arXiv:2601.04396v1 [cs.CE] 7 Jan 2026

存在模型形式不确定性时的推断：利用面向预测的方法改进不确定性刻画

Rebekah White、Rileigh Bandy 和 Teresa Portone

2025年1月

摘要

贝叶斯推断是一种常用的不确定性校准方法，但存在模型设定错误时，它可能会低估此类不确定性，从而影响其为决策提供依据的可靠性。近年来，统计学和机器学习领域提出了面向预测的推断方法，能够更好地校准不确定性，并适应模型设定错误的程度。然而，这些方法尚未应用于复杂科学应用场景；在这类场景中，相关现象由基于物理的模型控制。此类问题通常涉及高维时空数据的单次实现，以及非线性且计算成本高昂的参数到观测量映射。本文研究了具有这些相关特征的问题中的变分面向预测推断；具体而言，我们考虑了一个多项式模型，以及一个由对流–扩散方程控制的污染物输运问题。面向预测的损失函数定义为校准数据的对数预测概率。我们研究了模型设定错误和噪声增加所产生的影响，并使用蒙特卡罗抽样和逐分量核密度估计评估预测密度的近似效果。本文的一项创新之处在于，将面向预测的推断用于校准模型形式不确定性（MFU）表示；这些表示是嵌入控制方程中的基于物理的修正，旨在减少（但很少能完全消除）模型设定错误。计算结果表明，与标准推断相比，面向预测的框架能够提供更好的不确定性刻画，同时也适用于校准 MFU 表示。

1 引言

在无法获得观测数据的情境中，计算建模与仿真日益被用于辅助决策。然而，在能够获得观测数据的校准情境下，必须通过不确定性量化来建立对这类决策的信心。尽管贝叶斯推断为根据数据刻画不确定性提供了自然框架，但众所周知，当存在模型设定错误时，由于后验收缩的不良影响，它会低估不确定性 [13]。在复杂现象的基于物理的建模中，这种行为带来了重大挑战，因为知识或计算能力有限，往往必须采用简化假设。因此，模型设定错误通常难以避免；要利用建模与仿真辅助决策，推断方法就必须能够适应并应对这类模型设定错误。

为改进存在模型设定错误时的不确定性刻画，人们提出了其他概率框架。一类方法可统称为广义贝叶斯推断（GBI），其解通常称为吉布斯后验 [6, 38, 39, 11]。在这类方法中，放宽了模型预测与数据相符的要求：不再采用传统贝叶斯方法中的期望对数似然，而是考虑广义期望损失，其灵活性能够增强对模型设定错误的稳健性，并可根据具体问题进行调整。GBI 涵盖并关联多种方法，包括温度化后验 [9],Safe Bayes [8], 粗化后验 [23], 和近似贝叶斯计算（ABC） [17, 36]。此外，尽管广义解源于机器学习领域中不同的研究动机，但也可以将其视为随机预测器泛化误差的概率近似正确（PAC）贝叶斯上界的极小化解 [20, 1, 21, 19, 10]。尽管在有限数据条件下，GBI 能够改善存在模型设定错误时的不确定性表征，但未必能避免后验不必要的收缩，因此仍可能产生预测密度较差的后验分布 [22]。此外，在许多场景中（例如 ABC），该方法的稳健性可能取决于是否恰当设定稳健的汇总统计量，而这在实践中未必可行。

因此，本研究考虑采用能够抵御模型设定错误的替代推断方法，其解（后验分布）所提供的预测不确定性能够对校准数据赋予较高概率，同时又不会过于宽泛而失去信息价值

。为此，我们借鉴了统计学和机器学习领域近期提出的面向预测的框架 [22, 24, 35]。与 GBI 类似 [14], ，该推断问题被表述为一个无限维优化问题，并可通过变分近似求解；其目标函数由两部分构成：衡量对数据拟合程度的损失项，以及衡量对先验信念遵循程度的正则化项。不过，此处的损失函数经过专门设计，可直接针对预测不确定性对观测数据的刻画效果：为此，采用直接作用于预测密度的评分规则。尽管这一思路在概念上颇具吸引力，但将这些面向预测的推断技术应用于复杂科学问题的探索仍然有限；在这类问题中，相关现象由可能具有大规模的基于物理的模型所支配。因此，本研究旨在考察该方法在与此类问题相关的情境中的表现。具体而言，我们关注以下情形：1）观测数据由高维时空数据构成；2）参数到观测量的映射是非线性的，且由计算成本高昂的基于微分方程的模型定义；3）似然只能通过隐式方式获得。此外，基于物理的建模有一项独特优势：能够利用模型形式不确定性（MFU）的表示 [29, 4, 26, 25], ，即嵌入控制方程中的基于物理的修正，以帮助缓解模型设定错误。本研究的一项创新之处在于，探索将面向预测的推断应用于此类不确定性表示的确定，同时估计模型参数。

在将面向预测的推断应用于展示科学与工程应用相关特征的计算示例时，我们将数据的对数预测概率作为变分目标中的损失函数。我们考虑了一个多项式示例，其中对数预测损失可用闭式表示；此外，还考虑了一个由对流扩散偏微分方程（PDEs）描述的污染物输运问题。通过这些计算示例，我们探究了模型设定错误和噪声增加对预测性逆问题解的影响，并将其与标准变分推断进行比较。此外，我们还对预测密度的蒙特卡洛（MC）估计量的收敛性进行了数值评估，并与标准方法中采用的期望对数似然进行比较。我们进一步研究了逐分量密度近似的影响；对于无法获得显式似然的高维数据，通常需要采用这种近似。对于污染物输运问题，我们展示了如何通过面向预测的框架来构建 MFU 表示，从而获得更好的预测不确定性，并与标准方法相比得出偏差更小的参数估计值。总体而言，本研究初步探讨了将面向预测的推断应用于复杂现象的基于物理的模型及其所面临的挑战。本研究的主要贡献如下：

- 将面向预测的推断应用于具有现实工程和科学问题相关特征的问题。
- 探究高维数据和复杂基于物理的模型通常所需的各种近似所带来的影响。
- 创新性地将面向预测的框架应用于MFU表示及模型参数的校准。

本文其余部分的组织结构如下。首先，第2节回顾标准贝叶斯推断，并在第2.1节讨论模型设定错误情况下由此产生的影响，随后在第2.2节介绍广义贝叶斯推断及相关方法。在第3节中，我们回顾计算演示所采用的方法，其中第3.1节讨论面向预测的推断及相关近似，第3.2节概述MFU表征。计算结果见第4节，其中第4.1节考察一个多项式模型，而第4.2节采用基于偏微分方程的模型来描述污染物通过非均质多孔介质的迁移，并探究在面向预测的逆问题中采用MFU表征的影响。最后，第5节给出结论。

2 背景　　标准贝叶斯推断与广义贝叶斯推断　　e

我们首先在第 2.1 节介绍标准贝叶斯推断范式，并讨论在存在模型设定错误时进行推断所面临的相关挑战。随后，在第 2.2 节中，我们将探讨广义贝叶斯推断及相关方法，这些方法旨在更好地支持模型设定错误情形下的校准。尽管我们最终发现，这类广义方法并不能彻底解决模型设定错误存在时的不确定性校准问题，但仍会讨论这些方法，因为它们为第 3.1 节中介绍的较新面向预测的框架提供了重要背景。

2.1 模型错误设定存在时的标准贝叶斯推断及其影响

标准贝叶斯推断范式是一种广受欢迎的模型不确定性量化方法，因为它不仅能够估计参数 $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_{n_p}]^\top$，还能够给出相应的不确定性表征，为决策提供依据。首先，考虑一般形式下的参数到观测量映射 $f(\boldsymbol{x};\, \boldsymbol{\theta}) : \Theta \to \mathcal{Y}$，其中 $\boldsymbol{x}$表示离散化空间位置构成的向量，且假设 Θ为有限维。参数到观测量映射通常通过观测算子作用于底层模型（例如微分方程）的解来构造。推断旨在利用观测数据 $\boldsymbol{y} = [y_1, \ldots, y_{n_y}]^\top \in \mathcal{Y} \subset \mathbb{R}^{n_y}$为模型参数 $\boldsymbol{\theta}$提供信息。在构建贝叶斯反问题时，需要指定一个统计模型，用以表示数据收集过程。通常会采用以下加性高斯噪声假设：

$$\boldsymbol{y}(\boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\theta}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}}\right), \tag{1}$$

其中， $\boldsymbol{\epsilon}$表示噪声实现，其分布由噪声协方差 $\boldsymbol{\Gamma}_{\text{noise}}$定义。(1)的形式导出了如下似然设定：

$$\pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta}) \sim \mathcal{N}\left(f(\boldsymbol{x}; \boldsymbol{\theta}), \boldsymbol{\Gamma}_{\text{noise}}\right). \tag{2}$$

贝叶斯反问题的解是定义在模型参数上的后验概率密度函数，根据贝叶斯规则计算为

$$\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta})\pi_{\text{pri}}(\boldsymbol{\theta}), \tag{3}$$

其中， $\pi_{\text{pri}}(\boldsymbol{\theta})$是先验密度，反映了在收集任何数据之前人们对参数不确定性的信念。为了直观展示模型参数中的不确定性如何传播为模型预测的感兴趣量（QoI）中的不确定性，可以利用后验推前（pf）分布。形式化地定义如下：设感兴趣量定义为 $g(\boldsymbol{\theta}) : \Theta \to \mathcal{G}$，并定义 $G := g(\boldsymbol{\theta})$。 G 中的不确定性由 $\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{y})$诱导，并由推前测度刻画；我们假设该测度的密度存在，且相对于 $\mathcal{G}$上的勒贝格测度连续，其定义为

$$\pi_{\text{pf}}(z|\boldsymbol{y}) = \int_{\Theta} \delta(z - g(\boldsymbol{\theta}))\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}. \tag{4}$$

为了直观展示这一推前密度，可以考虑分量G_j 的边际 $100(1-\alpha)\%$ 可信区间，其定义为

$$C_\alpha^{(j)} = \left[z_{\frac{\alpha}{2}}^{(j)}, z_{1-\frac{\alpha}{2}}^{(j)}\right], \tag{5}$$

其中， $z_{\frac{\alpha}{2}}^{(j)}$ 和 $z_{1-\frac{\alpha}{2}}^{(j)}$ 分别表示边际预测分布的 $\frac{\alpha}{2}$ 分位数和 $1-\frac{\alpha}{2}$ 分位数。

尽管不正确的模型仍然可能有助于理解我们感兴趣的现象，但在推断中利用此类不完善模型时，往往会出现实际问题。似然设定（2）隐含地假设数学模型能够表示数据生成分布。然而，当存在模型设定错误时，这一假设便不再成立，并可能显著影响所得的不确定性刻画。继 [32], 之后，我们对模型设定错误作一般性定义，并讨论其在推断背景下的影响。考虑从未知的数据生成分布中独立同分布抽取的观测值 $y_{1:n}$ ，其联合密度为$\pi_g(y_{1:n})$。在实践中，推断使用参数族$\{\pi_p(y_{1:n}|\theta):\ \theta \in \Theta\}$ 旨在近似该分布。如果真实分布属于该参数族，则称模型设定正确，即存在 $\theta_0 \in \Theta$ 使得$\pi_g(y_{1:n}) = \pi_p(y_{1:n}|\theta_0)$；反之，如果对于 θ_0的所有可能取值，$\pi_g(y_{1:n}) \neq \pi_p(y_{1:n}|\theta_0)$，则称模型存在设定错误。在标准正则条件下，伯恩斯坦-冯·米泽斯定理表明，对于设定正确的模型，随着样本量增加，后验分布会收敛至以数据生成参数 θ_0 为中心的高斯分布。相反，当存在模型设定错误时，后验分布会集中于以下给出的 Kullback-Leibler（KL）散度的最小化者附近

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}}\, \mathcal{D}_{KL}\left(\pi_g(\cdot)\,||\,\pi_p(\cdot|\theta)\right). \tag{6}$$

更多详情请参见 [13] 。(6) 中给出的最小化解不一定唯一，而且通常并不对应能够重现观测数据的模型。尽管如此，随着

数据量不断增加，后验分布会集中于该值附近，使人们对错误模型的信心日益增强。在实践中，这种现象会导致传播得到的（通常是外推的）后验不确定性无法为决策提供可靠依据。

为说明模型设定错误及其对最终标准贝叶斯后验的影响，考虑如下情形：含噪数据根据以下真实数据生成过程确定

$$y^{\text{obs}} = \sin(x) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 0.4^2). \tag{7}$$

然而，由于推断模型存在模型设定错误，所假设的统计模型为

$$y = ax + b + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 0.4^2). \tag{8}$$

图1描绘了真实数据生成分布与后验推前的95%可信区间，其中后验分布是根据不断增加的数据量 $n = 15$、50、250计算得到的。从图1可以看出，随着样本量增大，后验分布及其推前因此变得越来越集中于模型设定错误的模型附近，从而使不确定性刻画变得不可靠。尽管在实践中，人们无法获得真实的

