面向神经偏微分方程代理模型直接学习校准感知不确定性

Carlos Stein Brito NightCity Labs，葡萄牙里斯本 carlos.stein@nightcitylabs.ai

arXiv:2602.11090v1 [cs.LG] 11 Feb 2026

摘要

神经偏微分方程代理模型常部署于数据有限或部分观测的场景中，在这些场景下，下游决策除了依赖较低的预测误差，也依赖经过校准的不确定性。现有方法通过集成重复训练、固定的随机噪声（如 dropout）或事后校准来获得不确定性。交叉正则化不确定性通过将梯度经由留出的正则化划分进行传递，在训练过程中学习不确定性参数。预测器在训练划分上针对拟合效果进行优化，而低维不确定性控制参数则在正则化划分上进行优化，以减少训练-测试失配，从而获得适应不同场景的不确定性，而无需针对每种场景进行噪声调优。该框架可以学习连续的噪声水平，位置包括输出头、隐藏特征内部，或谱模态等算子特定组件内部。我们在傅里叶神经算子中实例化这一方法，并在 APEBench 上针对观测比例和训练集大小进行参数扫描。在这些参数扫描中，学习得到的预测分布在留出划分上校准得更好；单步空间诊断还表明，最终的不确定性场集中在高误差区域。

1 引言

神经算子为快速偏微分方程代理建模提供了一条实用路径，并已成为数据驱动科学模拟的标准骨干网络（Li 等人，2021；Lu 等人，2021；Takamoto 等人，2022）。然而，在许多下游场景中，仅有原始预测精度并不足够：决策需要在分布偏移下数值稳定且考虑校准的不确定性估计。

主流基线方法包括 MC Dropout、深度集成（Gal & Ghahramani，2016年；Lakshminarayanan 等人，2017年）以及事后保形包装器（Angelopoulos & Bates，2021年；Ma 等人，2024年）。虽然这些方法提供了重要的基线，但它们缺乏直接机制，无法针对驱动神经偏微分方程流程中过拟合的训练-测试失配来优化不确定性。因此，固定的随机设置往往无法适应特定情境下的数据稀缺或稀疏性。

我们将不确定性表述为一个可学习的优化目标，并使其与正则化目标直接耦合。具体而言，我们保留一个预测不确定性分支，通过似然目标在训练划分上进行优化；此外，还引入一个独立的泛化不确定性分支，其参数**仅**依据在正则化划分上计算的梯度进行更新。由此，我们将交叉正则化（XReg）原则（Brito，2025）引入神经偏微分方程代理模型，将重点从噪声在架构中的放置方式转向目标路由学习。与传统的超参数调优不同，XReg 无需重新拟合预测器主干网络，即可利用正则化集在线更新不确定性参数。在全文中，“预测”不确定性和“泛化”不确定性是根据这种梯度路由定义的操作性标签，而非对物理过程的分解。我们的贡献如下：

• **神经偏微分方程代理模型的交叉正则化不确定性学习**。 我们将参数划分为预测器和不确定性控制参数，并通过训练/正则化划分对其进行训练：预测器参数

在训练划分上更新，而泛化不确定性控制项则根据正则化划分梯度在线更新（算法 1）。

• **基于留出梯度的自适应不确定性缩放**。正则化更新会根据在正则化划分上估计的训练-测试失配，直接调整不确定性幅度，从而在观测稀疏或数据不足时提高不确定性，并在监督信号更具信息量时降低不确定性。

• **实证校准与定位证据**。在 APEBench 上针对观测比例和训练集大小进行的参数扫描中，与采用匹配协议的 MC Dropout 和 3 成员集成相比，学习得到的预测分布改善了留出数据上的高斯混合校准；单步空间诊断还表明，不确定性集中在高误差区域。

2 相关工作

用于偏微分方程的神经算子代理模型。DeepONet 和傅里叶神经算子等算子学习架构，如今已成为偏微分方程代理建模的标准基线（Lu 等人，2021；Li 等，2021）。基准测试工作进一步阐明了这些方法的优势，以及它们在不同任务和分辨率下所面临的校准与泛化挑战（Takamoto 等人，2022）。

深度模型中的不确定性基线。常用的不确定性基线包括 MC Dropout 和深度集成（Gal & Ghahramani，2016年；Lakshminarayanan 等，2017年）。保形方法可提供有限样本校准保证，近期也已应用于算子学习场景（Angelopoulos & Bates，2021年；Ma 等，2024年）。我们关注的是训练集与正则化集目标之间的自适应耦合，它可以调整不确定性尺度，以适应特定状态失配。

验证梯度复杂度控制。先前关于交叉正则化（XReg）的研究提出直接利用留出梯度来学习正则化控制（Brito，2025）。我们将这一原理用于神经算子，在单个优化循环中分离预测不确定性与泛化不确定性。

3 XReg 入门

交叉正则化（XReg）（Brito，2025）通过将特征拟合与复杂度控制的梯度分别导向不同的数据划分，使二者相互分离。我们将不确定性分解为两个可操作的类别：预测不确定性，它吸收训练拟合所需的残差；以及泛化不确定性，它通过平滑模型来减少训练–留出失配。这种分解是相对于模型而言的，而非基于机制：预测不确定性在吸收残差变异（包括近似误差和优化误差）方面类似于偶然不确定性，而泛化不确定性则反映由证据有限所引发的认知不确定性。预测不确定性与泛化不确定性之间的区别，完全源于优化信号，而非来自不同的概率架构。

3.1 交叉正则化优化划分

将模型参数划分为确定性骨干参数 θ、预测噪声参数 ψ 和泛化噪声参数 ρ。XReg 采用交替更新：

$$(\theta,\psi)_{t+1} = (\theta,\psi)_t - \eta_{\theta,\psi}\nabla_{(\theta,\psi)}\mathcal{L}_{\text{train}}((\theta,\psi)_t,\rho_t), \tag{1}$$

$$\rho_{t+1} = \rho_t - \eta_\rho \nabla_\rho \mathcal{L}_{\text{reg}}((\theta,\psi)_{t+1},\rho_t). \tag{2}$$

