± 8.35
$\mathcal{M}_C, \mathbf{t}_{\text{amb}}$	176.7 ± 91.4	5.33 ± 2.62	5.80 ± 6.45
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{min}}$	188.6 ± 101.1	5.64 ± 2.78	6.58 ± 7.43
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}$	187.8 ± 106.8	5.53 ± 2.66	6.71 ± 7.99
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{min}}$	185.9 ± 101.7	5.62 ± 2.98	6.50 ± 7.54
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{amb}}$	180.5 ± 97.2	5.35 ± 2.62	6.13 ± 6.82
$\mathcal{M}_C, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	177.1 ± 94.1	5.26 ± 2.56	5.89 ± 6.60
$\mathcal{M}_C, \mathbf{t}_{\text{feels}}, \mathbf{t}_{\text{amb}}$	**173.4 ± 92.7**	**5.24 ± 2.71**	**5.61 ± 6.54**
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	182.3 ± 93.2	5.63 ± 3.16	6.05 ± 6.51
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{min}}$	182.3 ± 101.3	5.43 ± 2.79	6.30 ± 7.23
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}$	179.7 ± 93.2	5.53 ± 3.07	5.96 ± 6.54
$\mathcal{M}_C, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	178.9 ± 93.3	5.45 ± 3.01	5.90 ± 6.43
$\mathcal{M}_C, \mathbf{t}_{\text{max}}, \mathbf{t}_{\text{感受}}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\text{min}}$	178.5 ± 92.4	5.43 ± 2.83	5.87 ± 6.25

特征$\mathbf{t}_{\text{amb}}$ 显著改善了结果，DMA A 提高了 26.1%，DMA C 和 B 分别提高约 32% 和 33.5%。与相关性分析一致，$\mathbf{t}_{\text{amb}}$ 始终是影响最大的单一特征，而$\mathbf{t}_{\text{dew}}$ 的预测能力最弱，但仍带来了明显改善。

随后，评估扩展到特征组合，其动机是假设不同的温度指标能够捕捉独特的环境细微差异。例如，$\mathbf{t}_{\text{feels}}$ 纳入了湿度和风速等因素，而$\mathbf{t}_{\text{min}}$ 与$\mathbf{t}_{\text{max}}$ 的组合则有可能刻画供热管网的昼夜热力动态。对于 DMA A 和 DMA C，不同程度地纳入互补特征产生了最低误差，表明模型成功捕捉到了更细微的环境动态。然而，对于 DMA B，单一特征$\mathbf{t}_{\text{amb}}$ 优于复杂组合，尽管增加更多特征带来的性能下降很小（差异为 0.04% 至 0.15%）。尽管如此，需要指出的是，$\mathbf{t}_{\text{amb}}$ 为用气象特征增强内生特征提供了有力参照，而对内生特征的进一步增强仍能带来性能提升，尽管幅度较小且效果各异。事实上，同时加入全部四项主要温度特征后，相对于最优特征模型，性能出现了轻微下降。不过，这种表面上的冗余可能源于原始信号中趋势与季节信息的纠缠，从而掩盖了各变量的独特贡献。因此，在此阶段我们并不舍弃复杂特征集，而是将这一对比分析带入下一阶段。

3.1.3. 特征的季节性分解 - 实验*III*

表 8 对表 3 所述的使用原始特征的模型与加入分解分量的模型进行了全面比较。结果明确表明，采用分解特征集后，所有指标的误差均一致降低。对于每一种天气特征组合，使用分解特征的模型都显著优于相应的原始特征模型。这种提升幅度相当可观。例如，在 DMA C 中，表现最佳的原始特征模型的平均绝对误差为 173.4 千瓦时，而分解后的模型将其降至 152.6 千瓦时，相对提升近 12%。所有区域都取得了类似的改善，其中 DMA A 通过分解实现了 101.7 千瓦时的优异平均绝对误差。这种一致性验证了这一假设：明确分离信号分量能够简化模型的学习任务，使其可以有效地分别拟合趋势、季节性和残差驱动因素。

至关重要的是，分解改变了特征选择的格局。在实验II中，加入多个相关的温度变量往往会导致性能下降，尤其是在DMA B中。然而，经过分解后，

特征集 $[\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\text{feels}}, \mathbf{t}_{\text{amb}}]$ 成为 DMA B 和 C 的表现最佳者。这表明，尽管原始相关特征会引入噪声，但其分解分量包含独特且互补的信息，模型能够成功利用这些信息，有效释放这些复杂特征集的价值。

尽管最优配置各不相同，分解带来的一个关键结果是不同特征集之间的性能差距显著缩小。例如，在 DMA A 中，单个最佳特征（$\mathbf{t}_{\text{amb}}$）与复杂的最优特征集之间仅相差 1.1 千瓦时（102.8 对 101.7 千瓦时）。这一特点凸显了该框架在实际应用中的灵活性。这表明，虽然复杂组合在统计上略占优势，但分解使模型即使仅使用极少的数据，也能提取出大部分预测信号。在现实场景中，如果无法获取特定气象信息或信息有误，模型仅使用环境温度仍能保持很高的可靠性。不过，为了严格评估该框架在本次基准测试研究中的最大潜力，我们采用了针对每个区域确定的特定最优分解特征集：DMA A 使用 $[\mathcal{M}_C^d, \mathbf{t}^d_{\text{feels}}, \mathbf{t}^d_{\text{amb}}, \mathbf{t}^d_{\text{min}}]$，DMA B 和 C 使用 $[\mathcal{M}_C^d, \mathbf{t}^d_{\max}, \mathbf{t}^d_{\text{feels}}, \mathbf{t}^d_{\text{amb}}]$。

3.1.4. 时间与日历特征 - 实验 *IV*

特征工程分析的最后阶段评估了加入显式时间信息的影响。我们为实验III中表现最佳的分解模型加入了表示星期的周期性特征，并将其编码为正弦-余弦对（T_{cyc}）。比较结果见表9。

与时间序列预测中的普遍假设相反，加入显式周期性时间特征后，三个区域的性能均显著下降。DMA B和C的平均绝对误差分别增加了约6.5%和5.5%；Wilcoxon符号秩检验的p值（$p < 0.05$）证实，这种性能下降并非随机所致。DMA A的误差则从101.7千瓦时上升至103.0千瓦时。

通过模型架构（其中尺度图编码了时频信息），以及不含池化层（从而保留位置信息），再加上目标函数本身要求预测00:00至23:00的固定时间间隔，我们得以捕捉时间维度。尽管这有悖直觉，加入显式的 T_{cyc} 向量似乎引入了冗余信息，实际上起到了噪声的作用，导致泛化能力下降。因此，我们在最终模型配置中舍弃了显式周期性时间编码，转而采用更简单、更稳健的架构，直接从信号动态中提取时间上下文。

为应对非连续确定性变量建模这一挑战，我们研究了纳入显式节假日特征的效果。我们通过三种不同策略整合节假日信息。首先，使用一个分类的（$\mathbf{h}^{categorical}$）独热编码特征，标记一天中每个小时是否为假日。其次，为了捕捉节假日特有的行为，使用一个滞后（$\mathbf{h}^{lagged}$）特征，将前一周同一天的消耗值（$\mathbf{c}_{168}$）替换为上一次相同节假日的消耗值（例如，使用2018年圣诞节的数据来预测2019年圣诞节）。这保留了使用 $\mathbf{c}_{168}$ 的基本原则，即为模型提供特定日期的代表性需求行为。此动态替换考虑到节庆假日与典型每周模式存在显著差异。最后，采用一个组合（$\mathbf{h}^{lagged}, \mathbf{h}^{categorical}$）特征集，同时包含分类节假日特征和滞后节假日特征。

表10总结了这些方法在整个测试集上的总体定量表现。乍看之下，改进似乎微乎其微（例如，平均绝对误差减少了 < 1千瓦时）。然而，鉴于公共假日占全年数据的比例不到4%，总体指标会大幅稀释这些特征的影响。要真正评估其效用，还需对特定节假日时段进行细致分析。

图9详细展示了模型在各个公共假日上的性能。这些结果表明，没有哪一种节假日嵌入策略始终优于其他策略，最佳方法具有很强的随机性。然而，对于几乎所有节假日，至少有一种包含节假日信息的特征集优于不包含节假日信息的特征集。不同节假日之间的性能差异也并不一致。对于圣诞节和元旦等影响较大的节假日，基线模型的误差始终最高，往往无法预测需求波动。引入明确的节假日特征显著缓解了这一问题。例如，在 DMA A 和 C 中，类别嵌入（$\mathbf{h}^{categorical}$）大幅降低了圣诞节期间的平均绝对误差。在 DMA B 中，滞后特征（$\mathbf{h}^{lagged}$）表现更优。

此外，图10中的箱线图进一步证明了模型的稳定性。所有配置下，非节假日的误差分布几乎完全相同，这证实了添加稀疏节假日特征不会引入噪声，也不会降低常规运营日的模型性能。相比之下，在假日，对于所提出的方法，中位误差通常更低，四分位距也更小

表8：统　一比较所有 DMA 中采用原始特征与分解特征的模型性能 .

