

时间记忆多输出PINN动态不确定性量化验证

袁瑞阳, 吕 艳*

南京理工大学数学与统计学院, 江苏 南京

收稿日期: 2026年6月15日; 录用日期: 2026年7月8日; 发布日期: 2026年7月20日

摘 要

针对一维线性加性噪声随机热方程, 构建时间统计记忆增强的多输出物理信息神经网络(TC-MO-PINN)框架。在无显式时间记忆基线模型的基础上, 引入上一时间层均值记忆(B2)和均值-标准差联合记忆(B3), 结合教师继承、残差修正与时间连续性约束。以闭式解析均值、标准差和分位数为基准, 从分布中心、尺度校准、区间覆盖、分位数结构和误差传播稳定性多维度评价各模型。结果表明, 均值记忆有效降低分布中心误差, 均值-标准差联合记忆进一步增强均值传播稳定性; 去除解析统计矩锚定后, 联合记忆模型仍优于基线, 表明时间统计记忆对随机动力系统不确定性量化具有独立贡献。

关键词

随机偏微分方程, 不确定性量化, 物理信息神经网络, 多输出网络, 时间统计记忆

Time-Memory Multi-Output PINNs for Dynamic Uncertainty Quantification

Ruiyang Yuan, Yan Lv*

School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing Jiangsu

Received: June 15, 2026; accepted: July 8, 2026; published: July 20, 2026

Abstract

A time-statistical-memory enhanced multi-output physics-informed neural network (TC-MO-PINN) framework is proposed for a one-dimensional linear stochastic heat equation with additive noise. Starting from a memory-free baseline (B1), a previous-time mean memory model (B2) and a joint mean—standard-deviation memory model (B3) are introduced, combined with teacher model inheritance, residual correction, and temporal continuity constraints. Using closed-form analytical mean, standard deviation and quantiles as benchmarks, models are evaluated on distribution center accuracy, scale calibration, interval coverage, quantile agreement, and error propagation stability. Results show

*通讯作者。

that mean memory reduces distribution center error, while standard-deviation joint memory further suppresses late-time mean drift. A control experiment without analytical moment anchoring confirms that the joint memory model is still superior to the baseline, and time-statistical memory provides an independent contribution to stable uncertainty quantification of stochastic dynamical systems.

Keywords

Stochastic Partial Differential Equations, Uncertainty Quantification, Physics-Informed Neural Networks, Multi-Output Networks, Time-Statistical Memory

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随机偏微分方程(Stochastic Partial Differential Equations, SPDEs)描述受物理规律与随机扰动共同影响的动态系统, 广泛应用于热传导[1]、流体输运[2]等工程物理领域。其数值求解除均值预测外, 还需刻画方差、置信区间、分位数结构及其时间传播[3]。物理信息神经网络(PINNs)将控制方程嵌入损失函数, 为 PDE 求解提供数据驱动框架, 其性能受残差构造[4]、采样策略和损失权重影响[5]。标准 PINN 输出确定性点估计, 贝叶斯 PINN [6]、多输出 PINN [7]等方法从不同角度处理不确定性量化问题。对于演化型 SPDE, 上一时间层统计状态对经验分布传播的影响仍需检验[8]; 已有 RNN/LSTM-PINN 工作在确定性非稳态 PDE 中展示了时序建模能力, 但显式统计记忆在 SPDE 不确定性量化中的作用尚待分析。

本文面向解析可解的一维线性加性噪声随机热方程, 构建时间统计记忆增强的多输出 PINN 框架。无显式时间记忆模型 B1 作为基线; 均值记忆模型 B2 引入上一时间层均值信息, 用于检验分布中心记忆的作用; 均值-标准差联合记忆模型 B3 引入历史尺度信息, 用于检验尺度记忆对均值传播和分布校准的影响。数值实验以闭式均值、标准差和分位数为基准, 评价均值相对误差、标准差误差、95%覆盖率、QQ-RMSE、后期均值漂移和误差时间波动, 并以归一化 SPDE 残差作为稳定性诊断。为区分统计矩监督与时间统计记忆结构的作用, 进一步设置去除解析统计矩锚定的实验对照组。

2. 方法与理论框架

本节围绕一维线性加性噪声随机热方程建立解析基准和评价指标, 使模型结构、训练约束和结果评价一致。

2.1. 随机热方程与解析统计基准

本小节采用半群、温和解和模态分解建立解析统计基准[2]。考虑定义在 $(0, T] \times (0, 1)$ 区间上的一维线性加性噪声随机热方程。设 $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ 为满足通常条件的完备滤过概率空间, W_t 为其上一维标准 Wiener 过程。模型写为:

$$du(t, x) = \nu u_{xx}(t, x)dt + \sigma \sin(\pi x)dW_t, \quad x \in (0, 1), \quad 0 < t \leq T \quad (2-1)$$

其中, $u(t, x)$ 表示随机温度场, $\nu > 0$ 为扩散系数, $\sigma > 0$ 为噪声强度。满足齐次 Dirichlet 边界和初始条件:

$$u(t, 0) = u(t, 1) = 0, \quad u(0, x) = \sin(\pi x) \quad (2-2)$$

式(2-1)中的 dW_t 表示时间方向的 Wiener 增量, 由于空间区间为 $(0, 1)$ 且边界为齐次 Dirichlet 条件, 随机场

可自然视为 Hilbert 空间 $H = L^2(0,1)$ 中取值的随机过程。令

$$H = L^2(0,1), \quad A = \nu \frac{d^2}{dx^2}, \quad D(A) = H^2(0,1) \cap H_0^1(0,1), \quad g(x) = \sin(\pi x) \quad (2-3)$$

由于初值与噪声空间形状均位于第一正弦模态张成的一维子空间, 方程退化为 Ornstein-Uhlenbeck 型随机微分方程, 积分因子法直接给出显式解, 于是原随机热方程的显式解为:

$$u(t, x) = e^{-\kappa t} \sin(\pi x) + \sigma \sin(\pi x) \int_0^t e^{-\kappa(t-s)} dW_s \quad (2-4)$$

