基于生成对抗神经网络的干涸后传热系数预测

袁双楠[1,3]　王 凡[4]　宋美琪[1,3]　王云伟[4]　许 巍[1,3]　刘晓晶[2,3]

1（上海交通大学 国家电投智慧能源创新学院　上海 200240）

2（上海交通大学 机械与动力工程学院　上海 200240）

3（上海市数值反应堆技术融合创新中心　上海 200240）

4（国核自仪系统工程有限公司　上海 200241）

摘要　干涸后（Post-Dryout，PDO）传热系数是核电系统安全评估中的关键参数，其建模受限于实验数据稀缺和分布不均问题。为解决该问题，本文提出一种基于条件表格生成对抗网络（Conditional Tabular Generative Adversarial Network，CTGAN）的两阶段数据增强方法。首先，采用CTGAN对与努塞尔数（Nu_m）相关的无量纲输入特征进行建模，生成大量符合原始分布的输入数据；其次，利用训练好的监督回归模型（Backpropagation Neural Network，BPNN）生成对应输出，从而实现输入-输出的解耦建模；最后，将增强数据与真实样本联合训练预测模型，并在多个外部数据集上验证其外推能力。实验结果表明，所提方法在多个独立数据集上均实现了较原始模型更低的预测误差，CTGAN在Becker与SWE数据集上平均绝对百分比误差（Mean Absolute Percentage Error，MAPE）降幅达30%以上，整体优于表格变分自编码器（Tabular Variational Autoencoder，TVAE）、表格生成对抗网络（Tabular Generative Adversarial Network，TGAN）与条件表格生成对抗网络增强模型（Conditional Tabular Generative Adversarial Network Plus，CTABGAN+）等生成方法。相比传统的单阶段数据生成策略，本方法在数据外推能力和预测稳定性方面表现更优，验证了其在非平稳分布场景下的适应性与鲁棒性。

关键词　干涸后传热，数据增强，条件表格生成对抗网络，监督回归模型，传热系数预测

中图分类号　TL333

DOI: 10.3724/j.0253-3219.2026.hjs.49.250259

CSTR: 32193.14.hjs.CN31-1342/TL.2026.49.250259

Prediction of post-dryout heat transfer coefficient based on generative adversarial neural networks

YUAN Shuangnan[1,3]　WANG Fan[4]　SONG Meiqi[1,3]　WANG Yunwei[4]　XU Wei[1,3]　LIU Xiaojing[2,3]

1(*College of Smart Energy*, *Shanghai Jiao Tong University*, *Shanghai 200240*, *China*)

2(*School of Mechanical Engineering*, *Shanghai Jiao Tong University*, *Shanghai 200240*, *China*)

3(*Shanghai Digital Nuclear Reactor Technology Integration Innovation Center*, *Shanghai 200240*, *China*)

4(*State Nuclear Power Automation System Engineering Company*, *Shanghai 200241*, *China*)

Abstract　**[Background]** The post-dryout (PDO) heat transfer coefficient plays a critical role in thermal safety

国家自然科学基金(No.12427811)资助

第一作者：袁双楠，男，1999年出生，2022年毕业于南京工程学院，现为硕士研究生，研究领域为机器学习在干涸后传热领域的应用

通信作者：宋美琪，E-mail：songmeiqi@sjtu.edu.cn

收稿日期：2025-06-25，修回日期：2025-12-17

Supported by National Natural Science Foundation of China (No.12427811)

First author: YUAN Shuangnan, male, born in 1999, graduated from Nanjing Institute of Technology in 2022, master student, focusing on the application of machine learning in post-dryout heat transfer

Corresponding author: SONG Meiqi, E-mail: songmeiqi@sjtu.edu.cn

Received date: 2025-06-25, revised date: 2025-12-17

assessment in nuclear systems. However, accurate modeling of PDO heat transfer remains challenging due to experimental data scarcity and pronounced distribution shifts across datasets from different sources. **[Purpose]** This study aims to develop a robust data-driven prediction framework that can improve the generalization and extrapolation performance of PDO heat transfer coefficient models under limited and distribution-shifted data conditions. **[Methods]** A two-stage data augmentation and prediction approach based on Conditional Tabular Generative Adversarial Network (CTGAN) was proposed. In the first stage, CTGAN was trained to learn the joint distribution of dimensionless input features that are highly correlated with the Nusselt number (Nu_m), and synthetic feature samples were generated accordingly. In the second stage, a supervised regression model based on a backpropagation neural network (BPNN) was employed to predict the corresponding target values, thereby decoupling input feature generation from input-output mapping. The generated synthetic data were then combined with real experimental samples to train the final predictive model. Finally, the proposed method was evaluated on four independent post-dryout heat transfer datasets and compared with other data generation techniques. **[Results]** The experimental results show that the proposed two-stage approach significantly reduces prediction errors compared with models trained using only real data. CTGAN achieves the lowest mean absolute percentage error (MAPE) on the Becker and SWE datasets and outperforms other generation models, including tabular variational autoencoder (TVAE), tabular generative adversarial network (TGAN), and Conditional Tabular Generative Adversarial Network Plus (CTABGAN+). Compared with traditional one-stage data generation schemes, the proposed method exhibits superior extrapolation capability and robustness under distribution shift conditions. **[Conclusions]** The proposed CTGAN-based two-stage data generation framework effectively enhances the predictive accuracy and generalization ability of PDO heat transfer coefficient models, hence providinga reliable and extensible solution for data-scarce and distribution-shifted thermal-hydraulic modeling problems in nuclear engineering applications.

Key words Post-dryout heat transfer, Data augmentation, Conditional Tabular Generative Adversarial Network (CTGAN), Backpropagation Neural Network (BPNN), Heat transfer coefficient prediction

在核电站运行过程中，当堆芯功率异常升高或冷却剂流量意外降低时，冷却剂将发生沸腾汽化现象。随着包壳表面液膜被撕裂或完全蒸发，燃料棒表面将无法维持连续液膜覆盖，此时即发生干涸现象[1]。干涸点处传热性能急剧恶化，并伴随壁温的骤然升高[2]。然而在干涸后（Post-Dryout，PDO）区域，壁面温度可能因工况不同而持续上升或缓慢下降。因此，准确确定干涸后区域的传热过程有助于确定燃料棒温度上限，这对核反应堆的安全运行至关重要。在过去的几十年里，对PDO传热预测的方法主要有：经验关系式、半理论模型法和查表法[3]，但这些模型在某些工况下存在显著差异。现有模型的适用性大多局限于其所采用的实验数据库参数范围，而由于干涸后实验的实施难度，此类数据通常较为匮乏[3]。

近年来，已有大量研究者展开了关于核工程关键参数预测与数值建模方法的研究工作。例如，张凡等[4]基于人工神经网络与超参数优化方法，实现了堆芯中子学参数的高精度预测，验证了数据驱动方法在核工程复杂非线性问题中的应用潜力；袁硕等[5]针对燃料棒流致振动与微动磨蚀问题，系统分析了结构与流动耦合作用机理，为燃料组件安全性评估提供了重要依据。此外，沈书秋等[6]通过降阶模型实现了热管堆核热力耦合启堆瞬态模拟，朱鹏橙等[7]则针对核热推进反应堆开展了多目标优化设计研究。上述研究表明，结合先进建模方法对核工程关键过程进行高精度预测，已成为当前核工程研究的重要发展方向。

然而，与中子学参数或稳态热工参数相比，干涸后传热过程涉及高汽含率两相流动、液膜破裂及界面传热等复杂物理机制，其实验实施难度大、数据获取成本高，使得现有PDO传热模型大多依赖有限实验数据库，其适用范围和预测一致性仍存在明显局限。传统PDO预测方法主要包括经验关系式、半理论模型以及查表法，但在不同压力、质量流量及干度条件下，模型间预测结果常呈现显著差异。