4 交叉正则化不确定性框架

4.1 与骨干网络无关的表述

考虑一个通用的神经算子骨干网络 f_θ。我们为其增加两组随机参数集：

$$\psi \in \Psi \quad \text{(predictive-noise parameters)}, \qquad \rho \in \mathcal{R} \quad \text{(generalization-noise parameters)}.$$

对于输入 x，模型会生成预测矩

$$\mu_{\theta,\psi,\rho}(x), \qquad \sigma_{\text{pred},\psi}(x),$$

其中 ρ 可能通过内部特征和/或输出层的随机扰动影响输出分布。

4.2 噪声可注入的位置

该框架由训练/正则化优化划分所定义，因此不确定性路径可附加到模型的多个位置。本文重点关注输出头与内部特征实例化，并可选择将预测噪声与泛化噪声共置，同时通过梯度路由使二者保持区分。

对于本文使用的内部变体，随机性通过对所选潜在特征块以及（启用时）所选傅里叶模态通道施加乘性高斯分布扰动来实现，使用$h \mapsto h \odot(1+\sigma\epsilon)$。这种放置灵活性不需要贝叶斯权重后验，也避免将偶然不确定性与认知不确定性承诺于单一架构代理。

4.3 训练与正则化目标

训练目标。 训练更新在泛化采样下优化边际（混合）似然的蒙特卡洛近似：

$$\mathcal{L}_{\text{train}} = -\log \frac{1}{S}\sum_{s=1}^{S} p_{\text{hier}}(y \mid x, \omega_s),$$

with

$$p_{\text{hier}} = \exp\left(-\frac{1}{2}\left[\log(2\pi\sigma_{\text{pred}}^2) + \frac{(y-\mu)^2+\sigma_{\text{gen}}^2}{\sigma_{\text{pred}}^2}\right]\right).$$

在内部噪声变体中，σ_{gen} 隐含在采样得到的模型实例中，并通过评估时的 $\mathrm{Var}_\omega[\mu]$ 来衡量泛化贡献。训练更新优化（θ，ψ）；正则化参数 ρ仅通过 $\mathcal{L}_{\text{reg}}$进行优化。

正则化目标。正则化更新基于留出的一步样本对进行。本文使用两种估计量：

1. **混合负对数似然**:

$$\mathcal{L}_{\text{reg}}^{\text{mix}} = -\log \frac{1}{S}\sum_{s=1}^{S} p(y \mid \mu_s, \sigma_s^2),$$

   其中，每个蒙特卡洛样本对应一个高斯分布。

2. **矩匹配负对数似然**:

$$\mu_{\text{mm}} = \mathbb{E}[\mu_s], \quad \sigma_{\text{mm}}^2 = \mathbb{E}[\sigma_s^2+\mu_s^2] - \mu_{\text{mm}}^2,$$
$$\mathcal{L}_{\text{reg}}^{\text{mm}} = -\log p(y \mid \mu_{\text{mm}}, \sigma_{\text{mm}}^2).$$

仅输出头特殊情形（预测分支和泛化分支位于同一模型位置）。当两个分支都是输出头（没有内部层随机采样）时，这种不对称性显而易见：

$$\mathcal{L}_{\text{train}} = \frac{1}{2}\left[\log(2\pi\sigma_{\text{pred}}^2) \; + \; \frac{(y-\mu)^2+\sigma_{\text{gen}}^2}{\sigma_{\text{pred}}^2}\right],$$

而正则化目标使用总方差

$$\sigma_{\text{tot}}^2 = \sigma_{\text{pred}}^2 + \sigma_{\text{gen}}^2, \qquad \mathcal{L}_{\text{reg}} = \frac{1}{2}\left[\log(2\pi\sigma_{\text{tot}}^2) + \frac{(y-\mu)^2}{\sigma_{\text{tot}}^2}\right]$$

（或在启用内部随机噪声时使用其蒙特卡洛混合对应项）。这种目标函数与数据路由的不对称性，即使预测参数与泛化参数都位于输出头，也能产生不同的学习动态。

算法 1：交叉正则化双噪声训练

输入： $\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{reg}}, T, k_{\text{reg}}, \eta_\theta, \eta_\psi, \eta_\rho, S$。参数：骨干网络 θ、预测噪声参数 ψ、泛化噪声参数 ρ。输出：训练后的 (θ, ψ, ρ)。步骤 1：初始化 (θ, ψ, ρ)。步骤 2：对于 $t = 1, \ldots, T$，重复执行：(a) 采样训练小批量 $B_{\text{train}} \sim \mathcal{D}_{\text{train}}$。(b) 采样模型实例 $\omega_{1:S} \sim q_\rho$ 并计算 $\mathcal{L}_{\text{train}}(B_{\text{train}}; \theta, \psi, \omega_{1:S})$。(c) 更新 (d) 如果 $t \bmod k_{\text{reg}} = 0$：(i) 采样正则化小批量 $B_{\text{reg}} \sim \mathcal{D}_{\text{reg}}$。(ii) 采样 $\omega_{1:S} \sim q_\rho$ 并评估 $\mathcal{L}_{\text{reg}}(B_{\text{reg}}; \theta, \psi, \rho, \omega_{1:S})$。(iii) 更新$\rho \leftarrow \rho - \eta_\rho \nabla_\rho \mathcal{L}_{\text{reg}}$.步骤 3：返回 (θ, ψ, ρ)。

$$(\theta, \psi) \leftarrow (\theta, \psi) - \eta_{\theta,\psi} \nabla_{(\theta,\psi)} \mathcal{L}_{\text{train}}.$$