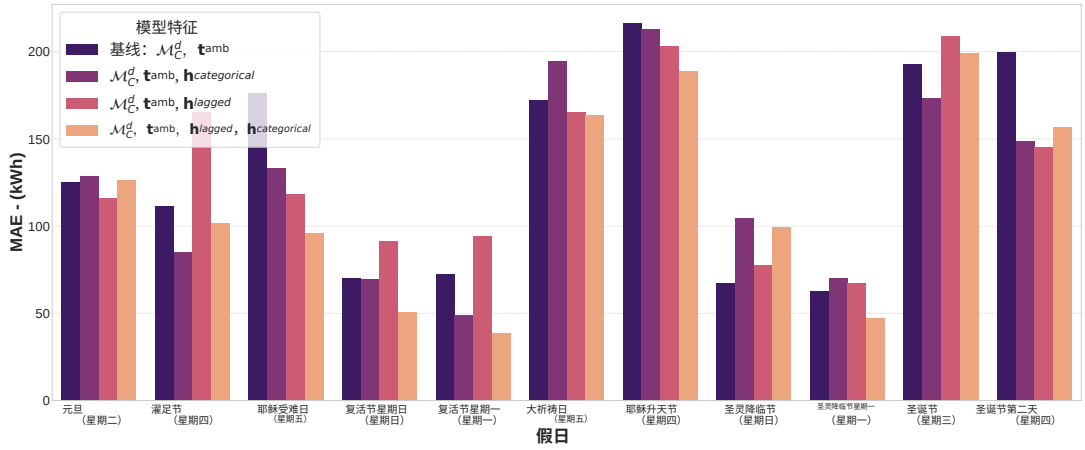
DMA	特征	均方误差 (千瓦时$^2 \times 10^4$)		平均绝对百分比误差 (%)		平均绝对误差 (千瓦时)	
		原始	分解后的	原始	分解后的	原始	分解后的
DMA A	$\mathcal{M}_C, \mathbf{t}_{\max}$	3.58	2.64	6.41	5.47	127.0	106.1
	$\mathcal{M}_C, \mathbf{t}_{感受}$	3.71	2.68	6.58	5.53	129.1	105.7
	$\mathcal{M}_C, \mathbf{t}_{\text{amb}}$	3.21	2.39	6.45	5.50	122.7	102.8
	$\mathcal{M}_C, \mathbf{t}_{\min}$	3.65	2.64	6.72	5.67	130.4	107.1
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{感受}$	3.33	2.69	6.22	5.68	123.4	108.6
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\text{amb}}$	3.17	2.52	6.21	5.55	120.1	105.1
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\min}$	3.23	2.55	6.28	5.48	121.1	104.9
	$\mathcal{M}_C, \mathbf{t}_{感受}, \mathbf{t}_{\text{amb}}$	3.03	2.56	6.18	5.66	117.6	106.3
	$\mathcal{M}_C, \mathbf{t}_{感受}, \mathbf{t}_{\min}$	3.60	2.98	6.46	6.04	126.2	113.9
	$\mathcal{M}_C, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\min}$	3.17	2.69	6.32	5.76	121.1	108.6
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{感受}, \mathbf{t}_{\text{amb}}$	3.10	2.47	6.24	5.49	118.9	104.7
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{感受}, \mathbf{t}_{\min}$	3.19	2.50	6.47	5.41	121.4	103.8
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\min}$	3.23	2.43	6.13	5.37	119.3	103.0
	$\mathcal{M}_C, \mathbf{t}_{感受}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\min}$	2.97	**2.45**	6.12	**5.37**	116.0	**101.7**
DMA B	$\mathcal{M}_C, \mathbf{t}_{\max}$	8.71	6.83	5.64	4.86	212.6	184.2
	$\mathcal{M}_C, \mathbf{t}_{感受}$	9.10	6.83	5.83	4.82	216.8	184.0
	$\mathcal{M}_C, \mathbf{t}_{\text{amb}}$	7.22	5.39	5.21	4.59	195.2	166.7
	$\mathcal{M}_C, \mathbf{t}_{\min}$	9.13	6.58	6.11	4.90	221.6	182.4
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{感受}$	7.95	6.03	5.38	4.79	201.8	175.7
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\text{amb}}$	7.25	5.81	5.45	4.80	196.2	173.4
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\min}$	7.72	6.35	5.37	4.84	199.9	180.1
	$\mathcal{M}_C, \mathbf{t}_{感受}, \mathbf{t}_{\text{amb}}$	7.59	5.74	5.47	4.71	200.3	172.7
	$\mathcal{M}_C, \mathbf{t}_{感受}, \mathbf{t}_{\min}$	8.17	6.87	5.45	5.13	204.5	188.0
	$\mathcal{M}_C, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\min}$	7.34	5.81	5.25	4.77	195.5	175.1
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{感受}, \mathbf{t}_{\text{amb}}$	7.74	**5.35**	5.37	**4.46**	199.5	**165.5**
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{感受}, \mathbf{t}_{\min}$	7.81	5.77	5.91	4.58	206.4	171.2
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\min}$	7.82	5.90	5.52	4.66	201.5	174.1
	$\mathcal{M}_C, \mathbf{t}_{感受}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\min}$	7.83	5.63	5.50	4.62	201.6	169.4
DMA C	$\mathcal{M}_C, \mathbf{t}_{\max}$	6.89	5.27	5.74	4.99	192.1	165.1
	$\mathcal{M}_C, \mathbf{t}_{感受}$	7.24	5.17	5.63	4.74	192.6	159.0
	$\mathcal{M}_C, \mathbf{t}_{\text{amb}}$	5.80	4.86	5.33	4.75	176.7	158.4
	$\mathcal{M}_C, \mathbf{t}_{\min}$	6.80	5.64	5.64	5.09	189.7	169.3
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{感受}$	6.71	4.75	5.53	4.78	187.8	156.4
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\text{amb}}$	6.13	5.17	5.35	5.11	180.5	166.8
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\min}$	6.58	5.04	5.64	4.86	188.6	161.9
	$\mathcal{M}_C, \mathbf{t}_{感受}, \mathbf{t}_{\text{amb}}$	5.61	5.00	5.24	4.93	173.4	161.9
	$\mathcal{M}_C, \mathbf{t}_{感受}, \mathbf{t}_{\min}$	6.50	5.34	5.62	5.08	185.9	166.2
	$\mathcal{M}_C, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\min}$	5.89	5.25	5.26	4.85	177.1	163.8
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{感受}, \mathbf{t}_{\text{amb}}$	5.96	**4.43**	5.53	**4.65**	179.7	**152.6**
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{感受}, \mathbf{t}_{\min}$	6.30	4.72	5.43	4.72	182.3	155.9
	$\mathcal{M}_C, \mathbf{t}_{\max}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\min}$	6.05	5.41	5.63	5.28	182.3	170.6
	$\mathcal{M}_C, \mathbf{t}_{感受}, \mathbf{t}_{\text{amb}}, \mathbf{t}_{\min}$	5.90	4.60	5.45	4.67	178.9	154.1

表 9：采用和未采用周期性时间特征（$+T_{\text{cyc}}$）的模型性能比较。将时间特征添加到表现最佳的分解模型中：$[\mathcal{M}_C^d, \mathbf{t}_{\text{feels}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{t}_{\text{min}}^d]$ 用于 *DMA A*，以及 $[\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{feels}}^d, \mathbf{t}_{\text{amb}}^d]$ 用于 *DMA B* 和 *DMA C*。

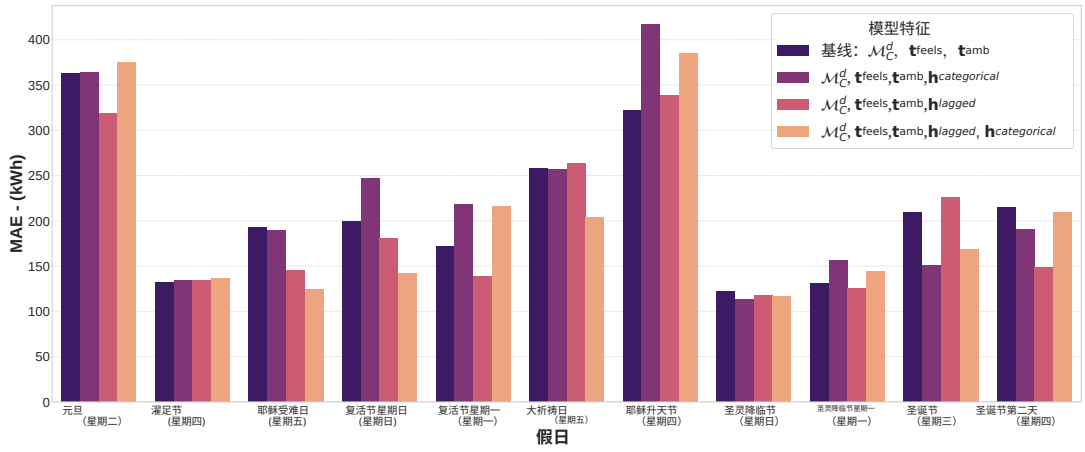
DMA	特征	均方误差 (千瓦时$^2 \times 10^4$)	平均绝对百分比误差 (%)	平均绝对误差 (千瓦时)	威尔科克森检验符号秩检验 p值
DMA A	$\mathcal{M}_C^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{t}_{\text{min}}^d$	2.45	5.37	101.7	> 0.05
	具有时间特征（$+T_{\text{cyc}}$）	2.44	5.50	103.0	
DMA B	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d$	5.35	4.46	165.5	< 0.05
	包含时间特征（$+T_{\text{cyc}}$）	5.93	4.93	176.3	
DMA C	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d$	4.43	4.65	152.6	0.035
	采用时间特征（$+T_{\text{cyc}}$）	5.06	4.90	161.0	

表10：不同节假日处理策略的性能比较。将模型表现最佳的特征与三种方法进行比较：将假日作为分类特征 $\mathbf{h}^{categorical}$ 添加、将特征 C_{168} 替换为上一次相同假日出现时的消耗值 $\mathbf{h}^{lagged}$，以及结合两种策略 $\mathbf{h}^{categorical}$ 和 $\mathbf{h}^{lagged}$。

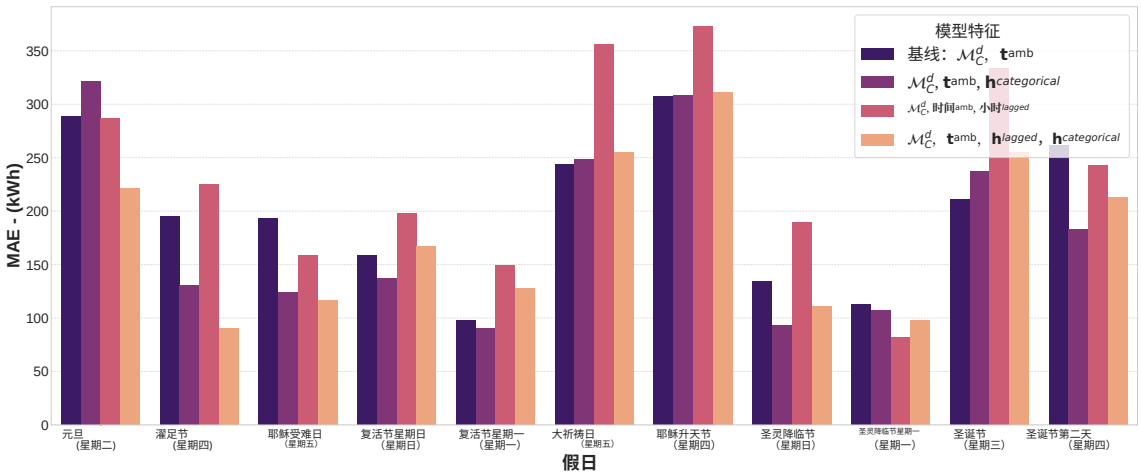
DMA	节假日处理方法	均方误差(×10^4 千瓦时^2)	平均绝对百分比误差 (%)	平均绝对误差 (千瓦时)
DMA A	$\mathcal{M}_C^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{t}_{\text{min}}^d$	2.45 ± 5.47	5.4 ± 3.3	101.73 ± 71.75
	$\mathcal{M}_C^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{t}_{\text{min}}^d, \mathbf{h}_{\text{分类的}}$	2.46 ± 5.58	5.5 ± 3.4	103.42 ± 71.32
	$\mathcal{M}_C^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{t}_{\text{min}}^d, \mathbf{h}_{\text{滞后}}$	2.51 ± 5.54	5.6 ± 3.3	104.84 ± 71.03
	$\mathcal{M}_C^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{t}_{\text{min}}^d, \mathbf{h}_{\text{滞后}}, \mathbf{h}_{\text{分类的}}$	**2.36 ± 5.42**	**5.3 ± 3.3**	**99。76 ± 72。71**
DMA B	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d$	5.35 ± 5.99	4.5 ± 2.3	165.48 ± 88.42
	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{h}_{\text{分类的}}$	5.63 ± 6.22	4.8 ± 3.0	171.97 ± 90.57
	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{h}_{\text{滞后}}$	**5.31 ± 6.00**	**4.4 ± 2.2**	**164。81 ± 88。36**
	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{h}_{\text{滞后}}, \mathbf{h}_{\text{分类的}}$	5.45 ± 6.22	4.7 ± 2.7	169.12 ± 87.10
DMA C	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d$	**4.43 ± 4.81**	**4.6 ± 2.3**	**152.60 ± 74.10**
	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{h}_{\text{分类的}}$	4.40 ± 4.75	4.8 ± 2.6	153.08 ± 72.30
	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{h}_{\text{滞后}}$	7.82 ± 7.86	6.3 ± 3.0	207.81 ± 101.98
	$\mathcal{M}_C^d, \mathbf{t}_{\text{max}}^d, \mathbf{t}_{\text{感受}}^d, \mathbf{t}_{\text{amb}}^d, \mathbf{h}_{\text{滞后}}, \mathbf{h}_{\text{分类的}}$	4.56 ± 5.21	4.8 ± 2.7	154.59 ± 75.79