近年来，人工神经网络（Artificial Neural Network，ANN）等数据驱动方法因其对复杂非线性关系的建模能力，逐渐应用于PDO建模[8]。Yu和Xu[9]提出了融合物理模型的机器学习预测框架，在沸腾传热研究中取得良好效果。然而，深度模型对训练数据数量的敏感性，成为其在数据稀缺领域推广的瓶颈[10-11]。

样本生成技术是解决样本匮乏的有效方法，通

过样本生成不仅可以扩大样本数量，而且可以丰富样本内容，提高模型的泛化能力[12]。自2014年Ian等[13]首次提出生成对抗网络（Generative Adversarial Networks，GAN）这一革命性生成框架以来，该技术经历了系列改进：针对原始GAN存在的模式崩溃和训练不稳定等缺陷。Arjovsky等[14]通过引入Wasserstein距离（W距离）替代传统交叉熵损失，开发出W-GAN框架，有效提升了训练收敛性和生成多样性；Xu等[15]于2019年进一步突破性地提出条件表格生成对抗网络（Conditional Tabular Generative Adversarial Network，CTGAN）模型，将GAN的应用领域从非结构化图像数据扩展至结构化表格数据生成；刘悦等[16]率先将CTGAN应用于船舶横摇阻尼预测领域，验证了其在工程领域的适用性。然而，传统单阶段生成方法在直接建模输入与输出联合分布时，常常难以保留其间复杂的非线性物理关系，限制了其在工程预测任务中的外推性能。

为解决该问题，Moon等[17]提出“两阶段生成框架”，通过解耦输入特征的分布学习与目标变量的映射建模，提升了生成数据的物理一致性与预测精度。

本文在Moon等工作的基础上，首次将两阶段生成策略应用于干涸后传热建模场景，构建了一种结构化数据增强与预测框架，研究流程如图1所示。具体而言，第一阶段使用CTGAN对斯皮尔曼（Spearman）相关性筛选出的高相关输入无量纲数进行建模与样本生成；第二阶段则通过监督回归模型生成目标输出，显式保留输入与目标变量之间的非线性物理映射。该方案有效解耦了“分布建模”与“物理建模”任务，在提升模型泛化能力的同时，也增强了其在外推场景下的稳定性与物理一致性。

主要贡献如下：1）提出一种基于CTGAN的两阶段数据生成方法，针对干涸后传热建模中输入-输出解耦难题进行结构性建模；2）在多个异构外部数据集上验证其在数据外推场景下的鲁棒性与稳定性；3）通过与表格变分自编码器（Tabular Variational Autoencoder，TVAE）、表格生成对抗网络（Tabular Generative Adversarial Network，TGAN）、条件表格生成对抗网络增强模型（Conditional Tabular Generative Adversarial Network Plus，CTABGAN+）等模型对比，定量分析其在生成质量、外推误差控制方面的优势；4）与传统单阶段CTGAN方法进行对比，揭示两阶段方案在保持物理一致性和预测性能方面的显著优势。5）与经验关系式进行对比，证明了本文方法的有效性。

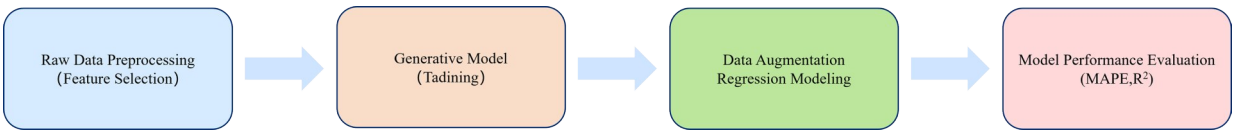


图1 研究流程示意图
Fig.1 Schematic diagram of the research process

1 数据预处理

本研究采用Becker等[18]提出的PDO实验数据库作为基础数据源。该数据库基于垂直向上圆管内均匀加热条件下的水工质实验，共包含8 519组数据样本。实验覆盖范围包括：压力P=17~21.5 MPa，质量流量G=497.8~3 500 kg·(m^2·s)$^{-1}$，管道直径D_h=2.5~24.7 mm，热流密度q=147~1 923 kW·m^{-2}，平衡含气率x_e=0.344~1.586。

$$x_e = \frac{H_B - H_L}{H_{VL}} \tag{1}$$

式中：H_B表示热平衡计算得到的主流焓；H_L表示饱和液相焓；H_{VL}表示蒸发潜热。

1.1 干涸后传热模型输入数据集建立

为建立干涸后传热预测模型，本研究根据Song等[19]的研究考虑了13个无量纲参数和努塞尔数

表1 特征选择前的无量纲参数
Table 1 The dimensionless numbers before feature selection

序号 Index	无量纲数 Dimensionless numbers	序号 Index	无量纲数 Dimensionless numbers
1	We_V	8	B_o
2	Re_{TP}	9	x_e
3	Fr_L	10	$\frac{c_{pL}}{c_{pV}}$
4	Pr_W	11	$\frac{\rho_L}{\rho_V}$
5	$\frac{\mu_V}{\mu_W}$	12	$\frac{\lambda_L}{\lambda_V}$
6	$\frac{\lambda_V}{\lambda_W}$	13	$\frac{\mu_L}{\mu_V}$
7	$\frac{c_{pV}}{c_{pW}}$	14	Nu_m

（Nu_m），详见表1。这些参数涵盖壁面温度特性的普朗特数Pr_W，两相流下的雷诺数Re_{TP}，流体力学特性的韦伯数We_V，流动状态的弗劳德数Fr_L及热力学特性的沸腾数B_o，同时还包括8项典型的热物性比参数x_e、$\frac{c_{pV}}{c_{pW}}$、$\frac{c_{pL}}{c_{pV}}$、$\frac{\lambda_V}{\lambda_W}$、$\frac{\lambda_L}{\lambda_V}$、$\frac{\mu_V}{\mu_W}$、$\frac{\mu_L}{\mu_V}$、$\frac{\rho_L}{\rho_V}$。其中$c$是比热、$\lambda$是热导率、$\mu$是黏度系数、$\rho$是密度。为明确物性参数的来源与使用，本研究对下标含义进行了规范说明：下标“TP”表示两相流状态下的饱和蒸汽特性参数，下标"V"特指饱和蒸汽相物性，下标“L”表示饱和液相物性。对于涉及壁面换热计算的情况，下标“W”用于标识基于壁面温度评估的蒸汽物性参数。

1.2 特征选择

为避免高维数据带来的维度灾难问题并降低计算成本，采用Spearman[20]相关性分析方法评估各特征与目标变量Nu_m之间的关联性。该方法通过计算秩相关系数，筛选出与Nu_m有显著相关性的无量纲数作为CTGAN模型输入特征。Spearman相关性分析具有以下优势：1）对异常值具有较强鲁棒性；2）不依赖于数据的特定分布形式；3）对样本量要求相对宽松；4）能够自动处理不同量纲的特征[21]。这些特性使其特别适用于本文数据分布未知或不符合正态分布假设的研究场景。Spearman相关性分析公式：

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}, \quad d_i = x_{i,r} - y_{i,r} \tag{2}$$

式中：N是观测点总数；r_s是Spearman相关系数；r_s=1表示两个变量完全正相关，r_s=−1表示两个变量完全负相关，r_s=0表示没有相关性；d_i为第i个观测点两变量秩之差；x_i，r和y_i，r分别为变量x和y的秩。

图2为各个特征对的无量数与Nu_m的Spearman相关系数柱状图。本文将相关系数绝对值（$|\rho|$）设为显著性阈值：将$|\rho|$>0.3设定为与Nu_m显著相关的无量纲数，$|\rho|$≤0.3则为相关性较弱的无量纲数。