图1：交叉正则化双噪声训练。训练更新在 $\mathcal{D}_{\text{train}}$ 上优化 $(\theta,\ \psi)$，周期性正则化更新则在 $\mathcal{D}_{\text{reg}}$ 上优化 ρ。

4.4 算法

每次迭代都在 $\mathcal{D}_{\text{train}}$ 上更新 $(\theta,\ \psi)$；每隔 k_{reg} 步，第二轮仅在 $\mathcal{D}_{\text{reg}}$ 上更新 ρ。这里，q_ρ 表示由 ρ 诱导的随机模型族，$\omega \sim q_\rho$ 表示采样得到的一个模型实例。在仅输出头变体中，ω 对应于采样得到的输出层泛化噪声实现；在内部噪声变体中，ω 对应于选定潜在特征块中的乘性扰动。在我们的主要实验中，$k_{\text{reg}} = 5$；实际开销及其对 k_{reg}的依赖关系总结于附录A.7，而降低正则化更新的频率可使训练成本接近标准训练（Brito，2025）。

4.5 本文中的实例化

在实证研究中，我们使用一个一维傅里叶神经算子骨干网络（Li etal.,2021）来实例化该框架。模型输出

$$\mu_\theta(x_t), \quad \log \sigma_{\text{pred},\psi}(x_t), \quad \log \sigma_{\text{gen},\rho}(x_t),$$

其中采用了输出头和内部层变体，并用于消融实验和参数扫描。

4.6 评估时的不确定性分解

在推理时，我们将不确定性分为预测项和泛化项：

$$U_{\text{pred}}(x) = \sigma^2_{\text{pred}}(x), \tag{3}$$

$$U_{\text{gen}}(x) = \text{Var}_{\omega \sim q_\rho} \left[\mu(x; \omega) \right], \tag{4}$$

$$U_{\text{tot}}(x) \approx U_{\text{pred}}(x) + U_{\text{gen}}(x), \tag{5}$$

其中，U_{gen} 在本文中有两种实现情况：

$$\text{head-only: } U_{\text{gen}}(x) = \sigma^2_{\text{gen}}(x), \qquad \text{internal stochastic: } U_{\text{gen}}(x) = \text{Var}_\omega[\mu(x; \omega)] \,.$$

我们将单步似然和高斯混合校准作为主要定量指标，并辅以空间定位诊断。

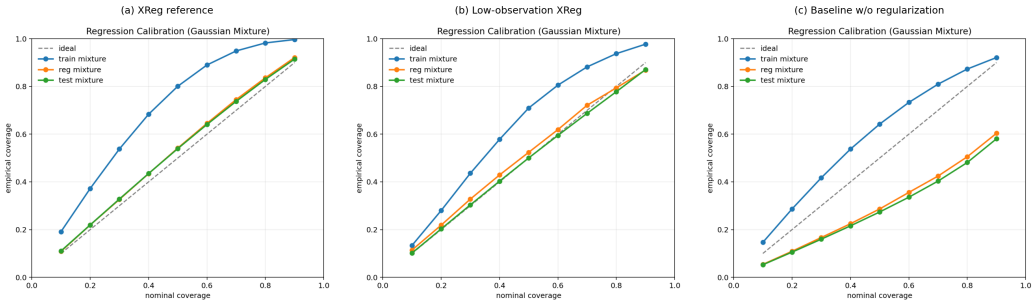


图 2：三个代表性设定下的回归混合校准。左：采用矩匹配正则化的 XReg参考设定。中：低观测XReg设定（观测到 30% 的数据）。右：未采用可学习正则化的模型。在所有面板中，XReg 在留出划分上的表现都更接近对角线，且在观测有限时提升最大。

5 实验设置

5.1 数据与配对构造

主要实验使用来自 APEBench 轨迹的单步教师强制样本对；汽车实验则在车身表面压力数据上采用 OTNO 流程。附录 A 记录了数据集来源、数据划分构建、掩蔽协议及指标定义。

5.2 图中使用的实验配置

正文图比较了四种受控设定：

- **低观测XReg设定：**部分观测下的输出头预测尺度与泛化尺度。
- **仅头部分解消融：**在输出头设置两个不确定性分支，用于检验仅凭目标函数和数据路由是否足以区分它们的作用。
- **无可学习正则化的基线：**采用相同的骨干网络族，不启用可学习的泛化分支。
- **内部噪声XReg参考设定：**启用内部泛化噪声注入和预测头，禁用头部级泛化分支，并采用矩匹配正则化。

这些设定用于检验分解可辨识性、校准行为，以及不确定性注入位置的影响。

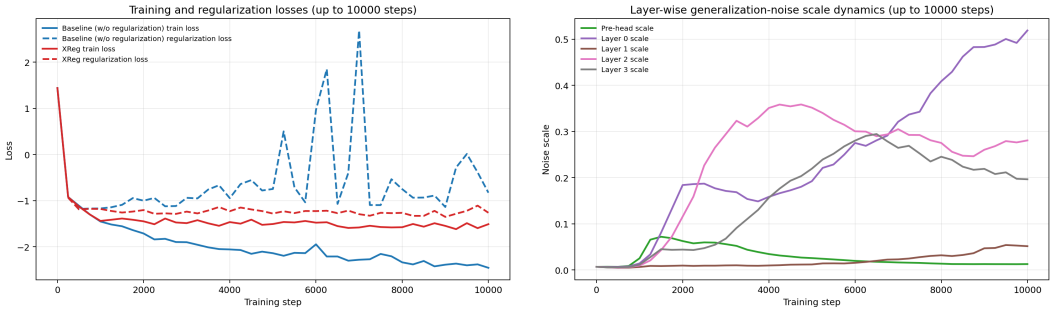
6 经验校准与动态特性

6.1 混合校准比较

我们优先进行校准分析，以验证该方法捕捉预测分布的能力，而不仅仅是评估平均误差。图 2 展示了匹配协议下的代表性情形；正式的指标定义（包括 $\mathrm{ECE}_{\mathrm{mix}}$ 和单步 NLL 归一化）见附录 A.6。