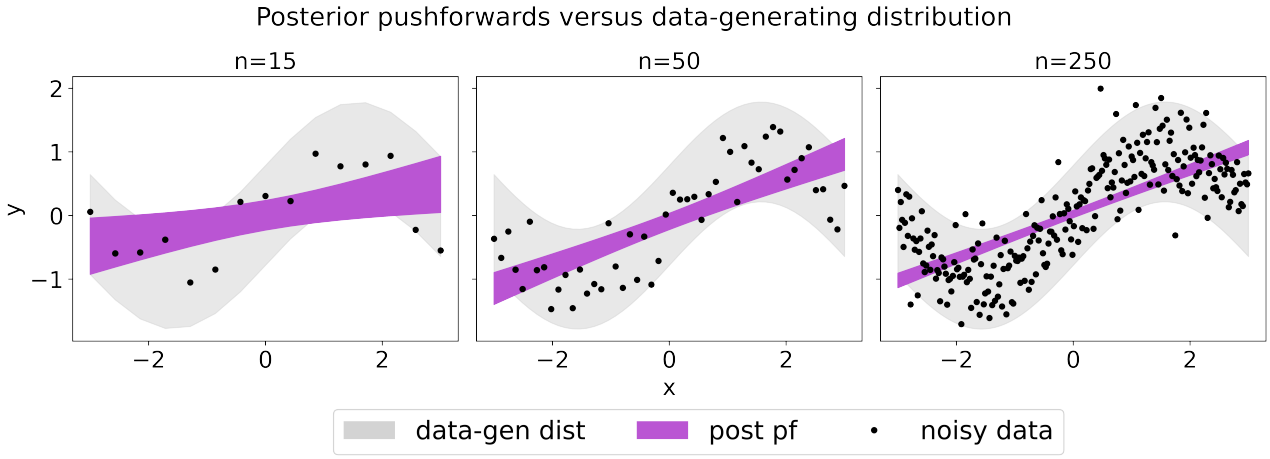


图1： 真实数据生成（正弦）分布与不同数据量 $y_{1:n}$ 下的后验推前所对应的95%边际可信区间的比较，其中 $n = 15$、50、250（从左到右），推断模型（线性）存在模型设定错误。

数据生成分布。为了进行直接比较，模型设定错误可通过传播的不确定性覆盖率较低以及非对称残差来表明。

2.2 广义贝叶斯推断及相关方法

在此，我们介绍**广义贝叶斯推断（GBI）**及相关方法的背景，这些方法旨在提高推断问题对模型设定错误的稳健性。首先，标准贝叶斯后验可以视为以下无限维优化问题的解：

$$\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{y}) = \underset{q \in \mathcal{P}(\Theta)}{\operatorname{argmin}} \mathbb{E}_{q(\boldsymbol{\theta})} \left[-\log(\pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta})) \right] + \text{KL}\left(q(\boldsymbol{\theta}) || \pi_{\text{pri}}(\boldsymbol{\theta}) \right), \tag{9}$$

其中，$\mathcal{P}(\Theta)$ 是 Θ [6, 14] 上所有概率测度构成的空间。目标函数的第一项通过衡量对数似然给出的损失来表示对数据的忠实度；第二项则表示对先验信念的忠实度。从这种以优化为中心的推断视角出发，可以使用其他损失函数 $l(\boldsymbol{\theta}, \boldsymbol{y})$ 替代对数似然，以增强对模型设定错误的稳健性，如下所示：

$$\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{y}) = \underset{q \in \mathcal{P}(\Theta)}{\operatorname{argmin}} \mathbb{E}_{q(\boldsymbol{\theta})} \left[-l(\boldsymbol{\theta}, \boldsymbol{y}) \right] + \text{KL}\left(q(\boldsymbol{\theta}) || \pi_{\text{pri}}(\boldsymbol{\theta}) \right), \tag{10}$$

[的工作 6],表明，更广义推断问题（10）的解可表示为

$$\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \exp\left(-l(\boldsymbol{\theta}, \boldsymbol{y}) \right) \pi_{\text{pri}}(\boldsymbol{\theta}), \tag{11}$$

这也称为吉布斯后验 [38], 其针对模型设定错误的稳健性已得到研究 [11]。请注意，无论这些解是标准或广义优化目标的解，还是通过变分方法等方式得到的近似解，我们都将（9）和（10）等优化问题的所有解称为后验分布 $\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{y})$。 [6] 的工作进一步表明，吉布斯后验形式的解能够提供"模型设定错误下后验信念的连贯、决策理论表述"，同时还通过引入缩放参数 α，在计算后验分布时提供一种将数据中的信息与先验信息进行缩放的方法，如下所示:

$$\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \exp\left(-\alpha \cdot l(\boldsymbol{\theta}, \boldsymbol{y})\right) \pi_{\text{pri}}(\boldsymbol{\theta}), \quad \alpha > 0. \tag{12}$$

GBI 涵盖或涉及多种替代方法。首先，如果将式（11）中的损失函数定义为对数似然，就可以得到标准贝叶斯后验。将对数似然用于缩放后的吉布斯后验（12）会得到一种**温度化后验**，它通过降低似然的权重来减弱模型设定错误的影响；有关温度化后验对模型设定错误的稳健性的更多细节，请参见 [9] 。**Safe Bayes**同样利用（12），但会在推断过程中，通过考察由后验期望损失 [8]衡量的预测性能，自适应地估计温度参数 α 。这种对温度调节参数进行自适应优化的效果是，当模型设定错误程度较高时， α 会降低，以减弱似然的影响。吉布斯后验还与一种旨在提高贝叶斯后验稳健性的替代方法有关，该方法称为**粗化** [23]。粗化后验并非以模型生成数据这一事件为条件，而是以这样的事件为条件：观测数据的经验分布，在某种统计意义上，与模型生成数据的经验分布相接近。吉布斯后验为粗化后验提供了大样本近似。

GBI 与机器学习（ML）领域发展的技术之间也存在联系，即可能近似正确（PAC）学习 [37], 。粗略来说，其目标是提供理论保证，确保基于有限数据训练的 ML 模型能够很好地泛化到新的、未见过的数据。该框架的贝叶斯扩展 **PAC-Bayes** [20, 1, 21, 19, 10], 研究随机预测器的泛化保证，而不对数据生成过程作出特定假设 [2]。这类方法为经验风险（泛化误差）建立上界；可以证明，吉布斯后验是该上界的最小化解 [21, 2]。尽管 PAC-Bayes 背后的动机未必关注所得后验分布的性质，但它与 GBI 的联系表明，探索计算上可处理的 PAC 上界或许能够改进 GBI 解的计算。

尽管其动机在于似然难以处理，而非似然不可靠，**近似贝叶斯计算（ABC）** [17, 36] 在概念上与 GBI 相似。ABC 通过灵活地界定模型评估结果（或汇总统计量）与观测数据之间的接近程度，无需评估似然即可从近似后验中获得样本。 [33] 的研究探讨了如何利用 GBI 设计能够抵御模型设定错误的 ABC 算法，方法是考虑基于 Wasserstein距离和软阈值化的距离度量。此外，选择汇总统计量也能增强对模型设定错误的稳健性 [16], ，但此类统计量的确定在很大程度上取决于具体应用。最后，我们注意到，近期研究通过考虑 KL散度之外的统计散度，进一步推广了式（10）[1] [14]。这种方法称为**广义变分推断（GVI）**，因为 $\mathcal{P}$（Θ）仅限于概率测度的一个变分族，因此这类推广可能无法像式（10）和（12）那样给出后验分布的闭式表示。GVI 倡导以优化为中心的贝叶斯推断视角，旨在解决模型和先验的设定错误，同时通过变分近似保持计算上的可处理性。

尽管上述方法背后的起源和动机可能各不相同，但与标准贝叶斯范式相比，它们都能在一定程度上抵御模型设定错误。然而，这些方法并不会直接惩罚预测不确定性表征，因此无法确保所得后验分布能够充分描述校准数据，也就是说，无法确保为其赋予较高概率。事实上，正如 [22], 所指出的，标准贝叶斯推断和广义贝叶斯推断都会随着数据量增加而导致后验收缩，因此可能产生质量较差的预测分布。因此，我们转而考虑面向预测的推断 [22], ，其中的损失项经过设计，旨在专门针对预测不确定性描述观测数据的效果。我们将在下一节详细讨论这一方法。

[1] [14] 的损失函数形式写为求和损失，因为它们考虑了数据的多个独立同分布实现，而（10）中的 $\boldsymbol{y}$ 是数据的一个多元实现。

3 方法

这里，我们介绍后续第4节计算结果中采用的方法。首先，我们将依据 [22, 24, 35, 15]的研究讨论面向预测的框架。然后，在第3.2节中，我们介绍有关 MFU 表示的相关背景，重点介绍校准此类表示的现有方法及其面临的挑战。

3.1 面向预测的推断

接下来，我们介绍面向预测的推断方法，该方法将在第4节的计算结果中使用。面向预测的优化目标的一般形式为

$$\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{y}) = \underset{q\in\mathcal{Q}(\Theta)}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{\theta},\boldsymbol{y}) + \mathrm{KL}\left(q(\boldsymbol{\theta})||\pi_{\text{pri}}(\boldsymbol{\theta})\right), \tag{13}$$

其中， $\mathcal{L}$ 是用于量化预测分布对观测（校准）数据的描述效果的损失函数，而$\mathcal{Q}(\Theta) \subset \mathcal{P}(\Theta)$是概率密度的一个变分族。感兴趣量上的预测分布的一般形式与第2.1节中的推前类似，其中同样有 $g(\boldsymbol{\theta}) : \Theta \to \mathcal{G}$，但噪声模型使得与观测相关的随机变量表示为 $G|\boldsymbol{\theta} \sim \pi(\cdot|\boldsymbol{\theta})$。相应的后验预测分布定义为

$$\pi_{\text{pred}}(z) = \int_{\Theta} \pi(z|\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad \text{for.} \tag{14}$$

要可视化预测分布，可以利用式（5）中定义的边际可信区间。在面向预测的推断中，我们关注后验预测分布对观测数据 $\boldsymbol{y}$的描述效果。因此，接下来我们根据后验预测分布下数据的概率，构造面向预测的损失函数:

$$\mathcal{L}(\boldsymbol{\theta},\boldsymbol{y}) = -\log \pi_{\text{pred}}(\boldsymbol{y}) = -\log \int_{\Theta} \pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta} = -\log \mathbb{E}_{q(\boldsymbol{\theta})}\left[\pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta})\right]. \tag{15}$$

对数评分规则用于衡量预测分布的准确性。请注意，也可以使用最大均值差异或累积排名概率评分等其他评分规则来计算面向预测的后验分布；更多详情请参见 [22, 35] 。我们将求解 (13) 称为执行**预测变分推断（PVI）**，将由此得到的解称为**面向预测的后验分布**[2]。

我们注意到，与 (9) 中的标准表述相比，(15) 中的表述实质上交换了对数与期望的顺序。尽管这看似只是一个细微差别，但它关系到关注点究竟是推断还是预测。在标准推断中，目标是确定最佳参数。这里，对数似然是一个评分，用于衡量参数 $\boldsymbol{\theta} \sim q(\boldsymbol{\theta})$ 对数据 $\boldsymbol{y}$ 的描述效果，并通过计算给定参数值所诱导的每个模型的平均评分来确定最佳解。相比之下，在面向预测的推断中，主要关注的量是后验预测分布 π_{pred}。评分是预测分布的对数概率，它考虑了整个分布 $q(\boldsymbol{\theta})$；因此，最佳解是平均模型的评分。我们也可以根据詹森不等式理解期望与对数的顺序所带来的影响:

$$-\log \int_{\Theta} \pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta} \leq -\int_{\Theta} \log \pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

其中，仅当模型设定正确时，该界才是紧的，这表明渐近地，在不存在模型设定错误的情况下，式（9）和（13）的解相互对应 [18, 24, 22]。为提供进一步的直观理解，图2展示了变分推断与预测变分推断之间的示例性比较，其中数据生成信号为二次函数，而逆模型被错误地设定为线性模型。图2描绘了标准贝叶斯推断所伴随的不确定性低估现象：后验预测分布为校准（观测）数据分配了较低的概率。相比之下，面向预测的方法为数据分配了较高的概率，从而得到足够宽广、能够涵盖数据，但又不会宽泛到失去信息意义的不确定性刻画。

本研究探讨了面向预测的推断是否适用于大规模工程应用。在这些应用中，模型通常计算成本高昂，且一般由偏微分方程支配。此外，数据通常

[2]请注意， [22] 将由此得到的后验分布称为预测导向（PrO）后验分布。

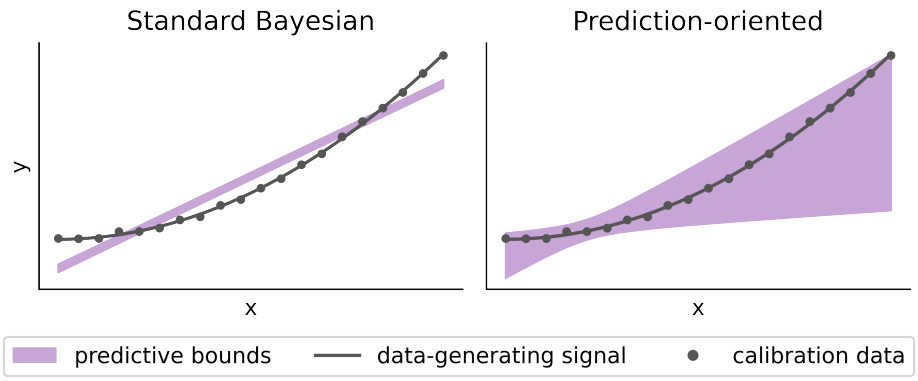


图2： 示意图比较了标准贝叶斯变分推断与预测变分推断：真实的数据生成信号为二次函数 $y = 3x^2 + 1$，而推断模型被错误地设定为线性函数$y = \theta_0 + \theta_1 x$。

有限，我们往往只能获取高维数据集的一次实现。这一研究重点与 [22]中考虑的统计模型和机器学习应用有显著差异。此外，我们旨在利用面向预测的后验分布，为模型形式表示和模型参数提供信息。因此，有必要考虑在此类情境下求解（13）的实际影响。首先需要注意的是，即使能够获得显式似然函数，（15）中定义的损失函数也始终必须进行近似。与标准变分推断相比，对数似然的蒙特卡洛近似在数值上稳定得多，而直接估计以下边际似然则不然：

$$\pi_{\text{pred}} \approx \frac{1}{n} \sum_{i=1}^{n} \pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta}^{(i)}), \quad \boldsymbol{\theta}^{(i)} \sim q(\boldsymbol{\theta}), \tag{16}$$

这在数值上可能比标准方法中的边缘对数似然更具挑战性。由于每次优化过程迭代时都要进行这种近似，因此可能需要大量模型评估。

除了可能需要大量前向模型评估之外，式 (16) 中的方法还依赖显式似然函数，而这在实践中并非总是可用。在这种情况下，可以考虑无似然方法或隐式似然方法。该方法并不直接评估似然，而是从预测分布中采样，并利用密度估计来评估式 (15) 中给出的损失函数；有关针对加性噪声和核密度估计（KDE）的详细方法说明，请参见算法 1。其中一个挑战是，

