感兴趣量预测的验证实验最优设计

Antonin Paquette-Rufiange[a], *, Serge Prudhomme[a], Marc Laforest[a]

[a]蒙特利尔理工学院数学与工业工程系，*H3T 1J4*，加拿大蒙特利尔

arXiv:2303.06114v1 [cs.CE] 10 Mar 2023

摘要

对物理系统中测得的感兴趣量进行数值预测，依赖于数学模型的使用，而这些模型应当经过验证，或至少不被证伪。模型验证通常涉及比较实验数据（来自关注系统的输出）与模型预测，两者均在特定验证场景下获得。验证实验的设计应与模型的目标直接相关，即在预测场景下预测某个关注量。本文探讨设计验证实验时出现的两个具体问题。第一个问题是在无法于受控环境中实施预测场景时，如何确定合适的验证场景。第二个问题涉及当关注量无法直接观测时，如何选择观测值。所提出的方法包括计算影响矩阵，以表征给定模型泛函的响应面。通过最小化影响矩阵之间的距离，可以选出最能代表预测场景的验证实验。我们通过两个数值示例说明这一方法。第一个示例考虑对一个简单模型进行验证，该模型基于描述物体自由落体运动的常微分方程，借此凸显验证实验选择的重要性。第二个数值实验聚焦于污染物输运，并展示关注量的选择对所实施验证实验的影响。

关键词： 验证、关注量、实验的最优设计、敏感性分析、不确定性量化

1. 引言

构建用于描述关注系统（无论是物理、生物还是社会系统）的模型，对于理解和预测其行为至关重要。数学模型可以源自物理定律和/或经验考量，既可针对系统的某些部分，也可针对整个系统。为了评估模型解释和预测给定关注量（QoI）的能力，必须对其进行验证，或至少确保其未被证伪。换言之，必须针对该关注量，精确量化模型与其旨在描述的现实之间的误差。验证过程包括将模型的输出与通过测试关注系统而获得的实验数据进行比较。过去几十年来，文献中提出了多种模型验证程序。例如，Oberkampf 和 Trucano [29]对确认与验证过程进行了详细综述；Roy 和 Oberkampf [35] 则深入探讨了这两个过程，重点讨论如何处理偶然不确定性和认知不确定性。Riedmaier 等人 [34] 最近发表的一篇综述文章，详细讨论了现有验证与预测方法的适用范围及其不足。Lee 等人 [23] 的论文全面综述了有关不确定性传播、参数校准以及

*通讯作者电子邮箱： antonin.paquette-rufiange@polymtl.ca(Antonin Paquette-Rufiange)

模型验证。由于当前采用的科学模型日益复杂，且对预测性、定量化结果的需求不断增长，模型验证在计算科学与工程领域变得愈发重要且不可或缺。然而，一个关键问题仍在于如何选择合适的验证实验，因为这些实验提供了用于比较模型预测的实验数据，进而直接影响模型是否被判定为有效。这一问题在验证实验数量或数据量有限时尤为重要。然而，在文献中，如何精确设计此类验证实验仍鲜有探讨，只有少数研究例外。

Oliver 等人 [30] 强调了预测建模中验证过程的重要性，而对于由多个耦合子模型构成的模型，这一点尤为重要。他们指出，在开展模型验证时有两个根本问题：1）如果无法复现作出预测时所处的条件，该如何验证模型；2）如果实际中无法观测到希望预测的感兴趣量，该如何验证模型。为回答这两个问题，作者估计影响各子模型的建模误差，并将其传播至预测场景下的感兴趣量。随后，他们提出了一系列指导原则，以确保校准和验证实验与预测目的相关 [30, 第3节.3]。这些原则大致可概括如下：如果感兴趣量对某些模型参数和/或某些建模误差敏感，那么校准和验证实验就应体现这些敏感度。不过，相关讨论基本上仍是定性的，分析也是事后进行的，也就是说，在完成校准和验证实验后，他们才进一步检验这些实验是否确实相关。

Hamilton 和 Hills 的论文 [17] 以及 Hills 后续的报告 [18] 也讨论了上述问题。作者构建了一个元模型，用于将建模误差外推至预测场景下的感兴趣量。更准确地说，他们找到了预测场景下感兴趣量的敏感度（某种基本效应）与验证场景下可观测量的敏感度之间的线性关系。因此，可以根据预测目标，为相关性更高的验证观测赋予权重。不过，评估仍然是在事后进行的。此外，作者没有讨论验证实验的设计。尽管如此，我们注意到，Hills [18, 第 1 节3] 列举了若干案例，说明验证实验可能无法有效评估模型能否用于预测特定预测场景下的感兴趣量。

其他研究也探讨了由多个子模型组成的数学模型的验证问题。其中，Hills 和 Leslie [19], 采用基于局部导数的指标，对各子模型的验证实验的重要性进行加权，以评估其对于使用完整模型进行数值预测的重要性。同样，Li 和 Mahadevan [24]提出了一种综合方法，以考虑低层级模型的有效性及其与系统级模型的相关性。具体而言，他们采用模型可靠性度量来评估低层级模型的有效性，并基于 Sobol 指数 [39] 进行敏感性分析，以量化这些低层级模型与完整系统模型之间的关系。

Farrell 等人 [14] 提出了所谓的奥卡姆-可信度算法，该算法能够从一组相互竞争的模型中识别出最简单且最可信的模型，并依据模型后验可信度对这些模型进行排序。Tan 等人 [42] 通过提出两项关于如何选择待执行验证实验的准则，进一步完善了奥卡姆-可信度算法。第一项准则规定，验证场景的敏感性指标必须与预测场景的敏感性指标接近。他们采用 Sobol 指数 [39]作为敏感性指标，因此将方差分解视为一种敏感性概念。提出验证场景后，还会对这一准则进行事后检验。第二项准则则检验，与校准场景相比，验证场景是否能为模型参数的后验分布带来信息增益。

Ao 等人 [3] 专门研究了寿命预测模型背景下验证实验的最优设计。他们的方法实际上与第 2.6 节所述的贝叶斯最优设计方法相似。然而，他们的方法依赖于一个用于表示关注系统的精细模型，因此需要关于关注系统的额外先验知识，而这并不总是

可获得。
我们还提及 Sunseri 等人 [41] 的论文，该论文研究了辅助参数对反问题解的影响。尽管他们的研究没有涉及验证问题，但其方法与本文相关。他们利用反问题解相对于辅助变量的梯度，确定了一些逐点和全局敏感性指标。若某个辅助参数具有较高的敏感性指标，这表明该参数值的微小扰动会显著影响反问题的解，因此应投入一定精力以降低该辅助参数的不确定性。

本研究旨在进一步完善上述论文提出的一些概念，尤其是运用敏感性指标评估验证实验的相关性。因此，本文的主要目标是阐述一种用于验证实验最优设计的新方法。所提出的方法主要基于两个不同优化问题的构建；这两个问题的构建方式使得模型在验证条件下的行为尽可能接近同一模型在预测条件下的行为。这样一来，人们便能对模型预测更有信心：验证实验的选择经过专门设计，以产生这样一种验证场景，其中可观测量的行为与预测场景中感兴趣量的行为相同。换言之，在验证场景中观察到的建模误差将与预测场景中预期的建模误差相似。

本文结构安排如下。第2节介绍精确描述验证过程所需的全部概念和符号，并详细说明模型中涉及的不同类型参数和不确定性来源。我们还将介绍校准与验证的方法，并简要概述校准实验的最优设计方法。第3节提出一种新颖的验证实验设计方法，专门用于预测可能无法观测到的关注量。我们将简要回顾主动子空间方法，这是用于执行敏感性分析的选定方法。我们将构建两个最优设计问题，以计算验证场景所需的控制参数和传感器参数。随后，我们将在第4节展示一些数值示例。首先，我们考虑一个玩具问题——物体自由落体问题，以说明验证过程，并着重强调设计最优验证实验的重要性。然后，我们研究第二个示例，即一个简化的污染物输运问题，以突出在设计最优验证实验时定义感兴趣量的重要性。第5节给出总结性评论，并提出一系列开放问题，探讨所提出的方法与更广泛的不确定性量化和模型验证过程之间的关系。

2. 术　语以及关于验证和预测建模的初步概念 g

本节旨在介绍一些初步概念和符号，以便准确描述验证过程。我们回顾一下，预测建模的首要目标是能够获得有关关注系统的定量预测。乍看之下，这项任务似乎相当容易完成，因为它只需评估模型的输出。然而，如前所述，数值预测要发挥作用，就必须确信模型对于预期用途仍然有效，并且利用模型得到的预测确实能准确描述现实。现在应该已经清楚，模型的有效性应根据人们希望利用模型在特定条件／工况下预测的感兴趣量来评估。在下文中，我们将尽量采用文献中通常使用的相同术语，但为强调某些模型参数和模型输出各自不同的作用，也会有少数例外。

2.1. 抽象模型

我们将一般的确定性模型 定义为三元组 $(r,\ p,\ u)$，其中 $p=(p_1,\ \ldots,\ p_d)\in\mathcal{P}$ 为参数，$u=(u_1,\ \ldots,\ u_s)\in\mathcal{U}$ 为状态变量，而 r 则包括平衡关系 以及

初始条件和/或边界条件

$$r(p, u) = 0. \tag{1}$$

为简单起见，假设参数空间$\mathcal{P} \subseteq \mathbb{R}^d$为有限维空间，状态空间$\mathcal{U}$则为巴拿赫空间或希尔伯特空间。平衡关系$r$可能涉及常微分方程或偏微分方程、代数方程，或这些方程的任意组合。平衡关系概括了一系列假设和科学认知，这些假设和认知有望促成一个足够准确的关注系统模型。平衡关系r必须具备一项重要性质，即适定性：对于每个$p \in \mathcal{P}$，都存在唯一且稳定的解$u \in \mathcal{U}$满足（1）。换言之，存在一个函数g，使得$u = g(p)$。

例如，我们考虑一个简单模型，将在第4.1节中用于说明我们的方法。该模型描述了一个球形抛射体在海平面的环境空气中垂直发射后的运动。物理系统的演化可以用线性常微分方程和初始条件来描述

$$m\frac{\mathrm{d}^2u}{\mathrm{d}t^2}(t) + 3\pi\ell\exp(\mu)\frac{\mathrm{d}u}{\mathrm{d}t}(t) = -mg, \qquad \forall t \in (0, T) \tag{2a}$$

$$u(0) = u_0, \tag{2b}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t}(0) = v_0, \tag{2c}$$

其中，$u(t)$表示抛射体在时间t时达到的高度，m和ℓ分别表示其质量和直径，u_0和v_0分别表示其初始位置和速度，g为引力常数，而μ为空气中运动的球形物体的黏度参数[43, p.488]。在此例中，问题的参数由$p = (m, \ell, u_0, v_0, g, \exp(\mu)) \in \mathcal{P} = (\mathbb{R}^+)^6$给出，状态变量为高度$u \in \mathcal{U} = C^1(\mathbb{R})$，并配备在$u$和$\mathrm{d}u/\mathrm{d}t$上的一致范数。平衡关系和初始条件集合可由初值问题（2）轻易得到。该模型通过对物理系统作出以下假设构建而成：

1. 假设抛射体的轨迹是一维的，因此只考虑竖直方向上的作用力。 2. 假设摩擦力与抛射体的速度成正比。 3. 空气的黏度和引力常数不随高度变化。

2.2. 参数类别

我们将问题的参数p分为两个不同的子集，即控制参数x和模型参数θ。我们还考虑一组额外的参数，称为传感器参数$z \in \Omega$，其中Ω是状态变量u的定义域。每类参数都将发挥特定作用，如下所述。引言中引用的论文也给出了类似的定义，尽管文献中对于统一术语尚未达成共识。

控制参数$x \in \mathcal{X}$表示在对关注系统进行实验时可以控制的参数。这些参数可以视为一组输入，它们能够唯一确定实验场景。这类参数的例子包括初始条件和边界条件中的参数、几何数据，以及实验装置可直接测量的物理性质（质量、长度、温度等）。对于第 2.1 节中的例子，控制参数为抛射体的质量和直径，以及其初始高度和速度，因此$x = (m, \ell, u_0, v_0)$。

剩余的模型参数$\theta \in \mathcal{T}$是完整描述模型（$\theta = p \setminus x$）所必需的参数。之所以这样命名，是因为这些参数明确源自人们为描述某一特定现象而定义模型时所采用的假设和简化。其他假设可能会产生一个具有不同模型参数的新模型，