可得到解析均值:

$$\mu(t, x) = \mathbb{E}[u(t, x)] = e^{-\kappa t} \sin(\pi x) \quad (2-5)$$

进一步,

$$\text{Var}[u(t, x)] = \sigma^2 \sin^2(\pi x) \mathbb{E}\left[\left(\int_0^t e^{-\kappa(t-s)} dW_s\right)^2\right] \quad (2-6)$$

$$\mathbb{E}\left[\left(\int_0^t e^{-\kappa(t-s)} dW_s\right)^2\right] = \int_0^t e^{-2\kappa(t-s)} ds = \frac{1 - e^{-2\kappa t}}{2\kappa} \quad (2-7)$$

从而得到解析方差:

$$v(t, x) = \text{Var}[u(t, x)] = \frac{\sigma^2}{2\kappa} (1 - e^{-2\kappa t}) \sin^2(\pi x) \quad (2-8)$$

解析标准差为:

$$s(t, x) = \sqrt{v(t, x)} = \left[\frac{\sigma^2}{2\kappa} (1 - e^{-2\kappa t}) \sin^2(\pi x) \right]^{1/2} \quad (2-9)$$

式(2-4)中的随机项是确定函数对 Wiener 过程的线性 Itô 积分, 因此任意固定时空点 (t, x) 处的解 $u(t, x)$ 服从高斯分布:

$$u(t, x) \sim \mathcal{N}(\mu(t, x), s^2(t, x)) \quad (2-10)$$

相应的理论 95%置信带为:

$$I_{0.95}(t, x) = [\mu(t, x) - 1.96s(t, x), \mu(t, x) + 1.96s(t, x)] \quad (2-11)$$

对任意分位水平, 解析分位数为:

$$q_p(t, x) = \mu(t, x) + z_p s(t, x), \quad z_p = \Phi^{-1}(p) \quad (2-12)$$

其中, Φ 为标准正态分布函数。若测试网格为 $\{(t_i, x_j)\}_{i=1, j=1}^{N_t, N_x}$, 模型预测均值为 $\hat{\mu}$ 、预测标准差为 $\hat{s}$, 则时间层 t_i 上的均值相对误差可定义为:

$$E_\mu(t_i) = \frac{\left(\sum_{j=1}^{N_x} |\hat{\mu}(t_i, x_j) - \mu(t_i, x_j)|^2 \right)^{1/2}}{\left(\sum_{j=1}^{N_x} |\mu(t_i, x_j)|^2 \right)^{1/2} + \varepsilon} \quad (2-13)$$

标准差误差可定义为:

$$E_s(t_i) = \frac{\left(\sum_{j=1}^{N_x} |\hat{s}(t_i, x_j) - s(t_i, x_j)|^2 \right)^{1/2}}{\left(\sum_{j=1}^{N_x} |s(t_i, x_j)|^2 \right)^{1/2} + \varepsilon} \quad (2-14)$$

若模型给出经验 2.5% 与 97.5% 分位数 $\hat{q}_{0.025}$ 和 $\hat{q}_{0.975}$, 则预测区间对真实高斯分布的覆盖率可按式计算:

$$\text{Coverage}_{0.95}(t_i) = \frac{1}{N_x} \sum_{j=1}^{N_x} \left[\Phi \left(\frac{\hat{q}_{0.975}(t_i, x_j) - \mu(t_i, x_j)}{s(t_i, x_j) + \varepsilon} \right) - \Phi \left(\frac{\hat{q}_{0.025}(t_i, x_j) - \mu(t_i, x_j)}{s(t_i, x_j) + \varepsilon} \right) \right] \quad (2-15)$$

式(2-15)表示预测区间对解析高斯分布的覆盖水平。数值接近 0.95 表明区间校准充分, 低于 0.95 表明区间偏窄, 高于 0.95 表明区间偏宽。对分位数结构, 可定义 QQ-RMSE:

$$E_{\text{QQ}}(t_i) = \left[\frac{1}{KN_x} \sum_{\ell=1}^K \sum_{j=1}^{N_x} \left(\hat{q}_{p_\ell}(t_i, x_j) - q_{p_\ell}(t_i, x_j) \right)^2 \right]^{1/2} \quad (2-16)$$

覆盖率和 QQ-RMSE 可以补充均值误差无法反映的概率校准信息。上述指标从均值、标准差、区间覆盖和分位数结构四个方面为数值实验评价提供基准。

2.2. 多输出分布表示

多输出分布表示将同一输入映射为多个通道值, 并把这些通道值视为有限经验样本, 再由样本均值、样本标准差和样本分位数近似真实分布统计量。给定参数 θ , 多输出网络定义为:

$$F_\theta(t, x) = \left(u_\theta^{(1)}(t, x), u_\theta^{(2)}(t, x), \dots, u_\theta^{(M)}(t, x) \right) \in \mathbb{R}^M \quad (2-17)$$

对固定 (t, x) , M 个输出通道诱导经验分布:

$$\hat{\mathbb{P}}_{\theta, t, x} = \frac{1}{M} \sum_{m=1}^M \delta_{u_\theta^{(m)}(t, x)} \quad (2-18)$$

其中, δ_y 表示集中在 y 点的 Dirac 测度。经验均值和经验标准差定义为:

$$\hat{\mu}_\theta(t, x) = \frac{1}{M} \sum_{m=1}^M u_\theta^{(m)}(t, x), \quad \hat{s}_\theta(t, x) = \left[\frac{1}{M} \sum_{m=1}^M \left(u_\theta^{(m)}(t, x) - \hat{\mu}_\theta(t, x) \right)^2 \right]^{1/2} \quad (2-19)$$

经验分位数定义为:

$$\hat{q}_{p, \theta}(t, x) = \text{Quantile}_p \left\{ u_\theta^{(m)}(t, x) \right\}_{m=1}^M \quad (2-20)$$

2.3. 离散随机驱动与残差约束

为使第 m 个输出通道在不同时间层上对应同一条随机驱动, 需要为每个通道指定一条离散 Brownian 路径。设时间网格为 $0 = t_0 < t_1 < \dots < t_N = T$, 步长为 $\Delta t = t_{n+1} - t_n$ 。第 m 个通道的 Wiener 增量定义为:

$$\Delta W_n^{(m)} = W_{t_{n+1}}^{(m)} - W_{t_n}^{(m)}, \quad \Delta W_n^{(m)} \sim \mathcal{N}(0, \Delta t) \quad (2-21)$$