算法 1： 对式 (15) 中损失函数进行无似然评估

输入： 观测数据 $\boldsymbol{y}^{\text{obs}} \in \mathbb{R}^{n_y}$、参数到观测量映射 $f(\boldsymbol{\theta})$: $\Theta \to \mathbb{R}^{n_y}$、噪声分布$\pi_\epsilon(\boldsymbol{\epsilon})$、采样密度 $q(\boldsymbol{\theta})$、核函数 K以及所需样本数量 $n > 0$**从预测分布中采样：**

对于 $i = 1, \ldots\ldots, n$ **执行**生成 $\boldsymbol{\theta}^{(i)} \sim q(\boldsymbol{\theta})$ 采样噪声 $\boldsymbol{\epsilon}^{(i)} \sim \pi_\epsilon(\boldsymbol{\epsilon})$ 将数据模拟为 $\boldsymbol{y}^{(i)} = f(\boldsymbol{\theta}^{(i)}) + \boldsymbol{\epsilon}^{(i)}$**结束 核密度估计：**
估计带宽 h[3]$\hat{\pi}_{\text{pred}}(\boldsymbol{z}) = \frac{1}{nh^{n_y}} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{z}-\boldsymbol{y}^{(i)}}{h}\right)$

输出： 损失函数评估 $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{y}^{\text{obs}}) = -\log \hat{\pi}_{\text{pred}}(\boldsymbol{y}^{\text{obs}})$

[3]虽然可以直接指定带宽，但通常会将其作为算法的一部分进行估计。本研究采用了 `scipy.stats.gaussian kde` 斯科特法则 [7, 34]的默认设置。

然而，应用这种方法时面临的问题是，数据通常具有高维特征，而密度估计在这种情况下难以扩展。因此，为了使方法具有可行性，我们利用以下逐分量近似来近似该密度

$$K\left(\frac{\boldsymbol{z}-\boldsymbol{y}^{(i)}}{h}\right)=\prod_{j=1}^{n_y}K_j\left(\frac{z_j-y_j^{(i)}}{h_j}\right), \tag{17}$$

其中，下标 $j=1$、.....、n_y 表示感兴趣的分量； $\boldsymbol{z}$ 或 z 表示函数的自变量； K 和 K_j 分别表示所考虑的多变量和单变量核，在下文的计算结果中，这些核均为高斯核。逐分量近似仍能确保预测分布在边际意义上描述数据，但它对所得后验分布的影响可能会影响不确定性外推的有效性。未来的工作可以探索适用于中等维度的其他密度估计方法，其优点包括无需显式似然函数，因此可适用于更广泛的问题。此外，密度估计能提供完整的预测分布；一旦构建完成，便可高效地评估不同数据收集策略下的预测分布，几乎不会增加额外计算成本；例如，这在最优实验设计中可能非常有用。

3.2 模型形式不确定性表示

本研究的一个新颖目标，是评估 PVI 在校准模型形式不确定性（MFU）表示方面的表现。模型设定错误通常源于简化的模型假设（例如闭合模型/本构模型、被忽略的相互作用、假设的各向同性），这些假设是由于对底层过程的了解有限，或难以以计算上可行的方式对这类复杂性进行建模而产生的。MFU 表示旨在更好地表示与这类假设相关的不确定性，从而提升模型的预测和外推性能 [26, 25, 27, 4, 5, 29]。这类表示被表述为对建模假设的参数化修改。首先考虑以下描述状态变量演化的标准模型，即可用一般数学形式描述这类表示： v

$$0=\mathcal{R}_a(v), \tag{18}$$
$$d=\mathcal{O}(v)+\epsilon_m,\quad \epsilon_m\sim\pi(\epsilon_m). \tag{19}$$

在式（18）中，控制方程以残差形式表示，其中 a 代表需要作出假设的模型方面；例如本构方程、耦合项，甚至是被忽略的项。随后，通过作用于状态变量的观测算子 $\mathcal{O}$ 确定校准数据 d，其中包含测量噪声 $\epsilon_m\sim\pi(\epsilon_m)$。

MFU 表示嵌入控制方程中的方式如下，具体体现为对式（18）–（19）的以下修改：

$$0=\mathcal{R}_{\xi_\gamma}(v), \tag{20}$$
$$d=\mathcal{O}(v(\xi_\gamma))+\epsilon_m,\quad \epsilon_m\sim\pi(\epsilon_m), \tag{21}$$

其中，MFU表征 ξ_γ 替代或补充 a，以表示假设中的不确定性。我们指出，将该表征直接嵌入控制方程，旨在使模型预测能够外推至观测数据范围之外。尽管许多现实工程学问题都需要进行外推，但现有的模型差异或模型形式误差建模方法 [12, 31, 30] 在将预测外推至与校准中所考虑情形有显著差异的状态或场景方面仍显不足。通过直接表示模型设定错误的潜在来源，MFU 表示能够减轻模型设定错误对推断的影响；然而，它们通常无法彻底消除模型设定错误。

虽然 MFU 表示为刻画源自模型形式的不确定性提供了一种方法，但长期以来，这类不确定性的校准一直颇具挑战。由于目标是刻画在功能上不可约的模型形式不确定性，因此当存在模型设定错误时，我们希望 MFU 表示的不确定性在无限数据的极限下不会集中。因此，标准贝叶斯范式和广义贝叶斯范式往往并不适用，现有方法转而采用层次方法。在这种方法中，MFU 表示的参数被建模为随机变量，其对应的概率分布本身也由参数化分布来描述。这种方法的基本思想是，即使超参数的不确定性趋于集中，只要对这些分布进行恰当设定，MFU 参数的不确定性描述就不会坍缩。然而，这种层次化方法在理论和实践上都面临若干挑战。首要问题是，

层次化方法往往会导致或加剧对预测不确定性有显著影响的参数的可辨识性问题。因此，要进行有意义的推断，就需要信息量很高的超先验。然而在实践中，我们很少能获得构建信息量高的超先验所需的信息或直觉。参见 [28]，了解将层次化推断应用于 MFU 校准所面临挑战的更深入讨论。

本研究探讨了面向预测的推断，将其作为校准此类 MFU 表征的一种替代方法；这种方法可能更直观，也能提供更多信息。由于存在模型设定错误时，面向预测的后验分布不会收缩，因此它为表征模型形式不确定性提供了一种自然范式。此外，这类框架通过确保预测分布为校准数据赋予较高概率，与利用 MFU 表征实现外推预测的目标相契合。从实践角度看，具有描述性但又不过于宽泛（缺乏信息）的后验预测分布，能够增强人们在外推此类不确定性时的信心。另一方面，MFU 表征可以通过减轻模型设定错误，助力面向预测的推断问题，并有望收紧不确定性界限。我们将在第4.2节探讨这种新型推断方法是否适用于构建 MFU 表征，以及 MFU 表征对逆问题解的影响。

4 计算结果

这里，我们展示将第3.1节提出的面向预测的框架应用于具有工程应用相关特征的系统所得的结果。首先，我们在第4.1节考察一个多项式示例，该示例的后验分布及其预测分布具有闭式表达式。因此，我们可以考察模型设定错误加剧以及噪声增加对所得解的影响。此外，在第4.1节中，我们还研究通过以下方法近似预测密度的影响：1）蒙特卡洛估计量；2）分量密度估计。对于具有高维数据的非线性物理模型，这类近似可能是必要的。在第4.2节中，我们将这一框架应用于模型形式不确定性表示。具体而言，我们采用污染物在某个区域内迁移的偏微分方程模型，其中模型设定错误源于对弥散过程了解不准确。我们将展示，与标准贝叶斯推断相比，MFU 表征与面向预测的推断相结合，如何能够更好地校准不确定性。

我们的推断算法和计算模型均使用 JAX Python 库（https:// docs.jax.dev/）实现，并采用 Optax（https://optax.readthedocs.io/）中的 Adam 实现，以学习率 10^{-2}最小化目标函数。除非另有说明，我们在 VI 和 PVI 中均采用多元正态分布作为变分族，初始猜测为 $\mathcal{N}(0,\ I)$，并使用 100 次模型评估来估计目标函数中的“与数据一致性”项——对于 VI，该项为平均对数似然；对于 PVI，则为对数预测密度。变分族的协方差通过其 Cholesky 因子的下三角元素进行参数化，并对对角元素应用 softplus 函数，以确保正定性。多项式示例使用 5000 个训练步骤，污染物输运示例使用 10000 个训练步骤；训练步数的选择以观察到损失函数收敛（即损失函数渐近稳定）为准。

4.1 多项式示例

本节将探讨多项式示例，借助解析公式研究标准变分推断（VI）和预测变分推断（PVI）问题的解，从而避免近似误差。我们首先介绍模型和逆问题表述，随后探讨模型设定错误程度和噪声增加对逆问题解的影响。接着，我们研究预测密度的各种近似对结果的影响，包括蒙特卡洛估计量的收敛性和分量密度估计。

首先，假设真实的数据生成过程由以下多项式模型给出

$$y^{\mathrm{obs}}(x_i) = 2x_i^p + 1 + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_{\mathrm{noise}}^2), \tag{22}$$

其中，真实的数据生成参数为 $\boldsymbol{\theta}^{dg} = [a^{dg},\ b^{dg}]^\top = [2,\ 1]^\top$，观测数据中的噪声为 $\sigma_{\mathrm{noise}} = 0.4$，而 $p \in (1, 3]$ 控制非线性程度，进而控制模型设定错误的程度。离散化点 x_i 由 $n_y = 40$ 个从 $[0,\ 2]$ 均匀间隔取值的点计算得到。我们将观测向量记为 $\boldsymbol{y}^{\mathrm{obs}} = [y_1^{\mathrm{obs}}, \ldots, y_{n_y}^{\mathrm{obs}}]$。

为了理解模型设定错误对标准逆 问题和面向预测的逆问题解的影响，我们假设存
rse 问题中，我们假设存在 i
是被忽略的非线性。因此，假定的推断模型表示为

$$f(\boldsymbol{\theta}; x) = ax + b, \quad x \in [0,\ 2], \tag{23}$$

其中，不确定参数由 $\boldsymbol{\theta} = [a,\ b]^\top$ 给出。统计模型的离散形式写为

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}}), \tag{24}$$

其中 $\boldsymbol{\Gamma}_{\text{noise}} = \sigma^2_{\text{noise}}\boldsymbol{I}$ 表示单位矩阵 $\boldsymbol{I} \in \mathbb{R}^{n_y \times n_y}$，且

$$\mathbf{A} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \tag{25}$$

其中 $\boldsymbol{x} = [x_1, \ldots, x_{n_y}]^\top$ 是模型进行评估的离散化位置。(24) 中统计模型的形式给出了似然

$$\pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{A\theta}, \boldsymbol{\Gamma}_{\text{noise}}). \tag{26}$$

给定 (24) 中的线性推断模型和高斯先验 $\pi_{\text{pri}}(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_{\text{pri}},\ \boldsymbol{\Gamma}_{\text{pri}})$，标准贝叶斯后验为高斯分布 $\pi_{\text{post}}(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_{\text{post}},\ \boldsymbol{\Gamma}_{\text{post}})$，其均值和协方差的闭式表达式为

$$\boldsymbol{\Gamma}_{\text{post}} = \left(\boldsymbol{A}^\top \boldsymbol{\Gamma}^{-1}_{\text{noise}} \boldsymbol{A} + \boldsymbol{\Gamma}^{-1}_{\text{pri}}\right)^{-1}, \tag{27}$$

$$\boldsymbol{\mu}_{\text{post}} = \boldsymbol{\Gamma}_{\text{post}} \left(\boldsymbol{A}^\top \boldsymbol{\Gamma}^{-1}_{\text{noise}} \boldsymbol{y}^{\text{obs}} + \boldsymbol{\Gamma}^{-1}_{\text{pri}} \boldsymbol{\mu}_{\text{pri}}\right). \tag{28}$$

请注意，将 $\mathcal{P}(\Theta)$ 限制为多元高斯族时，标准贝叶斯逆问题的这一解恰好就是变分推断解 (9)。

对于预测变分推断问题，我们采用 (15) 中给出的损失函数；在线性高斯情形下，对于 $\mathcal{Q} \sim \mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$，推前密度和预测密度的闭式表达式为

$$\pi_{\text{pf}}(\boldsymbol{z}) \sim \mathcal{N}\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T\right), \tag{29}$$

$$\pi_{\text{pred}}(\boldsymbol{z}) \sim \mathcal{N}\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T + \boldsymbol{\Gamma}_{\text{noise}}\right). \tag{30}$$

为了帮助区分模型参数与假定噪声模型各自带来的不确定性贡献，接下来的结果中我们同时评估后验预测分布和推前密度。鉴于我们的目标是利用模型预测的不确定性为决策提供依据，确保后验推前不确定性表征具有意义是关键目标。然而，在考量不确定性对观测数据的描述效果时，我们利用预测分布来纳入观测数据中的噪声。为直观呈现这些分布，我们绘制逐分量（边缘）95%可信区间，其计算方式为

$$\text{CI}_{95}(x_i) = (\boldsymbol{A\mu})_i \pm 1.96\sqrt{(\boldsymbol{\Gamma}_y)_{ii}}, \tag{31}$$

其中，$\boldsymbol{\Gamma}_y$ 对应于式（29）或式（30）中的协方差，具体取决于所考虑的是预测分布还是推前分布。

首先，随着模型设定错误程度增加，我们比较标准变分推断与预测变分推断（PVI）的结果。这里，先验分布定义为

$$\pi_{\text{pri}}(\boldsymbol{\theta}) \sim \mathcal{N}\left(\begin{bmatrix} 3 \\ 0.3 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2.4 \end{bmatrix}\right).$$