但不会改变控制参数。自然，参数 ρ的定义，以及模型参数 θ的定义，直接取决于建模者和/或用户将哪些量视为参数。例如，式（2）中的数值常数 π 显然不构成参数，但对于其他量，判断起来可能就不那么明确。一个经验法则是：如果某个量的值可能发生变化，或者在模型分析中具有一定意义，那么它就可以被视为模型参数。模型参数的例子包括表征材料属性的参数（如杨氏模量、泊松系数）或某些无量纲量（如雷诺数、普朗特数）。模型参数通过先验知识或校准过程确定。关于后者的更详细讨论将在第2.5节中进行。对于球形抛射体，模型参数为 $\theta = （g, \ \mu）$，因为它们源自建模过程中的假设（g 与高度无关，而 μ 与空气摩擦的子模型有关）。

最后一类参数是所谓的传感器参数 $z \in \Omega$。它们通常表示由自变量（例如空间或时间）定义的位置，在该位置上可以观测状态变量 u。在抛射体的例子中，传感器参数是我们希望评估高度的时间，因此 $z = t$。

2.3. 场景类别

控制参数和模型参数可以定义系统和模型运行的工况。特定的工况称为场景。为简化起见，这里我们将模型参数 θ 限定为验证场景和预测场景中的参数相同。对于抛射体的例子，这意味着预测和实验装置都处于海平面、环境大气中（引力常数 g 和黏度 μ相同）。换言之，只需控制参数 x 即可完整描述一个场景。我们将研究三类场景，重点关注前两类。

1. **预测场景：**定义在何种条件下（由 x_{pred}描述）对关注系统进行预测。2. **验证场景：**定义在何种条件下（由 x_{val}描述）开展验证实验。3. **校准场景：**定义在何种条件下（由 x_{cal}描述）执行校准实验。这里我们指出，模型校准本质上旨在估计模型参数 θ的不确定性。

一方面，验证实验和校准实验是在一组$\mathcal{X}_{\text{lab}} \subset \mathcal{X}$受控环境中进行的。另一方面，预测场景 x_{pred} 可能可以通过实验实现，也可能无法通过实验实现，因此 x_{pred} 可能不属于$\mathcal{X}_{\text{lab}}$。图1对各种场景作了概念性示意。

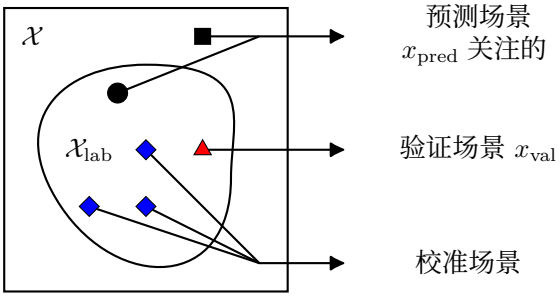


图1：各种场景示意图。预测场景可能可通过实验实现（黑色圆圈），也可能无法通过实验实现（黑色方块）。验证场景和校准场景均属于受控环境集合$\mathcal{X}_{\text{lab}}$。

2.4. 观测量与感兴趣量

我们再次强调，数学模型仅仅是对关注物理系统的一种抽象。事实上，我们只能获得关于关注系统的有限信息。我们将这些信息统称为实验观测量 y_{exp}。实验观测量 y_{exp} 可以通过各种实验设备和/或统计调查获取。重要的是，这些观测量是在特定场景和给定传感器参数下获得的，因此它们是x和 z 的函数，即 $y_{\text{exp}} = y_{\text{exp}}(x, z)$。我们还假设模型（$r, p, u$）足够丰富，因此每个实验观测量都可以抽象为参数 p和传感器参数 z 的一个泛函（状态变量 u 通过函数 $g(p)$隐式给出）。这些可观测量称为模型观测 $y := h_{\text{obs}}(p, u, z) = h_{\text{obs}}(p, z)$。下文中，“可观测量”一词将同时指实验观测量和模型观测。在讨论球形抛射体发射的示例中，实验观测量可以是抛射体的高度（由光学仪器或高度计获取），也可以是通过加速度计测得的抛射体加速度。无论考虑何种量，都会在给定的离散时间 $z = t_{\text{exp}}$进行测量。因此，相应的模型观测定义为 $y(t_{\text{exp}}) := u(t_{\text{exp}})$ 或 $y(t_{\text{exp}}) := \mathrm{d}^2u/\mathrm{d}t^2(t_{\text{exp}})$。

感兴趣量 扩展了模型可观测量的概念，指的是那些能够或不能够通过实验观测到的量。换言之，感兴趣量不一定取决于传感器位置，但可能会通过平均值或最大值隐式地依赖于独立的（空间或时间）变量。感兴趣量记作 $q := h_{\text{qoi}}(p, u) = h_{\text{qoi}}(p)$，因为它们是在特定场景下评估的。以下列出一些可能的感兴趣量示例：

- 子区域内的平均量 Ω （例如固体中的平均位移或应力）；
- 最小值或最大值及其位置（例如最大应力）；
- 统计量（例如核废料泄漏概率）。

一个重要概念是，感兴趣量对关注系统的关键特征进行了抽象，而决策过程将以这些特征为依据。图2概述了如何根据各项参数获得观测值和感兴趣量。以抛射体为例，一个感兴趣量可以是其随时间达到的最大高度，即 $q := \max_t u(t)$。值得注意的是，该量无法直接观测，因为我们需要事先知道真实抛射体达到最大高度的确切时间。

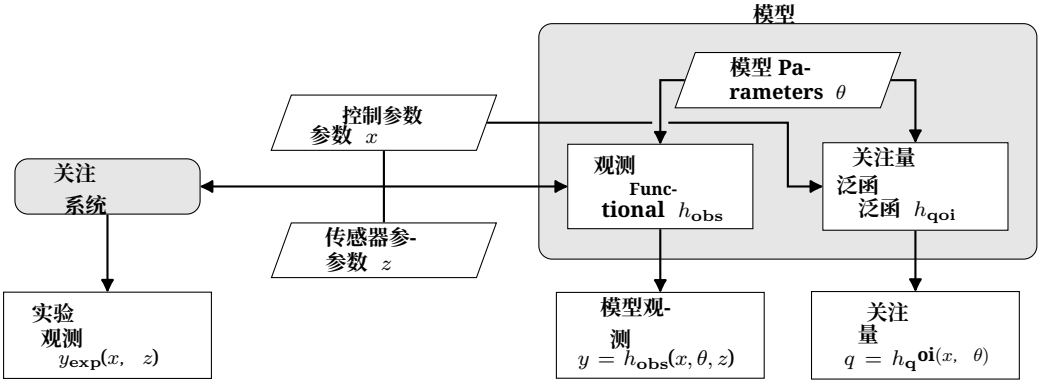


图2：各种参数与关注系统及模型之间相互作用的示意图。需要注意的是，模型参数 θ 不会影响关注系统，只会影响观测和感兴趣量泛函 h_{obs} 及 h_{qoi}。为简化起见，感兴趣量泛函 h_{qoi} 不显式依赖于传感器参数 z。

预测问题可以精确表述如下：对于给定的预测场景 x_{pred}，人们希望预测给定的感兴趣量 h_{qoi}。可观测泛函 h_{obs} 的定义将在校准和验证问题中发挥关键作用，以获得准确的预测。

模型输出（可观测量和感兴趣量）的准确性也取决于各种不确定性和误差的聚合，下一小节将对此进行说明。

2.5. 不确定性与建模误差

可以采用多种方法来表示参数、可观测量或感兴趣量等量的不确定性，如 [22, 34] 中所述。本文采用概率方法，将不确定量视为随机变量 A，其概率分布为 F_A，密度函数为 f_A。随机变量 A 可以是离散的，也可以是连续的。按照惯例，小写变量 a 表示随机变量 A（以大写字母表示）的一次实现。除了随机变量，我们还可能遇到随机过程 $A = A(b)$，通常由一族随机变量组成，每个变量都针对 b 的一个实例定义。

不确定性有时分为偶然 不确定性和认知 不确定性 [35]。本文假定所有不确定性都是偶然的。更确切地说，这意味着我们能够指定任意随机变量 A 的分布 F_A。相反，如果不确定性属于认知不确定性，我们便无法对其进行表征 F_A，例如参见 [35, 第 2] 节。因此，在本文中，认知不确定性可视为不确定性本身的不确定性。我们注意到，有些作者将认知不确定性定义为：若能获得更多信息和数据，这种不确定性最终可能会降低 [18, 27]。然而，这种对认知不确定性定义的细微差别与当前讨论无关，因此本文不予赘述。既然我们假设所有不确定性都是偶然的，那么说明如何确定这些分布便是合理的，以下将对此作出说明。

2.5.1. 参数的不确定性

作为第一步近似，我们假设控制参数 x 和传感器参数 z 不存在任何不确定性；但根据预测场景的性质和实际实验装置的情况，这一假设可能与事实相去甚远。不过，这一假设将简化最优验证实验的定义，也会使本文的整体阐述更为简洁。因此，本文仅将模型参数 θ视为不确定参数。

模型参数 θ 的不确定性量化并非易事，因为这类参数源于模型所依据的各种假设。考虑到关于每项假设的先验知识，可以确定其分布 F_Θ。以抛射体为例，我们对引力常数 g 和空气黏度 μ拥有相对丰富的先验知识，因此可以为它们指定方差较小的分布。然而，我们往往对模型参数的取值只有相当模糊的认识。此时，需要通过校准过程确定各参数的概率分布。这一专题已有大量文献详加论述。为求完整，并强调校准问题与验证问题之间的区别，我们将在第 2.6 节进一步详述模型参数的校准。

2.5.2. 观测量的不确定性与感兴趣量

实验观测的不确定性可能来自各种来源，也可能取决于情景 x 和传感器参数 z。因此，实验观测被视为随机过程 $Y_{\text{exp}}(x,\ z)$。能够表征可观测量的不确定性，对验证过程至关重要，正如第 2.7 节进一步解释的那样。我们列出两个可能的不确定性来源，但并非详尽无遗。第一个来源是测量设备固有且不可避免的变异性。不过，有时可以通过使用更精确的设备来降低可观测量的不确定性，从而使实验观测更接近真实 值。第二个不确定性来源是关注系统固有的随机性。对于固定的控制参数和传感器参数 x 和 z，重复进行同一实验可能会得到不同但准确的实验观测值。例如，如果相关现象对于控制参数的无穷小扰动表现出不稳定性，就可能出现这种情况。聚合这两个不确定性来源，可以确定

表 1：参数、可观测量和感兴趣量及其不确定性来源汇总

	定义	不确定性
控制参数 x	可控的参数 用于开展实验	None
模型参数 Θ	描述模型所需的其余参数 描述模型	先验知识或 后验密度
传感器参数 z	时空参数	None
实验观测 $Y_{\text{exp}}(x,\ z)$	控制参数 x 下的观测值 和传感器参数 z	经验密度
模型观测 $Y\ (x,\ z)$	$Y(x,\ z) \coloneqq h_{\text{obs}}(x,\ \Theta,\ z)$	不确定性传播
关注量 $Q\ (x)$	$Q(x) \colon = h_{\text{qoi}}(x,\ \Theta)$	不确定性传播

分布 $F_{Y_{\text{exp}}(x,\ z)}$，但这项任务未必简单。我们认为，对于某个实验，可以重复进行一定次数（对于非破坏性实验），从而确定经验分布 $F_{Y_{\text{exp}}(x,\ z)}$，该分布聚合了这两个不确定性来源。

模型观测 Y 和关注量 Q 的不确定性，是通过将模型参数 Θ 的不确定性经由各自的泛函传播得到的，即:

$$Y(x,z) \coloneqq h_{\text{obs}}(x,\Theta,z) \tag{3}$$
$$Q(x) \coloneqq h_{\text{qoi}}(x,\Theta) \tag{4}$$

只有对于非常简单的泛函和有限种分布 F_Θ （例如高斯分布），才能计算出精确的分布 F_Y 和 F_Q 。大多数情况下，需要使用广泛应用的蒙特卡洛方法、拉丁超立方方法或准蒙特卡洛方法 [40]等方法对分布 F_Θ 进行采样，以近似分布 F_Y 和 F_Q 。在进行任何验证步骤之前，先进行不确定性传播通常很有启发性。如果 Q 的预测具有过大的不确定性，那么无论模型是否被认为有效，该预测都可能毫无用处。在这种情况下，可能需要改进模型参数 Θ 的校准，以降低其不确定性，从而也有可能降低感兴趣量的不确定性。采用敏感性分析方法 [12, 37] 有助于识别对 Y 和 Q的不确定性贡献最大或最小的模型参数 Θ 。我们在表 1 中总结了迄今为止引入的各种量及其不确定性的来源。