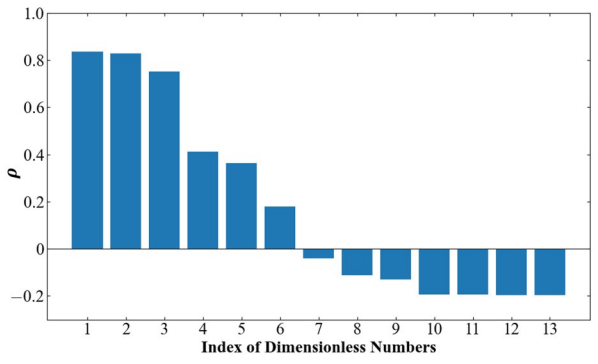


图2 13个输入特征与Nu_m的Spearman相关系数
Fig.2 Spearman correlation coefficients between 13 input features and Nu_m

最终选出的5个关键的无量纲数为：$\frac{\mu_V}{\mu_W}$、Fr_L、We_V、Re_{TP}、Pr_W。这些无量纲参数不仅在统计上与Nu_m呈现显著相关性，也具有明确的物理机制支撑。其中，$\frac{\mu_V}{\mu_W}$反映了壁面温度对蒸汽黏性影响，是衡量两相传热行为的重要间接指标；Fr_L和We_V则分别体现了液滴在重力场中的输运行为及其在蒸汽主流作用下的破碎特性，直接影响蒸汽-液滴界面的传热效率；Re_{TP}代表系统的对流能力，是干涸后传热中对流换热强度的核心量纲；而在干涸后区域，传热过程受到蒸汽-壁面传热和蒸汽-液滴界面传热的共同作用，因此Pr_W与Nu_m呈现出较高的斯皮尔曼相关性也是合理的。这些无量纲数覆盖了从壁面到界面的多尺度热传递过程，为后续两阶段生成模型提供了物理合理性支撑。

基于上述物理意义与统计显著性，这5个无量纲参数被选为后续数据生成模型的核心输入特征。

2 模型结构

本节介绍所提出模型的整体结构。模型主要由两大模块组成：1）数据生成模块：利用生成模型合成用于训练的无量纲特征；2）回归预测模块：结合真实数据与生成数据进行建模训练，完成干涸后传热系数预测任务。其中，数据生成模块由两部分组成：1）采用CTGAN模型生成5个高相关性的无量纲输入特征（如Fr_L、We_V等）；2）利用回归模型根据生成的输入特征生成对应的Nu_m输出值。

2.1 数据生成模型：CTGAN

标准生成对抗网络（Vanilla GAN）在数据生成任务中具有较强表现，但在处理结构复杂的表格数据时仍面临一定挑战。例如，当数据中各变量分布差异较大时，标准GAN难以全面建模；同时，对于以one-hot编码表示的分类变量，GAN往往难以生成满足编码约束的样本。为克服上述问题，CTGAN[15]引入了两项关键改进：

a）模式特定归一化（Mode-specific Normalization）：对于连续变量，CTGAN使用变分高斯混合（Variational Gaussian Mixture，VGM）模型估计其潜在的多重分布结构，并据此完成归一化映射。训练阶段，模型使用归一化后的编码值进行学习；数据生成后，再将结果逆变换回原始尺度，从而增强模型对多峰分布特征的刻画能力。

b）条件训练机制（Conditional Training）：为解决分类变量中类别频次不均的问题，CTGAN构建了

基于类别频率的条件向量机制。具体而言，该方法将分类变量编码为条件向量，并根据类别的对数频率进行采样，从而提高低频类别的出现概率，避免生成结果偏向于高频类别。这些条件向量与噪声向量共同输入生成器，以提升生成数据的多样性与结构一致性。

此外，CTGAN还融合了生成对抗网络领域的最新进展，例如采用Wasserstein生成对抗网络梯度惩罚版（Wasserstein Generative Adversarial Network with Gradient Penalty，WGANGP）[14]的Wasserstein损失函数与打包生成对抗网络（Packed Generative Adversarial Network，PacGAN）[22]的判别器架构，以提升模型训练的稳定性和生成数据的质量。图3展示了CTGAN相较于传统GAN的结构改进，CTGAN的损失函数定义如下：

$$L = E_{z\sim P_g}\left[D(G(z))\right] - E_{x\sim P_r}\left[D(x)\right] + \lambda E_{\hat{x}\sim P_{\hat{x}}}\left[\left(\left\|\nabla_{\hat{x}} D(\hat{x})\right\|_2 - 1\right)^2\right] \tag{3}$$

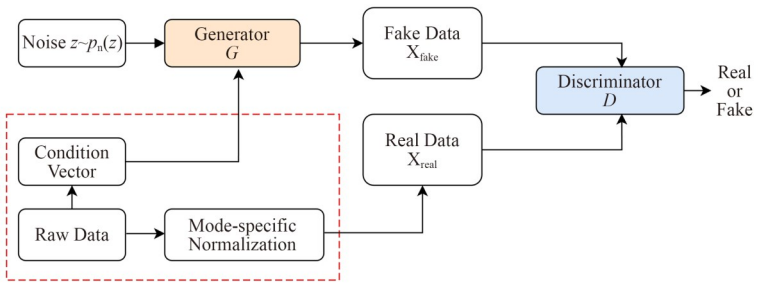


图3 CT-GAN模型结构示意图
Fig.3 Diagram of the model of CT-GAN

式中：前两项为WGAN基本损失；最后一项为梯度惩罚项，用于约束判别器在插值样本点上的梯度幅度；P_r和P_g分别表示真实数据与生成数据的分布；$\hat{x}$为真实样本的线性插值；λ为梯度惩罚系数。

本文基于CTGAN生成6个高相关性的无量纲输入特征，所有特征均为连续型浮点数，经过模式特定归一化后映射至[−1，1]区间。需要指出的是，该输入特征组合可根据实际应用场景灵活调整，以适配不同建模任务。

2.2 回归预测模型

本节将介绍干涸后传热系数预测模型的构建过程。图4展示了本方法的整体结构：首先使用真实的输入特征X_{real}训练CTGAN，然后利用训练好的CTGAN生成人工输入特征X_{fake}。CTGAN在训练过程中学习了X_{real}的分布，因此它可以生成与X_{real}相似的特征。第二阶段本文使用（X_{real}，Y_{real}）训练监督回归模型（Backpropagation Neural Network，BPNN），使其能够根据输入特征预测目标变量。将X_{fake}输入BPNN中，模型将输出对应的预测值，将其记为Y_{fake}。最后将真实数据X_{real}，Y_{real}与生成数据集X_{fake}，Y_{fake}合并，构建完整的训练集，基于该数据集训练一个BP神经网络，以构建最终的干涸后传热系数预测模型，该流程的完整实现步骤以伪代码形式给出，见算法1（表2）。

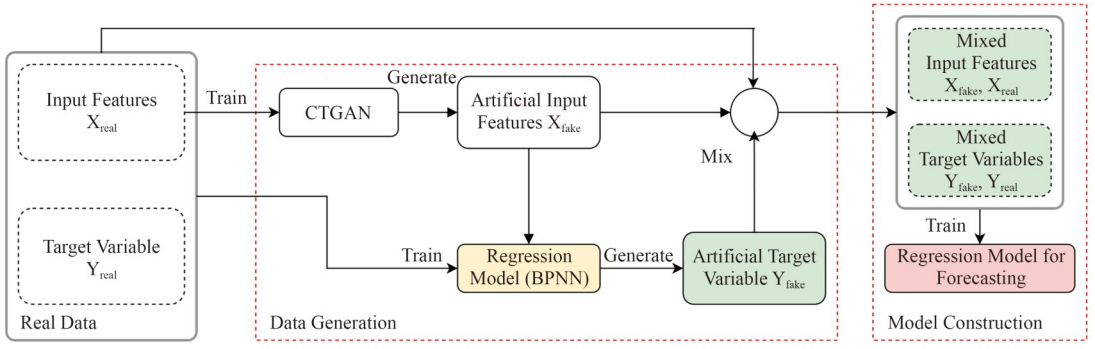


图4 拟议方案的总体结构
Fig.4 Overall structure of the proposed scheme

本文采用CTGAN与BPNN分别生成X_{fake}与Y_{fake}是为了减少CTGAN生成错误目标变量的概率。在CTGAN中，各变量维度在训练中具有相同的权重；因此，输入特征与目标变量的相对权重主要由其