在图3中，我们比较了随着模型设定错误程度增加（从左到右），标准变分推断（上排）和预测变分推断（下排）所对应的后验推前及预测密度的95%可信区间。从图3的上排可以看出，即使模型设定错误程度增加，标准变分推断仍会低估不确定性。也就是说，随着模型越来越难以拟合，预测不确定性并不会显著增加，而是均值发生偏移，以最小化式（9）中的损失函数。相比之下，图3下排中的预测变分推断结果表明，通过直接惩罚预测不确定性，面向预测的方法能够随着推断模型变得越来越难以拟合，通过增加不确定性来适应模型设定错误的程度。在无限数据的极限情况下，若考虑

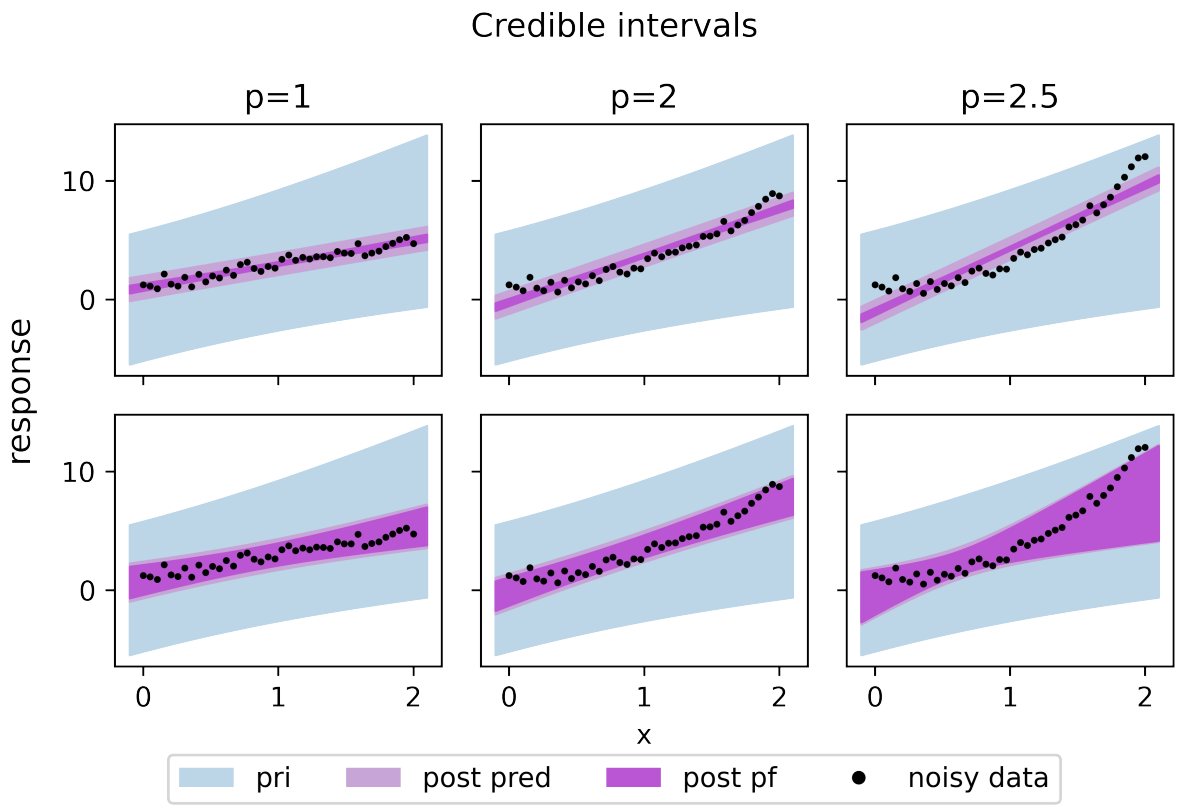


图3：比较不同程度模型设定错误 p 下，**标准变分推断（上排）**与**预测变分推断（下排）**对应的先验预测分布、后验预测分布和推前分布（pf）的95%可信区间（见式（1））。

数据各次实现上的平均损失，那么只有在不存在模型设定错误时，预测变分推断才会收敛到贝叶斯解 [22]。然而，正如图3第一列所示，对于有限数据，面向预测的方法得到的解与标准方法不同，而是具有更大的后验不确定性，因此预测不确定性也更大。

类似地，图4展示了随着模型设定错误程度逐渐增加（从左到右），标准变分推断（上排）与预测变分推断（下排）对应的先验与后验等高线。从图4上排可以看出，随着模型设定错误程度增加，后验均值逐渐偏离真实的数据生成参数 $\boldsymbol{\theta}_{dg} = [2,\ 1]^\top$，偏差越来越大。图4下排也呈现出类似趋势，但在模型设定错误程度最高时，后验均值的偏差较小（$p = 2.5$）。值得注意的是，采用以预测为导向的框架会增加后验不确定性，使后验预测分布能够更好地描述观测数据。一个关键区别是，对于标准变分推断，即使存在模型设定错误，后验分布最终仍会渐近收缩到一个流形上，该流形由参数空间的维度决定，并由参数到观测量映射提供信息。从图3和图4右上角的图中可以看出，这个流形可能无法很好地涵盖真实参数值，从而导致预测不确定性较差。然而，对于预测变分推断，尽管后验偏差会随着模型设定错误程度增加，但预测不确定性表征仍会为观测数据赋予较高概率。

接下来，我们探讨噪声水平升高对变分推断（VI）和预测变分推断（PVI）问题的解的影响。在标准贝叶斯范式下，噪声增大会有效降低数据所包含的信息量，从而减弱式（3）中似然的权重，并相对增强先验的影响。由于预测变分推断问题仍遵循类似的优化目标——其组成部分既体现对观测数据的拟合，也体现对先验信念的遵循——因此，随着噪声增大、先验的影响变强，参数不确定性也会相应增加。首先，来看图5，其中展示了随着噪声水平升高（从左到右），标准变分推断（上排）与预测变分推断（下排）对应的后验预测分布和推前密度的95%可信区间。请注意，观测数据（根据式（22）生成）以及式（26）中的假定似然模型

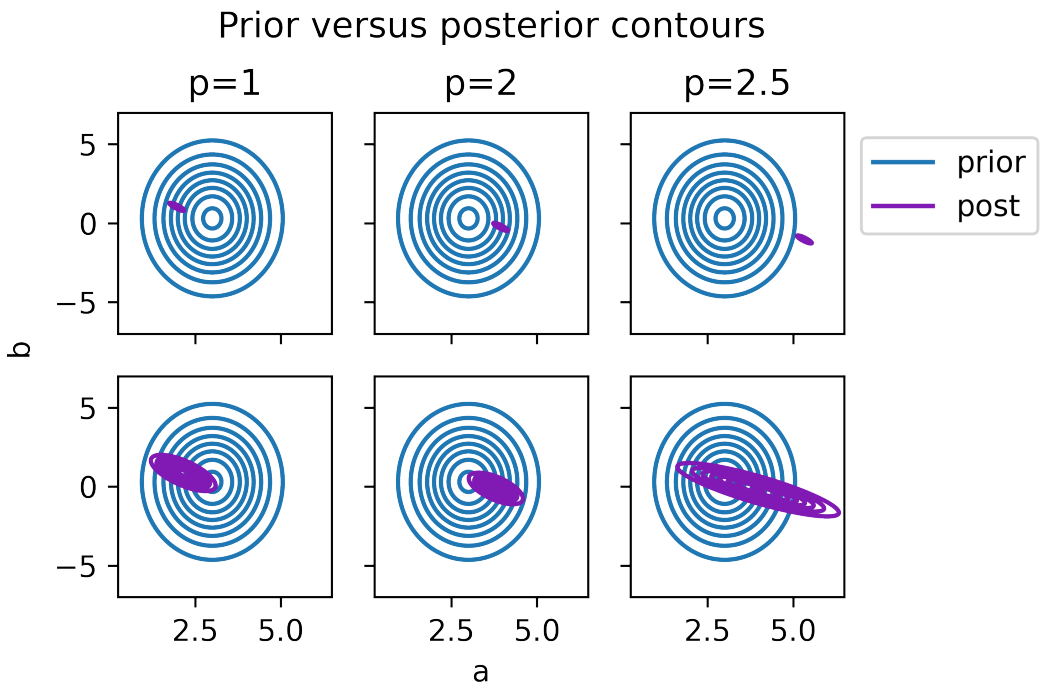


图4： 比较模型设定错误程度逐渐增加时对应的先验与后验等高线：**标准变分推断（上排）**与**预测变分推断（下排）**。模型设定错误程度由式 (22) 中 p 的递增取值表示。

都利用了增大的噪声。可以看到，随着噪声增大，变分推断和预测变分推断的预测不确定性均有所增加。为了直接比较噪声水平升高时的逆问题解，请看图5和图6，二者分别展示了相对于先验的后验边缘密度和等高线。从这些图中可以看到，随着噪声水平升高，后验不确定性也随之增加，因为后验逐渐趋近于先验；由于在式（22）中设置 $p = 2$ 所导致的模型设定错误，预测变分推断在每个噪声水平下的不确定性都更大。

现在，我们来考察预测密度近似所产生的影响。首先，假设可以获得显式似然函数，那么在变分优化问题的每次迭代中，计算式 (9) 中的标准变分推断损失或式 (15) 中的预测变分推断损失，都需要对期望对数似然或预测密度（见式 (16)）进行蒙特卡洛近似。这里，我们利用多项式示例中期望对数似然和预测密度的闭式表示，比较蒙特卡洛抽样所产生的误差。在图7中，我们比较了 100 组重复样本的蒙特卡洛近似标准变分推断损失（左图）与预测变分推断损失（右图）的百分比误差直方图，每组样本的大小为 $n = 10^2$，10_3，10_4。近似的预测变分推断损失定义为负对数预测密度，其中预测密度按照式 (16) 进行近似。类似地，标准变分推断损失定义为负期望对数似然，并采用以下蒙特卡洛估计量

$$\mathbb{E}[\log(\pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta}))] \approx \frac{1}{n}\sum_{i=1}^{n}\log(\pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{\theta}^{(i)})), \quad \boldsymbol{\theta}^{(i)} \sim q(\boldsymbol{\theta}). \tag{32}$$

在这两种情况下，$q \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$，其中 $\boldsymbol{I} \in \mathbb{R}^{2\times 2}$ 是单位矩阵。从图7可以看出，对于 $n = 10^2$，估计得到的预测变分推断损失的误差可能高达 $\approx 600\%$，而标准变分推断损失的误差最多为 $\approx 20\%$。此外，我们看到，需要 $n = 10^4$ 个蒙特卡洛样本，预测变分推断损失才能收敛到真实的（解析）值；对于许多依赖昂贵正向模拟的问题而言，这一要求是难以承受的。一般而言，即使控制了下溢/上溢数值问题[4]，，对预测分布（见式（16））进行近似所需的蒙特卡洛样本数量也可能显著多于 n 评估边缘对数似然所需的数量。这种不稳定性源于这样一个事实：似然函数在残差上呈

[4] ~~`logsumexp` 通过利用恒等式 $\log\left(\sum_i e^{w_i}\right) = m + \log\left(\sum_i e^{w_i - m}\right)$（其中 $m = \max_i w_i$ [3]）~~，在计算对数边际似然时提供数值稳定性。

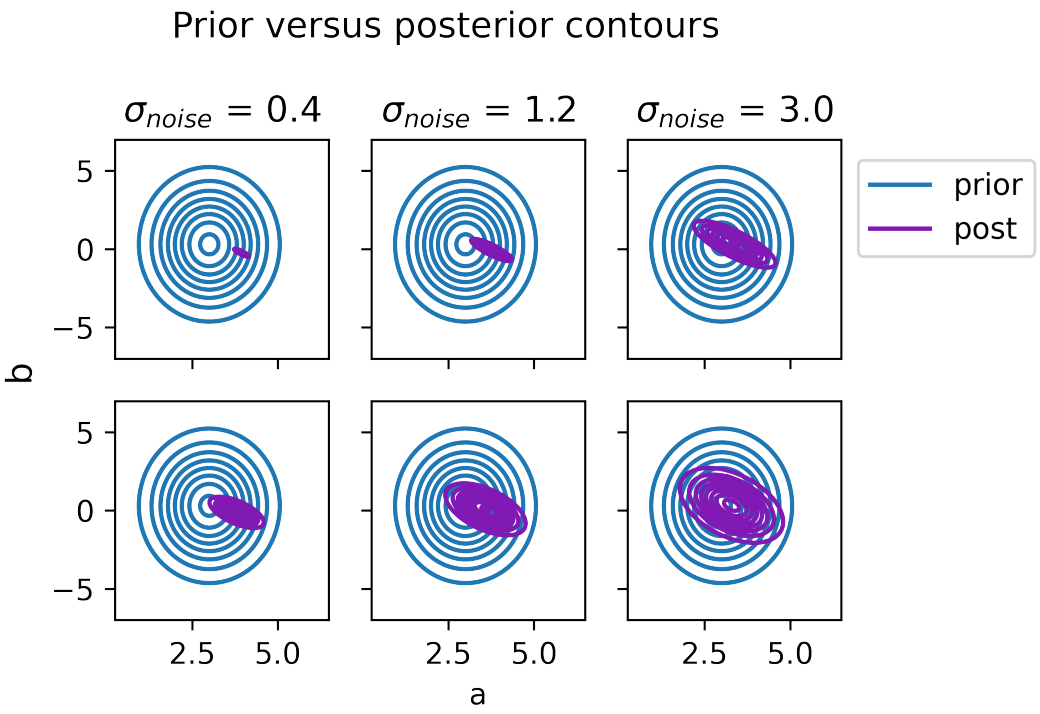


图5： 在固定模型设定错误程度 $p = 2$ 但噪声水平 σ_{noise} 逐渐增加的情况下，比较式(1)中**标准变分推断（上排）**与**预测变分推断（下排）**对应的先验与后验等高线。

指数形式 $f(\boldsymbol{\theta}) - y$，这意味着在参数空间的大部分区域内，似然几乎为0，而在该空间的一小部分区域内却非常大；对于高维数据，这种现象会进一步加剧。相比之下，对数似然关于残差呈二次形式，因此收敛更快。由于必须在优化问题的每次迭代中估计预测分布，因此，与标准变分推断相比，在直接利用显式似然时，解决预测变分推断问题可能需要数量显著更多的似然函数（进而也需要更多模型）评估n 。