2.5.3. 建模误差

作者认同一种普遍观点：无论模型多么复杂，都无法准确捕捉关注系统的行为，因为模型不过是对后者的抽象，并且往往依赖若干假设。因此，任何模型都存在一些建模误差，需要通过某种方式加以估计。此外，建模误差还应考虑实验观测和模型观测中不可避免的不确定性来源。结合参数不确定性与建模误差的一种方法是通过差异 E 来表示:

$$E(x,z) = Y_{\text{exp}}(x,z) - Y(x,z) = Y_{\text{exp}}(x,z) - h_{\text{obs}}(x,\Theta,z). \tag{5}$$

差异 $E(x,\ z)$ 可视为一个随机过程，因为对于控制参数 x 和传感器参数 z的每一种取值，它都对应一个随机变量。按照 Kennedy 和 O'Hagan [4, 21] 所采用的方法，这种差异 E 可视为实验误差与模型不充分性的聚合。差异 (5) 的一个重要性质是，它的定义独立于感兴趣量。还可以采用其他方法来量化建模误差 [18, 30]。为使论述简明，并且鉴于我们的重点是验证实验的设计，此处不再介绍这些方法。有关这些方法的概要，请参阅 [28, 34]。

2.6. 校准过程与校准实验的最优设计

这里简要概述校准过程，以及与校准实验最优设计相关的问题。尽管我们提出的验证实验最优设计方法本身并不直接涉及校准过程 本身，但简要讨论这一主题仍有助于将我们的方法与校准实验的最优设计问题区分开来。

模型参数 Θ 的校准（通常称为模型校准或参数识别）的主要目标，是利用实际实验观测量 y_{exp}（作为 Y_{exp} 的实现），对概率分布 F_{Θ} 进行刻画；该分布涵盖了模型参数的不确定性。换言之，目标是更新我们对模型参数的认知。校准有多种方法，例如最小二乘法（及其正则化版本）、最大似然估计（MLE）法或最大后验（MAP）法。在各种假设下，可以证明这些方法之间的等价性 [38]。这些方法的一个共同特点是，它们能提供模型参数 θ 的点估计量 θ 。这些方法存在两个密切相关的缺点。首先，要量化 Θ 的不确定性相对困难，且可能只有渐近结果适用 [13, 第 16 节4]。其次，如果模型不具备可辨识性，则无法运用这些渐近结果，点估计量 θ 也可能不够“稳定”。可辨识性条件较为苛刻，尤其难以验证 [9]。

另一种方法是利用贝叶斯分析的后验分布来更新模型参数的不确定性。与最大似然估计（MLE）和最大后验（MAP）方法一样，该方法需要定义似然函数 $L_{Y_{\text{exp}}=y_{\text{exp}},\ x_{\text{cal}},\ z_{\text{cal}}}(\theta)$，用于描述在给定控制参数、传感器和模型参数 x_{cal}、 z_{cal} 和 θ 的情况下，获得实验观测量 $Y_{\text{exp}} = y_{\text{exp}}$ 的概率。似然函数的定义并不唯一。下面我们将说明如何在假设误差为加性的情况下构造似然函数

$$y_{\text{exp}}(x_{\text{cal}}, z_{\text{cal}}) = y(x_{\text{cal}}, z_{\text{cal}}) + e(x_{\text{cal}}, z_{\text{cal}}) = h_{\text{obs}}(x_{\text{cal}}, \theta, z_{\text{cal}}) + e(x_{\text{cal}}, z_{\text{cal}}), \tag{6}$$

其中， $y_{\text{exp}}(x_{\text{cal}},\ z_{\text{cal}})$ 是随机变量 $Y_{\text{exp}}(x_{\text{cal}},\ z_{\text{cal}})$ 的一次实现，而 $e(x_{\text{cal}},\ z_{\text{cal}})$ 是描述误差的随机变量 $E(x_{\text{cal}},\ z_{\text{cal}})$ 的一次实现。注意， E 与式（5）中的相同。由于误差 $E(x_{\text{cal}},\ z_{\text{cal}})$ 实际上未知，因此可以选取合理的近似，使似然函数易于处理。一种典型的选择是均值为零、标准差为 $\sigma > 0$ 的高斯噪声：

$$E \sim \mathrm{N}(0, \sigma^2). \tag{7}$$

换言之，假设无论校准场景 x_{cal} 和传感器参数 z_{cal}如何，噪声都相同。由此选择得到的似然函数定义为

$$L_{Y_{\text{exp}}=y_{\text{exp}},x_{\text{cal}},z_{\text{cal}}}(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left(h_{\text{obs}}(x_{\text{cal}}, \theta, z_{\text{cal}}) - y_{\text{exp}}(x_{\text{cal}}, z_{\text{cal}})\right)^2}{2\sigma^2}\right]. \tag{8}$$

Kennedy 和 O’Hagan [21] 提出将差异表示为模型误差项（用高斯过程表示）与实验误差之和。然而，这种方法计算过程复杂，而且可能缺乏可辨识性，因为模型参数和模型误差都需要进行校准 [4]。

除似然外，还必须提供先验密度，记为 f_{Θ_0}，以表示当前对模型参数 Θ 所掌握的知识。利用似然函数和先验密度，可通过贝叶斯定理得到后验概率密度函数$f_{\Theta|Y_{\text{exp}}=y_{\text{exp}},\ x_{\text{cal}},\ z_{\text{cal}}}$ ，其形式如下：

$$f_{\Theta|Y_{\text{exp}}=y_{\text{exp}},x_{\text{cal}},z_{\text{cal}}}(\theta) = \frac{L_{Y_{\text{exp}}=y_{\text{exp}},x_{\text{cal}},z_{\text{cal}}}(\theta) f_{\Theta_0}(\theta)}{\int L_{Y_{\text{exp}}=y_{\text{exp}},x_{\text{cal}},z_{\text{cal}}}(\theta) f_{\Theta_0}(\theta) \mathrm{d}\theta}. \tag{9}$$

后验密度表示在给定实验观测值 Θ 的情况下，对模型参数 y_{exp} 的更新认知x_{cal}， z_{cal}。只有在少数情况下，后验密度才有解析表达式，例如模型关于模型参数呈线性，且选择形式为 (8) 的似然和高斯先验。否则，通常使用马尔可夫链蒙特卡罗方法或其变体等抽样方法进行近似 [16, 40]。

由于所有校准方法都需要使用实验观测 y_{exp}(x_{cal}， z_{cal})，因此应适当选择校准场景 x_{cal} 和传感器参数 z_{cal} ，以便更准确地确定模型参数 Θ的不确定性。据我们所知，最优设计校准场景的方法基本上有两种：一种基于费 Fisher 信息矩阵，另一种基于贝叶斯方法。

1. 第一种方法是优化与模型 [5, 6]相关的费 Fisher 信息矩阵的某种泛函。这种方法广泛用于多项式模型（相对于模型参数呈线性，而相对于控制参数和传感器参数可能呈非线性），因为费 Fisher 信息矩阵仅取决于控制参数和传感器参数 [5]。对费 Fisher 信息矩阵泛函的不同定义会导出不同的最优校准实验。例如，D-最优设计使模型参数估计量 θ 的方差最小化，而 G-最优设计则使预测方差的最大值最小化。线性模型的 D-最优问题可写为

$$(x_{\mathrm{cal}}, z_{\mathrm{cal}}) = \underset{(x,z)\,\in\,\mathcal{X}_{\mathrm{lab}}\times\mathcal{Z}_{\mathrm{lab}}}{\operatorname{argmax}} |\mathcal{I}(x,z)|, \tag{10}$$

其中 $\mathcal{I}$ 表示费 Fisher 信息矩阵。等价定理表明，这些不同的最优设计实际上彼此相关。关于这种方法及其对非线性模型的一些扩展的深入讨论，请感兴趣的读者参阅 Atkinson 等人 [6]的著作。

2. 第二种方法基于贝叶斯方法 [8, 25, 36] ，重点是在可能的观测值 Y 和模型参数 Θ的可能取值范围内，最小化后验分布$f_{\Theta|Y=y_{exp},\ x_{\mathrm{cal}},\ z_{\mathrm{cal}}}$ 的某种泛函（有时也最小化似然函数）。采用类似于费 Fisher 信息矩阵所用泛函的定义，可以得到类似的最优设计。例如，我们可以寻找控制参数 x_{cal} 和传感器参数 z_{cal}，使后验分布的协方差行列式最小

$$(x_{\mathrm{cal}}, z_{\mathrm{cal}}) = \underset{(x,z)\,\in\,\mathcal{X}_{\mathrm{lab}}\times\mathcal{Z}_{\mathrm{lab}}}{\operatorname{argmax}} \iint \left(\det(\mathrm{Cov}(\Theta|Y=y_{exp},x,z))\right)^{-1} \mathrm{d}F_{Y_{\mathrm{exp}}}\,\mathrm{d}F_{\Theta_0}. \tag{11}$$

我们指出，求解这类优化问题可能极具挑战性，因为通常需要进行嵌套积分。拉普拉斯近似后验分布等若干简化方法，可以使这些优化问题变得可处理 [8, 25]。

这两种方法的一个非常重要的特点是，确定最优校准实验并不依赖任何实验数据，因为这是一种先验分析。我们将在设计最优验证场景的方法中保留这一特点。整个校准过程总结于算法1中。

2.7. 验证过程

现在一切就绪，可以讨论我们关注的问题，即如何验证模型。再次强调，目标是在预测场景 x_{pred} 下评估关注量 $Q = h_{\mathrm{qoi}}$(x_{pred}， Θ) 预测的准确性。我们重申，式（5）中给出的差异 E 是一种有效的方式，可用于表征模型观测的准确性。实验观测和模型观测都应在特定场景（即所谓的验证场景 x_{val}）下进行，并采用给定的传感器参数 z_{val}。目前，假设我们已经确定了该验证场景 x_{val}、可观测量的类型（即泛函 h_{obs} 的选择）以及传感器参数的取值 z_{val}。正如引言中所述，人们已经提出了多种验证模型的方法。其中一些验证

算法1： 校准过程

输入：模型 (r, p, u)、观测泛函 h_{obs} 的选择、误差 $E(x, z)$ 的选择、先验密度 f_{Θ_0}、约束集 X_{lab} 和 Z_{lab} 1 通过校准实验的最优设计计算校准场景 x_{cal} 和传感器参数 z_{cal} ；2 实施实验以获得 $y_{\mathrm{exp}}(x_{\mathrm{cal}}, z_{\mathrm{cal}})$；3 计算后验密度 $f_{\Theta|Y_{\mathrm{exp}}=y_{\mathrm{exp}},x_{\mathrm{cal}},z_{\mathrm{cal}}}$ (9)，或通过抽样方法（例如马尔可夫链蒙特卡罗）进行近似；输出：后验密度 $f_{\Theta|Y_{\mathrm{exp}}=y_{\mathrm{exp}},x_{\mathrm{cal}},z_{\mathrm{cal}}}$

流程需要选择一种验证指标。例如，Roy 和 Oberkampf [35]采用了 Ferson 等人 [15]提出的面积验证指标

$$d(F_Y, F_{Y_{\mathrm{exp}}}) = \int_{-\infty}^{\infty} \left| F_Y(s) - F_{Y_{\mathrm{exp}}}(s) \right| \mathrm{d}s, \tag{12}$$

其中， F_Y 和 $F_{Y_{\mathrm{exp}}}$ 分别是模型分布和实验观测分布。他们为选择这一验证指标提供了若干理由，并介绍了多种使用方法，尤其是在存在不同验证场景时。另一种验证指标采用差异 E ，并对误差设定容差 ε

$$\gamma = F_{|E|}(\varepsilon) = \Pr(|E| < \varepsilon) = \Pr(|Y_{\mathrm{exp}}(x_{\mathrm{val}}, z) - h_{\mathrm{obs}}(x_{\mathrm{val}}, \Theta, z)| < \varepsilon). \tag{13}$$

如果 γ 的值高于某个阈值 η，则认为模型有效。Rebba 和 Mahadevan [33] 提出了这一特定度量，并将其命名为 模型可靠性度量。在他们的论文中，作者从频率学派和贝叶斯视角出发，将模型可靠性度量与点假设检验和区间假设检验进行了比较。他们发现，采用假设检验方法可能会导致模型是否有效的判断前后不一致，而模型可靠性度量则不会出现这种不一致。Li 和 Mahadevan [24] 将这一模型可靠性度量扩展到多元输出。Mullins 等人 [27] 进一步研究了模型可靠性度量在验证中的应用。在这篇论文中，作者指出，使用上述面积验证指标可能会产生误导。事实上，增大实验观测的不确定性会使 $F_{Y_{\mathrm{exp}}} \approx F_Y$ 对应的面积验证指标变小（模型有效），但由于实验数据的质量较低，我们本应对模型有效性的判断更缺乏信心。基于这些原因，本文余下部分将选择可靠性度量（13）作为验证指标。