(a) DMA A

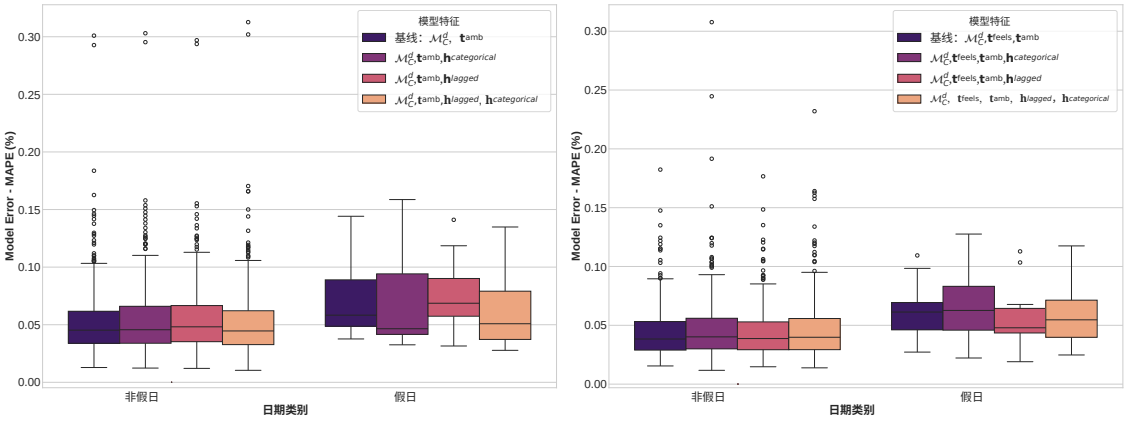


(b) DMA B



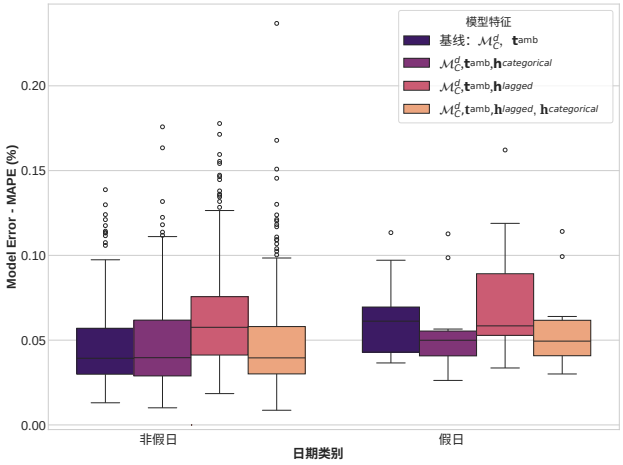
(c) DMA C

图9：比较丹麦公共假日中三个 DMA 的节假日特征不同嵌入表示的模型误差（平均绝对误差）：(a) DMA A、(b) DMA B 和 (c) DMA C。



(a) DMA A

(b) DMA B



(c) DMA C

图10：比较节假日样本与非节假日样本中节假日特征的不同嵌入表示，并与不含节假日特征的基线进行对照时，三个 DMA 的平均绝对百分比误差分布。(a)、(b) 和 (c) 中的箱线图分别显示了 DMA A、DMA B 和 DMA C 的模型误差。

基线方法。虽然这证实了模型能够积极地学习罕见事件，但也由于预测性能的随机性引发了疑问。这种随机性可能源于不同类型的节假日、事件发生的日期、发生季节等因素，需要进一步关注。

3.1.5. 与最先进方法的比较——实验 *V*

为严格评估我们所提出模型的性能，我们在全部三个 DMA 上开展了全面的基准测试。表 11 所列的定量结果明确表明，我们提出的方法性能优越。所提出的模型始终优于所有基线方法，在每个区域都取得了最低的误差指标。在 DMA A 中，所提出的框架实现了 99.76 千瓦时的平均绝对误差，与性能最佳的基线方法 TimeMixer（173.63 千瓦时）相比，误差大幅降低了约 42.5%。在其他区域也观察到类似的优势：相较于最强竞争方法，DMA B 的误差降低约 36%（164.81 对比 257.28 千瓦时），DMA C 的误差降低约 43%（152.60 对比 269.38 千瓦时）。

有趣的是，基准测试揭示了时间序列预测领域格局的变化。尽管 PatchTST 和 TimesNet 等基于变换器的模型表现良好，但 DLinear、TSMixer 和 TimeMixer 等更简单的基于多层感知机的架构始终是表现最强的基线方法，超越了季节性自回归积分滑动平均、XGBoost 等经典方法，也优于 Informer、Autoformer 等复杂的注意力机制。这进一步印证了近期文献中的发现，即质疑在某些时间序列任务中使用 Transformer 的必要性。然而，即便是这些先进的 Mixer 架构，在精度上仍难以匹敌我们提出的小波卷积神经网络框架。

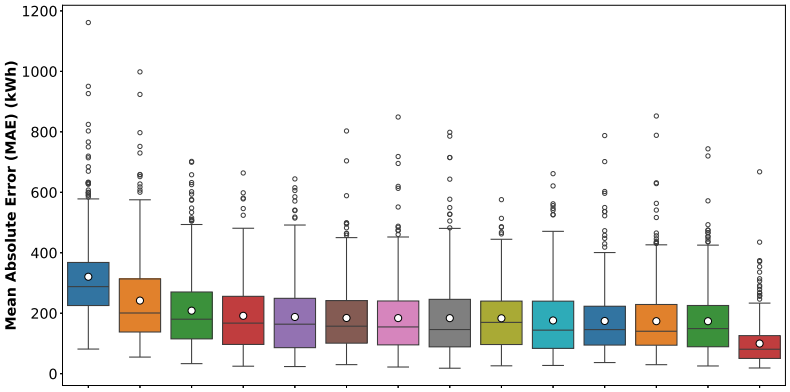
图 11 中的箱线图进一步证实了这些发现。我们的模型（最右侧）不仅平均误差（白点）最低，而且四分位距最小，对异常值的处理也更好。相比之下，Preformer 和 Autoformer 等模型表现出极大的方差，因此在电网运行中并不可靠，因为此类场景下最差情况下的性能至关重要。基础模型（Chronos-2 和微调后的 TTM）虽展现出一定潜力，但目前在这一特定高频领域仍落后于专门设计的架构。

表 11：针对 24 小时预测时域，对比现有统计方法、最先进方法、基础模型与我们提出的方法在小时热需求预测上的定量结果。MAE 的单位为千瓦时，MAPE 的单位为 %，MSE 的单位为 10^5千瓦时^2。

架构	DMA A			DMA B			DMA C		
	平均绝对误差	平均绝对百分比误差	均方误差	平均绝对误差	平均绝对百分比误差	均方误差	平均绝对误差	平均绝对百分比误差	均方误差
Preformer	320.68	21.37	1.69	627.20	21.12	6.37	552.04	19.34	4.99
Autoformer	241.86	11.52	1.03	385.94	12.51	2.62	387.36	11.97	2.61
Informer	208.51	9.42	0.79	306.92	9.30	1.75	312.38	9.23	1.77
TTM 微调版	191.47	10.08	0.73	320.25	8.28	1.98	301.66	8.49	1.71
LSTM	187.62	9.56	0.71	327.49	8.70	2.06	295.44	8.32	1.67
PatchTST	184.27	8.44	0.65	264.71	8.04	1.39	285.42	8.34	1.57
TimesNet	183.89	8.09	0.67	269.62	8.08	1.48	279.60	7.91	1.55
Chronos-2	183.71	9.38	0.73	353.80	8.89	2.60	277.50	7.81	1.57
XGBoost	183.25	9.81	0.67	318.31	8.58	1.85	286.56	8.37	1.52
SARIMAX	176.08	8.96	0.61	285.97	7.20	1.69	253.16	7.13	1.28
TSMixer	174.32	7.91	0.60	258.00	7.83	1.31	269.89	7.89	1.42
DLinear	173.71	7.93	0.60	257.88	7.90	1.32	270.02	7.91	1.45
TimeMixer	173.63	7.81	0.60	257.28	7.80	1.31	269.38	7.80	1.44
提出的	**99.76**	**5.26**	**0.24**	**164.81**	**4.44**	**0.53**	**152.60**	**4.65**	**0.44**

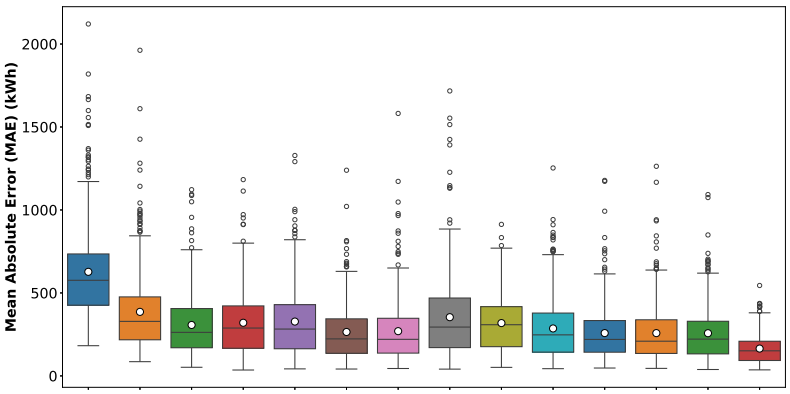
尽管这些汇总指标提供了清晰的性能排名，我们仍通过将代表性周的预测与真实值进行可视化，开展进一步的定性分析。为确定这些周，我们选取了代表集体误差谱低端与高端的情景。具体做法是对每个模型的周平均绝对误差进行归一化，并汇总这些分数。较低的总和表示总体上较易预测的周，而较高的总和则对应挑战所有架构的波动周。