6.2 优化动态与学习到的不确定性尺度

优化轨迹（图 3）展示了训练目标与正则化目标之间的耦合关系，以及不确定性幅度如何分布于各注入位置。



(a) 截至第 10，000 步的训练损失和正则化损失。(b) 截至第 10，000 步各层泛化噪声尺度的演变。

图 3：优化与不确定性分配动态。左：未采用学习式正则化的模型与 XReg 的训练/正则化损失对比（实线=训练，虚线=正则化）。XReg 显著缩小了训练集与正则化集之间的差距。右：头部之前的泛化噪声尺度及逐层泛化噪声尺度。不确定性在各层之间的分配并不均匀，表明模型学到了结构。

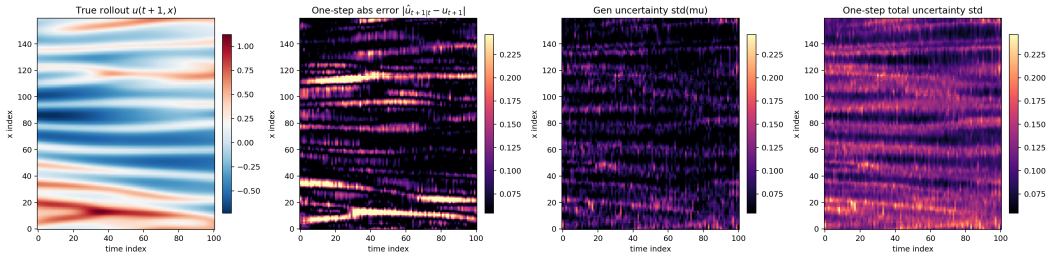


图 4：单步教师强制映射（时间索引 0 至 100）：真实场、绝对单步误差、泛化不确定性（$\mathrm{Std}_\omega[\mu]$）和总预测不确定性。误差图和不确定性图使用相同的数值范围（由总不确定性限值设定），以便直接进行空间比较。高误差区域与高预测不确定性区域相互重合，表明不确定性分配具有空间选择性。

6.3 空间分辨的单步不确定性跟踪

为检验不确定性分配的空间选择性，我们分析单步时空映射。

图 4 比较了匹配时空坐标处的单步误差和预测不确定性。不确定性场随空间变化，并集中在残差较大的区域。

6.4 OTNO 汽车表面压力实验

使用以最优传输神经算子（OTNO）为骨干网络的几何条件化汽车压力实验，评估模型在一维轨迹以外的迁移能力（Li 等，2025）。数据来源和预处理溯源记录于附录 A.3。

图5所示的运行采用了刻意设置的数据受限划分，其中包含 $n_{\text{train}} = 30$ 和 $n_{\text{reg}} = 50$；其余设置遵循附录 A.1 中的默认方案。预测不确定性通过输出头进行参数化，而泛化不确定性则注入内部层。

图5展示了三个互补视角：汽车表面上的绝对误差（左）、同一表面上的交叉正则化不确定性（中），以及包含均值和不确定性带的车顶线切片（右）。左侧和中间图中的空间对应关系表明，不确定性集中在高误差区域，同时在空间上具有选择性。切片视图显示了预测不确定性和泛化不确定性的共同贡献：在难以拟合的区域周围，不确定性包络变宽；而在拟合稳定的区域则变窄。图8报告了本次运行中更多的不确定性尺度变化轨迹。

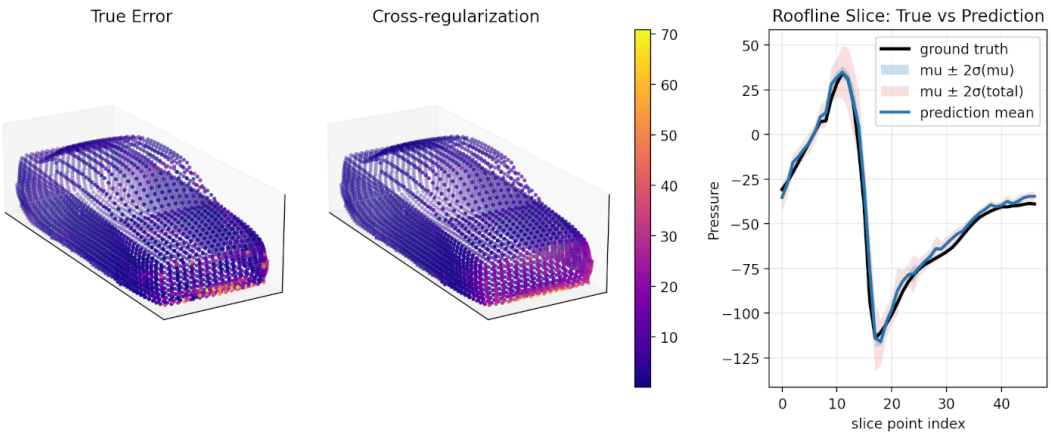


图5：采用交叉正则化不确定性的 OTNO 汽车压力实验（$n_{\text{train}} = 30$，$n_{\text{reg}} = 50$）。左：汽车表面的绝对预测误差。中：同一样本的学习得到的不确定性场。右：车顶线切片，显示真实值、预测均值和不确定性带。误差热点与不确定性热点在空间上相互对应，而切片显示，在残差较大的区域，不确定性经过校准后有所增大。