要计算概率 γ，需要知道差异 E 的分布。这里假设随机变量 $Y_{\mathrm{exp}}(x_{\mathrm{cal}}, z)$ 和 Θ 相互独立。这个假设是合理的，因为第一个量与实验观测有关，而第二个量与关于模型的各种假设有关。随后，我们可以使用前面提到的任意一种抽样方法，通过分别对 $Y_{\mathrm{exp}}(x_{\mathrm{cal}}, z)$ 的经验分布和模型参数 Θ 进行抽样，来估计 $\gamma = F_{|E|}(\varepsilon)$。

2.7.1. 验证场景

验证场景 x_{val}的选择、可观测量的类型以及传感器参数 z_{val} 的取值，会对验证过程的实际结果产生重大影响，这将在第4.1节中加以说明。我们很少能够在受控环境中重现我们希望进行预测时所处的条件，也就是说，预测场景与验证场景相同。如果无法做到这一点，那么我们必须针对预测场景设计一个具有信息量且相关的验证场景。此外，如果感兴趣量不可观测，我们就需要确定验证时应采用哪种类型的可观测量以及哪些传感器参数 z 。这些是本研究探讨的主要问题，并将在第3节中进一步讨论。

2.7.2. 验证工作流程

我们工作的另一个目标是尽可能消除验证过程中的任意决策。实际上，通过提供一种确定验证场景 x_{val}、可观测量类型和传感器参数 z_{val}的方法，我们离自动化验证过程又近了一步。我们对验证工作流程应有何种形式的设想总结于算法2中。

算法2： 验证过程

输入： 模型（r, p, u）、关注量 h_{qoi}、预测场景 x_{pred}，选择观测泛函 h_{obs}、误差 $E(x, z)$、先验密度 f_{Θ_0}，约束集 $\mathcal{X}_{\mathrm{lab}}$ 和 $\mathcal{Z}_{\mathrm{lab}}$、误差容限 ε、阈值 η

设定 $f_{\Theta} = f_{\Theta_0}$；
如果 校准实验可用 **则**
  模型参数校准 Θ (算法1);
  设定 $f_{\Theta} = f_{\Theta|Y_{\mathrm{exp}}=y_{\mathrm{exp}},\ x_{\mathrm{cal}},\ z_{\mathrm{cal}}}$；
结束
计算 $Q = h_{\mathrm{qoi}}(x_{\mathrm{pred}},\ \Theta)$;
如果 $Q = h_{\mathrm{qoi}}(x_{\mathrm{pred}}, \Theta)$ 的不确定性过大 **则**
  退出验证;
否则
  继续验证;
结束
计算验证场景 x_{val}、传感器参数 z_{val} 和观测类型泛函 $h_{\mathrm{obs},\ \mathrm{val}}$ (算法3);
执行实验以获取 $y_{\mathrm{exp}}(x_{\mathrm{val}},\ z_{\mathrm{val}})$;
计算验证指标 γ (13);
如果 $\gamma \geq \eta$ **则**
  模型未被判定为无效
否则
  模型无效
结束

验证过程的第一步是利用先验知识或通过校准过程（例如算法1中所述的过程），定义模型参数 Θ 的先验密度 f_{Θ_0}。第二步是进行合理性检查，即检查在模型参数 Θ 的不确定性已知的情况下，$Q = h_{\mathrm{qoi}}(x_{\mathrm{pred}}, \Theta)$ 的预测是否存在过大的不确定性。如果是，那么继续验证过程就无关紧要了，因为无论模型是否有效，该预测都没有实际用途。如果认为感兴趣量的预测不确定性可以接受，就需要设计适当的验证实验。随后进行验证实验，并估算验证指标，以确认其预测确实反映了现实。

最后需要说明的是，本文不讨论建模误差如何传播至预测场景下的感兴趣量，也不讨论如何决策模型是否有效。我们的主要目标是寻找最佳验证实验，以便更好地评估模型预测感兴趣量的能力。尽管如此，我们推测，下一节所述的最优验证实验设计方法能够补充算法2中介绍的验证工作流程，从而为评估模型有效性提供有用的信息。

3. 验证实验的最优设计

用于获取验证数据的验证场景必须能够代表预测场景，因为目标是判断模型是否适用于预测目的，更具体地说，是否适用于对感兴趣量 $Q = h_{\mathrm{qoi}}(x_{\mathrm{pred}},\ \Theta))$ 的预测。这里所说的能够代表，是指模型的各种假设在预测场景和验证场景中都应得到相似程度的满足

。在许多应用中，通常会基于无量纲数来设计验证实验。例如，在流体力学中，雷诺数常被用作选择验证条件的准则，以反映预测场景。然而，这种简单的选择不一定能为验证实验的设计提供完整的框架。首先，模型通常涉及的参数多于单位，因此需要同时考虑多个无量纲数。此外，仅凭量纲分析无法确定每个无量纲数的相对影响。实际上，量化无量纲数对感兴趣量的影响很重要，原因有以下两点：

1. 模型参数 Θ 中的不确定性会传播至无量纲参数；2. 验证与预测设置中的无量纲数很少能完全匹配。

尽管量纲分析是深入理解模型的有用工具，但对于验证而言，它似乎并不充分，且定义也不够明确。我们认为，在验证与预测场景中，模型相对于参数及其不确定性的行为应尽可能相似。换言之，参数敏感性分析应指导验证实验的最优设计。

3.1. 敏感性分析

敏感性分析是一个涵盖多种目标和方法的广阔领域，详尽综述可参见 [12, 32, 37]。敏感性分析所涉及的问题例如：

- 刻画参数与模型输出之间的关系。
- 识别影响较小的参数，这一主题在文献中通常称为因子固定。
- 识别对模型输出影响最大的参数，通常称为因子优先级排序。这里的一个目标是进一步降低这些参数所受的不确定性。

敏感性方法通常分为局部方法和全局方法。一方面，局部方法计算参数空间中特定点附近的敏感性（例如导数），且不考虑影响参数的不确定性。另一方面，全局方法会将参数的不确定性纳入考量。基于 Sobol' 的敏感性分析是一种全局方法，用于计算所谓的 Sobol' 敏感性指数，以量化模型输出的方差有多少可归因于单个参数或一组参数 [39]。

在我们的方法中，我们认为模型参数 Θ 具有不确定性，因此需要采用全局敏感性分析方法。此外，我们的目标是评估各个参数对模型输出（无论是可观测量还是感兴趣量，即前文所述的第一个目标）的影响，因此我们希望描述观测和感兴趣量泛函相对于参数的响应面。因此，我们的目标不同于 Sobol' 方法，例如将 $Y = h_{\text{obs}}(x, \Theta, z)$ 或 $Q = h_{\text{qoi}}(x, \Theta)$ 的方差分配给各个参数（或参数子集）。

3.2. 主动子空间方法

一种能够表征响应面并考虑参数不确定性的敏感性方法是主动子空间方法 [10, 11]。该方法计算模型泛函 h 相对于控制参数和模型参数的梯度，以量化它们的影响。对梯度外积关于分布 F_Θ 求期望，从而构建所谓的影响矩阵 M_h

$$\begin{aligned} M_h(x,z) &= \mathbb{E}\left(\nabla_{(x,\theta)} h(x,\Theta,z)\, \nabla_{(x,\theta)} h(x,\Theta,z)^T\right) \\ &= \operatorname{Cov}\left(\nabla_{(x,\theta)} h(x,\Theta,z)\right) + \mathbb{E}\left(\nabla_{(x,\theta)} h(x,\Theta,z)\right) \mathbb{E}\left(\nabla_{(x,\theta)} h(x,\Theta,z)\right)^T. \end{aligned} \tag{14}$$

接下来，对 $M_h(x, z)$ 进行特征值分解，以确定影响模型输出的主方向。该影响矩阵的特点是其梯度变化较大，

和/或期望梯度较大。从现在起，影响矩阵 M_h 将构成功能 h 响应面的定量描述。

给定两个影响矩阵 M_1 和 M_2，它们由不同的泛函 h、不同的场景 x和不同的传感器参数 z所表征，我们将这两个影响矩阵之间的距离定义为

$$d(M_1, M_2) = \|M_1 - M_2\|_2, \tag{15}$$

其中 $\|M\|_2$ 表示 M 的谱范数（也可以考虑其他范数）。该距离实质上度量了影响矩阵 M_1 与目标影响矩阵 M_2之间的误差。我们指出，使用主动子空间方法的前提是观测值和感兴趣量泛函具有足够的正则性。

在定义了泛函响应面的量化方法以及比较方法之后，我们转而关注当前问题：定义一种专门针对给定预测场景下感兴趣量预测而设计的验证实验。

3.3. 最优验证情景的选择

我们将受控实验集描述为一组场景 $\mathcal{X}_{\text{lab}} \subseteq \mathcal{X}$，在这些场景下，我们能够通过实验室实验来探查关注系统。需要指出的是，预测场景 x_{pred} 可能属于 $\mathcal{X}_{\text{lab}}$，也可能不属于。因此，我们针对最优验证场景提出两项要求：

1. 如果 $x_{\text{pred}} \in \mathcal{X}_{\text{lab}}$，那么验证场景应由 $x_{\text{val}} = x_{\text{pred}}$给出。2. 如果感兴趣量泛函 h_{qoi} 实际上是一个可观测量 h_{obs}，那么用于验证的可观测量和传感器参数应与感兴趣量所对应的可观测量和传感器参数一致。

这两项要求将指导我们定义两个最优设计问题，以确定目标导向验证实验。确定验证实验的第一步是确定验证场景 x_{val}。这里回顾一下，感兴趣量泛函 h_{qoi} 仅是控制参数 x 和模型参数 Θ的函数。此外，验证场景和预测场景中的模型参数 Θ 相同，因此我们有两个感兴趣随机量 $Q_{\text{val}} = h_{\text{qoi}}(x_{\text{val}}, \Theta)$ 和 $Q_{\text{pred}} = h_{\text{qoi}}(x_{\text{pred}}, \Theta)$。我们寻求使距离（15）最小化的验证场景 x_{val}，同时要求该场景能够在受控环境中实现

$$x_{\text{val}} \in \underset{x \in \mathcal{X}_{\text{lab}}}{\operatorname{argmin}}\, d(M_{h_{\text{qoi}}}(x), M_{h_{\text{qoi}}}(x_{\text{pred}})). \tag{16}$$

换言之，我们要寻找这样的验证场景：在预测场景中，Q_{val}（由影响矩阵 $M_{h_{\text{qoi}}}(x_{\text{val}})$ 编码）的响应面与 Q_{pred} 的响应面最为相似。这也就是说，导致模型建立的各种假设和简化，在验证设置和预测设置中具有相当的影响。我们注意到，如果 $x_{\text{pred}} \in \mathcal{X}_{\text{lab}}$，那么目标函数（16）非负，因此在 $x_{\text{pred}} = x_{\text{val}}$ 处可达到值为零的全局极小值。然而，全局极小值可能并不唯一。事实上，对于参数而言，线性泛函 h_{qoi} 无论这些参数取何值，都具有相同的影响矩阵 $M_{h_{\text{qoi}}}(x)$。因此，上文所述的第一项要求仅部分得到满足，因为优化问题（16）可能不存在唯一的极小值。

3.4. 观测量与传感器参数的最优选择

本节旨在定义一个问题，其解将给出观测泛函 h_{obs} 和传感器参数 z，以便全面表征最优验证实验。我们回顾一下，感兴趣量在预测场景或验证场景中不一定都是可观测量。尽管如此，我们仍希望验证模型对感兴趣量的预测。通过求解式（16）得到最优预测场景后，我们现在需要确定泛函 h_{obs} 和传感器参数 z，使其尽可能准确地模拟 $Q_{\text{val}} = h_{\text{qoi}}$（$x_{\text{val}}$，$\Theta$）的影响矩阵。同样，其依据是影响矩阵反映了模型中各种假设的影响。

算法3： 验证实验的最优设计

输入： 模型（r，p，u），关注量 h_{qoi}，预测场景 x_{pred}，模型参数 Θ，受控场景集合 $\mathcal{X}_{\mathrm{lab}}$，观测泛函集合 $\mathcal{H}_{\mathrm{lab}}$，受控传感器参数集合 $\mathcal{Z}_{\mathrm{lab}}$