这些增量在不同时间步和不同通道之间相互独立。相应白噪声离散近似为:

$$\xi_n^{(m)} = \frac{\Delta W_n^{(m)}}{\Delta t}, \quad \xi_n^{(m)} \sim \mathcal{N}\left(0, \frac{1}{\Delta t}\right) \quad (2-22)$$

时间方向采用前向差分近似:

$$D_t u_\theta^{(m)}(t_n, x) = \frac{u_\theta^{(m)}(t_{n+1}, x) - u_\theta^{(m)}(t_n, x)}{\Delta t} \quad (2-23)$$

空间二阶导数采用中心差分近似:

$$D_{xx} u_\theta^{(m)}(t_n, x) = \frac{u_\theta^{(m)}(t_n, x+h) - 2u_\theta^{(m)}(t_n, x) + u_\theta^{(m)}(t_n, x-h)}{h^2} \quad (2-24)$$

将随机热方程在时间区间 $[t_n, t_{n+1}]$ 上积分得:

$$u(t_{n+1}, x) - u(t_n, x) \approx \nu u_{xx}(t_n, x) \Delta t + \sigma \sin(\pi x) \Delta W_n \quad (2-25)$$

两边除以步长 Δt 后, 得到第 m 个输出通道的一步离散 SPDE 残差:

$$R_n^{(m)}(x) = D_t u_\theta^{(m)}(t_n, x) - \nu D_{xx} u_\theta^{(m)}(t_n, x) - \sigma \sin(\pi x) \xi_n^{(m)} \quad (2-26)$$

式(2-26)是 Euler-Itô 型离散残差。该残差使用上一时间层的扩散项和同一时间步的 Wiener 增量, 能够与随机方程的离散时间演化保持一致[3]。

2.4. 时间步耦合思想

若多输出网络仅以 (t, x) 为输入, 模型仅学习时空点到经验样本族的边缘映射。对于演化型随机方程, 当前状态由上一时间层状态和当前随机增量共同决定, 因此引入上一时间层统计记忆。对任意阶段模型 r , 定义

$$\hat{\mu}_r^-(t, x) = \hat{\mu}_r(\max(t - \Delta t, 0), x), \quad \hat{s}_r^-(t, x) = \hat{s}_r(\max(t - \Delta t, 0), x) \quad (2-27)$$

其中, 上标 “-” 表示上一时间步统计量。均值记忆提供分布中心的历史信息, 标准差记忆提供不确定性尺度的历史信息。TC-MO-PINN 采用递进结构: B1 为无显式时间记忆的多输出基线; B2 在 B1 基础上引入上一时间层均值记忆; B3 进一步引入上一时间层均值和标准差记忆, 并采用残差式小幅修正。B3-no-std 保持 B3 主体框架, 仅关闭标准差记忆通道, 用于检验该通道的边际作用。

2.5. B1/B2/B3 递进模型结构

2.5.1. B1: 无显式时间记忆的多输出基线

B1 是 MO-PINN 的基线模型。其输入为当前时空坐标, 输出为 M 个经验样本通道:

$$F_{\theta_1}^{B1}(t, x) = (u_{B1}^{(1)}(t, x), \dots, u_{B1}^{(M)}(t, x)) \quad (2-28)$$

B1 将标准确定性 PINN 的单一输出扩展为 M 维经验样本族, 并通过式(2-19)与式(2-20)计算经验均值、标准差和分位数。同时, B1 使用式(2-26)的离散随机残差以及初边值约束, 因此 B1 是带有多输出不确定性表示和物理约束的强基线模型。

2.5.2. B2: 上一时间层均值记忆模型

B2 的目标是在 B1 的基础上显式加入分布中心的历史信息。设 B1 训练后保持固定, 则 B2 的增强输入为:

$$z_{B2}(t, x) = (t, x, \hat{\mu}_{B1}^-(t, x)) \quad (2-29)$$

对应的多输出预测为:

$$F_{\theta_2}^{B2}(t, x) = G_{\theta_2}^{B2}(z_{B2}(t, x)) \quad (2-30)$$

式(2-29)中的 $\hat{\mu}_{B1}^-$ 由训练后的 B1 给出。B2 通过引入 B1 历史均值特征, 使模型在当前时空坐标之外利用上一时间层的分布中心信息, 从而检验均值记忆对动态统计传播的作用。

2.5.3. B3: 均值与标准差联合记忆模型

B3 在 B2 的基础上同时引入上一时间层经验均值和经验标准差, 其增强输入为:

$$z_{B3}(t, x) = (t, x, \hat{\mu}_{B2}^-(t, x), \hat{s}_{B2}^-(t, x)) \quad (2-31)$$

对于线性高斯随机热方程, 固定时空点处的分布由均值和方差决定。式(2-31)同时提供分布中心和分布尺度两类统计状态, 有助于模型区分均值传播误差与不确定性尺度误差。为避免新增输入使模型偏离 B2 已经形成的稳定解结构, B3 在 B2 输出附近学习小幅修正:

$$F_{\theta_3}^{B3}(t, x) = F_{\theta_2}^{B2}(t, x) + \alpha \Delta_{\theta_3}(z_{B3}(t, x)), \quad 0 < \alpha \leq 1 \quad (2-32)$$

其中, Δ_{θ_3} 为待学习的多输出增量, α 为增量尺度因子。为了控制修正幅度, 引入增量正则:

$$\mathcal{L}_{\Delta} = \mathbb{E}_{(t, x)} \left[\frac{1}{M} \sum_{m=1}^M \left| \Delta_{\theta_3}^{(m)}(z_{B3}(t, x)) \right|^2 \right] \quad (2-33)$$

B3 中的历史统计量还可进行有界缩放。设缩放函数为 $\gamma(t)$, 则

$$\tilde{\mu}_{B2}^-(t, x) = \gamma(t) \hat{\mu}_{B2}^-(t, x), \quad \tilde{s}_{B2}^-(t, x) = \gamma(t) \hat{s}_{B2}^-(t, x), \quad 0 \leq \gamma(t) \leq \gamma_{\max} \quad (2-34)$$