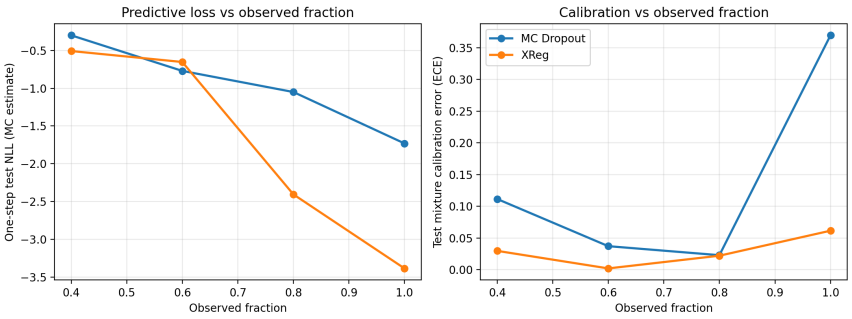


图6：观测比例扫描（0.4，0.6，0.8，1.0），比较 XReg 和 MC Dropout 在单步测试负对数似然（左，采用 MC 评估）与测试混合预期校准误差（右）上的表现。双面板视图将拟合质量与校准质量区分开来。在所有观测比例下，XReg 的校准效果始终更好（期望校准误差更低），同时保持了有竞争力的单步负对数似然。

6.5 扫描对比：XREG 与基线方法（损失 + 期望校准误差）

在匹配的观测设置下，XReg 与两个随机基线进行了比较：MC Dropout（$p = 0.1$）和深度集成（3个成员）。Dropout $p = 0.1$用作计算匹配基线；虽然可以针对每种情形分别调节 p，但这会违背本次比较的无手动调参目标。图6比较了 XReg 和 MC Dropout 的性能扩展。精确的观测比例数值见表1，训练集规模扫描数值见附录表2。深度集成的数值是根据每个扫描点上三个已训练成员汇总得到的预测混合分布计算的。

这些结果表明，交叉正则化不确定性尺度是直接从机制特定的训练-测试失配中学习得到的。根据构造，XReg 通过正则化划分梯度更新泛化不确定性，使不确定性尺度能够在训练过程中跟踪观测水平和数据机制。与固定比率丢弃法和由 3 个成员组成的深度集成相比，这种方法在不同数据条件和可观测性条件下实现了更一致的校准。

观测比例	方法	测试负对数似然（蒙特卡洛）↓	测试$\mathrm{ECE}_{\mathrm{mix}}$ ↓	正则化$\mathrm{ECE}_{\mathrm{mix}}$ ↓
0.4	XReg	**-0.5083**	**0.0296**	**0.0248**
0.4	MC Dropout	-0.2999	0.1115	0.1153
0.4	深度集成 (3)	-0.1002	0.1121	0.1125
0.6	XReg	-0.6533	**0.0020**	**0.0040**
0.6	MC Dropout	-0.7710	0.0370	0.0324
0.6	深度集成 (3)	**-0.8900**	0.0491	0.0522
0.8	XReg	**-2.4052**	**0.0220**	**0.0139**
0.8	MC Dropout	-1.0506	0.0228	0.0159
0.8	深度集成 (3)	-2.0772	0.0624	0.0530
1.0	XReg	-3.3849	**0.0614**	**0.0596**
1.0	MC Dropout	-1.7310	0.3696	0.3695
1.0	深度集成 (3)	**-3.5632**	0.2943	0.2972

表1：观测比例扫描的数值比较。XReg 在全部四种观测比例下均实现了最低的测试集混合期望校准误差和正则化混合期望校准误差，同时保持了有竞争力的单步似然。

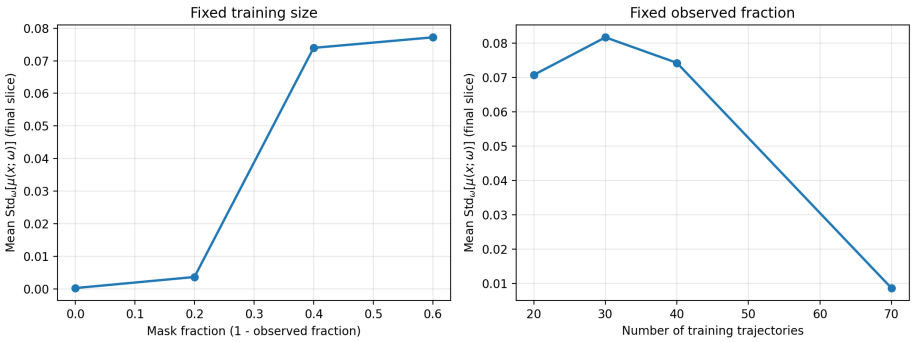


图 7：学习到的泛化不确定性随噪声/数据的变化情况，以最终切片上的平均 $\mathrm{Std}_\omega[\mu]$ 衡量。左：掩码比例增大（训练集大小固定为 $= 50$）意味着观测信号减少，并会使学习到的泛化不确定性增大。右：训练集规模增大（观测比例固定为 $= 0.7$）会降低学习到的泛化不确定性。这些趋势表明，XReg 会根据机制难度调整不确定性幅度。

6.6 泛化不确定性随噪声和数据可用性变化

图 7 单独展示了学习得到的泛化不确定性如何响应机制难度。观测比例降低（监督信号减弱）时，学习得到的泛化不确定性会上升；训练集大小增加（证据增多）时，则会下降。在交叉正则化更新下，正则化划分梯度会按照训练-测试失配所预测的方向设定不确定性尺度，从而产生机制依赖噪声水平。

6.7 主要发现与启示

在匹配协议下，交叉正则化不确定性提升了留出混合校准效果，同时保持了有竞争力的单步似然，这表明校准效果的提升并非由简单的方差膨胀所致。支持性诊断结果与所提出的机制一致：优化轨迹显示训练-正则化差距有所缩小；学习到的不确定性尺度在各层间呈现非均匀变化；单步空间图则显示，不确定性集中在单步误差较大的区域。