维度数量决定。多数情况下，输入特征维度远大于目标变量维度，因此CTGAN在训练时对输入特征的关注度远高于目标变量。若直接让CTGAN同时生成输入特征与目标变量，则可能导致生成结果中输入与输出之间的不一致性。

表2　两阶段数据生成与模型训练流程
Table 2　Procedure of two-stage data generation and model training

算法1 Algorithm 1
Input：Real input feature set $X_{real}=\{x_1,x_2,x_3,\cdots,x_n\}$ Real target variable set $Y_{real}=\{y_1,y_2,y_3,\cdots,y_n\}$
Output：Trained BPNN model Synthetic input feature set X_{fake} Synthetic target variable set Y_{fake}
1：Train CTGAN using the real input feature set X_{real}
2：Generate synthetic input samples X_{fake} using the trained CTGAN
3：Train an initial BPNN model using (X_{real}, Y_{real})
4：For each sample x_i in X_{fake} do
5：｜Predict the corresponding target value y_i using the ｜trained BPNN
6：｜Add the predicted value y_i to Y_{fake}
7：End for
8：Merge the datasets：
9：｜$X_{mix}=X_{real}\cup X_{fake}$
10：｜$Y_{mix}=Y_{real}\cup Y_{fake}$
11：Train the final BPNN model using (X_{mix}, Y_{mix})

3　实验设置

为验证本方法的有效性，本文开展了多组实验。在正式介绍实验内容前，先对所采用的对比模型与评估指标进行说明。

3.1　Becker数据集划分

为确保模型评估的客观性与可靠性，本研究针对Becker实验数据集采用了严格的数据划分策略。具体而言，本文对完成特征选择后的干涸后传热数据集进行随机划分，其中80%的样本被用作训练集，用于构建与训练预测模型；其余20%的样本作为独立测试集，用于模型性能评估。

此划分方式可有效模拟实际工程中“已知数据训练、未知数据预测”的应用场景，避免因数据泄漏导致的模型过拟合，从而更准确评估其在未见工况条件下的泛化能力。为保证实验结果的可复现性，本文在数据划分与模型训练过程中统一设置随机种子为42。

3.2　模型设置

为了对比本研究方法的性能，本文引入了三种生成模型作为对比，其中包括两种基于GAN的模型：TGAN[23]和CTABGAN+[24]；一种基于深度神经网络的生成模型：TVAE[15]。上述三种模型仅作为对比方法。本文以基于CTGAN的两阶段数据生成方法作为研究对象。所有模型均使用Python 3.9实现，其中TVAE与TGAN基于TensorFlow框架，CTGAN与CTABGAN+基于PyTorch框架进行训练。

对于GAN模型，本文采用相似的超参数。生成器和判别器均由两层构成，每层256个神经元，激活函数分别为生成器（Sigmoid）和判别器（ReLU）。训练的轮数（epoch）为1 000，学习率设为0.000 2，噪声维度为20，批处理大小为500。除了上述超参数外，其他参数设置参考了各模型原论文中的实验配置。在实验CTGAN时，本文的5个无量纲数均视为连续变量。TVAE相关的超参数（如优化器、激活函数等）均使用SDGym工具包中的默认设置。

作为回归器的BPNN，其超参数在所有实验中保持一致。具体设置如下：隐藏层层数为5，依据实验经验设定；每层神经元个数256；损失函数为预测值与真实Nu_m之间的均方误差（Mean Squared Error，MSE）；其他参数依据经验设置。

3.3　评估指标

为了比较各模型的预测精度，本文采用两种最常用的误差评估指标：平均绝对百分比误差（Mean Absolute Percentage Error，MAPE）和决定系数（R^2）。这两类指标在模型预测领域被广泛使用，具有较高的参考价值。具体而言，MAPE和R^2的计算公式分别如式（4）~（5）所示。

$$\mathrm{MAPE}=\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{4}$$

$$R^2=1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2} \tag{5}$$

式中：n表示样本总数；y_i表示真实数据；$\widehat{y_i}$表示模型预测值；$\bar{y}$是平均值。MAPE以真实值为分母计算相对误差，能够反映预测相对于真实值的偏离程度；R^2模型解释目标变量变异程度的比例。通过这两个指标可以全面反映模型在Nu_m预测任务中的性能表现。

4　实验与讨论

本文设计并实施了4组实验，用以评估所提出方法的有效性。首先，评估不同数据生成方案的效果；其次，测试使用这些生成数据集训练的预测模型性能；之后，比较提出的输入特征与目标变量分开生成的数据生成方案与传统同时生成输入特征与目标

变量的方案在效果上的差异；最后，探讨传统经验关系式法与本文方法预测的准确度对比。本文使用生成模型生成的数据量与原始训练数据量相同。

4.1 数据生成

在实验中，本文考虑将数据生成划分为两种类型：输入特征生成与目标变量生成。在§3.2选取的三种对比生成模型（TGAN、CTABGAN+和TVAE）基础上，进一步引入本文提出的基于CTGAN的两阶段数据生成方法，因此在输入特征生成阶段共考虑4种不同的生成模型。

在输入特征生成中通过主成分分析（Principal Component Analysis，PCA）方法对其外推能力进行对比分析。PCA是一种常见的可视化技术，通过从原始数据中提取主成分来降低数据维度，从而在低维空间中表示高维数据结构。当从生成数据中提取前两个主成分时，可将其映射至二维空间进行可视化。

为了进一步量化不同数据生成模型输入数据的外推能力，本研究引入了PCA的外推比例计算方法。提取PCA降维后的前两个主成分。随后，基于该二维空间构建其凸包（Convex Hull），并借助Delaunay三角剖分判断每一个生成样本是否落在该凸包内部。具体步骤如下：1）使用ConvexHull构建真实数据PCA映射空间的边界；2）使用Delaunay方法判定生成样本是否位于该凸包内部；3）将落在凸包之外的样本记为外推样本，计算比例：

$$\text{Extrapolation Ratio} = \frac{N_{\text{outside}}}{N_{\text{total}}} \tag{6}$$

式中：N_{outside}为落在真实数据凸包之外的生成样本数；N_{total}为生成样本总数。

该指标衡量的是生成数据分布外样本的占比，比例越高，表明该模型生成的数据越能跳出原始分布边界，具备更强的潜在外推能力。

对于目标变量，使用回归模型BPNN根据生成的输入特征进行输出预测，并采用直方图展示其分布情况，该分析在所有模型上均可进行。

图5展示了不同生成模型上所生成的输入特征在前两个主成分维度上的分布情况；表3展示了不同生成模型生成数据的外推比例。图6展示了对应的目标变量（即由BPNN生成的Nu_m）与真实输出数据的直方图。在图中x轴表示变量的取值范围，y轴表示该值的频数。