接下来，我们探讨如何利用逐分量密度估计，从样本中近似预测分布，而无需显式计算似然函数。密度估计颇具吸引力，因为它不依赖似然函数，并且能够评估多个可能数据对应的损失函数；然而，密度估计在维度较高时扩展性较差，因此通常会采用这类逐分量近似。因此，求解预测变分推断问题时，可能需要依赖对预测密度进行逐分量估计。对于线性高斯问题，我们可以对这一逐分量假设进行解析分析，从而无需考虑密度估计引入的误差。这里，我们假设式(30)中给出的预测密度近似为

$$\pi_{\text{pred}}(\boldsymbol{y}^{\text{obs}}) \approx \mathcal{N}\left(\mathbf{A}\boldsymbol{\mu}, \boldsymbol{\Gamma}_{\text{pred}}\right) \tag{33}$$

$$= \prod_{i=1}^{n_y} \mathcal{N}\left((\mathbf{A}\boldsymbol{\mu})_i, (\boldsymbol{\Gamma}_{\text{pred}})_{ii}\right). \tag{34}$$

其中， $\boldsymbol{\Gamma}_{\text{pred}} = \text{diag}\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T + \boldsymbol{\Sigma}_{\text{noise}}\right)$ 忽略了预测密度中的协方差。通过这一近似，我们可以将对数预测密度计算为平方误差项之和，这与标准变分推断公式中的常见做法一致[5]。

对于固定的噪声 $\sigma_{\text{noise}} = 0.4$ 和模型设定错误程度 $p = 2$，图8通过后验协方差的迹和行列式来比较后验等高线与参数不确定性。迹给出参数方差之和（忽略相关性），而行列式则与不确定性椭球的超体积成正比。从图8可以看出，参数不确定性（以协方差的迹和行列式刻画

[5]由于噪声相互独立，联合高斯似然可分解为 $\pi_{\text{like}}(\boldsymbol{y}^{\text{obs}}|\boldsymbol{\theta}) = \prod_{i=1}^{n_y} \pi_{\text{like}}(y_i^{\text{obs}}|\boldsymbol{\theta})$。

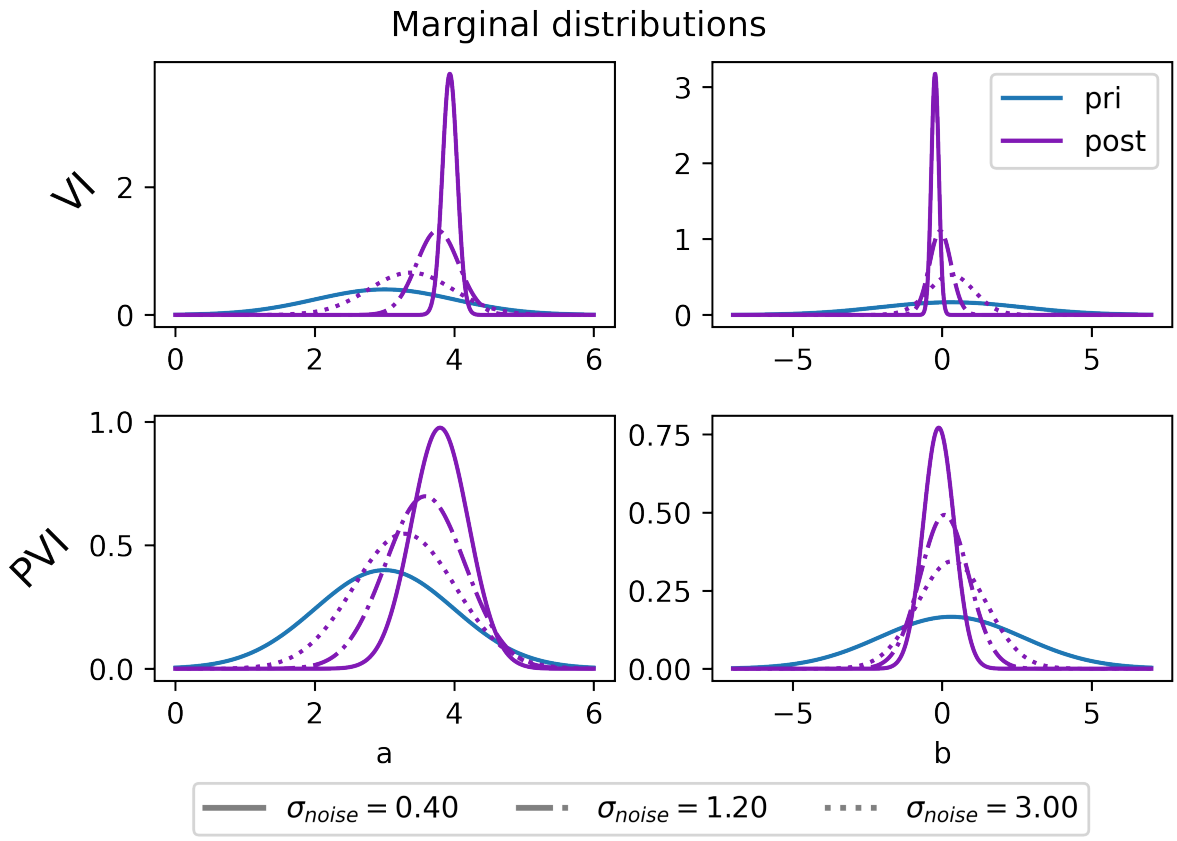


图6： 比较**标准变分推断（上排）**与**预测变分推断（下排）**对应的先验和后验边缘分布，展示噪声水平 σ_{noise} 在（1）中逐渐增加的情况（图例中以不同线型表示）。

）在损失函数中利用完整预测分布时会降低。直观而言，这是因为将协方差结构纳入预测分布后，逆问题受到更强的约束。图9展示了对预测分布的影响：其中比较了协方差迹所表征的预测不确定性（左），以及在式（15）的预测变分推断损失中利用完整多元预测密度（中）与逐分量近似（右）所对应的预测可信区间。从图9可以看出，逐分量假设会使后验预测分布的逐分量不确定性平均更大，这一点可由逐分量情形下后验预测分布的迹增加看出[6]。一般而言，忽略预测分布中的协方差结构会导致不确定性增加，因为无法利用底层相关结构来更好地约束参数。概括来说，如果只要求后验预测分布在边缘意义上能够良好描述数据，那么可能解的空间就不会受到那么多约束。然而，如果目标是外推不确定性，忽略相关性可能会影响所得不确定性表征的外推能力。未来的研究可以探讨这一假设对外推的影响，并研究能够在更高维度进行密度估计的方法。

[6]虽然可以通过协方差的行列式评估预测不确定性的大小，但由于数据维度为 40 维，这里的行列式非常小；而受参数影响的重要维度为 2 维，因此与迹的比较相比，这种比较的意义较小。

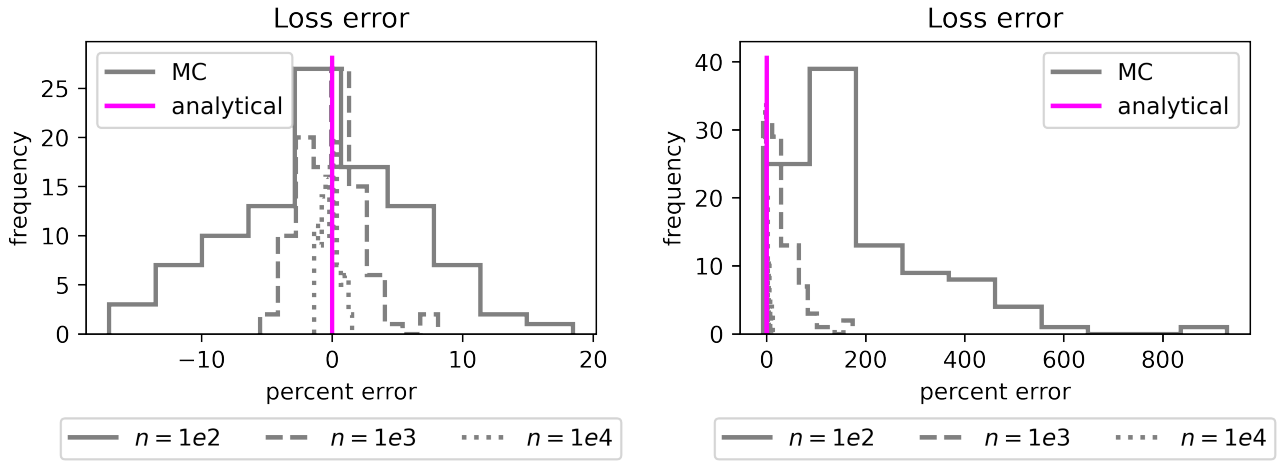
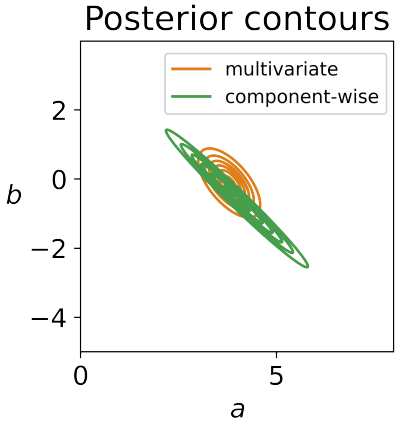


图7： 比较基于蒙特卡洛样本近似计算的**变分推断损失函数（左）**（使用式（32））与**预测变分推断损失函数（右）**（使用式（16））的百分比误差直方图；每种情况均使用 100 组重复蒙特卡洛样本，样本量为 $n = 10^2$，10^3，10^4。



	行列式	迹
多变量	**0.02674**	**0.435**
逐分量	0.02770	1.956

图8： 对于**预测变分推断**，比较在损失函数中采用完整多元预测分布与逐分量近似的效果。左图给出了后验分布的等高线；右侧表格列出了后验协方差的迹和行列式，以表征相应的不确定性。

	迹
多变量	**15.4**
逐分量	20.3

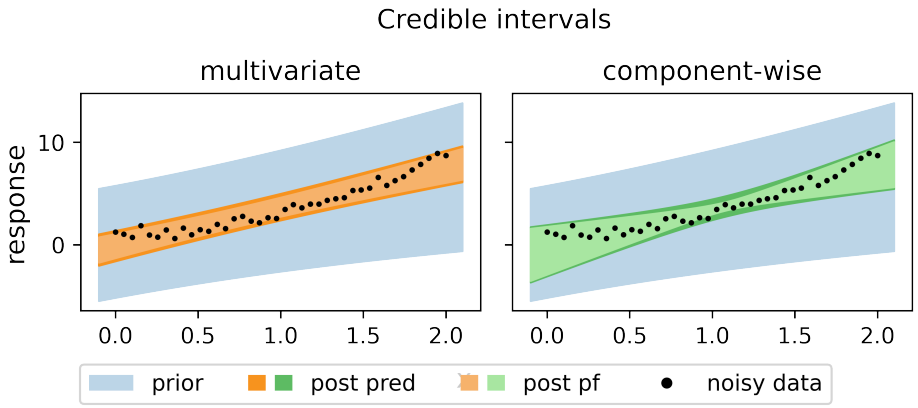


图9： 对于**预测变分推断**，比较了先验预测和后验预测分布及其推前分布的95%可信区间；这些分布对应于基于完整**多变量高斯预测分布（左）**与**逐分量近似（右）**构建的损失函数。后验预测协方差矩阵的迹列于表中。

4.2 非均质多孔介质中的污染物迁移

现在，我们探讨在校准 MFU 表示时如何使用面向预测的推断。在此，我们考虑污染物浓度场的演变，目标是根据特定时刻的含噪浓度测量值，推断模型中的不确定参数。假设污染物浓度 $c^{\mathrm{obs}}(x,\ t)$ 随时间 t 和空间 x 的真实演变由以下广义平流-扩散方程描述：

$$\begin{aligned}
\frac{\partial}{\partial t}c^{\mathrm{obs}}(x,t)+u\frac{\partial}{\partial x}c^{\mathrm{obs}}(x,t) &= \nu_p\frac{\partial^2}{\partial x^2}c^{\mathrm{obs}}(x,t)+\mathcal{D}c^{\mathrm{obs}}(x,t),\quad x\in[0,L_x=4], && (35)\\
c^{\mathrm{obs}}(x,0) &= \exp\left(-\frac{1}{2}\left(\frac{(x-0.85)^2}{(0.02)^2}\right)\right), && (36)\\
c^{\mathrm{obs}}(0,\ t) &= c^{\mathrm{obs}}(L_x,\ t), && (37)
\end{aligned}$$

其中，u 是地下水的体积速度，ν_p 是孔隙尺度扩散系数，$\mathcal{D}$ 是表示弥散效应的一般线性算子。由于该模型关于 c^{obs} 是线性的，且具有周期性边界条件，其解可表示为以下傅里叶级数

$$c^{\mathrm{obs}}(x,t)=\hat{c}_0^{\mathrm{obs}}(0)+2\Re\left[\sum_{k=1}^{\infty}\hat{c}_k^{\mathrm{obs}}(0)\exp\left(t\left[-\nu_p k_x^2-u(ik_x)+\lambda_k\right]\right)e^{i\frac{2\pi k}{L_x}x}\right],\quad k_x=\frac{2\pi k}{L_x}, \tag{38}$$

其中，$\hat{c}_k^{\mathrm{obs}}(0)$ 是初始条件的傅里叶系数，λ_k 是 $\mathcal{D}$的特征值，$\Re[z]$ 表示复变量 z的实部。我们将级数截断至 $k=257$ ，并在具有 $N_x=512$的均匀网格上求解。随后按如下方式生成合成观测数据：

$$y^{\mathrm{obs}}(x_i)=c^{\mathrm{obs}}(x_i,t=0.5,u^{dg}=1,\nu_p^{dg}=0.01)+e_i,\quad e_i\sim\mathcal{N}(0,\sigma_{\mathrm{noise}}^2), \tag{39}$$