关闭标准差记忆通道的 B3-no-std 可写为:

$$z_{B3\text{-no-std}}(t, x) = (t, x, \tilde{\mu}_{B2}^-(t, x), 0) \quad (2-35)$$

除将 $\tilde{s}_{B2}^-$ 标准差记忆通道置为零外, B3-no-std 保持教师模型继承、残差式修正、增量正则、随机驱动残差和评价指标不变。

2.6. 损失函数设计

由于自适应配置点、损失项信息表征和激活函数选择会影响残差训练稳定性。本文在 B1/B2/B3 比较中统一采样、网络规模和主要权重设置, 以减少训练策略差异对时间记忆机制判断的影响。TC-MO-PINN 的训练目标由随机动力学残差、均值方程残差、初边值条件、统计矩校准、时间连续性、蒸馏约束和增量正则组成[9], 总损失为:

$$\mathcal{L} = \lambda_r \mathcal{L}_r + \lambda_{\mu} \mathcal{L}_{\mu} + \lambda_b \mathcal{L}_b + \lambda_i \mathcal{L}_i + \lambda_m \mathcal{L}_m + \lambda_s \mathcal{L}_s + \lambda_d \mathcal{L}_d + \lambda_c \mathcal{L}_c + \lambda_{\Delta} \mathcal{L}_{\Delta} \quad (2-36)$$

其中, 各 λ 为非负权重, 具体地: λ_r 为路径级 SPDE 残差项权重, λ_{μ} 为均值方程残差项权重, λ_i 和 λ_b 分别为初值和边界条件损失权重, λ_m 和 λ_s 分别为均值锚定和标准差锚定权重, λ_d 为蒸馏约束权重, λ_c 为时间连续性约束权重, λ_{Δ} 为增量正则权重。路径级残差 $\mathcal{L}_r$ 基于式(2-26)定义为:

$$\mathcal{L}_r = \frac{1}{N_t N_x M} \sum_{n=0}^{N_t-1} \sum_{j=1}^{N_x} \sum_{m=1}^M \left| R_n^{(m)}(x_j) \right|^2 \quad (2-37)$$

该项约束每个输出通道在对应离散 Brownian 增量驱动下满足随机热方程的一步离散动力学。由于加性噪声具有零均值, 取期望后理论均值满足确定性热方程, 均值残差 $\mathcal{L}_{\mu}$ 定义为:

$$R_{\mu,n}(x) = D_t \hat{\mu}_{\theta}(t_n, x) - \nu D_{xx} \hat{\mu}_{\theta}(t_n, x), \quad \mathcal{L}_{\mu} = \frac{1}{N_t N_x} \sum_{n=0}^{N_t-1} \sum_{j=1}^{N_x} \left| R_{\mu,n}(x_j) \right|^2 \quad (2-38)$$

初值损失 λ_i 和边界损失 λ_b 分别约束式(2-2):

$$\mathcal{L}_i = \frac{1}{N_x M} \sum_{j=1}^{N_x} \sum_{m=1}^M \left| u_{\theta}^{(m)}(0, x_j) - \sin(\pi x_j) \right|^2, \quad \mathcal{L}_b = \frac{1}{N_t M} \sum_{n=1}^{N_t} \sum_{m=1}^M \left(\left| u_{\theta}^{(m)}(t_n, 0) \right|^2 + \left| u_{\theta}^{(m)}(t_n, 1) \right|^2 \right) \quad (2-39)$$

统计矩校准项利用这些解析统计量约束预测分布的中心和尺度:

$$\mathcal{L}_m = \frac{1}{N_t N_x} \sum_{n=1}^{N_t} \sum_{j=1}^{N_x} \left[\left| \hat{\mu}_{\theta}(t_n, x_j) - \mu(t_n, x_j) \right|^2 + \beta_s \left| \hat{s}_{\theta}(t_n, x_j) - s(t_n, x_j) \right|^2 \right] \quad (2-40)$$

其中, 均值锚定项 $\mathcal{L}_m$ 约束预测均值误差 $\sum (\hat{\mu}_{\theta} - \mu)^2$, 标准差锚定项 $\mathcal{L}_s$ 约束预测标准差误差 $\beta_s \sum (\hat{s}_{\theta} - s)^2$ 。

两者同时为解析基准实验提供统一的统计尺度参照, 使不同模型能够在同一分布基准下比较均值精度和不确定性校准。

时间连续性项 $\mathcal{L}_c$ 用于约束相邻时间层统计量的变化幅度:

$$\mathcal{L}_c = \frac{1}{(N_t - 1)N_x} \sum_{n=0}^{N_t-2} \sum_{j=1}^{N_x} \left| \hat{\mu}_\theta(t_{n+1}, x_j) - \hat{\mu}_\theta(t_n, x_j) \right|^2 \quad (2-41)$$

教师 - 学生蒸馏项用于保持学生模型不偏离教师模型的稳定解结构[10]:

$$\mathcal{L}_d = \mathbb{E}_{(t,x)} \left[\frac{1}{M} \sum_{m=1}^M \left| u_{\text{student}}^{(m)}(t, x) - \text{stopgrad} \left(u_{\text{teacher}}^{(m)}(t, x) \right) \right|^2 \right] \quad (2-42)$$

其中, $\text{stopgrad}(\cdot)$ 表示教师模型输出在学生模型训练中仅作为固定参考, 不参与参数更新。B1 主要依赖路径级残差 $\mathcal{L}_r$ 、均值残差 $\mathcal{L}_\mu$ 、初边值条件 λ_r, λ_b 和统计矩校准 $\mathcal{L}_m, \mathcal{L}_s$; B2 在此基础上加入上一时间层均值记忆和 B1 教师约束; B3 进一步加入上一时间层标准差记忆、残差式修正、增量正则 $\mathcal{L}_\Delta$ 和 B2 教师约束。

2.7. 评价指标与实验落地

数值实验的评价由上述指标得到。数值实验进一步报告后期均值漂移和误差时间波动两个辅助指标, 用于刻画时间推进过程中的误差稳定性。后期均值漂移定义为终止时间前一段后期时间层上均值相对误差的平均水平, 本文取 $t \geq 0.8T$ 的时间层作为后期区间。

$$E_{\text{late}}^\mu = \frac{1}{|\mathcal{T}_{\text{late}}|} \sum_{t_n \in \mathcal{T}_{\text{late}}} E_\mu^{\text{rel}}(t_n), \quad \mathcal{T}_{\text{late}} = \{t_n : t_n \geq 0.8T\} \quad (2-43)$$