如图12所示，高误差周的预测图表明，Preformer与Autoformer模型产生了过度平滑的预测，未能捕捉到至关重要的尖锐每日峰谷，而这些对于



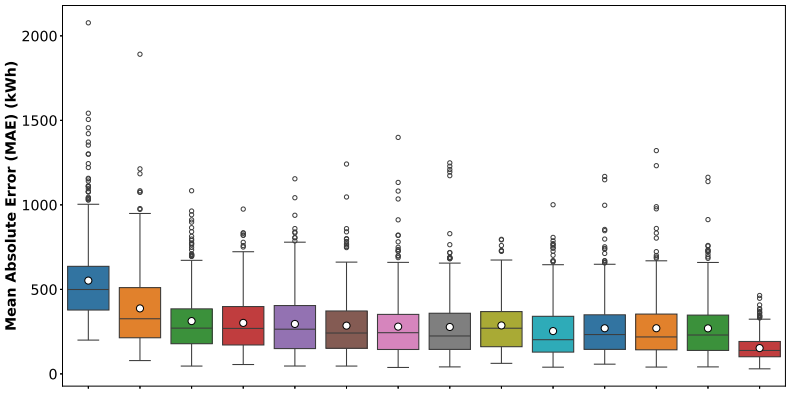
架构

(a) DMA A



架构

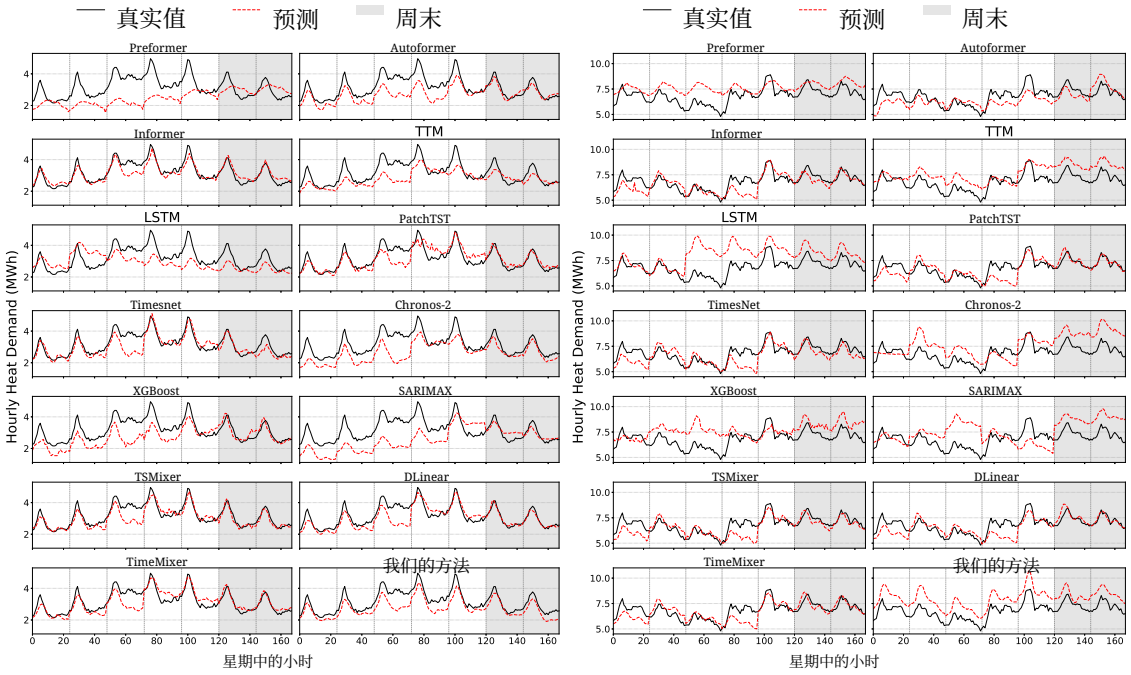
(b) DMA B



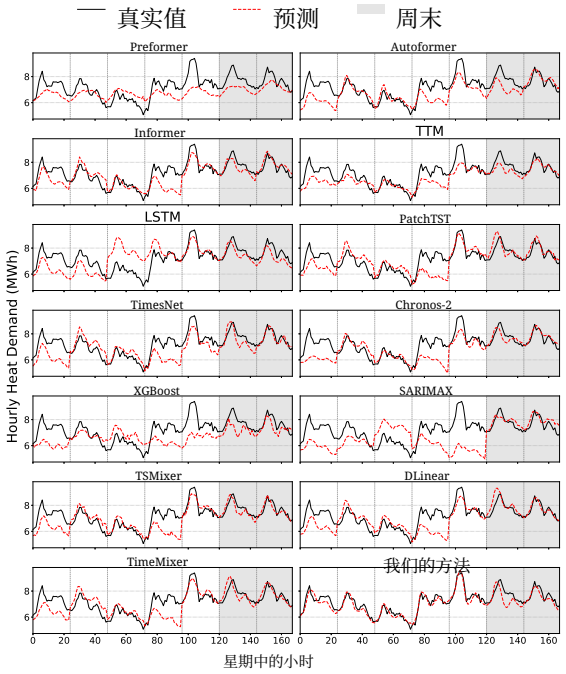
架构

(c) DMA C

图 11：分别对（a）DMA A、（b）DMA B 和（c）DMA C 的不同模型的平均绝对误差进行定量比较。与深度学习和统计基线方法相比，所提出的方法（最右侧）表现出更低的中位误差和更小的方差。

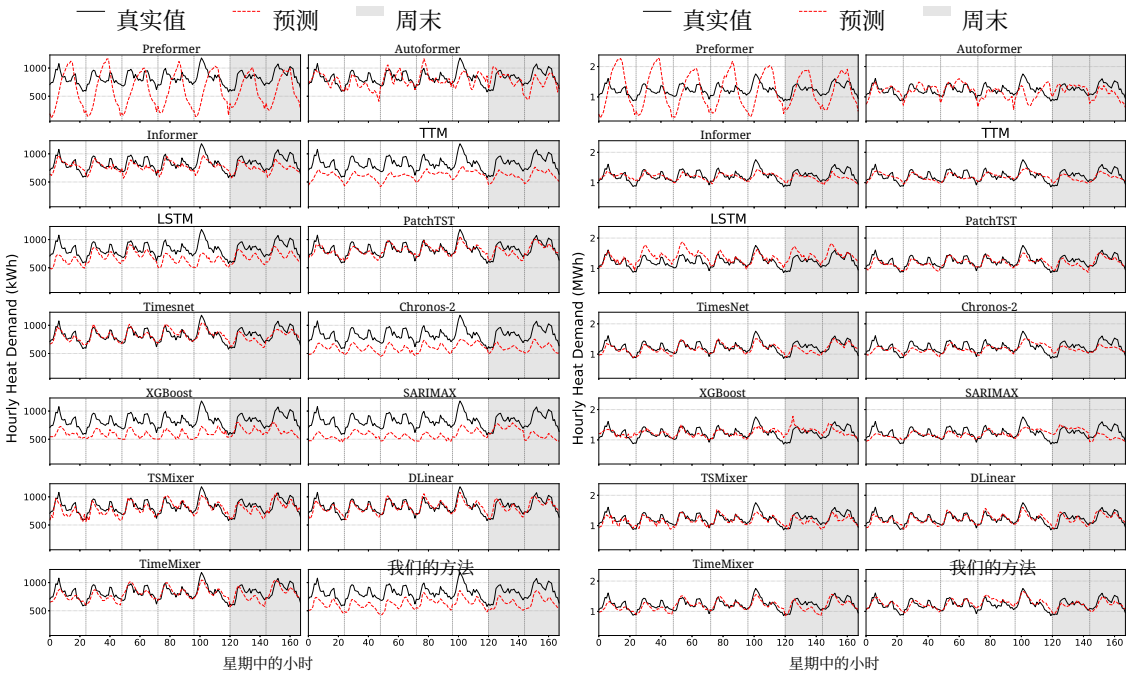


(a) DMA A (b) DMA B



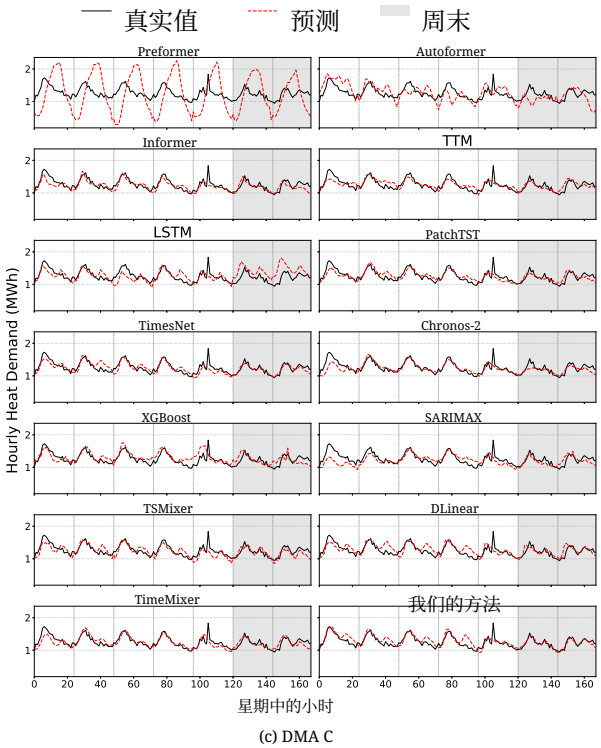
(c) DMA C

图12：三个区域中误差最高一周的定性预测比较：(a) DMA A（2019-04-08 至 2019-04-14）、(b) DMA B（2019-01-14 至 2019-01-20）和 (c) DMA C（2019-01-14 至 2019-01-20）。这些图展示了模型在需求波动剧烈时期的性能，凸显了我们的方法相较于基线方法的稳健性。



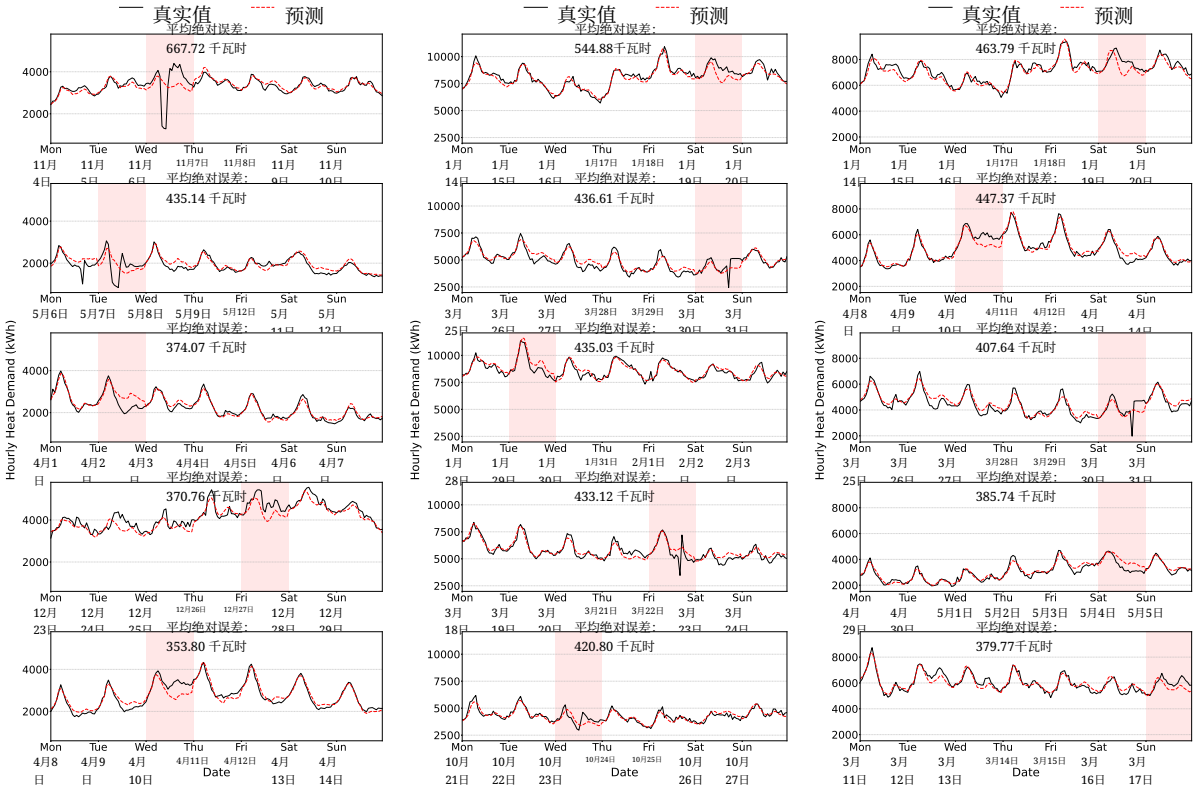
(a) DMA A

(b) DMA B



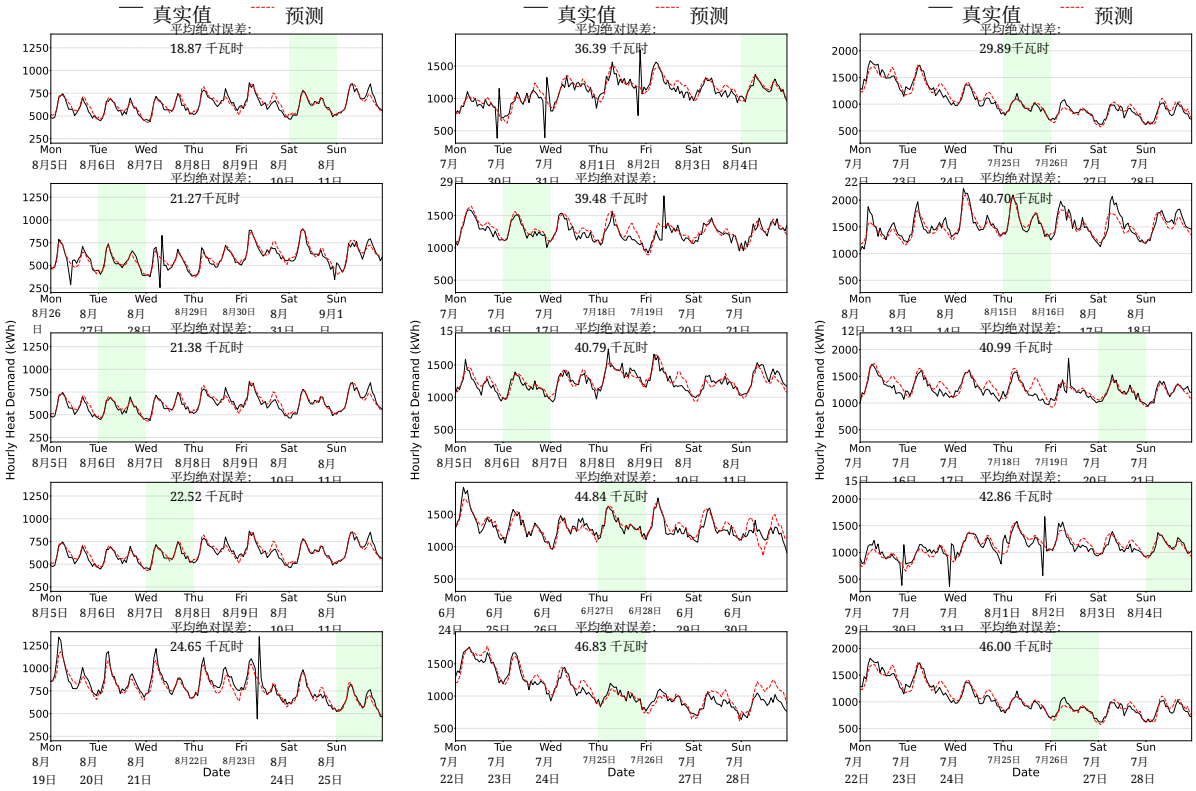
(c) DMA C

图13：三个区域中误差最低周的定性预测比较：(a) DMA A（2019-08-05至2019-08-11）、(b) DMA B（2019-08-05至2019-08-11）和 (c) DMA C（2019-07-15至2019-07-21）。这些图展示了需求稳定且可预测情况下的表现，此时所有模型的表现都相当不错。