1 通过求解（16）计算验证场景 x_{val}；

2 计算验证观测泛函 h_{obs} ,验证和验证传感器参数 z_{val}，通过求解式（17）；

输出： 验证场景 x_{val}，验证观测泛函 h_{obs} ,验证、验证传感器参数 z_{val}

利用这一特定泛函 h_{obs} 和传感器参数 z得到的验证过程结果，便可以合理地外推至感兴趣量。由于观测泛函和传感器参数 z 的选择受现有实验能力的限制，我们在此引入可能的观测泛函集合$\mathcal{H}_{\mathrm{lab}}$ 以及可能的传感器参数集合$\mathcal{Z}_{\mathrm{lab}}$ 。需要求解的最优设计问题如下:

$$(h_{\mathrm{obs,val}}, z_{\mathrm{val}}) \in \underset{(h_{\mathrm{obs}},z)\in\mathcal{H}_{\mathrm{lab}}\times\mathcal{Z}_{\mathrm{lab}}}{\operatorname{argmin}} d\left(\frac{M_{h_{\mathrm{obs}}}(x_{\mathrm{val}},z)}{\mathrm{Tr}(M_{h_{\mathrm{obs}}}(x_{\mathrm{val}},z))}, \frac{M_{h_{\mathrm{qoi}}}(x_{\mathrm{val}})}{\mathrm{Tr}(M_{h_{\mathrm{qoi}}}(x_{\mathrm{val}}))}\right). \tag{17}$$

由于观测泛函和感兴趣量泛函可能代表具有不同尺度和单位的不同物理量，因此需要对影响矩阵进行归一化。面向感兴趣量预测而设计的验证实验现已完全确定。该实验包括在场景 x_{val}下，以传感器参数 z_{val} 对关注系统进行探测，以获取与观测泛函$h_{\mathrm{obs,\ val}}$ 对应的实验观测。搜索最优实验的整个过程见算法3。

问题（16）和问题（17）的目标泛函值都能提供有关验证实验质量的有用信息。如果两个目标泛函值都很小，那么验证场景下可观测量 h_{obs} 的影响矩阵就能很好地模拟预测场景下感兴趣量的影响矩阵。因此，该验证实验与预测情境相关。如果问题（16）的目标泛函值较大，则可能需要考虑扩大受控场景集合 $\mathcal{X}_{\mathrm{lab}}$。此外，如果问题（17）的目标泛函值较大，则可以考虑采用其他类型的实验测量。若出现其中一种或两种情况，在将验证过程的结果外推至预测场景时都必须谨慎。

问题（16）和问题（17）的优化问题由求解器 NOMAD [7]求解，详情见附录 C。

4. 数值示例

我们通过两个示例来说明我们的方法：第一个是第 2.1 节介绍的抛射体问题，第二个是污染物输运问题。第一个问题用于说明整个验证过程，并强调验证实验设计的重要性。具体而言，我们将给出一个验证实验选择不当的示例，该实验会导致假阳性，也就是说，模型实际上无效，却未被判定为无效。相比之下，我们将展示，应用我们的方法设计最优验证实验可以得到真阴性。第二个问题将说明感兴趣量的定义对验证实验的重要性。

4.1. 抛射体问题

第一个示例考虑第 2.1 节中描述的关注系统，我们要确定球形抛射体垂直发射时达到的最大高度。该模型由线性

表2：预测场景 x_{pred}、受控环境 $\mathcal{X}_{\text{lab}}$ 和暂定验证场景 $x_{\text{val, tent}}$

	质量 m	直径 ℓ	初始位置 u_0	初始速度 v_0
预测场景 x_{pred}	0.05	0.01	1	100
受控环境 $\mathcal{X}_{\text{lab}}$	[1,5]	[0.05,0.1]	[0,2]	[10,120]
暂定验证场景 $x_{\text{val,tent}}$	2.5	0.05	1	20

普通　　微分方程 (2) 给出，可据此通过解析方法计算感兴趣量的值　　as

$$q(x,\theta) = \max_t u(t) = u_0 + \frac{mv_0}{3\pi\exp(\mu)\ell} - \frac{m^2 g}{(3\pi\exp(\mu)\ell)^2}\ln\left(1+\frac{3\pi\exp(\mu)\ell v_0}{mg}\right). \tag{18}$$

回顾一下，这里的控制参数为 $x=(m,\ \ell,\ u_0,\ v_0)$，而模型参数为 $\theta=(g,\ \mu)$。由于所提出的感兴趣量关于某些控制参数是非线性的，因此相关的影响矩阵 (14) 取决于场景 x。此外，我们将模型参数设为 $\Theta=(G,\ U)\sim(\mathrm{N}(9.81,\ 0.01_2),\ \mathrm{N}(-5\ln(10)\,0.5_2))$，其中 N 表示正态分布。这些数值的选取依据是：我们对引力常数 g 的取值比较了解，但对空气黏度 μ的了解不那么充分。

4.1.1. 最优验证情景

为了说明设计不佳的验证实验可能会导致假阳性，我们需要考虑一个特定的预测场景 x_{pred} 以及一个特定的受控环境$\mathcal{X}_{\text{lab}}$，该环境描述了可能的验证场景 x_{val}。预测场景、一个暂定验证场景以及受控环境均在表2中进行了描述。

暂定验证场景的选取使得在初始时刻取值的雷诺数 Re,

$$\mathrm{Re}(t) = \frac{\rho\ell}{\exp(\mu)}\left|\frac{\mathrm{d}u}{\mathrm{d}t}(t)\right|, \tag{19}$$

其中 ρ 为环境空气的密度，与预测场景 x_{pred}的雷诺数相匹配。这一理由与第3节开头的讨论相呼应，因此可将其视为一种可行的验证实验。然而，如图3所示，感兴趣量在 $x_{\text{val, tent}}$ 处的影响矩阵与感兴趣量在 x_{pred}处的影响矩阵差异颇大。受控环境 $\mathcal{X}_{\text{lab}}$ 由参数 x对应区间的笛卡尔积构成，定义见表2。现在，我们求解问题（16），以获得这一预测设置下的最优验证情景，并得到 $x_{\text{val}}=(1.000,\ 0.100, 0.705, 100.446)$。图3比较了预测场景 x_{pred}、暂定验证场景 $x_{\text{val, tent}}$以及上述最优验证情景 x_{val}下感兴趣量（18）的影响矩阵 M_h (14)。

我们观察到，与最大特征值 λ_1相关联的第一个特征向量 v_1在$x_{\text{val, tent}}$ 和 x_{val} 之间发生了显著变化，以便更接近预测场景中的特征向量（尤其是在 u_0、 v_0和 g 分量上）。第二个特征向量在暂定验证场景和最优验证场景之间也发生了显著变化。与预测场景相比，第二和第三个特征向量似乎互换了，但这可以用最优验证场景的第二和第三个特征值几乎相同来解释。除此之外，我们实际上观察到特征向量之间吻合良好。换言之，相较于暂定验证场景， x_{val} 处感兴趣量的影响矩阵更能模拟 x_{pred} 处感兴趣量的影响矩阵。

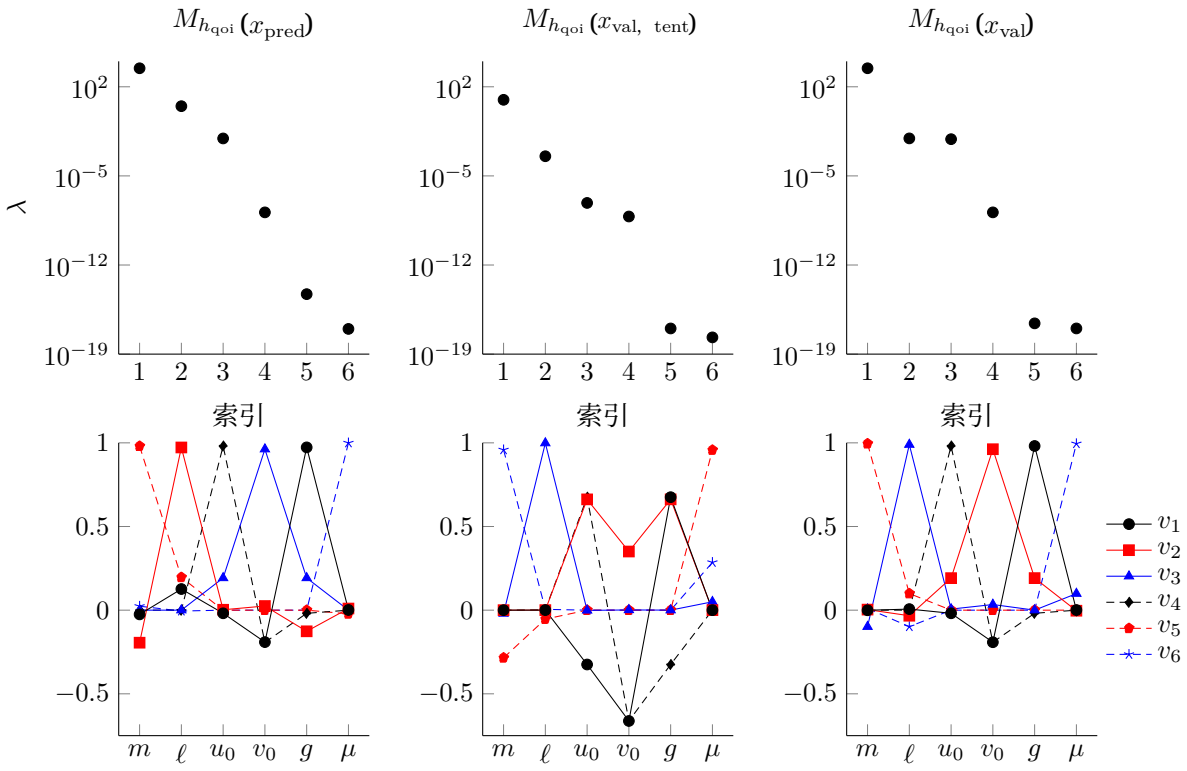


图3：预测场景 x_{pred} （左）、暂定验证场景 $x_{\mathrm{val,\ tent}}$ （中）和最优验证场景 x_{val} （右）下，与感兴趣量（18）的影响矩阵相关的特征值（上排）和特征向量（下排）。

4.1.2. 最优传感器与观测泛函

现在，我们要在观测泛函 h_{obs}（即泛函 $h_{\mathrm{obs}}(x_{\mathrm{val}}, \Theta, z) := u$和 $h_{\mathrm{obs}}(x_{\mathrm{val}}, \Theta, z) := \mathrm{d}^2u/\mathrm{d}t^2$）中寻找最佳者，这些泛函是时间 t的函数；同时还要寻找在验证实验中应采用的最佳传感器参数 $z := t$。为此，我们根据之前计算得到的验证场景 x_{val}，针对这两种观测泛函以及验证场景下的多个时间 t，计算问题（17）的目标泛函。结果如图4所示。

如预期，存在一个特定时刻 t，此时 $h_{\mathrm{obs}}(x_{\mathrm{val}}, \Theta, z) := u(t)$ 的影响矩阵与 $h_{\mathrm{qoi}}(x_{\mathrm{val}}, \Theta)$ 的影响矩阵完全相同。该特定时刻 t 对应抛射体达到最大高度的瞬间（如图5所示）。值得注意的是，图4显示，位置 u 和加速度 $\mathrm{d}^2u/\mathrm{d}t^2$ 对应的目标泛函随时间的变化截然不同。一方面，在所有时刻 t，加速度的影响矩阵都与验证场景下感兴趣量的影响矩阵有较大差异。另一方面，位置 u 在较小时刻 t 的目标泛函（17）相当大。这是因为发射后的位置主要受初始位置 u_0 和初始速度 v_0的影响，而验证场景下的感兴趣量并非如此，见图3下排最右侧的图。不过，问题（17）中位置的目标泛函迅速减小，这表明初始位置和速度（以及其他参数）的影响很快趋近于感兴趣量的影响。

为了呈现完整的验证过程，并说明可能得到错误验证结果，我们需要一些实验观测。我们将使用由下文称为精细模型 的模型所提供的生成数据或合成数据，该模型充当关注物理系统的替代模型。实际上，实验观测的具体来源在此并不直接相关，因为它们在最优验证实验的设计中不起任何作用。