该指标度量均值预测误差在长时间区间内的持续偏移, 数值越小, 表示分布中心在后期传播中越稳定。误差时间波动定义为全时间区间内均值相对误差序列的标准差, 用于衡量误差沿时间层推进时的起伏幅度。

$$\text{Std}_t(E_\mu^{\text{rel}}) = \left[\frac{1}{N_t} \sum_{n=1}^{N_t} \left(E_\mu^{\text{rel}}(t_n) - \bar{E}_\mu^{\text{rel}} \right)^2 \right]^{1/2}, \quad \text{TV}_t(E_\mu^{\text{rel}}) = \sum_{n=1}^{N_t-1} |E_\mu^{\text{rel}}(t_{n+1}) - E_\mu^{\text{rel}}(t_n)| \quad (2-44)$$

该指标越小, 模型在不同时间层上的均值误差更平稳。该指标与后期均值漂移结合, 可同时反映长时间区间的持续偏差和全时间区间的误差波动。路径级物理一致性由归一化 SPDE 残差评估。设时间层 t_n 上的残差均方根为:

$$\text{RMS}_R(t_n) = \left[\frac{1}{MN_x} \sum_{m=1}^M \sum_{j=1}^{N_x} \left| R_n^{(m)}(x_j) \right|^2 \right]^{1/2} \quad (2-45)$$

由于白噪声近似 $\xi_n^{(m)}$ 的标准差为 $1/\sqrt{\Delta t}$, 随机强迫项本身具有自然尺度。定义

$$R_{\text{scale}} = \sigma \left(\frac{1}{\Delta t} \right)^{1/2} \left(\frac{1}{N_x} \sum_{j=1}^{N_x} \sin^2(\pi x_j) \right)^{1/2} \quad (2-46)$$

归一化 SPDE 残差为:

$$E_R(t_n) = \frac{\text{RMS}_R(t_n)}{R_{\text{scale}} + \varepsilon} \quad (2-47)$$

式(2-47)将绝对残差转换为相对于随机强迫尺度的相对残差, 用于判断路径级残差是否处于合理范围。

3. 数值实验与机制验证

数值实验基于前文建立的解析基准, 对 TC-MO-PINN 在动态不确定性量化中的作用进行验证。

3.1. 建模实现与数值训练流程

3.1.1. 解析基准与离散随机驱动

数值实验基于第 2 节建立的解析基准, 实验参数取 $T=1$, $\nu=0.1$, $\sigma=0.3$ 。随机噪声采用固定离散 Brownian 路径表, 随机噪声采用固定离散 Brownian 路径表, 路径级物理约束按式(2-21)至(2-26)构建。

3.1.2. 多输出经验分布与训练约束

多输出模型在任意时空点给出 M 个通道值, 并将其解释为随机解的经验样本族。本研究主实验取 $M=300$ 。对应的经验均值和经验标准差为:

$$\hat{\mu}(t,x)=\frac{1}{M}\sum_{j=1}^Mu^{(j)}(t,x), \quad \hat{s}(t,x)=\left[\frac{1}{M}\sum_{j=1}^M\left(u^{(j)}(t,x)-\hat{\mu}(t,x)\right)^2\right]^{1/2} \tag{3-1}$$

网络结构中基础映射采用全连接前馈神经网络, 隐藏层由三层宽度为 128 的非线性层组成, 激活函数采用 tanh。B1 的输入为当前时空坐标 (t,x) , 输出为 M 个样本通道; B2 在 B1 的基础上增加上一时间层经验均值作为历史统计特征; B3 进一步增加上一时间层经验标准差, 并采用教师模型输出加小幅残差修正的形式。输入变量在进入网络前进行尺度归一化, 以减小不同时间、空间尺度对优化的影响。B2 和 B3 的新增输入通道在初始化时不破坏教师模型已有映射, 使递进训练能够从稳定基线出发, 再逐步学习时间统计记忆带来的修正。

3.2. 实验组设置与对照逻辑

实验分三层展开: 第一层比较 B1、B2 和 B3, 检验模型由无显式时间记忆、均值记忆到均值 - 标准差联合记忆的递进效果; 第二层比较 B3 与 B3-no-std, 在保持其他训练机制一致的条件下关闭标准差记忆通道, 评估其边际贡献; 第三层比较是否引入解析统计矩锚定, 考察解析均值与标准差监督对动态不确定性量化指标的影响。详细实验机制对比见表 1。

Table 1. Comparison of experimental group settings
表 1. 实验组条件对照

实验对象	输入或训练设置	核心机制	实验目标
B1	(t,x) 当前时空坐标	多输出残差 + 统计矩锚定	建立无显式时间记忆的多输出基线
B2	(t,x) 当前时空坐标 + 前层均值	均值记忆、教师模型继承	检验分布中心历史信息对动态传播的作用
B3	(t,x) 当前时空坐标 + 前层均值和 std	均值-std 联合记忆 + 残差修正	检验不确定性尺度记忆对均值传播和分布校准的作用
B3-no-std	(t,x) 当前时空坐标 + 前层均值, std = 0	保持 B3 主要机制, 关闭 std 记忆	检验标准差记忆通道的边际贡献
去除解析统计矩锚定	关闭解析均值/std 拟合约束	保留物理残差和时间统计记忆	检验训练阶段解析统计矩监督对动态 UQ 指标的影响

3.3. 实验结果与动态不确定性量化机制分析

3.3.1. 量化指标对比

最大均值相对误差衡量模型在全时间区间内对随机解分布中心的最大预测偏差。B2 与 B3 相较 B1 均呈递减趋势(B1 → B2 → B3), 表明上一时间层统计记忆有助于抑制均值传播漂移, B3 控制效果最稳定(详见表 2)。平均标准差误差与最大点态标准差误差用于评价模型对随机扰动累积尺度的恢复能力。B2 与