(a) DMA A (b) DMA B (c) DMA C

图 14：所有三个区域中最差预测排名前五的日期的定性预测比较：(a) DMA A、(b) DMA B 和 (c) DMA C。



(a) DMA A (b) DMA B (c) DMA C

图15：全部三个区域中最佳预测排名前五的日期的定性预测比较：(a) DMA A、(b) DMA B和(c) DMA C。

运营规划。其他方法，如SARIMAX和XGBoost，虽然能够捕捉一定的波动性，但往往存在显著的相位滞后，或预测与实际需求之间偏差较大，对非高峰需求预测过高、对峰值预测不足。现代基于多层感知机的架构（DLinear、TimeMixer）以及所提出的方法能够捕捉相位对齐和波动性。与所展示的模型相比，所提出的方法展现出捕捉行为变化的能力，如12a和12b所示，这些变化偏离了滞后特征的趋势。这种稳健性归因于趋势分量的时频与解耦表示，使模型能够基于外生驱动因素预判爬坡事件，而残差分量则能够重建高频波动性。

在图13中，所展示的是误差较低的一周，其特点是需求模式更为规律且可预测，模型之间的性能差距自然缩小。然而，即便在这些有利条件下，细微差异依然存在。基线模型仍倾向于产生平滑伪影或轻微滞后期。而我们的方法则与真实值保持最紧密的拟合，成功区分了随机噪声（在夏季生活热水负荷曲线中较为常见）与结构信号。

为评估该框架的运行限制，我们考察图 14 和图15中所提出模型平均绝对误差最高和最低的具体日期。最坏情况（图 14）尤其能说明问题。最大的误差往往对应突发的运行异常，例如 DMA A 在11月6日，星期三观测到的需求骤降，曲线几乎垂直（可能是传感器故障或维护停机）。在这种情况下，模型预测的是预期的 正常需求曲线，而非异常。关键在于，这种行为体现了模型的稳健性，并不意味着实际控制回路出现故障；预测预期负荷可确保供热保障。在系统故障导致运行状况不稳定时，略微高估需求在运行层面比低估更安全，因为低估可能导致热量输送不足。相反，图15证实，在正常运行期间，模型能够高精度地捕捉典型的双峰，为电网自动优化提供可靠依据。

3.1.6. 时间与日历特征 - 实验 *VI*

3.1.7. 纵向稳定性与多数据集评估——实验 *VI*

为评估所提出框架的时间鲁棒性及其在丹麦和德国不同区域供热系统中的泛化能力，我们采用滚动块窗口策略进行了纵向评估。这项实验基于布伦讷斯莱夫的多区域数据集，以及弗伦斯堡和奥尔堡两座城市的两个外部公开数据集开展，旨在模拟真实部署场景：随着新的历史数据不断产生，模型会定期重新训练。

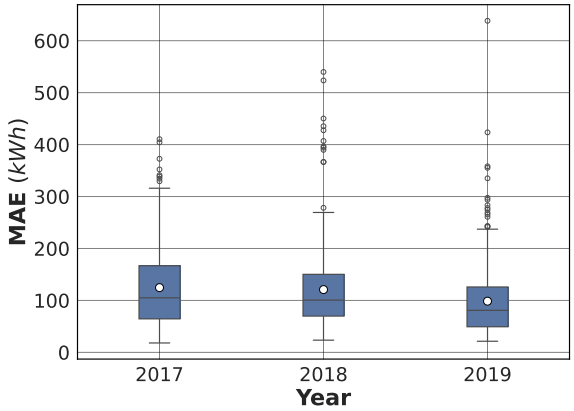
图16展示了三个私人区域（DMA A、B、摄氏度）的逐年性能。随着数据量增加，观测到的趋势始终稳定。例如，在 DMA A 中，2017 至 2019 年间的平均绝对误差中位数基本保持不变，最后一年四分位距略有缩小。这表明模型能够有效利用不断累积的历史数据来改进其内部表征。DMA B 和摄氏度也表现出类似的稳定性，各评估年份的误差分布都较为集中。

对奥尔堡住宅数据集的评估结果如图17a所示，证实小波卷积神经网络架构能够有效泛化至不同的聚合层级，即使面对住宅特有的负荷曲线，也能保持较高的精度。通过长时间跨度的弗伦斯堡数据集，进一步验证了该框架的稳健性；该数据集涵盖七年（2017—2024年）。如图17b所示，在这一较长时间范围内，预测准确率始终十分稳定，平均绝对误差中位数保持在约8兆瓦时，尽管年度天气条件存在显著差异，且十年间可能发生基础设施变化；不过，五年后预测性能略有下降。

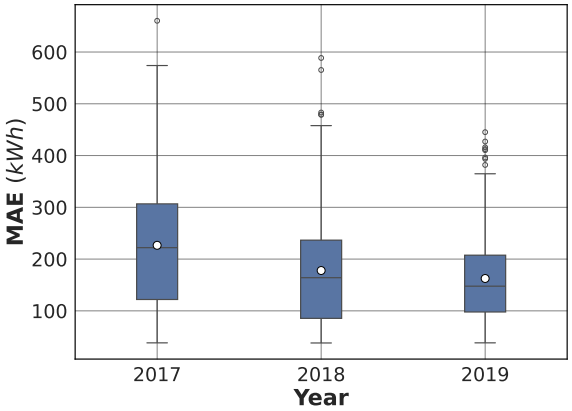
综合来看，这些结果表明，所提出的框架可应用于更广泛的场景，作为一种稳健、达到工程级标准的解决方案，能够在地理位置和结构各异的供热网络中，于多年运行周期内保持高性能。

4. 结论

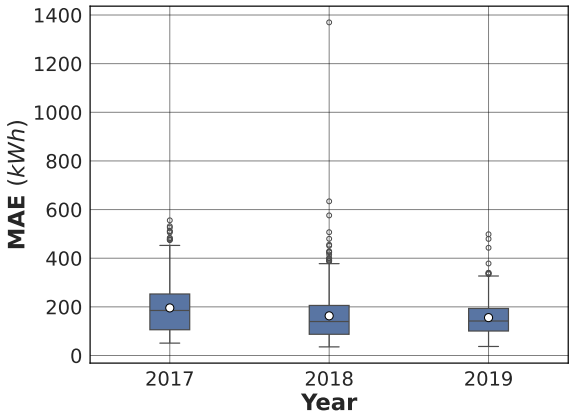
在本研究中，我们提出了一种新颖的深度学习框架，用于区域供热系统中日前多步小时热需求预测。我们的方法不同于传统时域模型，它使用连续小波变换，将多变量时间序列输入（包括历史消耗量、天气数据和日历特征）转换为二维时频表示。这些类似图像的尺度图



(a) DMA A

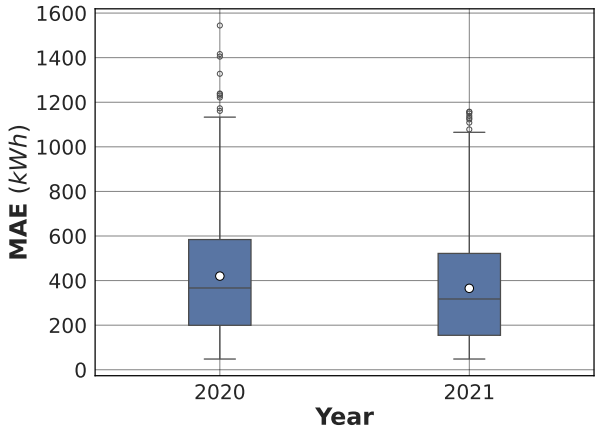


(b) DMA B

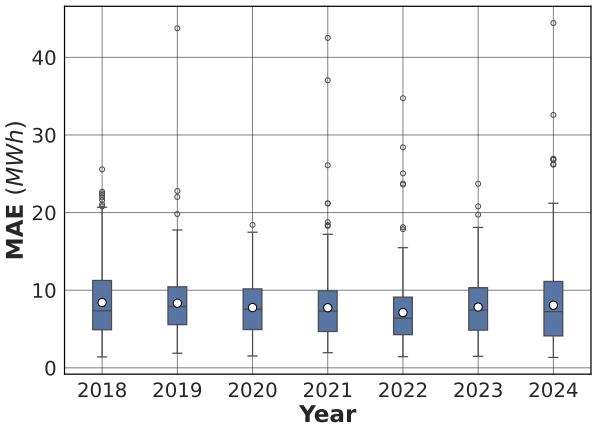


(c) DMA C

图16：三个 DMA（上排）和两个城市级数据集（下排）在不同年份的平均绝对误差（MAE）分布。



(a) 奥尔堡



(b) 弗伦斯堡

图17：三个 DMA（上排）和两个城市级数据集（下排）在不同年份的平均绝对误差（MAE）分布。

随后由基于卷积的模型进行处理，该模型擅长提取有意义的模式，同时应对时间序列数据中的非平稳性。我们在三个不同的真实分区计量区域以及两个公开基准数据集（弗伦斯堡和奥尔堡）上进行了全面评估，结果表明该框架有效且稳健。定量结果显示，我们的模型始终优于多种基准方法，包括经典统计模型、成熟的机器学习方法、先进的深度学习架构和基础模型。值得注意的是，与最强基线 TimeMixer 相比，所提出的框架将平均绝对误差降低了约 36% 至 43%。此外，采用滚动窗口策略进行的七年纵向分析进一步证实了模型的时间稳定性及其保留历史信息的能力。

从应用角度来看，该框架具备出色的峰值跟踪保真度和较低的误差方差，这对能源系统的智能控制具有直接意义。通过最大限度地降低预测不确定性，运营商可以在不冒供热不足风险的前提下，安全地缩小供水温度的安全裕度。这一能力通过最大限度地减少输配损耗并促进可变可再生热源的深度整合，直接助力供热行业脱碳。此外，我们在基准测试中观察到的轻量级架构的竞争力表明，实现高精度并非必须依赖庞大的计算规模；更小、更具资源效率的模型为在资源受限的工业环境中实现可扩展部署提供了一条有前景且切实可行的路径。

除新颖架构之外，本文的一项关键贡献是系统阐明了不同输入特征的作用。通过一系列广泛的消融研究，我们量化了自回归、气象和日历数据的影响。我们证明，尽管日消费滞后值提供了强有力的基线，但加入周滞后（$\mathbf{c}_{168}$）能够稳定周末等过渡时期的预测，同时也使模型能够动态应对罕见的节庆活动。我们的分析证实，环境温度是影响最大的单一外生变量；此外，我们还指出，将其他温度变量与环境温度结合，并进行分解，可为实现最优预测提供一条途径。