7 讨论：相对于模型的不确定性

神经偏微分方程代理模型中的不确定性具有操作性，因此取决于模型。预测不确定性由不可约的观测噪声（包括掩蔽和测量噪声）与源自近似局限、优化误差和数据稀缺的残差结构共同构成。泛化不确定性反映拟合模型对训练特定模式的敏感性，这些模式无法迁移到留出数据，因而类似于认知不确定性成分。本文通过使用正则化划分梯度优化不确定性控制参数来学习这一不确定性，无需对权重设定先验或进行后验近似。附录B讨论了其对混沌状态的影响，并阐明了它与传统超参数优化的区别。

8 结论

我们提出了适用于神经偏微分方程代理模型的交叉正则化不确定性学习方法：不确定性参数通过正则化划分梯度在线训练，而预测器参数则在训练划分上进行训练。这种训练/正则化路由在操作层面实现了预测不确定性（拟合时的残差吸收）与泛化不确定性（训练集与留出集之间的不匹配控制）的分离，且无需集成或贝叶斯后验近似。从实证结果来看，在部分可观测性和数据有限的情况下，所得预测分布能够改善单步高斯混合校准；空间诊断则表明，不确定性集中于误差较高的区域。要将这些保证从单步教师强制评估扩展到经过校准的长时域自回归滚动预测，还需要通过累积的模型误差传播不确定性；这一情形留待未来研究。

一个相关方向是在模型内部引入更丰富的预测噪声。在内部层注入预测噪声，原则上可以诱导相关的或多模态的输出不确定性。沿这一方向扩展当前的实现，或许能够缩小单模型不确定性学习与科学决策中所采用的更丰富不确定性族之间的差距。

参考文献

Anastasios N. Angelopoulos 和 Stephen Bates。保形预测与无分布假设不确定性量化简介。*arXiv* 预印本 *arXiv:2107.07511*，2021年。

Carlos Stein Brito。交叉正则化：通过验证梯度实现自适应模型复杂度。*ICML*，2025年。

Yarin Gal 和 Zoubin Ghahramani。将 dropout 作为贝叶斯近似：在深度学习中表示模型不确定性。载于 *ICML*，2016年。

Felix Koehler、Simon Niedermayr、R¨udiger Westermann 和 Nils Thuerey。APEBench：偏微分方程自回归神经模拟器的基准。神经信息处理系统进展，38，2024年。

Balaji Lakshminarayanan、Alexander Pritzel 和 Charles Blundell。使用深度集成进行简单且可扩展的预测不确定性估计。载于 *NeurIPS*，2017年。

Xutong Li、Zongyi Li、NikolaKovachki 和 Anima Anandkumar。通过最优传输进行几何算子学习。*arXiv* 预印本 *arXiv:2507.20065*，2025年。

Zongyi Li、Nikola Kovachki、Kamyar Azizzadenesheli 等。用于参数化偏微分方程的傅里叶神经算子。国际学习表征会议，2021年。卢璐、Pengzhan Jin、George Pang 等。基于算子的通用逼近定理，通过 DeepONet 学习非线性算子。《自然·机器智能》，2021年。

马子奇、David Pitt、Kamyar Azizzadenesheli 和 Anima Anandkumar。通过共形预测对算子学习进行校准不确定性量化。机器学习研究汇刊，2024年。

Albert Matveev、Sanmitra Ghosh、Aamal Hussain、James-Michael Leahy 和 Michalis Michaelides。用于傅里叶神经算子的轻量级扩散乘子与不确定性量化。*arXiv* 预印本 *arXiv:2508.00643*，2025年。

PhysicsX。Dinozaur 数据集仓库。https://huggingface.co/datasets/ PhysicsX/DINOZAUR，2025年。

Makoto Takamoto 等。PDEBench：科学机器学习的广泛基准测试。载于*NeurIPS* 数据集与基准，2022年。

A 附录：数据集、数据划分与可复现性

A.1 默认训练配置

除非扫描轴中明确覆盖，否则运行使用：

- 一个包含 12 个傅里叶模态、隐藏层宽度为 8 并使用 GELU 激活函数的 4 层 FNO 主干网络，
- 批次大小 16，
- 优化总步数 30，000，
- 评估间隔为 250 步，
- 每训练 5 步进行一次正则化更新，
- 10 用于训练、正则化和测试评估的蒙特卡洛样本，
- 学习率 10^{-3} 用于主干网络权重， 10^{-2} 用于内部预测噪声尺度， 10^{-2} 用于内部泛化噪声尺度，以及 10^{-3} 用于头部级泛化噪声尺度，
- 初始化日志 $\sigma_{\text{预测}} = -5$，日志$\sigma_{\text{生成}} = -5$。

A.2 APEBENCH 数据与任务构建

所有主要实验均基于 APEBench（Koehler 等，2024）的 diff ks 设置（一维轨迹数据，包含 160 个空间点）。我们从每条轨迹中构造单步监督样本对（x_t， x_{t+1}），用于教师强制训练和评估。除非另有说明，训练时域为 10，测试时域为 200。由于模型经过训练以量化单步不确定性，我们直接基于这些样本对报告单步指标。长时域滚动校准是未来的研究方向，且并非易事：经过反复的自回归传播并伴随预测误差累积后，不确定性仍须保持校准。

A.3 OTNOCAR 数据来源

汽车压力实验采用 DINOZAUR 数据发布版本（Matveev 等，2025；PhysicsX，2025）中的 ShapeNetCar 子集，并使用为这些实验准备的本地 data/shapenetcar 划分。在此设置中，每个样本均为一个汽车表面点集，包含由几何／流场导出的输入以及标量表面压力目标。我们的 OTNO 模型遵循 neuraloperator 的实现流程；OTNO 指最优传输神经算子 （Li 等，2025）。

A.4 训练集/正则化集/测试划分

优化方案采用明确的三路划分：

- **训练集：** 用于优化确定性模型权重和预测不确定性参数；
- **正则化集（正则化集）：** 仅用于通过正则化目标更新泛化噪声参数；
- **测试划分：** 留作最终的一步预测和校准报告。