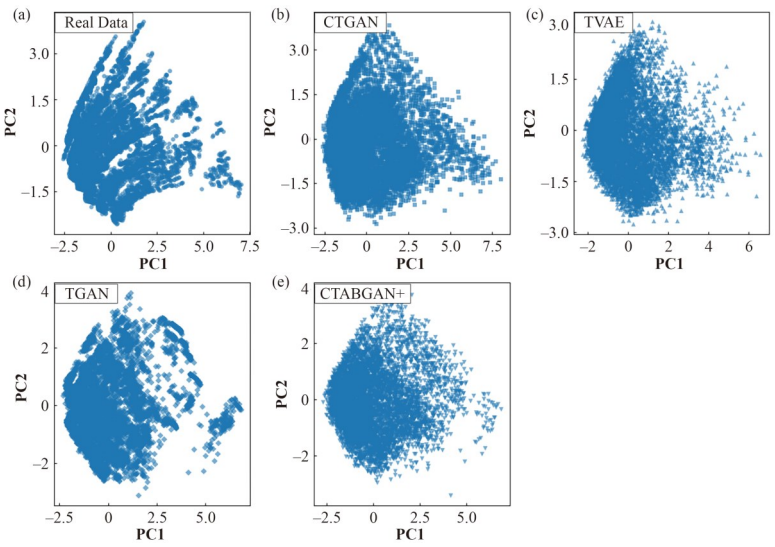


图5 真实数据(a)、CTGAN (b)、TVAE (c)、TGAN (d)和CTABGAN+ (e)生成模型生成数据的输入特征PCA结果
Fig.5 PCA results of input features from real data (a), CTGAN (b), TVAE (c), TGAN (d) and CTABGAN+ (e) data generated by different models

表3 不同模型生成数据的数据外推比例
Table 3 Extrapolation ratio of data generated by different models

生成器名称 Name of the generator	外推比例 Extrapolation ratio / %
CTGAN	3.54
TVAE	0.24
TGAN	3.52
CTABGAN+	1.75

图5(a)展示了真实数据的分布，表现出明显的非对称弯曲条带结构，主方向沿PC1正方向，右侧逐渐稀疏，有明显“尾部结构”，显示原始数据分布复杂，非高斯分布。(b)~(e)这些生成模型的输入特征接近真实分布，特别是CTGAN，其生成的数据在分布范围和覆盖程度方面表现最优，与真实数据的相似度最高，外推比例较高（3.54%），说明该模型具备

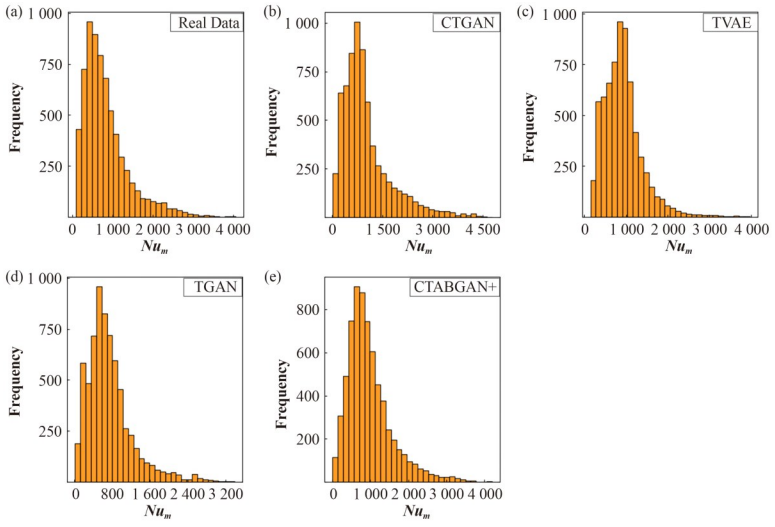


图6 真实数据(a)与BPNN生成目标变量CTGAN (b)、TVAE (c)、TGAN (d)和CTABGAN+ (e)的直方图分布
Fig.6 Histogram distributions of real data (a) and target variables CTGAN (b), TVAE (c), TGAN (d) and CTABGAN+ (e) generated by BPNN

较强的边界探索能力。TVAE分布密集、中心聚集，几乎完全局限在原始数据主密度区域；外推比例最小，说明该模型过于保守，几乎没有生成边界样本。TGAN分布密度较广、形状较散，虽然生成样本有明显扩散趋势，部分甚至超出CTGAN生成范围，但非自然的条带状结构（如右下方）可能是训练不稳定所致。CTABGAN+分布形式居中，比TVAE更广但比CTGAN更收敛；外推比例适中。虽然保留了主分布的弯曲趋势，但是未能覆盖尾部，属于一个相对平衡但偏保守的生成策略。

综合对比CTGAN、TVAE、CTABGAN+和TGAN的结构与训练机制可以看出，CTGAN在外推能力上表现最优（3.54%），主要得益于其以下特点：1）支持关闭最小最大值限制，使生成器可以输出训练边界之外的样本；2）引入条件向量与多模态归一化方法，使模型能在多样化条件下建模数据分布尾部；3）采用WGAN-GP与PacGAN等稳定训练机制，有效提升边界区域样本的学习效果。相比之下，CTABGAN+虽在稀疏区域建模上有所加强，但其保守的剪裁机制限制了模型的泛化能力；TVAE与TGAN则在外推样本多样性与分布控制方面表现不如CTGAN，并且TGAN采用传统的交叉熵作为损失函数，极易出现训练不稳定的问题，严重影响生成数据的质量。

4.2 预测性能分析

在第一组实验中，本文采用4种不同的数据生成模型对Becker数据集进行输入特征的扩充，以增强训练样本规模并提升预测模型的泛化能力。为评估数据增强策略的有效性，本文对基于扩充数据集训练的回归模型在预测精度方面开展了对比分析。

作为对比基准，采用原始数据集训练构建一个BPNN。为全面评估预测模型的外推能力，不仅在Becker数据集[25]的测试集上进行了测试，同时将训练好的模型应用于HER[26]、BIS[27]和SWE[28]三个独立数据集上验证其外推能力。需要强调的是，CTGAN模型仅在Becker训练集上进行训练，HER、BIS和SWE数据集仅用于外推测试过程，未参与任何模型训练或参数更新，以避免数据泄露问题。

图7给出了不同数据集输入特征的PCA结果。可以观察到，BIS样本点集中在右上角（PC1>5），几乎没有与Becker分布重合；且分布稀疏，属于典型的极端外推。HER与SWE相对外推程度较低。

表4列出了各模型在4个数据集上的MAPE表现，每组结果为重复10次实验的平均值，最优结果已加粗标出。结果表明，基于CTGAN的数据增强模型在所有数据集上的预测精度均优于未扩充数据的BPNN基准模型，特别是在Becker和SWE数据集上显著降低了MAPE，表明其生成数据对提升预测能力具有积极作用。尽管TVAE在HER上表现最佳，但在其他数据集波动较大，稳定性较差。TGAN与CTABGAN+整体表现逊于CTGAN，生成数据的质量和分布拟合程度较低。

BIS是一个极端外推场景，尽管CTGAN的MAPE达106.29%，但相比原始模型仍有提升，显示其一定的缓解能力，然而在主成分空间偏离训练分布较远时仍存在明显局限性。