未来的工作将着重提升模型对罕见且非重复事件期间需求的预测能力，这是本研究发现的一项关键挑战。当前的节假日特征表示虽有益处，但对不同事件的效果各异。

本研究发现了若干值得进一步探索的方向。首先，由于罕见且非重复事件在历史训练样本中的表示稀疏，对其进行预测仍是一项挑战。未来的工作应研究专门设计的架构，例如记忆增强网络或小样本学习框架，明确用于建模标准回归模型往往会平滑掉的高影响异常。此外，在特征变换阶段，可探索自适应表示或多小波表示，研究针对不同特征动态选择小波基能否较本文采用的固定方法带来边际收益。其次，从数据角度来看，增强模型对情境的感知能力具有巨大潜力。整合更精细的外部特征，例如工业运营日程、学校日历和市政活动日志（如音乐会），有望弥合统计模式与现实因果关系之间的差距，使模型能够预判由特定文化或运营活动引起的需求变化。最后，尽管我们的纵向分析证明了模型的稳定性，但要充分刻画长期概念漂移，仍需获取更大规模的数据集。同时，还应对新兴基础模型开展更深入、更全面的评估，以确定如何最好地将其泛化能力与领域专用架构（如本文提出的架构）的精确性相协调。

术语表

CRediT 作者贡献声明

AdithyaRamachandran: 概念化、数据整理、形式分析、调查、方法、验证、可视化、撰写——初稿、撰写——审阅与编辑。**Satyaki Chatterjee:** 概念化、撰写——审阅与编辑。**Thorkil Flensmark B. Neergaard:** 数据整理。**Maximilian Oberndoerfer:** 资源、数据整理。**Andreas Maier:** 资源、监督、撰写——审阅与编辑。**Siming Bayer:**概念化、经费获取、项目管理、资源、监督、撰写——审阅与编辑。

缩略语

AIC	赤池信息准则	MAE	平均绝对误差
CNN	卷积神经网络	MAPE	平均绝对百分比误差
CWT	连续小波变换	ML	机器学习
DHS	区域供热系统	MLP	多层感知机
DL	深度学习	MLR	多元线性回归
DMA	分区计量区域	MSE	均方误差
DNN	深度神经网络	ReLU	修正线性单元
DWT	离散小波变换	季节性自回归积分滑动平均	季节性 ARIMA
FFNN	前馈神经网络	SARIMAX	带外生回归变量的 SARIMA
GPR	高斯过程回归	SVR	支持向量回归
KAN	科尔莫戈罗夫-阿诺德网络	TFT	时序融合变换器
LSTM	长短期记忆网络	TTM	微型时间混合器

致谢

作者感谢其研究合作伙伴 Diehl Metering GmbH（Maximillian Oberndorfer、Lena Yoshi- hara）和 Brønderslev Forsyning A/S 提供领域专业知识并给予持续支持。

资金来源

本研究由巴伐利亚州经济事务、媒体、能源和技术部资助，资助编号为 DIK0325/01。

利益冲突声明

我没有需要声明的事项。

数据声明

数据和代码库已在GitHub仓库中公开提供——https://github.com/lme-dpui/heat-demand- forecasting。

参考文献

[1] B. Rismanchi，区域能源网络（DEN）的当前全球现状与未来发展，《可再生能源与可持续能源评论》75（2017年）571–9。doi:https://doi.org/10.1016/j.rser.2016.11.025。[2] D. Stecher、M. Neumayer、A. Ramachandran、A. Hort、A. Maier、D. B¨ucker、J. Schmidt，创建带标签的区域供热数据集：从异常检测到故障检测，《能源》313（2024年）134016。doi:https://doi.org/10.1016/j.energy.2024.134016。[3] K. Keramidas、F. Fosse、A. Diaz Vazquez、P. Dowling、R. Garaffa、J. Despr´es等，《2021年全球能源与气候展望》：迈向气候中和，政策评估，国别报告，卢森堡（2021年）。doi:https://doi.org/10.2760/410610。[4] S. Chatterjee、S. Bayer、A. K. Maier，使用粒子群优化增强的支持向量回归模型预测户级热消耗，载于：NeurIPS 2021应对气候变化机器学习研讨会，2021年。URL https://www.climatechange.ai/papers/neurips2021/42 [5]

L.-S. Schneider、P. Krauss、N. Schiering、C. Syben、R. Schielein、A. Maier，计量学中的数据驱动建模——简要介绍、当前进展与未来展望，tm-Technisches Messen 91（9），（2024年）480–503。[6] L. Miu、D. Nazare、D. Diaconu，国家长期战略中的区域供热（2022年10月）。URL https://climatedialogue.eu/outputs/district-heating-national-long-term-strategies.html [7]

57

S. V. Pedersen，附件：通过在多载能能源系统和热力网络中实施热泵实现灵活性——最终报告，技术报告HPT-AN57-1，热泵中心（2024年）。doi:https://doi.org/10.23697/3tf9-r091。[8] A. Golla、J. Geis、T. Loy、P. Staudt、C. Weinhardt，区域供热网络的运行策略：数据驱动热负荷预测的应用，《能源信息学》3（1）（2020年）22。doi:https://doi.org/10.1186/s42162-020-00125-5。[9] M. Stienecker，预测热需求对多能源系统日前优化调度与实时控制的影响，《能源》297（2024年）131156。doi:https://doi.org/10.1016/j.energy.2024.131156。[10] D. Guericke、A. Schledorn、H. Madsen，面向电力市场参与的热生产优化，载于：R. Garay-Martinez、A. Garrido-Marijuan（编），《低温区域供热手册》，Springer International Publishing，Cham，2022年，第179–193页。[11] S. Werner，区域供热与供冷，载于：地球系统与环境科学参考模块，Elsevier，2013年。doi:https://doi.org/10.1016/B978-0-12-409548-9.01094-0。[12] H. Anand、R. Nateghi、N. Alemazkoor，自下而上预测：利用智能电表数据进行负荷预测的应用与局限，《以数据为中心的工程》4（2023年）e14。doi:https://doi.org/10.1017/dce.2023.10。[13] E. Dotzauer，区域供热系统负荷预测的简单模型，《应用能源》73（3）（2002年）277–284。doi:https://doi.org/10.1016/S0306-2619(02)00078-8。[14] S. Grosswindhager、A. Roither-Voigt、M. Kozek，区域供热网络系统热负荷的在线短期预测，第31届国际预测研讨会论文集，布拉格，捷克共和国（2011年1月）。[15] T. Fang、R. Lahdelma，多元线性回归模型与SARIMA模型在区域供热系统热需求预测中的评估，《应用能源》179（2016年）544–552。doi:https://doi.org/10.1016/j.apenergy.2016.06.133。[16] S. Idowu、S. Saguna、C. Åhlund、O. Schel´en，应用机器学习：区域供热系统中的热负荷预测，《能源与建筑》133（2016年）478–488。doi:https://doi.org/10.1016/j.enbuild.2016.09.068。[17] D. Petkovi´c、M. Proti´c、S. Shamshirband、S. Akib、M. Raos、D. Markovi´c，区域供热系统中热负荷最具影响参数的评估，《能源与建筑》104（2015年）264–274。doi:https://doi.org/10.1016/j.enbuild.2015.06.074。[18] M. Proti´c、S. Shamshirband、D. Petkovi´c、A. Abbasi、M. L. Mat Kiah、J. A. Unar、L. ˇZivkovi´c、M. Raos，使用结合离散小波变换算法的支持向量机预测区域供热系统中用户热负荷，《能源》87（2015年）343–351。doi:https://doi.org/10.1016/j.energy.2015.04.109。[19] P. Potoˇcnik、P. ˇSkerl、E. Govekar，基于机器学习的区域供热系统多步热需求预测，《能源与建筑》233（2021年）110673。doi:https://doi.org/10.1016/j.enbuild.2020.110673。[20] G. Suryanarayana、J. Lago、D. Geysen、P. Aleksiejuk、C. Johansson，利用深度学习与先进特征选择方法进行区域供热网络热负荷预测，《能源》157（2018年）141–9。doi:https://doi.org/10.1016/j.energy.2018.05.111。[21] F. Yao、W. Zhou、M. A. Ghamdi、Y. Song、W. Zhao，一种用于区域供热系统短期热需求预测的集成D-CNN-LSTM方法，《能源报告》8（2022年）98–107。doi:https://doi.org/10.1016/j.egyr.2022.08.087。[22] P. Xue、Y. Jiang、Z. Zhou、X. Chen、X. Fang、J. Liu，使用机器学习算法对区域供热系统热负荷进行多步超前预测，《能源》188（2019年）116085。doi:https://doi.org/10.1016/j.energy.2019.116085。[23] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、L. Kaiser、I. Polosukhin，注意力机制就是你所需要的一切（2023年）。arXiv:1706.03762。URL https://arxiv.org/abs/1706.03762 [24]

H. Zhou、S. Zhang、J. Peng、S. Zhang、J. Li、H. Xiong、W. Zhang，Informer：超越高效Transformer的长序列时间序列预测，载于：第35届美国人工智能协会人工智能会议，美国人工智能协会2021，线上会议，第35卷，美国人工智能协会出版社，2021年，第11106–11115页。[25] H. Wu、J. Xu、J. Wang、M. Long，Autoformer：用于长期序列预测的带自相关的分解Transformer模型，载于：神经信息处理系统进展，2021年。[26] H. Wu、T. Hu、Y. Liu、H. Zhou、J. Wang、M. Long，TimesNet：面向一般时间序列分析的时间二维变化建模，载于：国际学习表征会议，2023年。[27] M. Gong、Y. Zhao、J. Sun、C. Han、G. Sun、B. Yan，基于Informer的区域供热系统负荷预测，《能源》253（2022年）124179。doi:https://doi.org/10.1016/j.energy.2022.124179。[28] L. Frison、S. G¨olzh¨auser、M. Bitterling、W. Kramer，评估不同人工神经网络预测方法以优化区域供热网络运行，《能源》307（2024年）132745。doi:https://doi.org/10.1016/j.energy.2024.132745。[29] T. Jiang、W. Liu、J. Lin、X. Jin、Z. Liu，基于异构模型融合的区域供热系统逐小时热负荷预测模型研究，《热力工程案例研究》73（2025年）106500。doi:https://doi.org/10.1016/j.csite.2025.106500。URL https://www.sciencedirect.com/science/article/pii/S2214157X25007609

[30] T. Quanwei, X. Guijun, X. Wenju, Cakformer：基于自相关和科尔莫戈罗夫-阿诺德网络的长期热负荷预测Transformer模型，《能源》324 (2025) 135460. doi:https://doi.org/10.1016/j.energy.2025.135460. URL https://www.sciencedirect.com/science/article/pii/S0360544225011028

[31] Z. Chen, Y. Yang, C. Jiang, Y. Chen, H. Yu, C. Zhou, C. Li, 通过多尺度融合集成深度学习增强区域供热网络中的工业热负荷预测，《专家系统与应用》272 (2025) 126783. doi:https://doi.org/10.1016/j.eswa.2025.126783. URL https://www.sciencedirect.com/science/article/pii/S0957417425004051

[32] J. Yang, M. Peng, J. Zhu, X. Liang, H. Qi, C. Ben, 基于机理的数据融合建筑热负荷预测：面向区域供热系统的堆叠集成方法，《应用热工程》279 (2025) 127552. doi:https://doi.org/10.1016/j.applthermaleng.2025.127552. URL https://www.sciencedirect.com/science/article/pii/S1359431125021441

[33] Y. Zhang, X. Tian, Y. Zhao, C. Zhang, Y. Zhao, J. Lu, 一种基于先验知识的时间序列模型，用于区域供热系统热需求预测，《应用热工程》252 (2024) 123696. doi:https://doi.org/10.1016/j.applthermaleng.2024.123696. URL https://www.sciencedirect.com/science/article/pii/S1359431124013644

[34] M. Schaffer, T. Tvedebrink, A. Marszal-Pomianowska，安装于丹麦住宅建筑的智能热量表三年小时数据 [数据集], 版本 1.1.0（2022年5月）。doi:10.5281/zenodo.6563114. URL https://doi.org/10.5281/zenodo.6563114

[35] Y. Huang, Y. Zhao, Z. Wang, X. Liu, H. Liu, Y. Fu，基于主动深度学习的可解释区域供热负荷预测，《应用能源》350 (2023) 121753. doi:https://doi.org/10.1016/j.apenergy.2023.121753. URL https://www.sciencedirect.com/science/article/pii/S0306261923012170