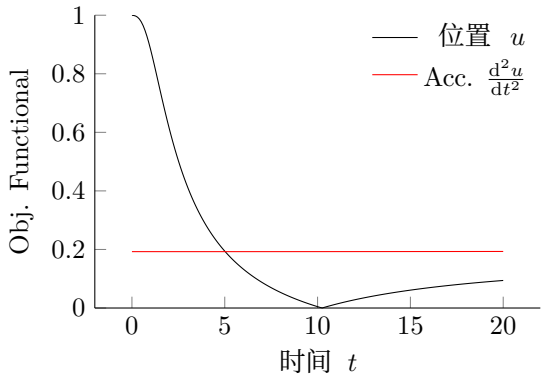


图4：与观测泛函 $h_{\mathrm{obs}}(x_{\mathrm{val}},\ \Theta,\ z) := u$ 和 $h_{\mathrm{obs}}(x_{\mathrm{val}},\ \Theta,\ z) := \mathrm{d}^2u/\mathrm{d}t^2$相关的问题（17）的目标泛函。

4.1.3. 制造数据

精细模型由以下非线性常微分方程 [43, 第 7 节6] 和初始条件给出

$$m\frac{\mathrm{d}^2\tilde{u}}{\mathrm{d}t^2}(t) + \frac{\rho\pi\ell^2 c_D}{8}\frac{\mathrm{d}\tilde{u}}{\mathrm{d}t}(t)\left|\frac{\mathrm{d}\tilde{u}}{\mathrm{d}t}(t)\right| = -mg, \qquad \text{for } t \in (0,T), \tag{20a}$$

$$\tilde{u}(0) = u_0, \tag{20b}$$

$$\frac{\mathrm{d}\tilde{u}}{\mathrm{d}t}(0) = v_0. \tag{20c}$$

根据 [44], 摩擦系数 c_D 的定义，其取决于雷诺数 Re (19)，并且对于层流状态（即 Re $< 2\times10^5$）定义如下:

$$c_D = \frac{24}{\mathrm{Re}}\left(1 + 0.15\times\mathrm{Re}^{0.681}\right) + \frac{0.407}{1+8710\times\mathrm{Re}^{-1}}. \tag{21}$$

当雷诺数非常低（Re $\ll 1$）时，该系数可由 $c_D = 24/\mathrm{Re}$ 很好地近似。采用这一近似，实际上可恢复线性模型（2）。

精细模型使我们能够获得任意时刻 t 的位置 $\tilde{u}$ 测量值。通过抽样获得这些伪实验观测 $Y_{\mathrm{exp}}(x,z)$ 的实现值 $y_{\mathrm{exp}}(x,z)$,

$$Y_{\mathrm{exp}}(x,z) \sim \tilde{u}(x,z)\,\mathrm{N}(1,0.05^2). \tag{22}$$

这种噪声选择是任意的，对应于标准差为 5% 的乘性高斯噪声。这里我们假设，在求解常微分方程（20）时，离散化误差得到了精确控制。最后，对于这个精细模型，我们分别将引力常数、空气黏度和空气密度的值设为 $g = 9.81$、exp(μ) $= 1.8\times10^{-5}$ 和 $\rho = 1.2$。

4.1.4. 验证分析

现在，我们比较模型（2）和模型（20）在预测场景 x_{pred}、暂定验证场景 $x_{\mathrm{val,tent}}$ 以及最优验证情景 x_{val} 下得到的轨迹。此外，我们在三个场景各自的最优传感器参数 z_{val} 处计算差异（ 5 ），该参数对应于抛射体到达最大高度的时刻。图5展示了差异 E 的轨迹和分布 $F_{|E|}$ 。需要指出的是，对于预测场景，由于没有实验数据，通常无法获得差异 E 的分布和物理系统的轨迹（红色曲线）。在这里，我们之所以能够绘制这些图，是因为测量值是借助模型（20）获得的。

我们观察到，模型（2）未能准确捕捉抛射体在预测场景下达到的最大高度，因为红色曲线与蓝色曲线之间存在显著差异，如

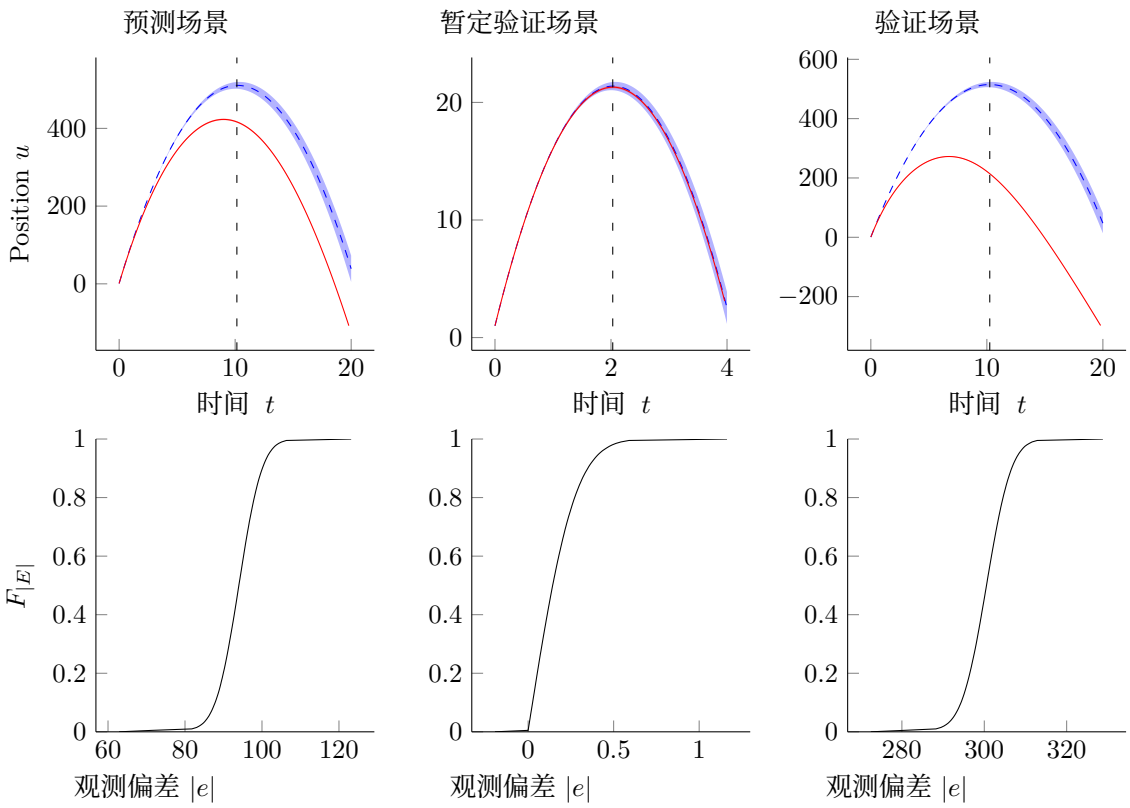


图5：预测场景（左列）、暂定验证场景（中列）和最优验证情景（右列）中的位置 u （上排）及差异（下排）。实验观测以红色表示，模型观测以蓝色表示。阴影区域表示实验观测和模型观测的95%置信区间。差异的绝对值（5）是在黑色虚线所示的传感器位置处计算的。

图5左上图所示。实际上，模型中的摩擦力被低估了。模型与关注系统之间的这种不匹配反映在差异的分布 $F_{|E|}$ 中：绝对误差小于80的概率接近于零。换言之，几乎可以确定最大位置的绝对误差会介于80和110之间。因此，在预测场景 x_{pred} 下用于预测最大高度的模型可视为无效。尽管实际中无法获得这一信息，验证过程应事先告知我们模型预测的准确性。

在暂定验证场景$x_{\text{val, tent}}$中，我们观察到模型与关注系统的抛射体位置高度一致。这可以解释为，在此情况下，摩擦力相较于惯性力和重力可以忽略不计。由于模型与关注系统之间的唯一区别在于摩擦力的建模方式，因此，在最优传感器点 $z = t \approx 2$处计算的差异 E仍然很小。其分布 $F_{|E|}$ 表明，观测误差几乎肯定小于0.5。根据验证指标的容差 ε 和阈值 η ，该模型很可能会被判定为有效。因此，该模型不适用于预测场景中的感兴趣量预测，而采用这一暂定验证场景进行验证过程则会导致假阳性。

在最优验证情景 x_{val} 下，我们观察到模型与关注系统之间存在显著偏差。最优验证情景似乎捕捉到了模型中摩擦力被低估这一事实。根据此前计算出的最优传感器点，差异表明位置误差几乎可以肯定介于 280 和 310 之间。模型很可能无法通过验证测试，而这更能反映模型在预测场景下的真实预测能力

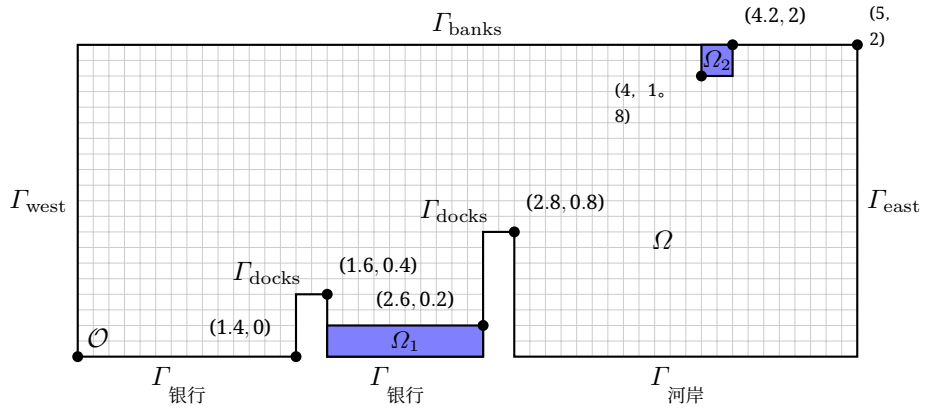


图6：污染物输运模型的区域 Ω 及其边界。我们要获取污染物平均浓度的区域由 Ω_1 和 Ω_2 标出（分别用于定义感兴趣量）。网格用于标示可以观测污染物平均浓度的区域 Ω_{obs} 。

预测场景下模型的。

这里我们想强调，所提出的最优验证实验设计方法并不能保证验证过程的结果总能给出正确答案。它只能确保验证场景下观测的影响矩阵尽可能与预测场景下感兴趣量的影响矩阵相当。在本例中，最优验证实验的条件使模型能够捕捉摩擦力相对于其他力的重要性，这与预测场景中的情况一致。换言之，由于验证设置中摩擦力未被正确建模，因此可以推断预测场景中也可能存在同样的问题，而本例中确实如此。

4.2. 污染物输运问题

第二个示例是一个虚构河流中的污染物输运问题，我们希望分析上游新建工厂的设计所带来的影响。我们考虑两个感兴趣量，它们都表示区域内某一特定区域的污染物平均浓度。我们希望验证的模型是一个二维线性稳态扩散-平流方程，该方程描述河流中污染物的浓度 $\phi = \phi(z)$：

$$-\nabla \cdot (\exp(k)\nabla\phi) + \nabla \cdot (v\phi) = 0, \qquad \text{in } \Omega, \tag{23a}$$

$$\phi = \phi_D, \qquad \text{on } \Gamma_{\text{west}}, \tag{23b}$$

$$-n \cdot \exp(k)\nabla\phi = 0, \qquad \text{on } \partial\Omega \setminus \Gamma_{\text{west}}, \tag{23c}$$

其中， n 表示区域边界的外法向单位向量，$\exp(k)$为扩散系数， v 为平流速度。为简便起见，我们假设速度场 v 已知且给定，因此在本研究中不将其视为参数（见附录 A）。

控制参数 x 由刻画狄利克雷条件 ϕ_D的参数组成，稍后将对此进行描述；模型参数仅为 $\theta = k$。这里我们假设扩散参数 k 已知，因此 $\Theta \sim \delta(k_0)$，其中 δ 为狄拉克测度， $k_0 = -2$为 $\ln(10)$。传感器参数为笛卡尔坐标 $z = (z_1, z_2)$。图6展示了几何结构示意图。区域在西侧和东侧边界分别设有入口和出口。两个码头阻碍流体绕过第一个感兴趣区域，该区域记为 Ω_1。第二个感兴趣区域 Ω_2 如图6所示。需要指出的是，区域 Ω 的具体尺寸以及模型参数 k 的具体取值均为任意选定，因为此示例仅用于说明。

为了验证模型，我们设想可以进行一项实验：在上游注入少量且可控的污染物。因此，狄利克雷条件 ϕ_D 用于描述

表3：污染物输运模型预测场景 x_{pred} 及受控环境 $\mathcal{X}_{\text{lab}}$ 的描述。

	位置 z_0	**长度** L	**强度** c
预测场景 x_{pred}	0.75	0.6	2
受控环境 $\mathcal{X}_{\text{lab}}$	[0.1,1.9]	[0.05,0.1]	[0.5,1]