进一步来看，BIS数据集可视为一种极端分布外工况，其样本在PCA空间主要分布于PC1>5的区

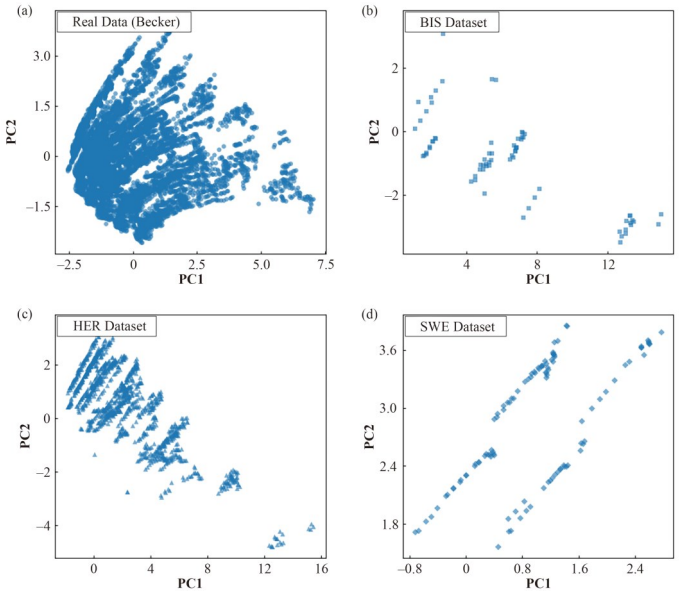


图7 各个数据集输入特征的PCA
Fig.7 PCA of the input features across all datasets

表4 预测结果MAPE的比较
Table 4 MAPE comparison of forecasting results

	Becker (Test set)	BIS	HER	SWE
BPNN (without data augmentation)	3.78%	201.39%	18.88%	13.75%
CTGAN	3.27%	106.29%	18.43%	11.94%
TVAE	3.58%	71.48%	17.83%	13.59%
TGAN	3.96%	92.65%	22.86%	15.09%
CTABGAN+	3.56%	161.04%	18.25%	13.01%

域，与Becker训练集的主密度区域几乎没有重叠，相当于完全落在Becker数据凸包之外。为了进一步从输入特征维度验证这一分布偏移现象，图8给出了两者在Fr_{L}特征上的频数分布直方图。可以看出，Becker数据集的Fr_{L}主要集中在较低区间，而BIS数据集的Fr_{L}分布显著更高且更宽，两者几乎不存在重叠区，说明BIS工况在关键参数上远离训练集的物理分布区间，因而在BIS数据集上表现出较大的预测误差。

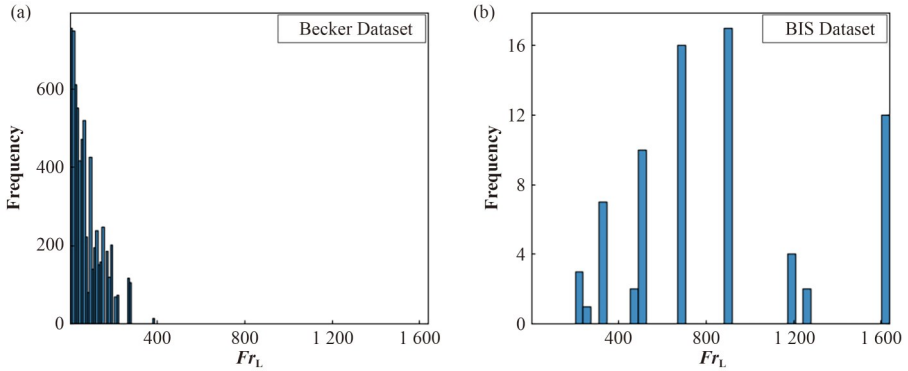


图8 Becker数据集(a)与BIS数据集(b)的Fr_{L}频数分布直方图
Fig.8 Histograms of the Fr_{L} distributions for the Becker (a) and BIS (b) datasets

综合分析结果表明，引入数据生成策略确实提升了预测效果。尤其是CTGAN，在大多数数据集上取得了较低的MAPE，说明它在捕捉原始数据整体分布方面更有优势，能生成更贴近真实的数据。

表5显示，相比于不进行数据扩充的BPNN，加入CTGAN合成数据后模型的R^2指标表现更好。而其他生成模型在某些数据集上的表现略有下降。并且，在外推最困难的BIS数据集中，所有方法的R^2表现均不理想，部分方法甚至为负值，说明这些模型在此类分布外数据上几乎完全失效。相比之下，CTGAN的R^2为0.427 0，为唯一取得正值的方法，尽管其仍明显低于其他测试集，但显示出一定程度的分布外泛化能力。

表5　预测结果R^2的比较
Table 5　R^2 comparison of forecasting results

	Becker (Test set)	BIS	HER	SWE
BPNN (without data augmentation)	0.994 4	−1.165 1	0.847 9	0.769 6
CTGAN	0.995 0	0.427 0	0.873 8	0.781 0
TVAE	0.993 7	0.193 2	0.867 3	0.708 1
TGAN	0.993 1	−0.263 3	0.845 1	0.560 6
CTABGAN+	0.994 2	−0.245 8	0.902 0	0.682 9

CTGAN相较于其他生成模型表现更优，主要源于其设计上对表格数据分布特性的更精准建模能力。具体而言，CTGAN采用了模式特定归一化和条件向量生成机制，能够有效处理变量分布不均、特征间强非线性关系等问题，进而更好地拟合高维输入空间中真实数据的联合分布；此外，基于Wasserstein损失函数提高了模型训练的稳定性。无量纲物理参数之间存在高度非线性耦合关系的工程场景尤为关键。尽管CTABGAN+同样采用了模式特定归一化和条件向量生成机制，但其整体优化目标更偏向于分类型变量建模，在面对本研究中全连续变量且分布跨度较大的数值特征时，其判别器与生成器的收敛稳定性相较CTGAN仍存在一定不足。

4.3　两阶段数据生成方法与单阶段数据生成方法相比较

在本实验中，将两阶段数据生成方案与传统单阶段数据生成方法进行比较。单阶段数据生成方法是输入特征和目标变量同时生成。为保证对比的公平性，单阶段方法同样采用CTGAN作为数据生成模型。

表6是两种方法在各个数据集上的MAPE和R^2。实验结果表明，单阶段数据生成方法在各个数据集上的MAPE和R^2均不及本文的方法。尤其是在外推数据集（BIS、HER和SWE）上，单阶段方法出现了极端负R^2，表明模型在训练集分布外的样本上拟合完全失效，相比之下，本文方法在上述数据集上仍保持了一定程度的预测能力，显示出更强的分布外泛化性能。

表6　两阶段数据生成方法与单阶段数据生成方法的比较
Table 6　Comparison between two-stage and one-stage data generation methods

数据集 Data set	两阶段数据生成方法 Two-stage data generation method		单阶段数据生成方法 One-stage data generation method	
	MAPE	R^2	MAPE	R^2
Becker (Test set)	3.27%	0.995 0	8.33%	0.970 6
BIS	106.29%	0.427 0	1 248.89%	−70.553 2
HER	18.43%	0.873 8	103.42%	−1.036 5
SWE	11.94%	0.781 0	56.03%	−0.591 7

产生上述差异的主要原因在于单阶段数据生成方法是同时生成输入特征和目标变量，这在实际中容易引发一个关键问题：输入特征与目标变量之间的物理或统计对应关系难以准确建立。因为在CTGAN训练中，各变量在损失函数中被赋予同等权重，这会导致维度较少的目标变量在生成过程中被弱化，从而导致“输入正确但输出偏差大”的情况。

4.4　两阶段数据生成方法与经验关系式对比

为进一步验证本方法的有效性，本文将其与两种经典的干涸后传热经验关系式（Mikropol'skii等[29]和Groeneveld等[30]）在4个测试集上进行了对比评估。对每种方法计算其预测结果的MAPE与R^2，经验关系式的适用范围如表7所示。根据表7可知，BIS与SWE两个数据集的关键参数（如x_e、D_h、G）均在Groeneveld关系式的适用区间内，因此经验关系式可在该类“非极端外推”场景中提供更稳定的预测结果。相反，在Becker测试集的高干度区（$x_e > 1$）超出经验关系式适用范围时，经验关系式的预测精度显著下降，而两阶段数据生成方法仍能保持相对较