其中 $\sigma_{\mathrm{noise}}=0.005$，数据生成参数由 $\boldsymbol{\theta}^{dg}=[u^{dg},\ \nu_p^{dg}]$ 给出。数据在 $[0,\ 4]$, 中的 $n_y=64$ 个等间距点处采集，观测数据向量仍记为 $\boldsymbol{y}^{\mathrm{obs}}=[y_1^{\mathrm{obs}},\ldots,y_{n_y}^{\mathrm{obs}}]$。数据生成特征值由直接数值模拟得到，该模拟针对非均质多孔介质中的详细平流-扩散方程，具体如 [27] 所述。

为展示如何将面向预测的后验分布与 MFU 表示结合起来，以提高计算模型的预测能力，我们首先假设在校准过程中忽略数据生成过程中存在的弥散效应。基于这一假设，可通过传统平流-扩散方程定义如下推断模型：

$$\begin{aligned}
\frac{\partial}{\partial t}c^{\mathrm{inv}}(x,t)+u\frac{\partial}{\partial x}c^{\mathrm{inv}}(x,t) &= \nu_p\frac{\partial^2}{\partial x^2}c^{\mathrm{inv}}(x,t),\quad x\in[0,L_x=4], && (40)\\
c^{\mathrm{inv}}(x,0) &= \exp\left(-\frac{1}{2}\left(\frac{(x-0.85)^2}{(0.02)^2}\right)\right), && (41)\\
c^{\mathrm{inv}}(0,t) &= c^{\mathrm{inv}}(L_x,t). && (42)
\end{aligned}$$

推断问题的目标是根据含噪观测数据推断不确定参数 $\boldsymbol{\theta}=[u,\ \nu_p]^\top$ 。与之前一样，式 (40)–(42) 可表示为傅里叶级数解：

$$c^{\mathrm{inv}}(x,t)=\hat{c}_0^{\mathrm{inv}}(0)+2\Re\left[\sum_{k=1}^{\infty}\hat{c}_k^{\mathrm{inv}}(0)\exp\left(t\left[-\nu_p k_x^2-u(ik_x)\right]\right)e^{i\frac{2\pi k}{L_x}x}\right],\quad k_x=\frac{2\pi k}{L_x}, \tag{43}$$

相应的统计模型定义为

$$\boldsymbol{y}=c^{\mathrm{inv}}(\boldsymbol{x},t=0.5;\boldsymbol{\theta})+\boldsymbol{\epsilon},\quad \boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\boldsymbol{\Gamma}_{\mathrm{noise}}), \tag{44}$$

其中 $\boldsymbol{\Gamma}_{\mathrm{noise}}=\sigma_{\mathrm{noise}}^2\boldsymbol{I}$ 和 $\boldsymbol{x}=[x_1,\ldots,x_{n_y}]^\top$，由此得到似然

$$\pi_{\mathrm{like}}(\boldsymbol{y}|\boldsymbol{\theta})\sim\mathcal{N}\left(c^{\mathrm{inv}}(\boldsymbol{x},t=0.5,\boldsymbol{\theta}),\boldsymbol{\Gamma}_{\mathrm{noise}}\right). \tag{45}$$

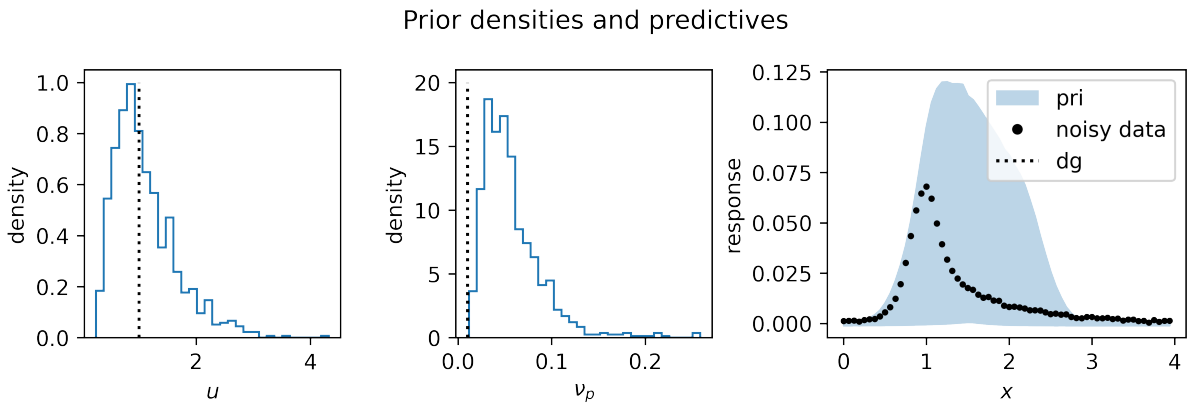


图10： 模型参数 u（左）和 ν_p （中）的先验密度，这些密度由（46）中给出的定义取指数得到；以及相应的先验预测95%可信区间与含噪观测数据（右）的对比。虚线表示数据生成（dg）值。

由于这两个参数均为正值，我们通过它们的对数值设定先验分布：

$$\pi_{\text{pri}}(\log(u)) \sim \mathcal{N}(\log(1), (0.5)^2), \tag{46}$$
$$\pi_{\text{pri}}(\log(\nu_p)) \sim \mathcal{N}(\log(0.05), (0.5)^2), \tag{47}$$

其中，标准差的选取使得 u 和 ν_p 的指数化密度在物理上合理；指数化先验及相应的先验预测95%可信区间如图10所示。

首先，我们比较在标准变分推断与预测变分推断逆问题中采用（43）所定义的推断模型所得结果。由（43）可知，参数到观测量映射关于不确定参数是非线性的。因此，预测密度不存在闭式表示。为此，我们在计算（15）中的损失函数时，采用预测分布的分量级核密度估计近似；详见算法 1 及（17）中的近似。图11展示了标准变分推断（左图）与预测变分推断（右图）对应的后验预测分布和推前分布的95%可信区间。从图11左图可以看出，标准推断低估了不确定性，后验预测分布未能以较高概率覆盖观测数据。相比之下，图11右图显示，预测变分推断得到的后验分布，其推前可信区间涵盖了观测数据，同时又比先验信念更具信息量。图12直接比较了模型参数对数值的标准后验分布（左图）和面向预测的后验分布（右图）等高线与先验分布。可以看出，标准贝叶斯推断导致不确定性表征显著集中，而面向预测的后验不确定性表征并未因模型设定错误而坍缩。图13进一步比较了标准后验分布和面向预测的后验分布的边缘分布，以及真实的数据生成值（水平线）。从图13可以看出，标准变分推断（上排）导致模型参数的估计值集中在有偏估计附近，即远离真实数据生成值。相比之下，预测变分推断的解在参数估计值上的偏差显著较小，尽管 ν_p 的估计值仍存在一定偏差。总体而言，结果表明，即使存在严重的模型设定错误，与标准变分推断相比，预测变分推断也能提供更好的不确定性表征和偏差更小的参数估计值。

现在，我们研究引入MFU表征对标准逆问题和面向预测的逆问题的解的影响。这里，MFU表征描述了弥散效应；定性而言，众所周知，这些效应会对污染物的演化产生非局部影响，但这种非局部性的确切性质尚不清楚。因此，我们在控制对流扩散方程中提出如下弥散的分数阶导数MFU表征：

$$\frac{\partial}{\partial t}c^{\text{mfu}}(x,t) + u\frac{\partial}{\partial x}c^{\text{mfu}}(x,t) = \nu_p\frac{\partial^2}{\partial x^2}c^{\text{mfu}}(x,t) + \nu\frac{\partial^\alpha}{\partial x^\alpha}c^{\text{mfu}}(x,t), \quad x \in [0, L_x = 4], \tag{48}$$
$$c^{\text{mfu}}(x,0) = \exp\left(-\frac{1}{2}\left(\frac{(x-0.85)^2}{(0.02)^2}\right)\right), \tag{49}$$
$$c^{\text{mfu}}(0,\ t) = c^{\text{mfu}}(L_x,\ t)。 \tag{50}$$

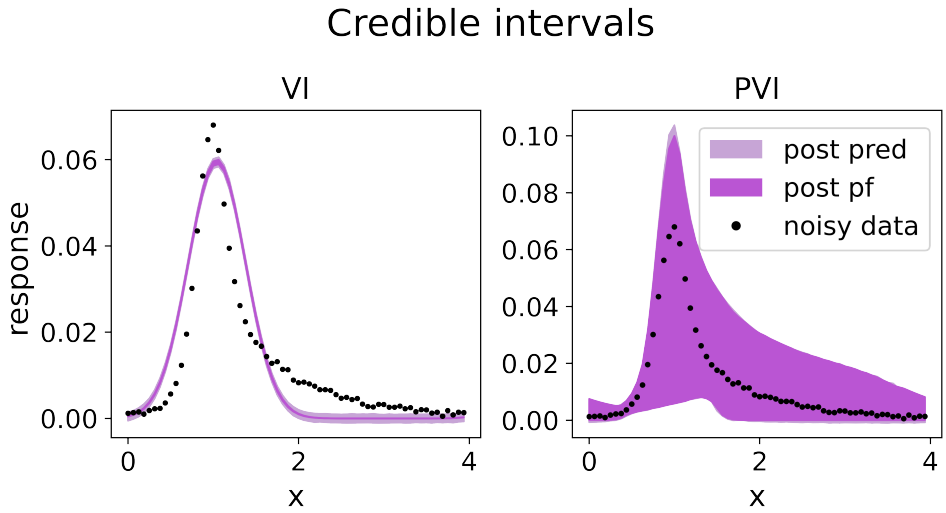


图11： 对比式（43）所定义推断模型中**标准变分推断（左）**与**预测变分推断（右）**对应的后验预测分布与推前分布的95%可信区间。

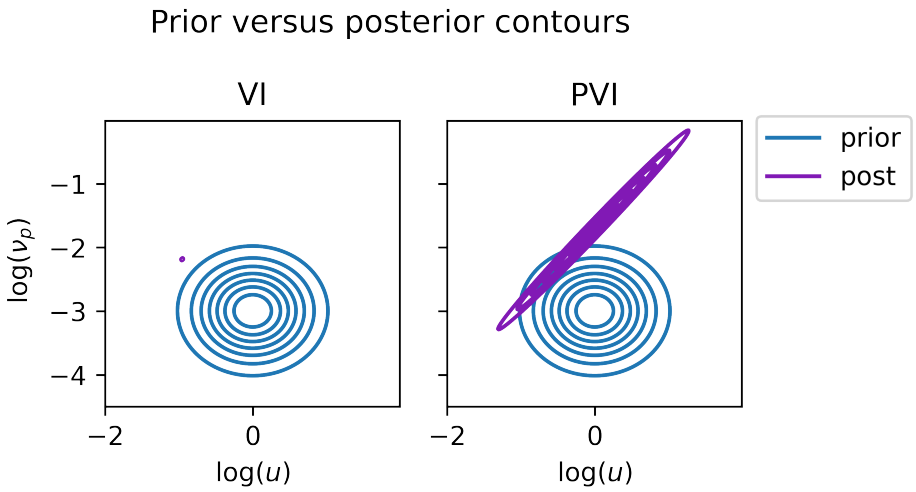


图12： 对比式（43）所定义校准模型中**标准变分推断（左）**与**预测变分推断（右）**对应的变换后参数的先验与后验等高线。

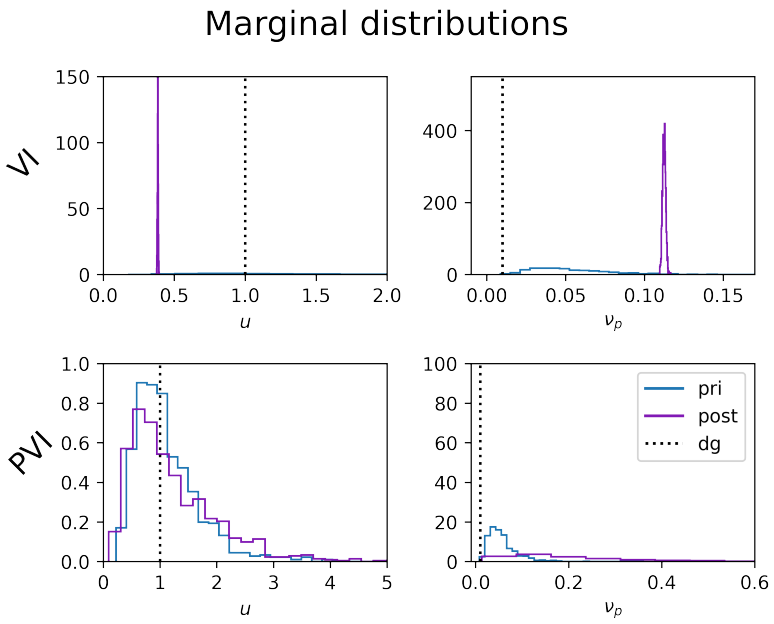


图13： 校准模型（定义见 (43)）中先验分布和后验边缘分布与数据生成（dg）值的比较：**标准变分推断（上排）与预测变分推断（下排）**。

这里，MFU参数为 $[\nu,\ \alpha]$，其中 ν 是模型色散系数，$\alpha \in$ （1，2）是分数阶幂；当其取非整数值时，会对解产生非局部效应，解表示为

$$c^{\mathrm{mfu}}(x,t) = \hat{c}_0^{\mathrm{mfu}}(0) + 2\Re\left[\sum_{k=1}^{\infty} \hat{c}_k^{\mathrm{mfu}}(0)\exp\left(t\left[-\nu_p k_x^2 - u(ik_x) + \nu(ik_x)^\alpha\right]\right) e^{i\frac{2\pi k}{L_x}x}\right], \quad k_x = \frac{2\pi k}{L_x}. \tag{51}$$

请注意，一般而言，特征值并不直接对应 MFU 表示，即 $\lambda_k \neq \nu(ik_x)^\alpha$。因此，尽管 MFU 表征可以通过表示非局部弥散行为来缓解模型设定错误，但无法将其彻底消除。