B3 均低于 B1，说明时间统计记忆可降低不确定性尺度恢复的整体偏差。最低 95%覆盖率反映预测区间在全时间区间内的覆盖下限。B3 相较 B1 有所提升，递进模型减小了低覆盖偏差。最大 QQ-RMSE 衡量经验分布分位数结构与解析高斯分布之间的偏差。B3 与 B1 相近，表明降低均值误差的同时保持了经验分布分位数结构的稳定性。

Table 2. Comparison of indicators B1, B2 and B3
表 2. B1、B2 与 B3 指标对比

模型	最大均值相对误差	平均标准差误差	最大点态标准差误差	最低 95%覆盖率	最大 QQ-RMSE
B1	0.002847	0.003214	0.012640	0.936036	0.011977
B2	0.002114	0.002904	0.010747	0.943421	0.011680
B3	0.001903	0.002893	0.010741	0.943434	0.011680

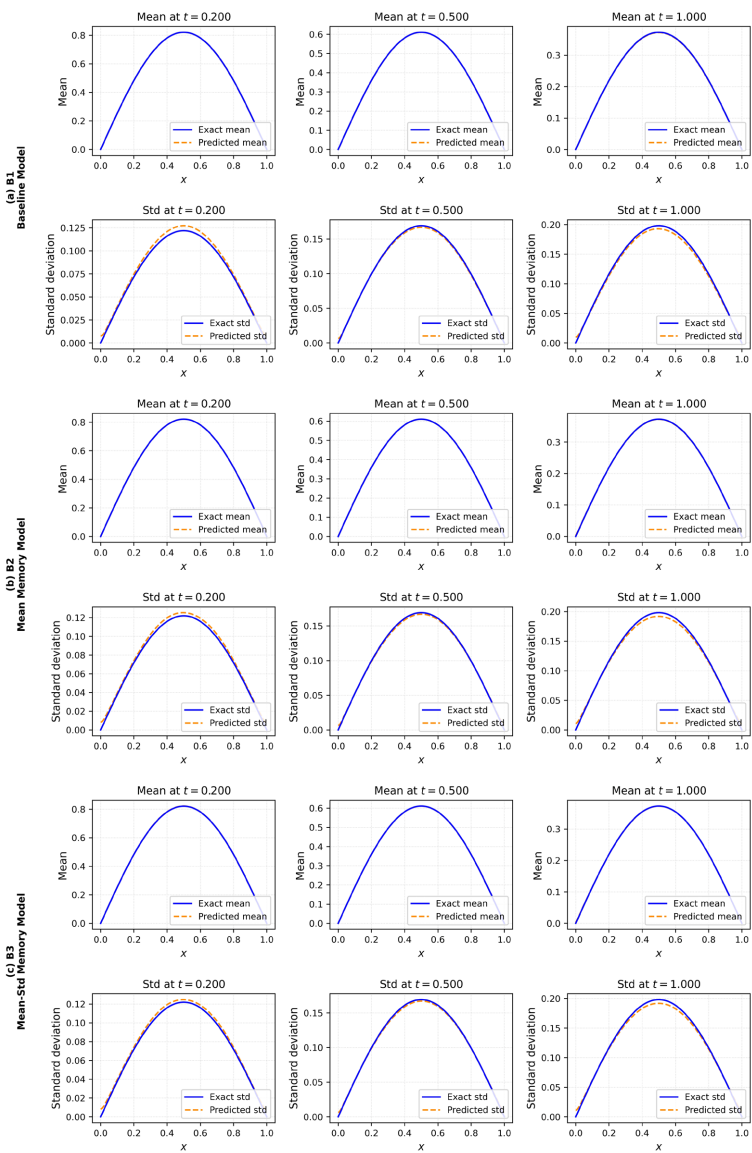


Figure 1. Comparison of mean and standard deviation curves of B1, B2 and B3 at representative time slices
图 1. B1、B2 与 B3 在典型时间片上的均值和标准差曲线对比

图 1 说明多输出样本可以有效表示随机解的主要分布形态。随着时间统计记忆的引入, 预测曲线与解析曲线更加接近: B1 在后期时间层的均值曲线偏差更明显, B2 引入上一时间层均值后减弱了分布中心漂移, B3 在保持分布校准水平的同时进一步改善均值传播稳定性。

3.3.2. 误差传播稳定性分析

为进一步说明时间步统计记忆对动态传播的作用, 表 3 给出了后期均值误差和误差时间波动指标。这些指标直接检验模型误差是否随时间推进而累积或波动。

Table 3. Comparison of error propagation stability metrics
表 3. 误差传播稳定性指标对比

模型	后期均值相对误差	均值误差时间标准差	均值误差总变差	后期 QQ-RMSE
B1	0.002225	0.000688	0.004662	0.007318
B2	0.001404	0.000426	0.003278	0.007024
B3	0.001209	0.000343	0.004052	0.007013

后期均值相对误差方面, B3 在 $t \geq 0.8T$ 区间低于 B1, 说明联合统计记忆改善了时间传播后期的误差累积。均值误差总变差方面, B2 略低于 B3; B3 引入标准差记忆后前期存在局部扰动, 但后期时间层的均值误差和时间波动均优于 B2, 表明 B3 的主要收益集中于后期传播稳定性。