表7 干涸后传热模型的应用范围
Table 7 Post-dryout prediction models and application ranges

预测模型 Prediction model	P / MPa		G / kg·(m²·s)⁻¹		D_h / mm		x_e	
	Min	Max	Min	Max	Min	Max	Min	Max
Mikropol'skii	4.25	22	700	2 000	8	24	0.06	1.0
Groeneveld	6.88	21.5	700	5 300	1.5	25.0	0.1	0.9
Becker	17.0	21.5	497.8	3 500	2.5	24.7	0.344	1.586
BIS	16.8	21.9	1 350	3 400	—	—	0.1	1.0
HER	17.0	21.5	700	3 500	5	20	0.1	1.0
SWE	20.68	20.68	949.4	1 356.2	10.4	10.4	0.08	0.98

表8 两阶段数据生成方法与经验关系式在MAPE指标下的比较
Table 8 Comparison between the two-stage data generation method and empirical correlations in terms of MAPE

	Becker (Test set)	BIS	HER	SWE
两阶段数据生成方法 Two-stage data generation method	3.27%	106.29%	18.43%	11.94%
Mikropol'skii	40.40%	70.55%	75.71%	56.34%
Groeneveld	22.25%	25.38%	37.73%	26.22%

表9 两阶段数据生成方法与经验关系式在R^2指标下的比较
Table 9 Comparison between the two-stage data generation method and empirical correlations in terms of R^2

	Becker (Test set)	BIS	HER	SWE
两阶段数据生成方法 Two-stage data generation method	0.995 0	0.427 0	0.873 8	0.781 0
Mikropol'skii	0.818 0	0.732 8	0.613 7	0.030 4
Groeneveld	0.904 4	0.924 7	0.883 0	0.751 6

高的准确性，体现出其在复杂工况中的灵活性。实验结果如表8和表9所示。在Becker测试集，HER和SWE三个数据集中，两阶段数据生成方法相较于经验关系式表现出更低的MAPE和更高的R^2，体现其更强的拟合能力和一定的外推能力。然而在BIS数据集这一极端外推场景下，两阶段数据生成法的预测误差MAPE高达106.29%，为三种方法中最高。这说明该方法在与训练数据分布差异极大的样本上仍存在外推能力不足，尤其在生成样本无法覆盖原始物理机制变化区域时，回归模型的拟合能力会显著下降。

综合Becker、HER和SWE数据集的预测结果可见，本文方法在训练集附近及中等分布偏移工况下均能稳定保持较小误差，具有较好的可靠性。需要指出的是，BIS数据集属于典型的极端分布外工况，其输入特征在训练集分布范围之外，因此模型在BIS上预测性能下降属于深度学习方法的普遍局限性，而非模型本身不稳定所致。

5 结语

本文针对干涸后传热系数建模中面临的数据匮乏与分布外推难题，提出了一种基于CTGAN的两阶段数据生成与回归预测框架。该方法通过在输入特征空间引入条件生成模型进行结构化样本增强，并结合监督学习回归器实现对目标变量的预测，有效解耦了输入分布建模与输入-输出物理映射两项任务。

实验结果表明，相较于单阶段数据生成方法，本文提出的两阶段方案在多个数据集上均展现出更优的泛化性能，特别是在Becker、HER和SWE数据集上，其MAPE降幅显著、R^2提升明显。其中，CTGAN能够较好拟合高相关输入特征的真实分布，为后续回归预测提供更高质量的数据支撑。在面对极端外推的BIS数据集时，尽管整体外推精度仍有限，但相较无数据增强模型依然取得了大幅度误差缓解。

从定量评价结果来看，CTGAN在各类模型中整体表现最优。相较未进行数据扩充的BPNN基准模型，CTGAN在4个数据集上的MAPE平均降低约19.1%，并在BIS极端外推数据集中实现由负R^2向正值（−1.165 1提升至0.427 0）的显著转变，体现出更强的分布外泛化能力。与单阶段数据生成方法相比，两阶段数据生成方案在所有测试数据集上均表现出明显优势，其中MAPE平均降幅超过75%，R^2指标由普遍负值提升为稳定正值，表明该方法在复杂分布外工况下具有更好的稳定性与鲁棒性。进一步与经典干涸后传热经验关系式对比发现，本文方法

在Becker、HER与SWE数据集上的MAPE指标显著优于传统经验公式，而在极端工况数据集BIS上，经验公式仍表现出更低的误差水平。尽管如此，本文方法在BIS数据集上的R^2仍保持为正值，证明其在完全分布外样本上的趋势建模能力。此外，通过与TVAE、TGAN、CTABGAN+等模型的对比分析，本文验证了所提方法在生成数据质量、分布匹配度和预测稳定性方面的综合优势。该框架具有良好的可扩展性，可适用于其他存在强输入-输出非线性关系、样本采集困难的科学建模任务，如热工水力分析、材料性能评估等工程领域。

未来研究将进一步探索：1）结合物理引导的生成机制，以提升生成样本的物理一致性；2）引入更复杂的学生模型如集成回归网络，增强模型对高维输入特征的表达能力；3）在联邦学习框架下集成多源机构数据，提升小样本跨域建模的实用性与隐私安全性。

针对BIS数据集所代表的极端分布外工况，后续工作将重点围绕提升该类工况下的预测性能展开。具体而言，可通过引入少量具有代表性的BIS工况样本，采用迁移学习或分阶段微调策略，对生成模型与回归模型进行联合校正；同时，针对BIS数据中主导干涸后传热机理的特征子空间，对生成样本施加加权约束或条件化生成机制，以提高模型在极端外推场景下的稳定性与预测精度。

作者贡献声明 袁双楠负责模拟计算以及论文撰写；王凡提供技术支持；宋美琪负责基金支持、文章修改；王云伟、许巍负责提供技术支持；刘晓晶负责基金支持。

参考文献

1 Xia Z H, Cheng X. Investigation on development of new mechanistic model for post dryout heat transfer[C]// Volume 6: Thermal-Hydraulics and Safety Analysis. August 4-8, 2024. Prague, Czech Republic. American Society of Mechanical Engineers, 2024: V006T07A060. DOI: 10.1115/icone31-135118.

2 Rensch N, Köckert L, Badea A F, *et al*. Experimental investigation of post-dryout heat transfer with R-134a at high pressures[J]. Nuclear Science and Engineering, 2025, **199**(10): 1659 – 1678. DOI: 10.1080/00295639.2024.2403890.

3 Köckert L, Liu W, Cheng X. Post-dryout heat transfer in circular tubes using R-134a: experiment and correlation assessment[J]. Heat and Mass Transfer, 2024, **60**(8): 1453 – 1466. DOI: 10.1007/s00231-024-03498-5.

4 张凡, 张俊达, 孙启政, 等. 基于神经网络超参数优化方法的堆芯中子学参数预测研究[J]. 核技术, 2025, **48**(10): 100606. DOI: 10.3724/j. 0253-3219.2025. hjs. 48.240219. CSTR: 32193.14. hjs. CN31-1342/TL. 2025.48.240219.

ZHANG Fan, ZHANG Junda, SUN Qizheng, *et al*. Prediction of core neutronic parameter based on neural network hyper parameter optimization method[J]. Nuclear Techniques, 2025, **48**(10): 100606. DOI: 10.3724/j.0253-3219.2025. hjs. 48.240219. CSTR: 32193.14. hjs. CN31-1342/TL.2025.48.240219.

5 袁硕, 柴翔, 刘晓晶, 等. 新型搅混翼对燃料棒流致振动导致的微动磨蚀影响研究[J]. 核技术, 2025, **48**(3): 030604. DOI: 10.11889/j.0253-3219.2025.hjs.48.230466. CSTR: 32193.14.hjs.CN31-1342/TL.2025.48.230466.