采用 MFU 表示后，推断问题的目标是确定模型参数和 MFU参数 $\boldsymbol{\theta} = [u,\ \nu_p,\ \nu,\ \alpha]$的分布。我们为 ν 和 α 设定先验分布，以满足其取值范围的约束。对于 ν，我们使用其对数值：

$$\pi_{\mathrm{pri}}(\log(\nu)) \sim \mathcal{N}(\log(0.1), (0.5)^2), \tag{52}$$

而对于 α，我们为 $\alpha - 1$ 的 probit[7] 变换设定先验，此处将其参数化为 $g(\alpha) \equiv \Phi^{-1}(\alpha - 1)$，其中$\Phi^{-1}(\cdot)$ 是标准正态分布的逆累积分布函数，提供

$$\pi_{\mathrm{pri}}\left(g(\alpha)\right) \sim \mathcal{N}(g(1.5), (0.8)^2). \tag{53}$$

这些针对 ν 和 α 变换后取值的先验，分别在区间 (0，.25] 和 (1，2) 内产生非零概率密度。u 和 ν_p 的先验定义与式（46）和（47）相同。图14给出了 MFU 参数先验密度的直方图，以及相应的 95%可信区间。

为了计算面向预测的后验分布，在评估式（15）中的损失函数时，我们再次利用预测分布的分量逐项KDE近似。图15展示了标准变分推断（左图）与预测变分推断（右图）对应的后验预测分布和推前分布的 95%可信区间。从图15左图可以看出，即使纳入了 MFU表征，标准变分推断仍会低估不确定性。不过，对比图15和图11可以发现，尽管 MFU表征未能完整描述数据生成过程的物理机制（定义见式（35）–（37））， 但将其纳入模型有助于提升模型的预测性能，从而更充分地刻画预测不确定性。

[7]这里， `prob it` 变换将 $\alpha - 1 \in$ (0,1) 中的值映射到实数轴上，从而可以设置高斯先验。

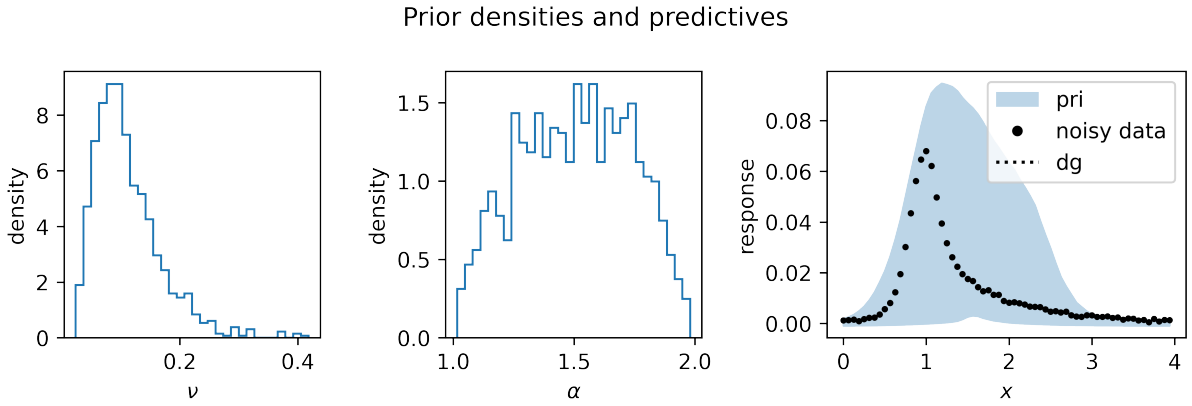


图14： 式（52）和（53）分别定义的 MFU 参数 ν（左）和 α（中）的先验密度，以及相应的先验预测95%可信区间与含噪观测数据的对比（右）。

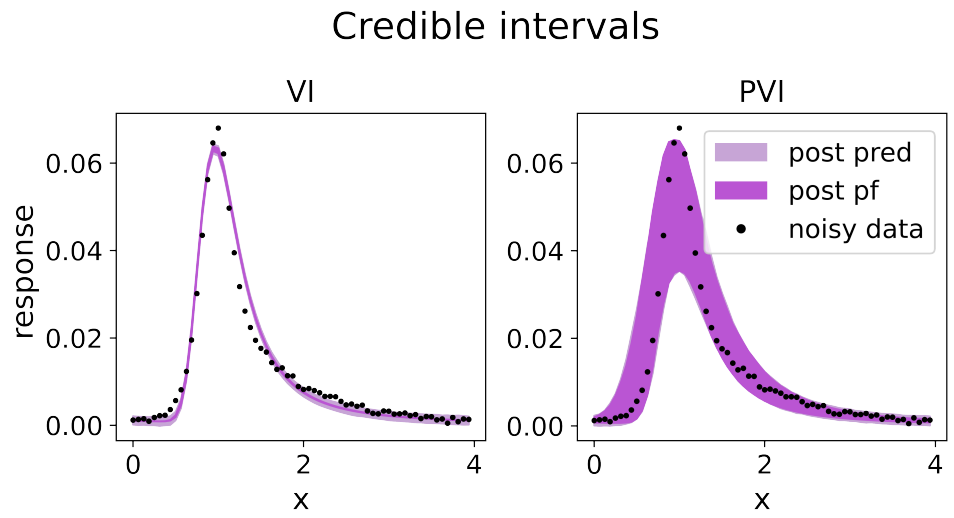


图15： 针对式（51）中定义的推断模型，对比**标准变分推断（左）**与**预测变分推断（右）**对应的后验预测分布和推前分布的95%可信区间。

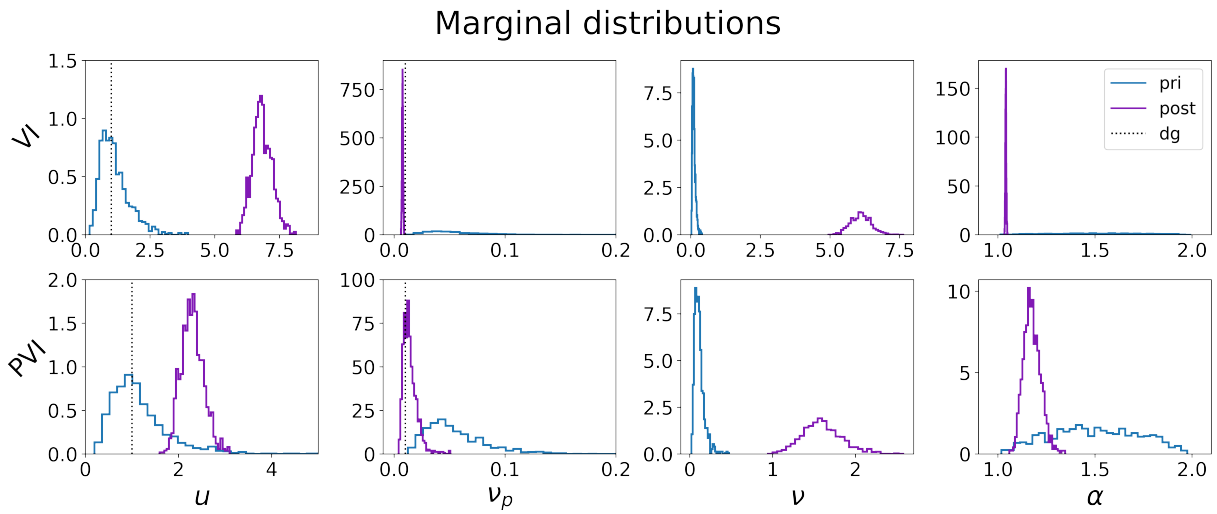


图16：校准模型（51）中，**标准变分推断（上排）**与**预测变分推断（下排）**的先验和后验边缘分布与数据生成（dg）值的比较。

为了解引入 MFU 表征对所得参数估计值的影响，请考察图16所示的模型参数先验和后验边缘分布与真实数据生成（dg）值的对比，其中标准变分推断（上排）与预测变分推断（下排）分别列出。请注意，MFU 参数没有数据生成值，因为它们并不存在于数据生成模型中。将图16与图13对比可见，引入 MFU 表征会影响变分推断和预测变分推断下模型参数的不确定性刻画及偏差。值得注意的是，ν_p 的偏差有所降低；直观而言，这是合理的，因为 MFU 项在一定程度上考虑了弥散（或反常扩散）。相比之下，与图13左侧图示相比，u 的估计偏差增大，这表明 MFU 项的引入和校准会影响模型参数的估计。比较图16的上、下两排可见，预测变分推断能够提供偏差较小的模型参数估计值，即 u；同时也能得到后验分布受到先验支持更充分的 MFU 参数，即 ν。此外，比较变分推断和预测变分推断框架下 ν_p 和 α 的后验分布可知，预测变分推断框架能够减小后验分布的收缩程度。

我们进一步在图17和图18中分别比较了 VI 与 PVI 的变换参数成对先验与后验等高线。可以看到，与标准推断相比，面向预测的后验分布偏差更小、不确定性更大。归根结底，尽管纳入 MFU 表征可以构建预测能力更强的模型，但仅凭这一点并不能确保不确定性得到有意义的校准。事实上，纳入 MFU 表征可能会因参数空间维度增加，以及由此引入的模型参数间潜在虚假依赖关系，而给求解逆问题带来挑战。这些挑战与模型设定错误持续存在的问题相叠加，可能导致参数密度产生偏差。然而，图15至图18所示的结果表明，借助面向预测的推断框架，可以减小偏差并改善最终的不确定性表征。

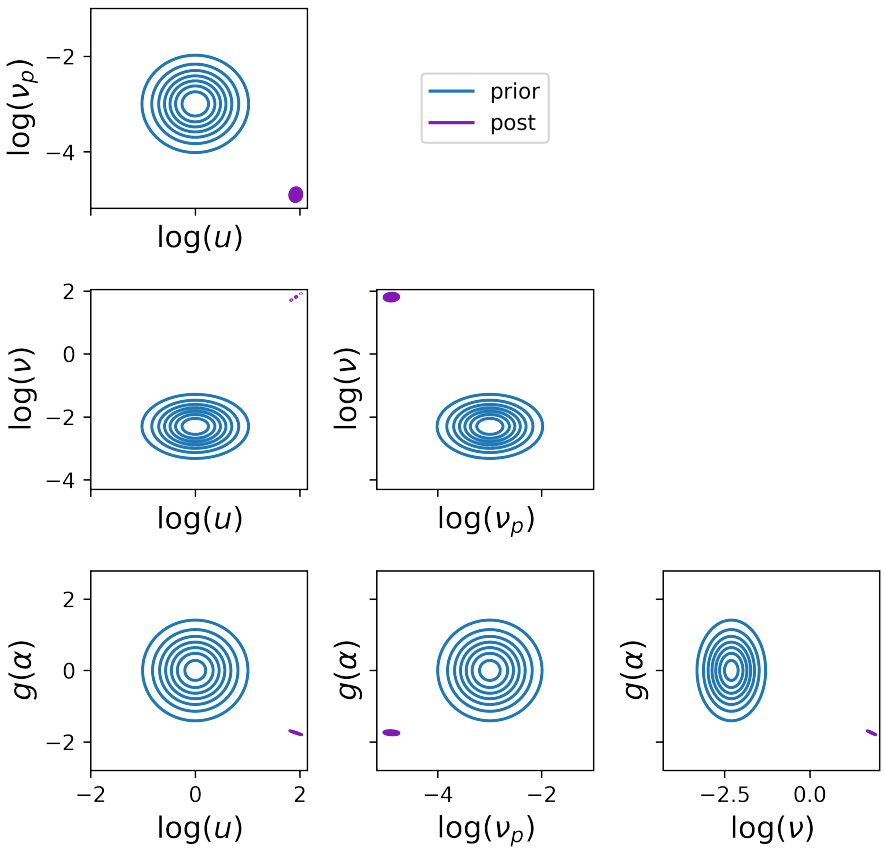


图17：式（51）中定义的校准模型对应的变换参数的先验与后验等高线两两比较**标准变分推断**。

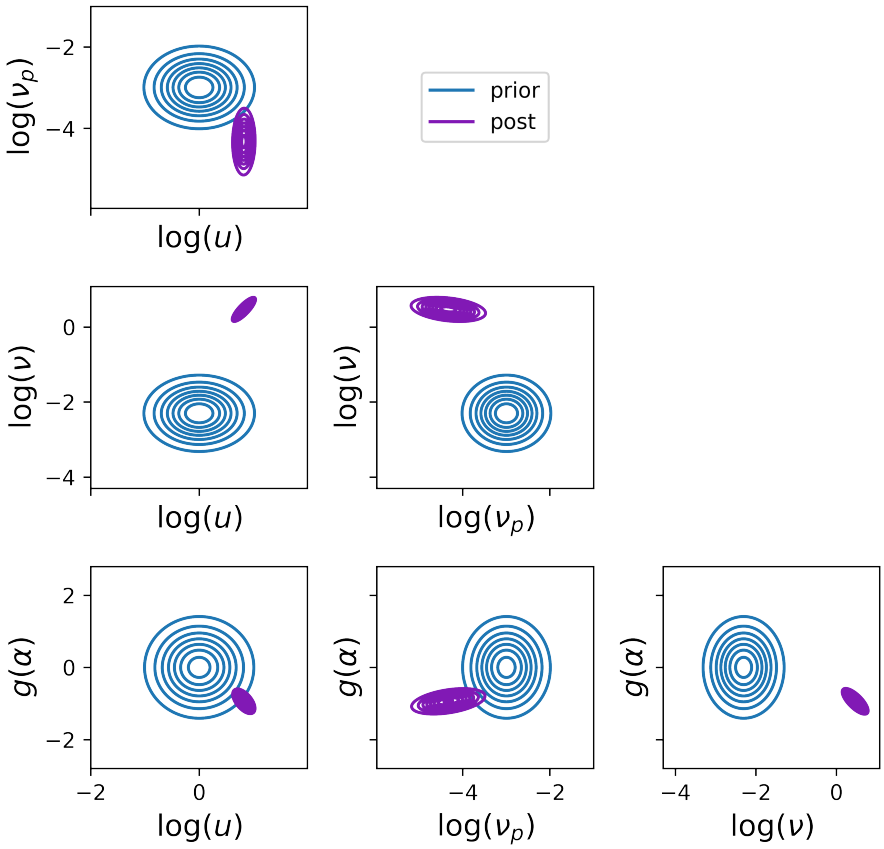


图18： 校准模型（51）中对应于**预测变分推断**的变换参数的先验与后验等高线的两两比较。

5 结论

本研究首次探索了面向预测的贝叶斯推断在校准模型设定错误的基于物理模型中的应用，并考察了是否纳入模型形式不确定性（MFU）表征的情形。该以优化为核心的逆问题框架采用变分近似，其中面向预测的损失函数定义为校准数据在对数预测分布下的概率。与标准变分推断（VI）所采用的基于平均对数似然的损失函数（即证据下界（ELBO）损失）相比，这一转变所得到的后验分布能够更好地描述数据中的预测不确定性，尤其是在存在模型设定错误的情况下；而标准方法通常会低估不确定性。