工厂的污染物释放和验证实验。我们将 ϕ_D 参数化为以 $z_2 = z_0$ 为中心、长度为 L 、强度为 c 的光滑化函数，如下所示：

$$\phi_D(z_0, L, c, z_2) = \begin{cases} c \exp\left(\left(\left| \frac{z_2 - z_0}{L} \right| - 1 \right)^{-1} \right), & \text{if } \left| \frac{z_2 - z_0}{L} \right| < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{24}$$

本应用中的控制参数由 $x = (z_0, L, c)$ 给出。这种参数化足够丰富，既能描述工厂的污染物释放，也能涵盖广泛的验证场景，同时涉及的参数数量有限。此外，由于上述光滑化函数具有良好的正则性，因此颇具吸引力，因为我们需要计算泛函相对于控制参数 x 的梯度（如附录B所述）。

现在，我们可以明确预测场景 x_{pred} 和受控环境 $\mathcal{X}_{\text{lab}}$。预测场景描述了拟建工厂上游可能出现的污染物分布，并列于表3。受控环境 $\mathcal{X}_{\text{lab}}$ 包括所有实际可执行的验证场景。这里假设可以沿边界 Γ_{west} 任意位置以受控方式注入污染物，而长度 L 和强度 c 只能取有限的值。三个控制参数 z_0、 L和 c 的取值区间也列于表3。

我们还考虑以下两个感兴趣量：

$$h_{\text{qoi},i}(x, \theta) = \frac{1}{|\Omega_i|} \int_{\Omega_i} \phi(x, \theta, z) \, \mathrm{d}z, \qquad \text{for } i = 1, 2,$$

其中， Ω_i、 $i = 1,2$ 是图6所示的两个子区域。

4.2.1. 最优验证场景

现在，预测场景 x_{pred}、感兴趣量 $h_{\text{qoi},i}$以及受控环境 $\mathcal{X}_{\text{lab}}$均已给定，我们可以针对两个感兴趣量计算最优验证场景。因此，我们针对每个感兴趣量分别求解问题（16）。预测场景 x_{pred}下的污染物浓度，以及针对第一个和第二个感兴趣量的最优验证场景，均显示在图7中。

我们观察到，最优验证场景对我们希望预测的具体感兴趣量高度敏感，因为污染物的位置 z_0 在两个感兴趣量之间差异很大。这一结果很重要，因为它突显了这样一个事实：验证场景应由预测场景与希望预测的感兴趣量共同决定。从预测场景可以看出，对于第一个感兴趣量，污染物进入区域 Ω_1 ；而对于第二个感兴趣量，几乎没有污染物到达第二个区域 Ω_2 。我们在此重申，目标并非寻找满足 $h_{\text{qoi}}(x_{\text{val}}, \Theta) \approx h_{\text{qoi}}(x_{\text{pred}}, \Theta)$ 的验证场景 x_{val} ，而是寻找满足 $M_{h_{\text{qoi}}}(x_{\text{val}}) \approx M_{h_{\text{qoi}}}(x_{\text{pred}})$ 的场景。

影响矩阵 $M_{h_{\text{qoi}}}(x_{\text{val}})$ 和 $M_{h_{\text{qoi}}}(x_{\text{pred}})$ 的特征值及特征向量如图8所示。模型参数 $\theta = k$ 不存在不确定性，这意味着影响矩阵 $M_{h_{\text{qoi}}}$ 应为秩一，因为在这种情况下，公式（14）中的协方差为零。数值结果清楚地体现了这一点：每个影响矩阵的第一个特征值都比其他特征值高出数个数量级。

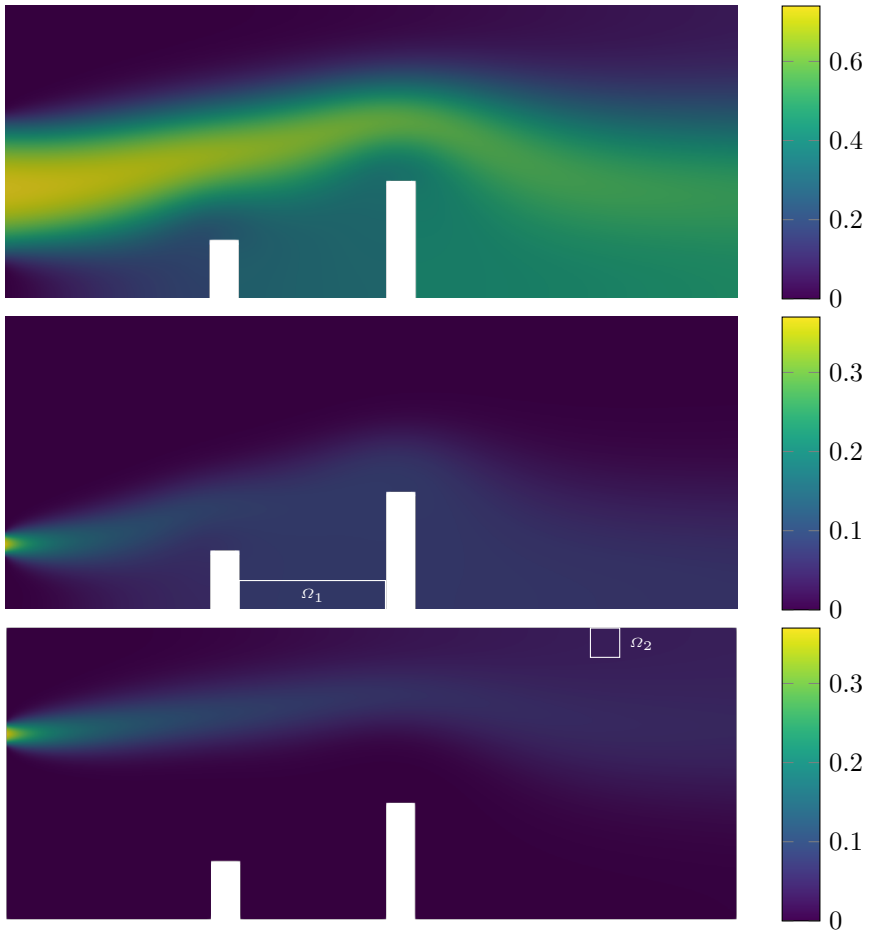


图7：预测场景和最优验证场景下污染物的浓度：（上）预测场景，（中）与 $h_{\text{qoi},1}$ 相关联的最优验证场景，（下）与$h_{\text{qoi},2}$ 相关联的最优验证场景。

关于第一个感兴趣量，我们观察到，验证场景中的特征值幅值略小于预测场景中的特征值幅值。这可以解释为：由于受控环境 $\mathcal{X}_{\text{lab}}$的约束，验证设置中的污染物总量较少。我们还观察到，最优验证场景的特征向量往往会尽可能贴近预测场景的特征向量。造成这一差异的原因可能是，验证场景中注入污染物的数量少于预测场景，因此注入区域的大小 L 应有更大的影响。最优验证场景确实捕捉到了这样一个事实：污染物源的强度 c 和扩散率 k 对确定 Ω_1中的污染物平均浓度影响不大。

关于第二个感兴趣量，我们观察到验证场景下影响矩阵的特征值与预测场景下的特征值高度吻合。与第一个感兴趣量一样，验证场景的特征向量也倾向于尽可能贴近预测场景的特征向量。

4.2.2. 传感器布置

针对所考虑的每个关注量计算出最优验证情景后，我们着手开展方法的最后一步，即传感器的最优位置布设。对于污染物

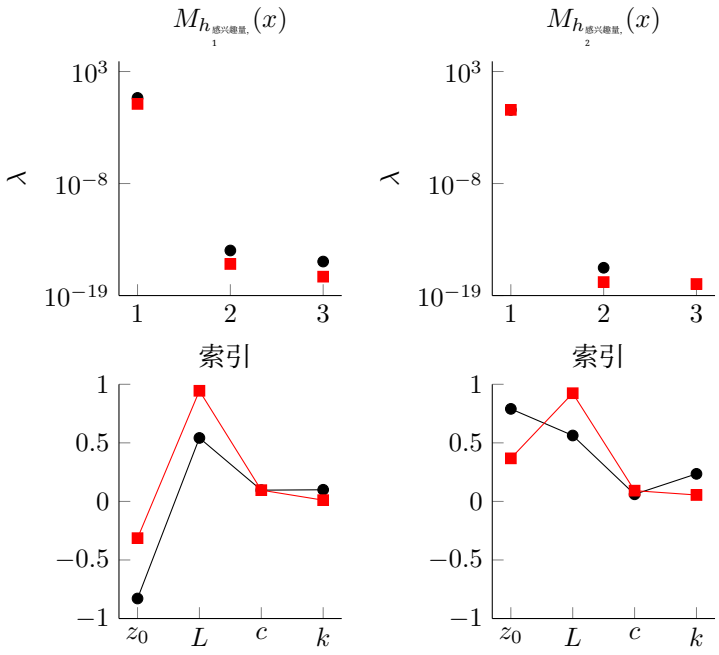


图8：预测场景（黑色圆圈）和验证场景（红色方块）对应影响矩阵的特征值（上排）和特征向量（下排）。左列和右列分别显示与第一个和第二个感兴趣量相关的敏感性指标。为清晰起见，每个影响矩阵仅显示第一个（也是唯一相关的）特征向量。

输运问题，我们仅考虑一种观测泛函 h_{obs}

$$h_{\text{obs}}(x,\theta,z) = \frac{1}{|\Omega_{\text{obs}}(z)|}\int_{\Omega_{\text{obs}}(z)} \phi(x,\theta,w)\,\mathrm{d}w, \tag{25}$$

其中， $\Omega_{\text{obs}}(z)$表示一个以 z 为中心、边长和宽度均为0.1的方形区域。现在的目标是在区域 Ω 内确定方形区域 $\Omega_{\text{obs}}(z)$的位置 z ，使得 h_{obs} 的影响矩阵与验证场景下感兴趣量的影响矩阵尽可能匹配。我们不直接求解优化问题（17），而是绘制方形区域在 Ω内不同位置时问题（17）中的目标函数值。具体而言，方形区域取为图6所示的网格单元。图9展示了两个感兴趣量各自在其对应最优验证场景下的结果。图中方块所示位置为传感器的最佳位置，此处目标函数值最接近于零。

对于第一个感兴趣量，我们清楚地看到，最好将传感器布置在 Ω_1 内；这是显而易见的结果，因为我们最终要预测 Ω_1 中污染物的平均浓度。正如图 A.10 所示，最佳观测区域似乎延伸至发生回流的区域。在一定程度上，该区域还向上游延伸至第一个码头上方，并向下游延伸至第二个码头正上方。此外，我们发现，将传感器布置在区域的上半部分，对第一个感兴趣量的预测毫无帮助。事实上，这一结果颇为直观，因为在图7所示的最优验证情景下，只有极少量污染物流入这些区域，而确有一定量的污染物到达 Ω_1。最后，与第二个码头下游区域的平均浓度相关的影响矩阵，与影响矩阵 $M_{h_{qoi,1}}(x_{\text{val}})$ 有一些相似之处，但与两个码头之间区域的影响矩阵相比，相似程度没有那么高。我们还注意到，下游污染物羽流中目标函数的波动，可能是梯度计算过程中累积数值误差所致（见附录B）。

对于第二个感兴趣量，在这种情况下，将传感器放置在感兴趣区域内同样是最优选择。污染物传感器也可以最优地放置在 Ω_2 的正上游和正下游。有趣的是，将传感器直接放在污染物羽流内部似乎并非最优选择。

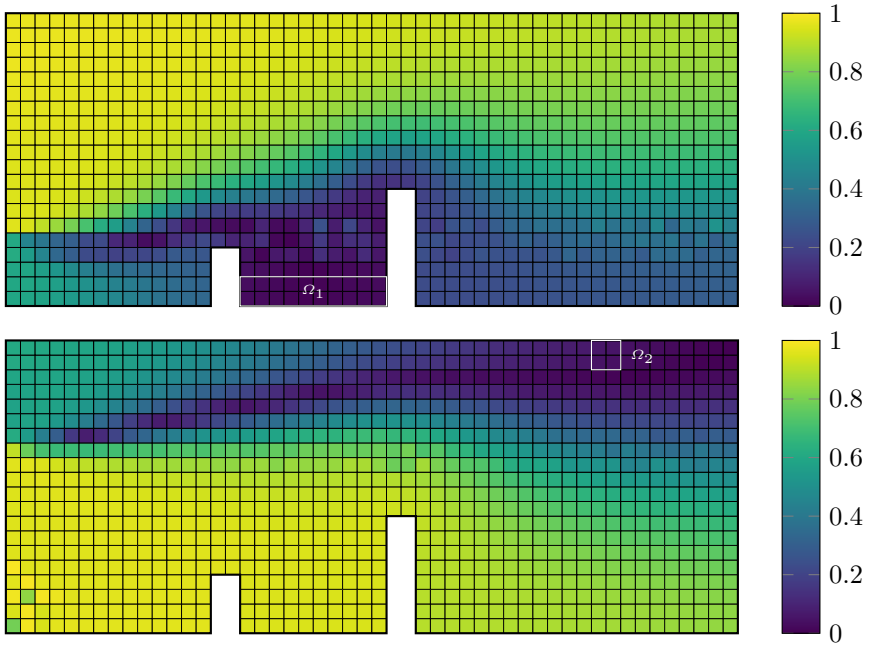


图9：问题（17）中目标函数的值，分别对应第一个感兴趣量（上）和第二个感兴趣量（下）　　.