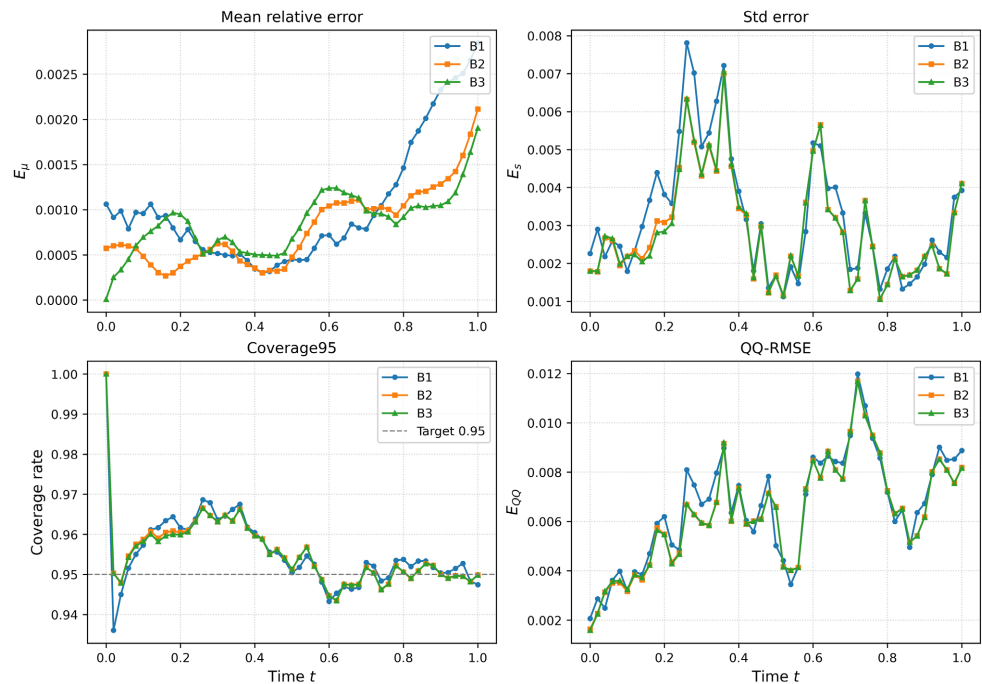


Figure 2. Temporal evolution of core error and distribution calibration metrics for B1, B2 and B3
图 2. B1、B2 与 B3 核心误差和分布校准指标的时间演化

图 2 中均值相对误差曲线表明, B1 在后期时间层出现更明显上升, 而 B2 与 B3 将后期误差保持在较低水平; 标准差误差曲线显示, 时间步统计记忆降低了不确定性尺度恢复中的整体偏差; 覆盖率曲线显示, B2 与 B3 在最低覆盖率附近更接近理论覆盖水平 0.95; QQ-RMSE 曲线说明 B3 在改善均值传播稳定性的同时保持了分位数结构。

3.4. B3-No-Std 消融实验

B3-no-std 消融模型用于隔离标准差记忆通道的边际作用。该模型保持 B3 的教师模型继承、教师模型残差修正、增量正则、蒸馏约束和连续性约束不变，仅将上一时间层标准差记忆通道置零。因此，B3 与 B3-no-std 的差异直接反映标准差记忆输入对结果的影响。

Table 4. Ablation results comparison between B3 and B3-no-std

表 4. B3 与 B3-no-std 消融结果对比

模型	最大均值相对误差	后期均值相对误差	均值误差时间标准差	平均标准差误差	最低 95% 覆盖率	最大 QQ-RMSE
B3	0.001903	0.001209	0.000343	0.002893	0.943434	0.011680
B3-no-std	0.002155	0.001344	0.000391	0.002867	0.943433	0.011673

表 4 显示，关闭标准差记忆后，B3-no-std 的最大均值相对误差从 B3 的 0.001903 升至 0.002155，后期均值相对误差由 0.001209 升至 0.001344，均值误差时间标准差由 0.000343 升至 0.000391。该结果表明，标准差记忆有助于 B3 在均值误差和后期漂移控制方面保持更稳定表现。同时，B3-no-std 在平均标准差误差和最大 QQ-RMSE 上与 B3 接近，部分指标略低。该结果表明，标准差记忆的主要贡献体现在维持分布校准水平的同时改善均值传播稳定性。B3-no-std 消融进一步明确了主模型的机制来源：时间步统计记忆的有效性首先体现在 B2 相对 B1 的整体改善，B3 则在保持分布校准水平的同时进一步压低最大均值误差和后期漂移。

3.5. 解析统计矩锚定的作用与局限

为检验训练阶段是否依赖解析统计矩监督，设置去除解析统计矩锚定的对照实验。该对照关闭训练目标中的解析均值拟合、解析标准差拟合、末端均值锚定和覆盖率约束，仅保留路径级 SPDE 残差、均值方程残差、初边值条件、教师继承、连续性约束和时间统计记忆。

Table 5. Results of the control experiment of analytical statistical moment anchoring

表 5. 解析统计矩锚定对照实验结果

训练设置	模型	最大均值相对误差	平均标准差误差	最低 95%覆盖率	最大 QQ-RMSE
含解析统计矩锚定	B1	0.002847	0.003214	0.936036	0.011977
含解析统计矩锚定	B3	0.001903	0.002893	0.943434	0.011680
去除解析统计矩锚定	B1	0.003545	0.003382	0.939226	0.010934
去除解析统计矩锚定	B3	0.001579	0.003051	0.933429	0.010974

表 5 显示，去除解析统计矩锚定后，B3 的最大均值相对误差为 0.001579，低于 B1 的 0.003545，表明时间统计记忆增强结构在无解析统计矩监督条件下仍能改善均值传播误差。B3 的平均标准差误差由含锚定设置下的 0.002893 变为 0.003051，最低 95%覆盖率由 0.943434 变为 0.933429，表明解析统计矩锚定主要增强概率区间下界稳定性和尺度校准质量。

图 3 中均值相对误差曲线表明，B1 在去除解析统计矩锚定后后期误差持续累积，至 $t = 1.0$ 时约达 0.0028，而 B2 与 B3-no-std 在整个时间区间内维持更平缓的误差轨迹，终端误差约为 0.0021。结合表 5，去除锚定后 B1 最大均值相对误差提升 24.5%，而具有时间记忆的模型均值误差未见明显上升，说明时间统计记忆通过在时间步间传递历史统计量作为软约束，在无外部解析监督条件下仍能抑制均值误差的时序累积。同时，B2 与 B3-no-std 在前期($t < 0.6$)误差相近，后期出现一定分化，标准差记忆通道对后期误

差控制有辅助作用, 但均值记忆本身已足以在去锚定场景下产生显著正则化效果。上述结果表明, 解析统计矩锚定主要改善最低 95%覆盖率和平均标准误差, 而时间统计记忆对均值传播稳定性的贡献在关闭外部统计矩监督后依然可观测。