YUAN Shuo, CHAI Xiang, LIU Xiaojing, *et al*. Study on the influence of novel mixing vanes on fretting wear caused by fuel rods flow-induced vibration[J]. Nuclear Techniques, 2025, **48**(3): 030604. DOI: 10.11889/j.0253-3219.2025. hjs. 48.230466. CSTR: 32193.14. hjs. CN31-1342/TL.2025.48.230466.

6 沈书秋, 熊进标, 刘晓晶, 等. 基于降阶模型的热管堆核热力耦合启堆瞬态模拟研究[J]. 核技术, 2025, **48**(6): 060013. DOI: 10.11889/j.0253-3219.2025.hjs.48.240471. CSTR: 32193.14.hjs.CN31-1342/TL.2025.48.240471.

SHEN Shuqiu, XIONG Jinbiao, LIU Xiaojing, *et al*. Transient simulation of nuclear thermal-hydraulic coupling startup of heat pipe reactor based on reduced-order models[J]. Nuclear Techniques, 2025, **48**(6): 060013. DOI: 10.11889/j.0253-3219.2025.hjs.48.240471. CSTR: 32193.14.hjs.CN31-1342/TL.2025.48.240471.

7 朱鹏橙, 赵鹏程, 刘晓晶, 等. 核热推进反应堆多目标优化设计方法研究[J]. 核技术, 2025, **48**(11): 110601. DOI: 10.3724/j. 0253-3219.2025. hjs. 48.240286. CSTR: 32193.14.hjs.CN31-1342/TL.2025.48.240286.

ZHU Pengcheng, ZHAO Pengcheng, LIU Xiaojing, *et al*. Multi-objective conceptual design method for nuclear thermal propulsion reactor[J]. Nuclear Techniques, 2025, **48**(11): 110601. DOI: 10.3724/j. 0253-3219.2025. hjs. 48.240286. CSTR: 32193.14. hjs. CN31-1342/TL. 2025.48.240286.

8 Zhu F D, Chen J C, Han Y G, *et al*. A deep learning method for estimating thermal boundary condition parameters in transient inverse heat transfer problem[J]. International Journal of Heat and Mass Transfer, 2022, **194**: 123089. DOI: 10.1016/j. ijheatmasstransfer. 2022. 123089.

9 Yu D L, Xu C, Hu C J, *et al*. Universal approach to predicting full-range post-dryout heat transfer under uniform and non-uniform axial heat fluxes[J]. Nuclear Engineering and Design, 2022, **393**: 111790. DOI: 10.1016/j.nucengdes.2022.111790.

10 Yang J F, Hu Y B, Zhang J, *et al*. Identification of marine oil spill pollution using hyperspectral combined with thermal infrared remote sensing[J]. Frontiers in Marine Science, 2023, **10**: 1135356. DOI: 10.3389/fmars. 2023. 1135356.

11 Fan J C, Zhang S, Wang X Z, *et al*. Multifeature semantic complementation network for marine oil spill localization and segmentation based on SAR images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, **16**: 3771 – 3783. DOI: 10.1109/JSTARS.2023.3264007.

12 刘川, 马玉娟, 范剑超. 基于小样本生成对抗网络的光学遥感影像溢油语义分割[J]. 海洋环境科学, 2025, **44**(1): 107 – 115. DOI: 10.13634/j.cnki.mes.2025.01.004.
LIU Chuan, MA Yujuan, FAN Jianchao. Small sample generation adversarial network semantic segmentation based on optical remote sensing oil spill images[J]. Marine Environmental Science, 2025, **44**(1): 107 – 115. DOI: 10.13634/j.cnki.mes.2025.01.004.

13 Ian G, Jean P A, Mehdi M, *et al*. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, **27**(9).

14 Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[C]//International Conference on Machine Learning, 2017.

15 Xu L, Skoularidou M, Cuesta-Infante A, *et al*. Modeling tabular data using conditional GAN[C]//Neural Information Processing Systems, 2019.

16 刘悦, 李敏, 徐娜, 等. 基于离群修正CTGAN的船舶横摇阻尼智能预测模型[J]. 中国造船, 2024, **65**(4): 66 – 76. DOI: 10.3969/j.issn.1000-4882.2024.04.006.
LIU Yue, LI Min, XU Na, *et al*. Intelligent prediction model for ship roll damping based on outlier-modified CTGAN[J]. Shipbuilding of China, 2024, **65**(4): 66 – 76. DOI: 10.3969/j.issn.1000-4882.2024.04.006.

17 Moon J, Jung S, Park S, *et al*. Conditional tabular GAN-based two-stage data generation scheme for short-term load forecasting[J]. IEEE Access, 2020, **8**: 205327 – 205339. DOI: 10.1109/ACCESS.2020.3037063.

18 Becker K M, Djursing D, Lindberg K, *et al*. Burnout conditions for round tubes at elevated pressures[M]// Proceedings of the International Symposium on Two-Phase Systems. Amsterdam: Elsevier, 1972: 55 – 73. DOI: 10.1016/b978-0-08-017035-0.50009-4.

19 Song M Q, Liu X J, Cheng X. A new correlation for post-dryout heat transfer in upward vertical flow[J]. Nuclear Engineering and Design, 2022, **392**: 111747. DOI: 10.1016/j.nucengdes.2022.111747.

20 Wissler C. The spearman correlation formula[J]. Science, 1905, **22**(558): 309 – 311. DOI: 10.1126/science. 22. 558.309.

21 de Winter J C F, Gosling S D, Potter J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data[J]. Psychological Methods, 2016, **21**(3): 273 – 290. DOI: 10.1037/met0000079.

22 Lin Z N, Khetan A, Fanti G, *et al*. PacGAN: the power of two samples in generative adversarial networks[J]. IEEE Journal on Selected Areas in Information Theory, 2020, **1**(1): 324 – 335. DOI: 10.1109/jsait.2020.2983071.

23 Xu L, Veeramachaneni K. Synthesizing tabular data using generative adversarial networks[EB/OL]. 2018: arXiv: 1811.11264. https://arxiv.org/abs/1811.11264.

24 Zhao Z L, Kunar A, Birke R, *et al*. CTAB-GAN+ : enhancing tabular data synthesis[J]. Frontiers in Big Data, 2024, **6**: 1296508. DOI: 10.3389/fdata.2023.1296508.

25 Becker K M, Ling C H, Hedberg S, *et al*. An experimental investigation of post dryout heat transfer: KTH-NEL--33 (v. 1, 2) [R]. Royal Institute of Technology (Sweden, Stockholm), Department of Reactor Technology, 1983.

26 Herkenrath H, Mörk-mörkenstein P, Jung U, *et al*. Heat transfer in water with forced circulation in the 140-250 bar pressure range- results of investigations[R]. 1967.

27 Bishop A A, Sandberg R O, Tong L S. Forced-convection heat transfer to water at near-critical temperatures and supercritical pressures[R]. Westinghouse Electric Corporation, Pittsburgh, Pa. Atomic Power Div, 1964.

28 Swenson H S, Carver J R, Szoeke G. The effects of nucleate boiling versus film boiling on heat transfer in power boiler tubes[J]. Journal of Engineering for Power, 1962, **84**(4): 365 – 371. DOI: 10.1115/1.3675052.

29 Miropolskiy Z L. Heat transfer in film boiling of a steam-water mixture in steam-generator tubes[J]. Teplonergetika, 1963, **10**: 49 – 52.

30 Groeneveld D C. Post-dryout heat transfer: Physical mechanisms and a survey of prediction methods[J]. Nuclear Engineering and Design, 1975, **32**(3): 283 – 294. DOI: 10.1016/0029-5493(75)90099-0.