我们首先将这种预测变分推断（PVI）方法应用于一个线性高斯多项式问题；在该问题中，闭式表达式给出了逆问题解及其对应的预测性能，且不存在近似误差。在此，我们探讨了模型设定错误和噪声增加对逆问题解的影响，并发现，与标准贝叶斯推断不同，PVI 解会随模型设定错误的程度而调整。不过，与标准推断一致，噪声增加会降低损失函数相对于先验的重要性，使面向预测的后验分布随着噪声水平上升而越来越趋近于指定的先验分布。此外，我们还探讨了以下因素的影响：1）预测分布的蒙特卡洛（MC）近似，该方法需要显式似然；2）采用逐分量密度估计的近似方法，该方法适用于无似然方法和多元数据。对于蒙特卡洛近似，我们发现，每次优化迭代可能都需要进行数量大到难以实现的模型评估，才能达到足够的统计收敛。采用逐分量密度估计可以实现对面向预测的损失函数的可行评估。然而，忽略预测密度的协方差结构会改变逆问题的约束条件，可能导致不确定性增大；这一点可由后验协方差的迹和行列式以及边际预测不确定性体现。

随后，我们将 PVI 框架应用于异质多孔介质中污染物迁移的偏微分方程模型。这类模型提供非线性的参数到观测量映射，无法对预测密度进行闭式表征。在此，我们将校准数据模拟为高维空间数据的一次实现。我们考虑两种情景：1）忽略污染物弥散导致模型设定错误；2）通过纳入 MFU 表征缓解（但未完全消除）模型设定错误，并比较两种情景下标准变分推断和 PVI 得到的解。对于严重的模型设定错误（情景 1），我们发现，与标准变分推断相比，PVI 得到的解提供了合理得多的后验预测不确定性；标准变分推断则显著低估了不确定性。然而，进一步纳入 MFU 表征并随后使用 PVI 进行校准（情景 2），会使预测不确定性更加集中，同时仍对校准数据赋予较高概率，这凸显了此类方法如何改善模型的预测性能。此外，在估计模型参数以及 MFU 表征的参数时比较变分推断与 PVI，我们发现 PVI 能减小参数估计值的偏差，并防止后验分布过度集中。总而言之，最终目标是校准参数不确定性，使其传播至下游决策中的关注量时，既不会产生过度自信的预测，也不会宽泛到毫无信息价值。就描述观测到的关注量而言，我们发现，PVI 提供的后验不确定性表征比标准贝叶斯推断更有意义。

在大规模物理建模的背景下，仍有若干值得探索的研究方向，可进一步拓展此方法的适用性。未来工作可探索旨在提升应用此面向预测的方法时计算可处理性的途径，具体而言，是开发样本效率更高且可扩展的密度估计流程。能够实现更具表达力的密度近似的方法（如三角传输或归一化流）或可改善后验近似。最后，鉴于引入 MFU 表示的一个关键目标是提升外推能力（超出校准区间），如何在构建面向预测的损失函数时利用外推目标，也是值得研究的方向。

致谢

本研究由桑迪亚国家实验室的实验室定向研究与开发计划（项目 233072）资助。该实验室是一所多任务实验室，由桑迪亚国家技术与工程解决方案有限责任公司管理和运营；该公司是霍尼韦尔国际公司的全资子公司，代表美国能源部国家核安全管理局依据合同 DE-NA0003525 开展工作。本文由桑迪亚国家技术与工程解决方案有限责任公司的一名雇员撰写。该雇员拥有本文的全部权利、所有权

和利益，并对本文内容负全部责任。美国政府保留一项非独占、已付清、不可撤销、全球范围的许可；出版商接受本文发表，即表示承认美国政府保留此许可，可为美国政府目的发表或复制本文的已发表版本，或允许他人这样做。美国能源部将根据《美国能源部公共访问计划》向公众开放这些由联邦资助的研究成果：https://www.energy.gov/ downloads/doe-public-access-plan。

参考文献

[1] Pierre Alquier。随机化经验风险最小化器的 PAC-贝叶斯界。《统计数学方法》，17(4):279–304，2008年12月。https://doi.org/10.3103/S1066530708040017.[2] Pierre Alquier、James Ridgway 和 Nicolas Chopin。吉布斯后验分布变分近似的性质。《机器学习研究杂志》，17(236):1–41，2016年。http://jmlr.org/papers/v17/15-290. html.[3] Jax 作者。jax.scipy.special.logsumexp — Jax 文档。https://docs.jax.dev/en/latest/_autosummary/jax.scipy.special.logsumexp.html。访问日期：2025-12-08。[4] Rileigh Bandy 和 Rebecca Morrison。弹簧-质量-阻尼系统中的模型形式不确定性量化。载于第*41*届 *IMAC* 结构动力学会议暨展览论文集，第3卷模型验证与不确定性量化，第9–19页。施普林格，2023年。[5] Rileigh Bandy、Teresa Portone 和 Rebecca Morrison。针对自适应振动隔离循环测试挑战的随机模型修正。载于第*42*届 *IMAC* 结构动力学会议暨展览论文集，第3卷模型验证与不确定性量化，第53–62页。施普林格，2024年。https: // doi.org/10.1007/978-3-031-68893-5_8.[6] Pier G. Bissiri、Chris C. Holmes 和 Stephen G. Walker。更新信念分布的通用框架。《英国皇家统计学会杂志》。*B*辑（统计方法学），78(5):1103–1130，2016年。https: // doi.org/10.1111/rssb.12158.[7] SciPy 开发者。SciPy 参考文档：scipy.stats.gaussian kde。https://docs.scipy.org/doc/scipy/ reference/generated/scipy.stats.gaussian_kde.html#ra3a8695506c7-1。访问日期：2026-01-06。[8] Peter Gr¨unwald。安全贝叶斯方法：通过可混合性间隙学习学习率。载于第*23*届国际算法学习理论会议论文集，第169–183页，柏林、海德堡，2012年。施普林格出版社。https://doi.org/10.1007/978-3-642-34106-9_16.[9] Peter Gr¨unwald 和 Thijs van Ommen。贝叶斯推断在模型设定错误的线性模型中不一致，并提出修正方案。《贝叶斯分析》，12，2014年12月。https://doi.org/10.1214/17-BA1085.[10] Benjamin Guedj。PAC-贝叶斯学习入门，2019年。arXiv 预印本：https://arxiv.org/abs/1901. 05353.[11] Wenxin Jiang 和 Martin Tanner Martin。高维分类与数据挖掘中的变量选择吉布斯后验。《统计年刊》，36:2207–2231，2008年10月。https://doi.org/110.1214/ 07-AOS547.[12] Marc C. Kennedy 和 Anthony O'Hagan。计算机模型的贝叶斯校准。《英国皇家统计学会杂志》：*B*辑（统计方法学），63(3):425–464，2001年1月。https://doi.org/10. 1111/1467-9868.00294.[13] Bas J.K. Kleijn 和 Aad W. van der Vaart。模型设定错误下的伯恩斯坦-冯·米塞斯定理。《电子统计学杂志》，6:354–381，2012年1月。https://doi.org/10.1214/12-EJS675.[14] Jeremias Knoblauch、Jack Jewson 和 Theodoros Damoulas。以优化为中心看贝叶斯定理：回顾并推广变分推断。《机器学习研究杂志》，23(132):1–109，2022年。http://jmlr.org/papers/v23/19-1047.html.[15] Jinlin Lai 和 Yuling Yao。预测变分推断：学习预测最优的后验分布，2025年。arXiv 预印本：https://arxiv.org/abs/2410.14843.[16] John R. Lewis、Steven N. MacEachern 和 Yoonkyung Lee。贝叶斯限制似然方法：贝叶斯回归中对非充分统计量进行条件化（附讨论）。《贝叶斯分析》，16(4):1393 – 2854，2021年。https://doi.org/10.1214/21-BA1257。

[17] Jean-Michel Marin、Pierre Pudlo、Christian P. Robert 和 Robin J. Ryder。近似贝叶斯计算方法。*Statistics and Computing*，22:1167–1180，2012年11月。https://doi.org/10.1007/ s11222-011-9288-2.[18] Andres R. Masegosa。模型设定错误下的学习：在变分方法和集成方法中的应用，2020年。arXiv 预印本：https://arxiv.org/abs/1912.08335.[19] David A. McAllester。PAC-贝叶斯模型平均。载于第十二届计算学习理论年会论文集，COLT '99，第164–170页，纽约，纽约州，美国，1999年。计算机协会。https://doi.org/10.1145/307400.307435.[20]David A. McAllester。若干PAC-贝叶斯定理。机器学习，37(3):355–363，1999年12月。https://doi.org/10.1023/A:1007618624809.[21] David A. McAllester。PAC-贝叶斯随机模型选择。机器学习，51(1):5–21，2003年4月。https://doi.org/10.1023/A:1021840411064.[22] Yann McLatchie、Badr-Eddine Cherief-Abdellatif、David T. Frazier 和 Jeremias Knoblauch。预测导向后验分布，2025年。arXiv 预印本：https://arxiv.org/abs/2510.01915.[23] Jeffrey W. Miller 和 David B. Dunson。通过粗化实现稳健贝叶斯推断。美国统计学会会刊，114(527):1113–1125，2019年7月。https://doi.org/10.1080/01621459.2018.1469995.[24] Warren R. Morningstar、Alex Alemiand 和 Joshua V. Dillon。PAC^m-Bayes：缩小设定错误的贝叶斯情形下的经验风险差距。载于第*25*届国际人工智能与统计学会议论文集，第151卷机器学习研究论文集，第8270–8298页。PMLR，2022年3月28日至30日。https://proceedings.mlr.press/v151/morningstar22a/morningstar22a.pdf.[25] Rebecca E. Morrison、Todd A. Oliver 和 Robert D. Moser。表示模型不充分性：随机算子方法。*SIAM/ASA*不确定性量化期刊，6(2):457–496，2018年1月。https: //doi.org/10.1137/16M1106419.[26] Todd A. Oliver、Gabriel Terejanu、Christopher S. Simmons 和 Robert D. Moser。验证未观测量的预测。计算机辅助应用力学与工程，283:1310–1335，2015年1月。https://doi.org/10.1016/j.cma.2014.08.023.[27] Teresa Portone。基于缺失微观结构信息表示模型形式不确定性。博士论文，德克萨斯大学奥斯汀分校，2019年。https://dx.doi.org/10.26153/tsw/10112.[28] Teresa Portone 和 Rebekah D. White。模型形式不确定性分层贝叶斯推断中的理论与方法挑战。桑迪亚技术报告SAND2025-10780，桑迪亚国家实验室，美国，2025年9月。https://doi.org/10.2172/2587508.[29] Teresa Portone、Rebekah D. White 和 Joseph L. Hart。量化模型形式不确定性对模型预测的敏感性，2025年。arXiv 预印本：https://arxiv.org/abs/2509.08708.[30] Khachik Sargsyan、Xun Huan 和 Habib N. Najm。用于贝叶斯模型校准的嵌入式模型误差表示。国际不确定性量化期刊，9(4):365–394，2019年。https://doi.org/10.1615/int.j.uncertaintyquantification.2019027384.[31] Khachik Sargsyan、Habib N. Najm 和 Roger Ghanem。物理模型的统计校准。国际化学动力学期刊，47(4):246–276，2015年。https://doi.org/10.1002/kin.20906.[32] Andrea Scarinci、Michael Fehler 和 Youssef Marzouk。利用输运-拉格朗日距离进行模型设定错误下的贝叶斯推断：在地震反演中的应用，2021年。arXiv 预印本：https://arxiv.org/abs/2105.07027.[33] Sebastian M. Schmon、Patrick W. Cannon 和 Jeremias Knoblauch。近似贝叶斯计算中的广义后验分布，2021年。arXiv 预印本：https://arxiv.org/abs/2011.08644.[34] David W. Scott。多元密度估计：理论、实践与可视化。John Wiley & Sons，奇切斯特，纽约，1992年。http://dx.doi.org/10.1002/9780470316849。

[35] Zheyang Shen、Jeremias Knoblauch、Sam Power 和 Chris J. Oates。《基于预测的最大均值差异不确定性量化》，2025年。arXiv 预印本：https://arxiv.org/abs/2410.11637。[36] Scott A. Sisson、Yanan Fan 和 Mark A. Beaumont。《近似贝叶斯计算概述》，2018年。arXiv 预印本：https://arxiv.org/abs/1802.09720。[37] Leslie. G. Valiant。《可学习性理论》。《*ACM*通讯》，27(11):1134–1142，1984年11月。https://doi.org/10.1145/1968.1972。[38] Tong Zhang。《统计估计的信息论上界与下界》。《*IEEE*信息论汇刊》，52(4):1307–1321，2006年5月。https://doi.org/10.1109/TIT.2005.864439。[39] Tong Zhang。从 ϵ熵到KL熵：密度估计的最小信息复杂度分析。《统计年刊》，34，2007年3月。https://doi.org/10.1214/009053606000000704。