[36] Y. Huang, P. Zhang, Z. Lu, Z. Ni，采用稀疏动态图学习的多尺度时间表征，用于区域供热负荷预测，《能源》333 (2025) 136838. doi:https://doi.org/10.1016/j.energy.2025.136838. URL https://www.sciencedirect.com/science/article/pii/S0360544225024806

[37] Y. Huang, Y. Zhao, Z. Wang, X. Liu, Y. Fu，稀疏动态图学习在区域供热负荷预测中的应用，《应用能源》371 (2024) 123685. doi:https://doi.org/10.1016/j.apenergy.2024.123685. URL https://www.sciencedirect.com/science/article/pii/S0306261924010587

[38] D. Ge, B. Liu, Y. Cheng, Z. Zhu，重新思考时间序列预测中的分解：从时间模式中学习，《专家系统与应用》287 (2025) 128044. doi:https://doi.org/10.1016/j.eswa.2025.128044. URL https://www.sciencedirect.com/science/article/pii/S0957417425016653

[39] J. Hu, Y. Hu, W. Chen, M. Jin, S. Pan, Q. Wen, Y. Liang，面向长期时间序列预测的吸引子记忆：混沌视角，arXiv abs/2402.11463 (2024)。URL https://api.semanticscholar.org/CorpusID:267750974

[40] Y. Zhao, Y. Shen, Y. Zhu, J. Yao，利用注意力神经网络预测小波变换时间序列，2018年IEEE国际数据挖掘会议（ICDM）（2018）1452–1457。URL https://api.semanticscholar.org/CorpusID:57363599

[41] S. Chatterjee, A. Ramachandran, T. F. Neergaard, A. K. Maier, S. Bayer，基于层构时间集成多分辨率表征的热需求预测，载于：NeurIPS 2022机器学习应对气候变化研讨会，2022年。URL https://www.climatechange.ai/papers/neurips2022/46

[42] A. Ramachandran, A. Maier, S. Bayer，通过注意力机制推进热需求预测：机遇与挑战，载于：NeurIPS 2024机器学习应对气候变化研讨会，2024年。URL https://www.climatechange.ai/papers/neurips2024/26

[43] Y. Liang, H. Wen, Y. Nie, Y. Jiang, M. Jin, D. Song, S. Pan, Q. Wen，时间序列分析基础模型：教程与综述，载于：第30届ACM SIGKDD知识发现与数据挖掘会议论文集，KDD '24，计算机协会，纽约，美国纽约州，2024年，第6555–6565页。doi:https://doi.org/10.1145/3637528.3671451.

[44] V. Ekambaram, A. Jati, P. Dayama, S. Mukherjee, N. H. Nguyen, W. M. Gifford, C. Reddy, J. Kalagnanam，微型时间混合器（TTMs）：用于增强多变量时间序列零样本/少样本预测的快速预训练模型，载于：神经信息处理系统进展（NeurIPS 2024），2024年。

[45] A. F. Ansari, O. Shchur, J. K˙uken, A. Auer, B. Han, P. Mercado, S. S. Rangapuram, H. Shen, L. Stella, X. Zhang, M. Goswami, S. Kapoor, D. C. Maddix, P. Guerron, T. Hu, J. Yin, N. Erickson, P. M. Desai, H. Wang, H. Rangwala, G. Karypis, Y. Wang, M. Bohlke-Schneider，Chronos-2：从单变量预测迈向通用预测（2025）。URL https://arxiv.org/abs/2510.15821

[46] S. F. GmbH，弗伦斯堡市2014年以来的区域供热网络数据2016 [数据集]（2019年1月）。doi:10.5281/zenodo.2562658. URL https://doi.org/10.5281/zenodo.2553967

[47] ´H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horanyi, 等，Mu˜noz Sabater，ERA5全球再分析，《英国皇家气象学会季刊》146, [访问日期：1月4日 2026]（2020年6月）。doi:10.1002/qj.3803.

[48] 丹麦气象研究所，基础知识 — 免费数据文档，[访问日期：1月4日 2026]（2026年）。URL https://www.dmi.dk/friedata/dokumentation/basics

[49] R. Osczevski, M. Bluestein，新的风寒等效温度图表，《美国气象学会通报》86 (10) (2005) 1453–1458. doi:10.1175/BAMS-86-10-1453. URL https://journals.ametsoc.org/view/journals/bams/86/10/bams-86-10-1453.xml

[50] Y. Zhao, Y. Shen, Y. Zhu, J. Yao，利用注意力神经网络预测小波变换时间序列，载于：2018年IEEE国际数据挖掘会议（ICDM），2018年，第1452–7页。doi:https://doi.org/10.1109/ICDM.2018.00201.

[51] L. P. A. Arts, E. L. van den Broek，用于实时、高质量、抗噪声时频分析的快速连续小波变换（FCWT），《自然·计算科学》2 (1) (2022) 47–58. doi:https://doi.org/10.1038/s43588-021-00183-z.

[52] S. Liu, H. Ji, M. C. Wang，适用于具有趋势的季节性时间序列的无池化卷积神经网络预测，《IEEE神经网络与学习系统汇刊》31 (8) (2020) 2879–2888. doi:10.1109/TNNLS.2019.2934110.

[53] F. R. Alharbi, D. Csala，基于带外生因素的季节性自回归综合移动平均（SARIMAX）预测模型的时间序列方法，Inventions 7 (4) (2022年)。doi:10.3390/inventions7040094。URL https://www.mdpi.com/2411-5134/7/4/94 [54] T. Chen, C. Guestrin, XGBoost：一种可扩展的树提升系统，载于：第22届ACM SIGKDD知识发现与数据挖掘国际会议论文集，KDD '16，计算机协会，纽约，NY，美国，2016年，页码785–794。doi:10.1145/2939672.2939785。URL https://doi.org/10.1145/2939672.2939785 [55] S. Hochreiter, J. Schmidhuber，长短期记忆网络，Neural Computation 9 (8) (1997) 1735–1780。doi:10.1162/neco.1997.9.8.1735。[56] D. Du, B. Su, Z. Wei，Preformer：一种利用多尺度分段相关性进行长期时间序列预测的预测型Transformer，载于：ICASSP 2023 - 2023 IEEE声学、语音与信号处理国际会议（ICASSP），2023年，页码1–5。doi:10.1109/ICASSP49357.2023.10096881。[57] Y. Nie, N. H. Nguyen, P. Sinthong, J. Kalagnanam，时间序列抵得上64个词：使用Transformer模型进行长期预测，载于：国际学习表征会议，2023年。[58] A. Zeng, M. Chen, L. Zhang, Q. Xu, Transformer模型对时间序列预测有效吗？，2023年。[59] S. Chen, C.-L. Li, N. Yoder, S. ¨O. Arik, T. Pfister，TSMixer：一种用于时间序列预测的全多层感知机架构，Trans. Mach. Learn. Res. 2023 (2023)。URL https://api.semanticscholar.org/CorpusID:257482532 [60] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, J. Zhou，TimeMixer：用于时间序列预测的可分解多尺度混合，载于：国际学习表征会议（ICLR），2024年。

附录

A. 架构设计选择

A.1. 池化层

为通过实证验证不采用池化层这一架构决策，我们开展了消融实验，比较模型在不同丢弃率下采用和不采用池化层时的性能。图 A.18 所示的结果表明，与不采用池化层的配置相比，引入池化层会降低预测准确率。在所有测试的丢弃概率下，不采用池化层的模型均取得了更低的平均值和平均绝对误差中位数，四分位范围也更低。

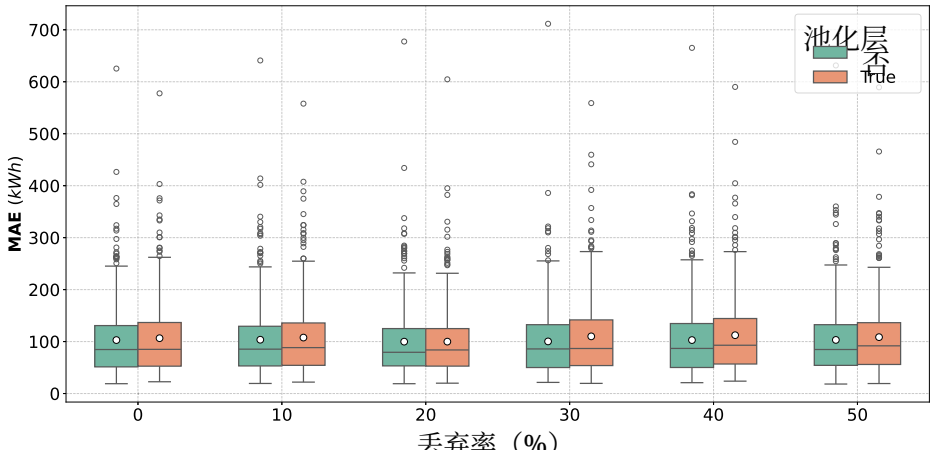


图 A.18：不同丢弃概率下采用丢弃率的影响，以及池化层对预测性能影响的定量比较。

在标准计算机视觉任务中，池化用于实现平移不变性，确保无论物体在图像中的位置如何，都能识别该物体。然而，在通过时频标度图进行时间序列预测时，横轴表示需求事件发生的确切时间或相位，因此特征的空间位置具有关键的语义意义。引入池化层会进行空间下采样，从而降低时间分辨率。相位相干性的丧失会削弱模型区分需求发生在时间 t 还是 $t+1$ 的能力。因此，不采用池化层可使所提出的架构保留平移等变性，并维持精确多步回归所需的高保真特征定位能力。

A.2. 母小波的选择

为了评估该框架对母小波选择 $\psi(t)$的敏感性，我们评估了多个不同小波族的性能，包括墨西哥帽小波、Morlet小波和高斯导数。如图 A.19 所示，所有候选方案的误差分布都非常稳定，平均绝对误差的变化通常低于 2%。此外，特征缩放会对信号振幅进行标准化，并保留同一小波族（例如，高斯族）母小波的相对结构模式，从而消除所捕获信息中的细微差异。

表 A.12 中提及的候选母小波作为超参数优化过程的一部分进行选择，并且所有特征均使用同一个固定小波，而非为各个特征指定不同的母小波。因此，我们选择性能最佳的母小波，包括“morl”和“gaus8”。

A.3. 超参数搜索空间

为使所提出的小波卷积神经网络架构适应热需求数据的具体特征，我们针对关键的架构和训练超参数进行了网格搜索。鉴于输入尺度图具有高维特性，搜索重点是在模型容量（深度和宽度）与正则化（丢弃率）之间取得平衡，以防止过拟合。

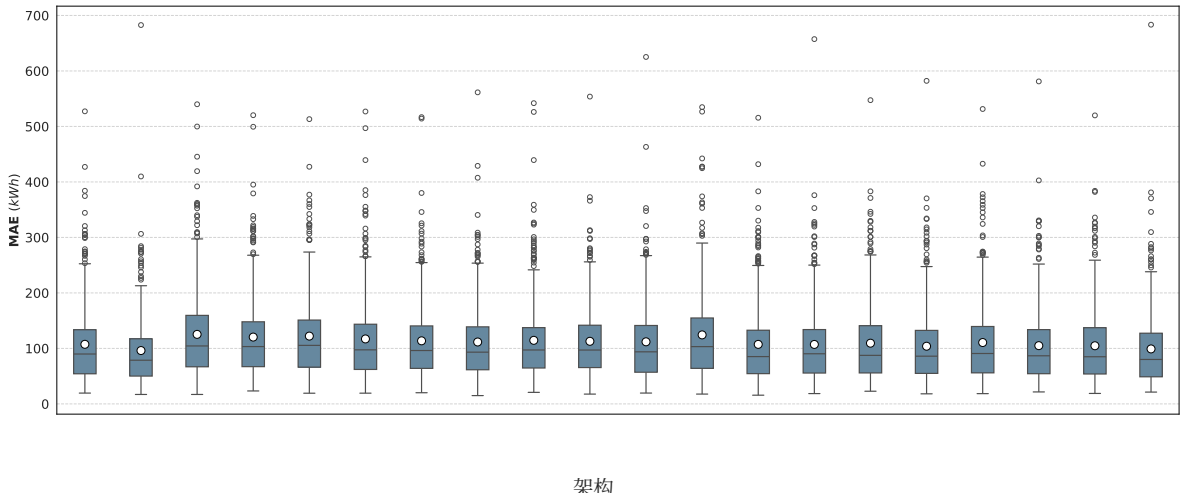


图 A.19：定量　　比较不同候选母小波对 DMA A 连续小波变换预测性能影响的定量评估，以平均绝对误差为指标用于 DMA A

搜索空间包括批量大小、卷积特征提取器的深度、是否采用池化操作、全连接层的宽度和深度，以及丢弃率的不同取值。表 A.12 详细列出了开发阶段探索的完整搜索网格。最终配置根据验证集上均方误差损失的最小值选定。