这一划分是该方法主张的核心：预测不确定性与泛化不确定性通过数据路由和目标路由加以分离，参数命名也体现了这种分离。

A.5 观测掩蔽协议

空间掩码由输入字段和目标字段的观测比例控制。我们采用受控遮挡协议：每次运行时，掩码使用固定随机种子采样一次，并在整个训练过程中重复使用，以从方法比较中消除掩码随机性。在所有扫描图中，所比较方法（XReg 与 MC Dropout）之间的观测设置保持一致，因此差异可归因于不确定性学习策略，而数据可见性保持固定。逐步随机掩码是未来工作中一个相关的扩展方向。

A.6 评估指标与日志记录规范

混合校准通过以下方式衡量

$$\mathrm{ECE}_{\mathrm{mix}} = \frac{1}{|\mathcal{A}|} \sum_{\alpha \in \mathcal{A}} \left| \widehat{\mathrm{Coverage}}(\alpha) - \alpha \right|,$$

其中 $\mathcal{A} = \{0.1, \ldots, 0.9\}$。对于每个 α，Coverage(α) 是根据高斯混合预测累积分布函数计算得出的经验中心区间覆盖率，并在时空单步目标上取平均。单步负对数似然按空间目标点报告。

我们报告：

- 通过蒙特卡洛预测评估得到的单步测试负对数似然，
- 正则化和测试数据划分上的混合校准误差（期望校准误差），
- 标称覆盖率网格点上的校准曲线，
- 不确定性分解统计量，包括标准差$_\omega[\mu]$ 和逐层泛化噪声尺度。

A.7 计算开销

XReg 会为 ρ 增加一次周期性正则化更新。若一次训练更新的成本为 C_{train}，一次正则化更新的成本为 C_{reg}，且每隔 k_{reg} 步进行一次正则化，则近似开销系数为

$$\text{overhead} \approx 1 + \frac{1}{k_{\text{reg}}} \frac{C_{\text{reg}}}{C_{\text{train}}}.$$

在我们的默认设置（$k_{\text{reg}} = 5$）下，由于 ρ更新较为稀疏且维度较低，因此开销适中。增大 k_{reg}（降低正则化更新频率）还能进一步降低成本。先前的实证报告表明，实际设置下的成本可达到标准训练成本的大约 $1.1\times$（Brito，2025），与常规训练流程相当；这类流程通常也包含周期性验证检查。

B 附录：关于解释与优化的扩展讨论

本节进一步讨论两个次于主要经验性主张的议题。

确定性情形下的模型相对不确定性。 在“内在噪声”存在争议的领域，模型相对解释尤其重要。混沌系统便是一个例子：支配动力学可能是确定性的，但有限数据和有限容量模型仍会产生较大的预测残差。在这种情况下，预测不确定性反映未解释的残差结构，而泛化不确定性则反映对特定训练样本的过拟合。

自适应为何仍能保持稳定。 自适应机制在优化过程中，根据观察到的训练–留出失配学习不确定性幅度。在我们的设置中，这一过程仍保持稳定，因为正则化变量是低维噪声控制量，而非完整的特征映射；正则化步骤会调整不确定性尺度，而不会在留出数据上重新拟合整个预测器。

与超参数优化的关系。XReg 在单次训练过程中直接学习不确定性控制参数。经典的超参数调优会优化外部

元参数（例如通过外循环搜索选定的固定 λ），而 XReg 更新控制不确定性的模型参数。在算法 1 中，$(\theta,\ \psi)$ 在训练数据上更新，而 ρ 则根据正则化集梯度在线更新。

C 附录：OTNO 汽车实验不确定性诊断

本附录报告图5中 OTNO 汽车实验在优化期间的不确定性尺度轨迹及最终不确定性大小。其目的是展示在数据有限的汽车场景中，预测分支和泛化分支均保持活跃且未退化。

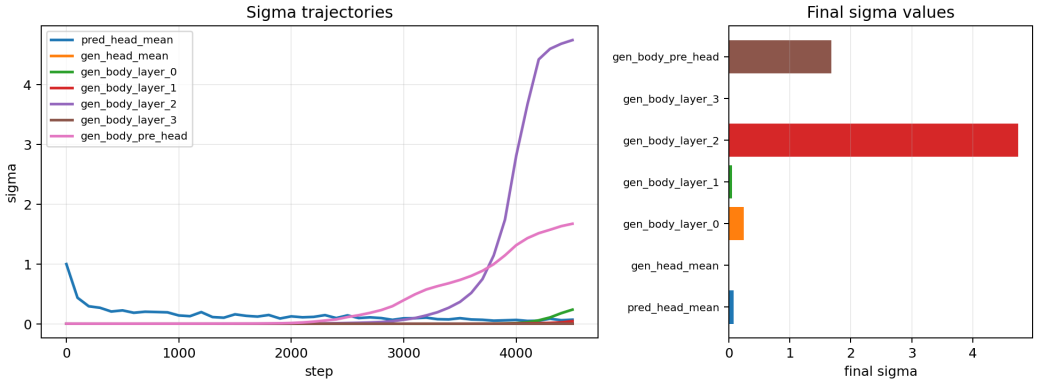


图8：图5所示汽车实验的 Sigma 诊断。左：优化期间学习到的不确定性尺度轨迹。右：各不确定性分量的最终值。该次运行中的预测尺度和泛化尺度均未退化，部分泛化分量明显增大，且没有任何分量塌缩至接近零的值。

D 附录：参数扫描校准拼图

本附录提供 XReg 和 MC dropout 参数扫描中所有运行的完整校准拼图；这些扫描结果已在正文表格和汇总图中概述。这些面板构成了完整的可视化审计轨迹，补充了标量 ECE 汇总。