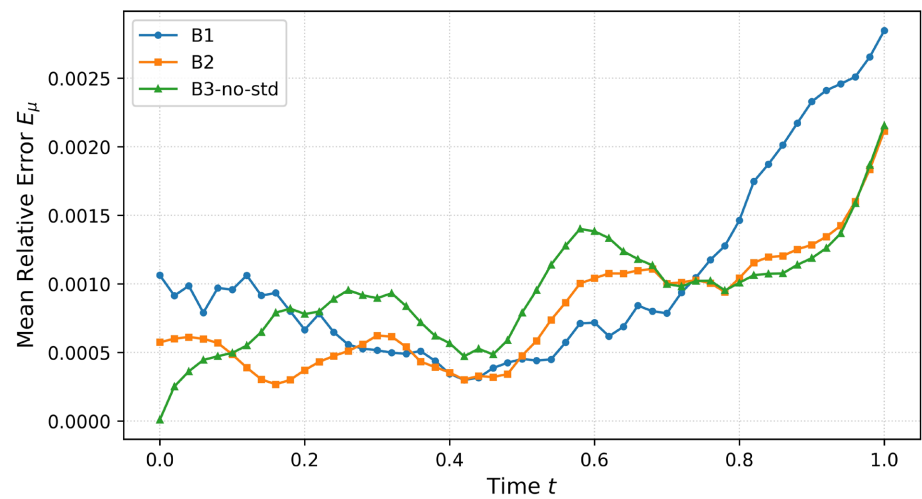


Figure 3. Mean error evolution without analytical statistical moment anchoring
图 3. 无解析统计矩锚定下的均值误差演化

对于路径级稳定性, 归一化 SPDE 残差诊断显示, 各模型最大残差均处于同一数量级, 说明递进统计记忆未引入显著的路径级不稳定性。表 6 显示, B2/B3 的残差方差高于 B1, 反映时间统计记忆增强了多输出通道在随机增量驱动下的路径分散程度, 说明统计记忆输入在提升分布中心和尺度恢复精度的同时保持了路径级数值稳定性。

Table 6. Supplementary diagnosis of normalized SPDE residuals
表 6. 归一化 SPDE 残差补充诊断

模型	残差方差诊断	最大残差诊断
B1	0.035188	4.975340
B2	0.289935	5.002459
B3	0.289861	5.001751
B3-no-std	0.289897	5.002505

4. 讨论与总结

本文将时间相关 SPDE 的动态不确定性量化分解为经验分布表示、时间统计记忆输入与路径级物理约束三个环节, 在具有解析参照的随机热方程上完成闭环验证。实验结果表明, 引入时间统计记忆的 B2 与 B3 相较基线 B1 均降低最大均值相对误差, B3 在均值传播稳定性和后期漂移控制上优势更显著; 消融实验显示, 标准差记忆通道形成更稳定的结构约束, 去除解析统计矩锚定后 B3 仍保持均值传播改进, 表明时间统计记忆具有独立贡献; 归一化 SPDE 残差诊断确认递进模型未引入显著路径级数值不稳定性。

RNN/LSTM 等循环结构将历史信息压缩为高维隐状态[11], 物理含义由数据驱动学习决定, 适合时间依赖复杂、非马尔可夫效应显著的问题, 但可解释性较弱[12]。TC-MO-PINN 直接将上一时间层的均值场与标准差场作为显式记忆输入, 记忆通道的物理含义与 UQ 评价指标对应, 作用可通过消融实验独立

分解, 更适合作为机制可解析、诊断可控的 SPDE 不确定性量化基准框架。

综上, 本文选择显式统计记忆, 是面向具有可靠统计参照的基准场景, 构建结构清晰、作用可分解的 SPDE 不确定性量化框架。随机热方程实验表明, 将前一时间层的分布中心与不确定性尺度作为模型输入, 可提升演化型 SPDE 中均值传播的稳定性, 并改善若干边缘分布量化指标。

参考文献

- [1] Da Prato, G. and Zabczyk, J. (2014) Stochastic Equations in Infinite Dimensions. 2nd Edition, Cambridge University Press. <https://doi.org/10.1017/cbo9781107295513>
- [2] Prévôt, C. and Röckner, M. (2007) A Concise Course on Stochastic Partial Differential Equations. Springer.
- [3] Lord, G.J., Powell, C.E. and Shardlow, T. (2014) An Introduction to Computational Stochastic PDEs. Cambridge University Press. <https://doi.org/10.1017/cbo9781139017329>
- [4] Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019) Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *Journal of Computational Physics*, **378**, 686-707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- [5] Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. and Yang, L. (2021) Physics-Informed Machine Learning. *Nature Reviews Physics*, **3**, 422-440. <https://doi.org/10.1038/s42254-021-00314-5>
- [6] Yang, L., Meng, X. and Karniadakis, G.E. (2021) B-PINNs: Bayesian Physics-Informed Neural Networks for Forward and Inverse PDE Problems with Noisy Data. *Journal of Computational Physics*, **425**, Article 109913. <https://doi.org/10.1016/j.jcp.2020.109913>
- [7] Yang, M. and Foster, J.T. (2022) Multi-Output Physics-Informed Neural Networks for Forward and Inverse PDE Problems with Uncertainties. *Computer Methods in Applied Mechanics and Engineering*, **402**, Article 115041. <https://doi.org/10.1016/j.cma.2022.115041>
- [8] Zhu, Y. and Zabarar, N. (2018) Bayesian Deep Convolutional Encoder-Decoder Networks for Surrogate Modeling and Uncertainty Quantification. *Journal of Computational Physics*, **366**, 415-447. <https://doi.org/10.1016/j.jcp.2018.04.018>
- [9] Cipolla, R., Gal, Y. and Kendall, A. (2018) Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7482-7491. <https://doi.org/10.1109/cvpr.2018.00781>
- [10] Gou, J., Yu, B., Maybank, S.J. and Tao, D. (2021) Knowledge Distillation: A Survey. *International Journal of Computer Vision*, **129**, 1789-1819. <https://doi.org/10.1007/s11263-021-01453-z>
- [11] Rao, C., Ren, P., Wang, Q., Buyukozturk, O., Sun, H. and Liu, Y. (2023) Encoding Physics to Learn Reaction-Diffusion Processes. *Nature Machine Intelligence*, **5**, 765-779. <https://doi.org/10.1038/s42256-023-00685-7>
- [12] Chen, Y., Dong, B. and Xu, J. (2021) Meta-MgNet: Meta Multigrid Networks for Solving Parameterized Partial Differential Equations. *Journal of Computational Physics*, **455**, Article ID: 110996. <https://doi.org/10.1016/j.jcp.2022.110996>