A.4. 模型收敛

所提出模型在全部三个区域上的学习曲线均显示出收敛性和训练稳定性。如图 A.20 所示，训练损失和验证损失均在前 20 个训练轮次内迅速下降，随后进入稳定的平台期。值得注意的是，在整个优化过程中，验证损失始终与训练损失密切跟随，没有出现发散迹象，表明模型未对训练集过拟合。

A.5. 训练与推理效率

为评估小波卷积神经网络框架的运行可行性，我们对三个 DMA 在 5 次迭代中的计算成本进行了评估，结果记录于表 A.13。

结果表明，该框架在不同数据集上均保持稳定的效率。由于收敛速度不同，各 DMA 的训练时间略有差异。不过，所有模型均在 3.5 分钟内收敛，从而支持快速重新训练。

对于部署而言，最关键的是所有区域的推理延迟始终很低，平均约为 **0.23 毫秒/样本**。这一亚毫秒级响应时间证实，所提出的框架非常适用于实时控制应用，能够同时处理多个区域供热行业中的高吞吐量预测任务。

所有实验均在一台配备 NVIDIA Quadro RTX 8000 GPU（48 GB 显存）的工作站上进行。

B. 超参数调优

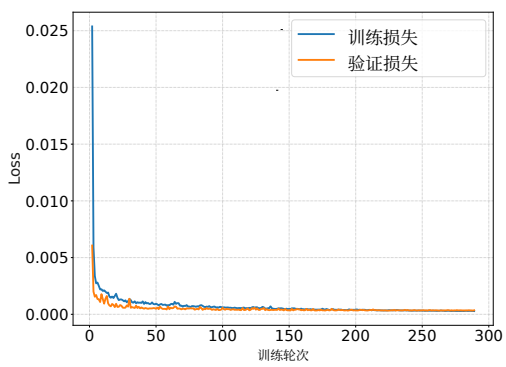
为确保比较评估公平且严谨，正文中讨论的所有基准模型均进行了超参数优化。我们对这些模型进行了调优，使其在特定的热需求预测任务上达到最佳性能。本节详细介绍了所探索的搜索空间、优化流程以及每个模型的最终配置。

表 A.12：所提出的小波卷积神经网络框架的超参数搜索空间及选定值。

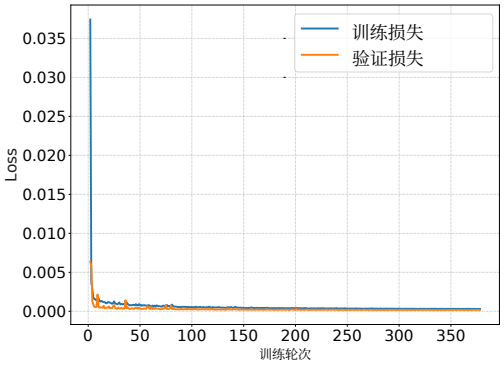
超参数	搜索空间 /网格	选定值
训练动态		
学习率 (η)	{0.01, 0.001, 0.0001}	0.001
批量大小	{16, 32, 64, 256}	32
特征变换		
母小波候选项	mexh, morl gaus1-gaus8, cgau1-cgau8 cmor0.1-1.0 ，shan1.5-1.0	Morlet小波（DMA A，摄氏度） gaus8（DMA B）
卷积块		
滤波器配置	{[16, 32, 64], [32, 64, 128]}	[32, 64, 128]
池化操作	{真、假}	假
线性头		
隐藏层维度（神经元）	[1024, 1024] [2048, 1024] [2048, 2048] [4096, 1024] [4096, 2048] [4096, 4096] [4096, 2048, 1024]	[1024, 1024]
正则化		
丢弃率（p）	{0.0、0.1、0.2、0.3、0.4、0.5}	0.1

表 A.13：每个 DMA 的计算效率指标。数值表示 5 次迭代的均值 ± 标准差。推理延迟以毫秒/样本报告。

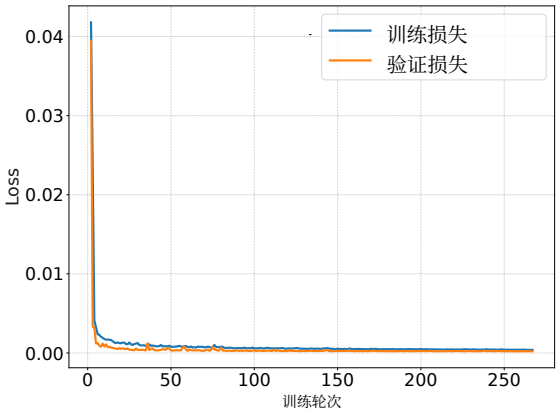
数据集	训练时间（分钟）	推理延迟（ms/样本）
DMA A	2.76 ± 0.26	0.23 ± 0.14
DMA B	3.10 ± 0.46	0.24 ± 0.17
DMA C	2.48 ± 0.11	0.22 ± 0.14
平均值	**2.78 ± 0.24**	**0.23 ± 0.15**



(a) DMA A



(b) DMA B



(c) DMA C

图 A.20：所提出的小波-CNN模型的训练与验证损失曲线示意图，模型最多训练1000个训练轮次，耐心值为50个训练轮次。(a)、(b)和(c)中的箱线图分别展示了DMA A、DMA B和DMA C的损失曲线。

B.1. 统计与机器学习基线

SARIMAX模型使用pmdarima库的自动 ARIMA函数实现，通过逐步搜索策略寻找最优参数。该函数通过最小化赤池信息准则（AIC）自动搜索最优模型阶数。搜索配置了以下约束：

- **季节周期（m）：**设为24，以捕捉小时数据中的日周期性。
- **非季节阶数：**自回归（p）、移动平均（q）和差分（d）分量的最大阶数分别设为6、6和1。
- **季节阶数：**季节性自回归（P）、季节性移动平均（Q）和季节性差分（D）分量的最大阶数分别设为 1、1 和 1。

通过对最关键的超参数进行网格搜索，对 XGBoost 模型进行了优化。该模型采用递归重拟合策略，在每个 24 小时预测窗口中使用最新的实际值进行重新训练。搜索空间和最终选定的超参数详见表 B.14。

B.1.1. 深度学习模型

我们将一个经典的长短期记忆网络作为神经网络基线。该架构由多个长短期记忆网络层堆叠而成，后接一个用于最终预测的全连接（线性）层，该层包含 24 个神经元。超参数搜索空间及最终选定的配置详见表 B.15。

表 B.14：XGBoost 模型的超参数调优。

超参数	搜索空间	选定值
估计器数量	{200, 400, 600, 800, 1000, 1500, 2000}	200
学习率	{0.1、0.01、0.001、0.0001}	0.1
最大深度	{2, 4, 6, 8, 10, 12, 24}	4
子采样	[0.6、0.7、0.8、0.9、1.0]	0.8

表 B.15：长短期记忆网络模型的超参数调优。

超参数	搜索空间	选定值
LSTM 层	{1, 2, 4}	4
LSTM 隐藏单元	{16, 32, 64, 128}	32
学习率	{1e-4, 5e-4, 1e-3}	1e-3

为确保比较公平且严谨，并保证计算资源相同，所有基于 Transformer 和多层感知机的基线方法均采用统一的实验方案进行基准测试。我们使用 Optuna 框架，对每种架构进行系统的超参数搜索。每个模型的搜索预算统一为 50 次试验，以避免无意中带来调参优势。优化目标是使验证集均方误差损失最小化。

所有基线模型均采用与所提出框架相同的计算约束进行训练：批量大小固定为 32，训练轮次最多为 1000，并将早停耐心值设为 50 个训练轮次，以防止过拟合。表 B.16 全面总结了这一过程，详细列出了各主要超参数所探索的具体搜索空间，以及为每个模型选定的最终最佳配置。对于输入序列长度和标签长度的组合，我们分析了 (24, 12)、(48, 24)、(96, 48)、(168, 96) 和 (336, 168) 这些取值，并为基于 Transformer 的模型选择了 (96, 48)。

表 B.16：深度学习基线方法的超参数搜索空间及最终配置。所有模型均使用 Optuna 进行调参。搜索空间反映了在 50 次试验的优化过程中遍历的网格。

超参数	搜索空间	Informer	Autoformer	PatchTST	DLinear	TimesNet	TimeMixer
		通用训练参数					
学习率	{0.01, 0.001, 0.0001}	0.0001	0.0001	0.0001	0.001	0.0001	0.001
丢弃率	[0.05, 0.3]	0.1	0.1	0.1	0.1	0.1	0.1
		通用架构参数					
d_{model}	{32, 64, 256, 512, 1024}	512	1024	512	–	1024	32
n_{heads}	{4, 6, 8, 10}	6	8	8	–	8	–
d_{ff}	{128, 256, 512, 1024, 2048, 4096}	2048	4096	2048	–	4096	256
e_{layers}	{1, 2, 4, 6, 8}	4	2	2	1	2	2
d_{layers}	{1, 2, 4}	2	1	–	–	1	–
		模型特定参数					
移动平均	{13, 25, 49}	–	25	–	25	–	25
补丁/段长度	{8, 16, 24}	–	–	16	–	–	16
因子	{7, 9}	9	–	–	–	–	–
前k	[3, 7, 10]	–	5	–	–	5	5
通道独立	{0, 1}	0	0	1	1	0	0
		TimeMixer					
下采样层	[1, 3]	–	–	–	–	–	2
下采样窗口	{2, 3, 5}	–	–	–	–	–	2

B.2. 基础模型

为评估预训练模型这一新兴范式，我们评估了针对该任务微调的两个基础模型。首先，我们使用 TTM-r2 变体 [44], , 该模型在多样化的能源时间序列语料库上进行了预训练。按照作者的建议，我们在数据集上对模型进行 1000 个训练轮次的微调，学习率为 1×10^{-5}，以调整其表征，同时避免灾难性遗忘。使用的上下文长度为 1024。同样，我们评估 Chronos-2 [45],

训练轮次为 1000，学习率为 1×10^{-5}。我们采用基础变体，并使用相同的方案对其进行微调，以确保与领域专用模型进行公平比较。

C. 节假日

表 C.17 列出了数据集中具有代表性的年份——2019年——丹麦观测到的公共假日的完整清单。该表包括具体日期、假日所在的星期、丹麦语官方名称，以及便于理解的英文翻译。这些信息可作为参考，帮助了解我们在不同节假日处理策略实验评估中分析的具体假日。

表 C.17：丹麦假日的英文翻译及星期（2019年）。

日期（2019年）	星期	丹麦语名称	英文翻译
2019年-01-01	星期二	元旦	元旦
2019年-04-18	星期四	濯足节	濯足节
2019年4月19日	星期五	耶稣受难日	耶稣受难日
2019年4月21日	星期日	复活节	复活节星期日
2019年4月22日	星期一	复活节星期一	复活节星期一
2019年5月17日	星期五	大祈祷日	大祈祷日
2019年5月30日	星期四	耶稣升天节	耶稣升天节
2019年6月9日	星期日	圣灵降临节	圣灵降临节主日（圣灵降临节）
2019-06-10	星期一	圣灵降临节后的星期一	圣灵降临节星期一
2019-12-25	星期三	圣诞节	圣诞节
2019-12-26	星期四	圣诞节第二日	圣诞节第二天（节礼日）