训练集大小	方法	测试负对数似然（蒙特卡洛）↓	测试ECE_{mix} ↓	正则化ECE_{mix} ↓
20	XReg	-0.3277	**0.0965**	**0.0867**
20	MC Dropout	**-0.4343**	0.1254	0.1197
20	深度集成 (3)	0.7512	0.1344	0.1361
30	XReg	-0.5962	**0.0273**	**0.0210**
30	MC Dropout	**-0.6894**	0.0857	0.0815
30	深度集成 (3)	0.1574	0.1157	0.1131
40	XReg	-0.6671	**0.0144**	**0.0118**
40	MC Dropout	**-0.8858**	0.0375	0.0410
40	深度集成 (3)	-0.7355	0.0760	0.0727
70	XReg	-1.4173	**0.0118**	**0.0128**
70	MC Dropout	-1.1361	0.0666	0.0654
70	深度集成 (3)	**-2.0064**	0.0569	0.0555

表 2：观测比例固定为 0。7 时训练集规模扫描的数值比较。XReg 在所有四种训练集规模下的测试/回归混合 ECE 都显著更低，包括基线单步 NLL 表现有竞争力的情形。

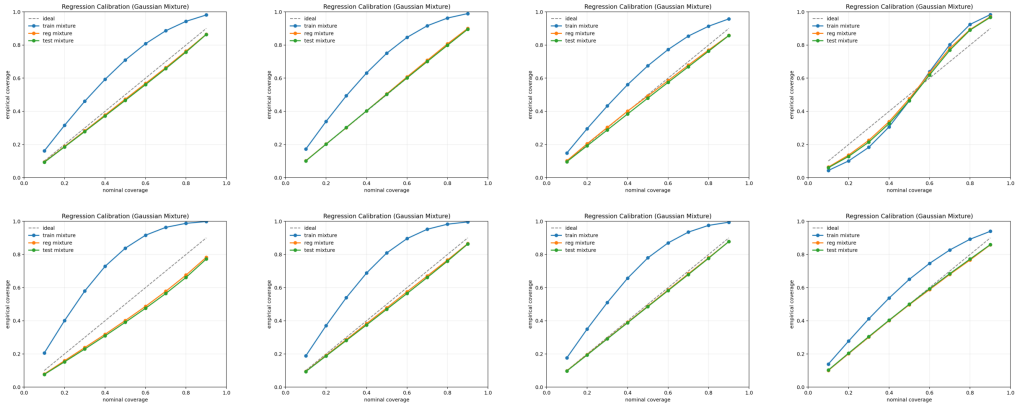


图 9：完整 XReg 扫描校准拼图（正文报告的 1 万步扫描中全部八次 XReg 运行的回归混合校准）。这些面板作为原始可视化审计轨迹纳入其中：它们展示了各次运行的曲线形状差异，而这些差异无法通过标量 ECE 汇总得到充分体现。

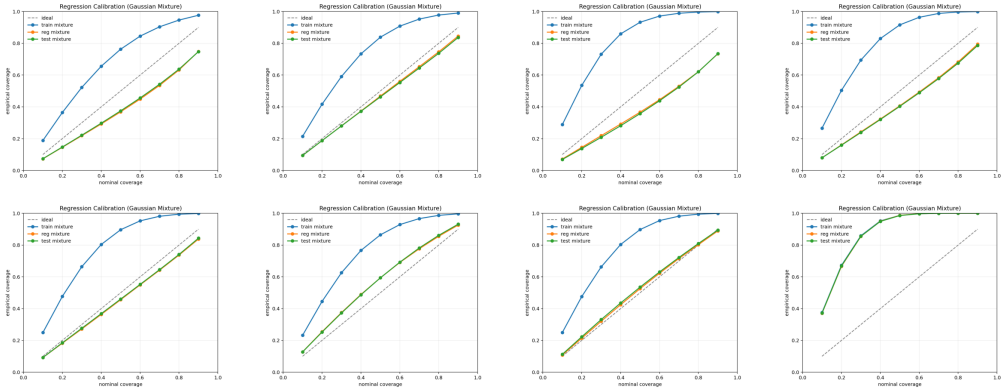


图 10：完整 8 次运行扫描的 MC dropout 扫描校准拼图（观测比例轴和训练规模轴，dropout $p = 0.1$）。与图 9 一同，这些面板提供了多面板校准对比，补充正文中的聚合损失/ECE 表。

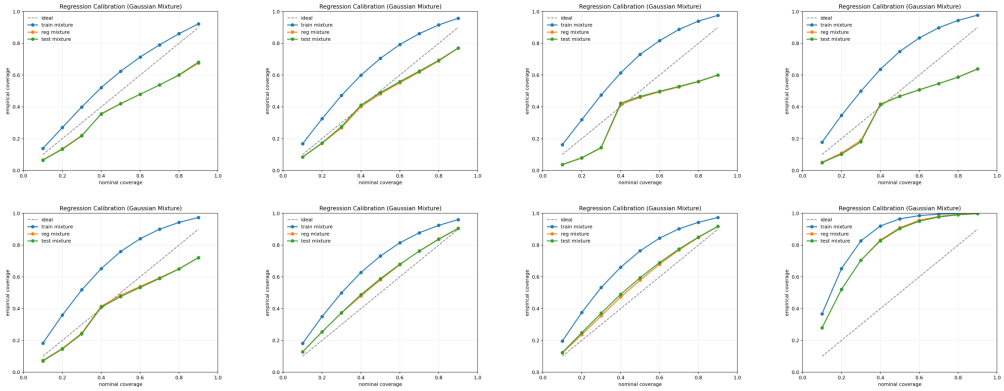


图 11：完整 8 点扫描的深度集成扫描校准拼图（观测比例轴和训练规模轴），每个点均作为 3 成员聚合预测混合进行评估。此视图与图 9 和图 10 相辅相成，为集成基线提供完整面板的可视化审计。