更好的做法是探测感兴趣量附近的污染物羽流边界区域，因为该处的浓度变化更大，因此测量值对污染物释放位置的变化更敏感。

5. 结论

本文探讨了如何设计针对感兴趣量预测的最优验证实验。据我们所知，文献中很少考虑这一问题。因此，我们详细描述了一个数学模型、各输入参数的作用，以及使用该模型进行预测时的目标。我们还对影响模型参数、实验观测值和计算观测，以及感兴趣量的偶然不确定性进行了统一处理。在此框架下，我们引入了利用主动子空间方法计算得到的影响矩阵，用以定量描述与各个模型泛函相关的响应面。随后，我们提出了一种方法，通过比较影响矩阵来设计最优验证情景并确定最优观测值。该方法实质上包括求解两个优化问题：1）第一个最优设计问题（16）用于计算最优验证情景；2）第二个优化问题（17）用于确定在关注系统上应进行的最佳测量。

该方法在两个数值示例中进行了测试。简单抛射体问题的结果表明，最优验证场景能够恢复预测场景下与感兴趣量相关的影响矩阵的关键特征。我们对给定时间段内通过两种不同可观测泛函得到的影响矩阵进行了比较，并据此确定了一个最优观测泛函和传感器参数。污染物输运问题的结果突出表明，最优验证实验确实取决于我们希望预测的感兴趣量。结果也证实，对于不可观测的感兴趣量，仍可确定具有相似影响矩阵的观测泛函。该数值示例还表明，所提出的方法可切实用于具有工程意义的应用。

本文在多个方面作出了一些假设和简化，这在一定程度上限制了预测问题的适用范围。若放宽控制参数和传感器参数均无不确定性的假设，或许能够得到更稳健的验证实验，尤其是在这些控制参数由边界条件和/或初始条件构成时。此外，使用主动子空间方法定义影响矩阵，要求观测泛函和感兴趣量泛函具有一定的正则性。对于缺乏正则性的模型，可以考虑采用响应面的变差函数分析 [31] 来计算影响矩阵。

另一方面，对抛射体问题的研究表明，仅靠量纲分析不足以选出最优验证实验。然而，通过量纲分析转换参数空间，可能会得到模型泛函影响矩阵的另一种表示形式。这种新的表示形式或许能提供有关模型的额外信息。例如，我们或许能够识别出指导验证实验设计的重要无量纲量。这个具体问题是目前正在进行的研究课题。

此外，敏感性分析除了可用于设计验证实验，也可用于分析现有数据集。通过描述某个数据集下模型的影响矩阵，例如可以从该数据集中找出可用于验证的特定数据。还可以找出旨在测试广泛预测设置的对抗性验证场景。同样，将影响矩阵应用于交叉验证方法也可能颇具启发性。

最后，量化模型误差的方法依赖于以某种方式比较现实与模型可观测量。分析实验选择对预测场景下感兴趣量的建模误差所产生的影响，或许会很有意义。所提出的方法或许能更好地揭示建模误差，我们将在未来的研究中对此进行探讨。

致谢

APR 感谢*Fonds de recherche du Qu´ebec - Nature et technologies*提供的资金支持。SP 和 ML 感谢加拿大自然科学与工程研究委员会（NSERC）发现计划 [的资助，资助编号为 RGPIN-2019-7154、RGPIN-2018-06592]。

参考文献

[1] M. Alnæs、J. Blechta、J. Hake、A. Johansson、B. Kehlet、A. Logg、C. Richardson、J. Ring、M. Rognes 和 G. Wells。FEniCS项目1.5版。《数值软件档案》，3，2015年。[2] M. S. Alnæs、A. Logg、K. B. Ølgaard、M. E. Rognes 和 G. N. Wells。统一形式语言：用于偏微分方程弱形式的领域专用语言。《*ACM*数学软件汇刊》，40(2)，2014年。[3] D. Ao、Z. Hu 和 S. Mahadevan。寿命预测模型验证实验的设计。《可靠性工程与系统安全》，165：22–33，2017。[4] P. D. Arendt、D. W. Apley 和 W. Chen。模型不确定性的量化：校准、模型差异与可辨识性。《机械设计杂志》，134(10)，2012。[5] A. C. Atkinson。最优设计。载于*Wiley StatsRef*：在线统计学参考资料，第1–17页。约翰·威立父子有限公司，2015年。[6] A. C. Atkinson、A. N. Donev 和 R. D. Tobias。最优实验设计（使用 *SAS*）。牛津统计科学丛书第 34 卷。牛津大学出版社，2007年。[7] C. Audet、S. L. Digabel、V. R. Montplaisir 和 C. Tribes。算法1027：NOMAD 4版——使用MADS算法进行非线性优化。《*ACM*数学软件汇刊》，48(3)：22，2022年。

[8] J. Beck、B. M. Dia、L. F. Espath、Q. Long 和 R. Tempone。快速贝叶斯实验设计：采用基于拉普拉斯的重要性抽样估计期望信息增益。《计算力学与工程方法》，334：523–553，2018年。[9] D. Cole。参数冗余与可辨识性。CRC出版社，2020年。[10] P. G. Constantine。主动子空间：参数研究降维的新思路。SIAM焦点丛书第2卷。工业与应用数学学会，2015年。[11] P. G. Constantine、E. Dow 和 Q. Wang。主动子空间方法的理论与实践：在克里金曲面中的应用。《*SIAM*科学计算杂志》，36(4)：A1500–A1524，2014年。arXiv：1304.2070。[12] S. Da Veiga、F. Gamboa、B. Iooss 和 C. Prieur。敏感性分析基础与趋势。计算科学与工程。工业与应用数学学会，2021年。[13] A. DasGupta。统计与概率渐近理论。施普林格统计学丛书。施普林格纽约，2008年。[14] K. Farrell、J. T. Oden 和 D. Faghihi。用于原子系统粗粒化模型自适应选择、校准与验证的贝叶斯框架。《计算物理学杂志》，295：189–208，2015年。[15] S. Ferson、W. L. Oberkampf 和 L. Ginzburg。热挑战问题的模型验证与预测能力。验证挑战研讨会，197(29)：2408–2430，2008年。[16] J. Goodman 和 J. Weare。具有仿射不变性的集合采样器。《应用数学与计算科学通讯》，5(1)：65–80，2010年。[17] J. R. Hamilton 和 R. G. Hills。验证实验与应用之间的关系。《数值传热学*B*辑：基础》，57(5)：307–332，2010年。[18] R. G. Hills。面向目标应用的验证结果汇总。技术报告 SAND2013- 7424，桑迪亚国家实验室，美国，2013年。[19] R. G. Hills 和 I. H. Leslie。工程与科学模型的统计验证：从验证实验到应用。技术报告 SAND2003-0706，桑迪亚国家实验室，美国，2003年。[20] M. Hinze、R. Pinnau、M. Ulbrich 和 S. Ulbrich，编。偏微分方程约束优化。数学建模：理论与应用第23卷。施普林格，2009年。[21] M. C. Kennedy 和 A. O' Hagan。计算机模型的贝叶斯校准。《皇家统计学会会刊：*B*辑（统计方法学）》，63(3)：425–464，2001年。[22] G. J. Klir。不确定性与信息：广义信息论基础。Wiley-Interscience，2006年。[23] G. Lee、W. Kim、H. Oh、B. D. Youn 和 N. H. Kim。统计模型校准与验证综述——从不确定性结构的角度出发。《结构与多学科优化》，60(4)：1619–1644，2019年。[24] C. Li 和 S. Mahadevan。校准、验证与相关性在多层次不确定性集成中的作用。《可靠性工程与系统安全》，148：32–43，2016年。[25] Q. Long、M. Scavino、R. Tempone 和 S. Wang。基于拉普拉斯近似，快速估计贝叶斯实验设计的期望信息增益。《计算力学与工程方法》，259：24–39，2013年。[26] A. Montoison、P. Pascal 和 L. Salomon。NOMAD.jl：用于约束黑箱求解器NOMAD的Julia接口，2020年。

[27] J. Mullins、Y. Ling、S. Mahadevan、L. Sun 和 A. Strachan。概率模型验证中偶然不确定性与认知不确定性的分离。《可靠性工程与系统安全》，147：49–59，2016年。[28] J. G. Mullins、B. B. Schroeder、R. G. Hills 和 L. G. Crespo。预测中不确定性与模型形式误差集成方法综述。载于报告编号：*SAND2016-4585C*，2016年。桑迪亚国家实验室，美国。[29] W. L. Oberkampf 和 T. G. Trucano。验证与确认基准。用于核反应堆安全应用的计算流体力学代码基准测试，238(3)：716–743，2008年。[30] T. A. Oliver、G. Terejanu、C. S. Simmons 和 R. D. Moser。未观测量预测的验证。《计算力学与工程方法》，283：1310–1335，2015年。[31] S. Razavi 和 H. V. Gupta。全面、稳健且高效的全局敏感性分析新框架：1. 理论。《水资源研究》，52(1)：423–439，2016年。[32] S. Razavi、A. Jakeman、A. Saltelli、C. Prieur、B. Iooss、E. Borgonovo、E. Plischke、S. Lo Piano、T. Iwanaga、W. Becker、S. Tarantola、J. H. Guillaume、J. Jakeman、H. Gupta、N. Melillo、G. Ra- bitti、V. Chabridon、Q. Duan、X. Sun、S. Smith、R. Sheikholeslami、N. Hosseini、M. Asadzadeh、A. Puy、S. Kucherenko 和 H. R. Maier。敏感性分析的未来：系统建模与政策支持的一门关键学科。《环境建模与软件》，137：104954，2021年。[33] R. Rebba 和 S. Mahadevan。模型可靠性评估的计算方法。《可靠性工程与系统安全》，93(8)：1197–1207，2008年。[34] S. Riedmaier、B. Danquah、B. Schick 和 F. Diermeyer。模型验证、确认与不确定性量化的统一框架及综述。《计算工程方法档案》，28(4)：2655–2688，2021年。[35] C. J. Roy 和 W. L. Oberkampf。科学计算中验证、确认与不确定性量化的综合框架。《计算力学与工程方法》，200(25)：2131–2144，2011年。[36] E. G. Ryan、C. C. Drovandi、J. M. McGree 和 A. N. Pettitt。贝叶斯最优设计现代计算算法综述。《国际统计评论》，84(1)：128–154，2016年。[37] A. Saltelli、M. Ratto、T. Andres、F. Campolongo、J. Cariboni、D. Gatelli、M. Saisana 和 S. Tarantola。全局敏感性分析：入门指南。约翰·威立父子有限公司，2007年。[38] D. S. Sivia 和 J. Skilling。数据分析：贝叶斯教程。牛津科学出版物。牛津大学出版社，第2版，2006年。[39] I. Sobol。非线性数学模型的全局敏感性指数及其蒙特卡罗估计。《数学与计算机模拟》，55(1)：271–280，2001年。[40] T. Sullivan。不确定性量化导论，第63卷，收录于应用数学文本。施普林格国际出版，2015年。[41] I. Sunseri、J. Hart、B. van Bloemen Waanders 和 A. Alexanderian。受偏微分方程约束的反问题的超微分敏感性分析。反问题，36(12)：125001，2020年。[42] J. Tan、B. Liang、P. K. Singh、K. A. Farrell-Maupin 和 D. Faghihi。迈向最优预测性多尺度模型的选择。《计算力学与工程方法》，402，2022年。[43] F. M. White。流体力学。麦格劳-希尔机械工程丛书。麦格劳-希尔，第7版

’ 2009。

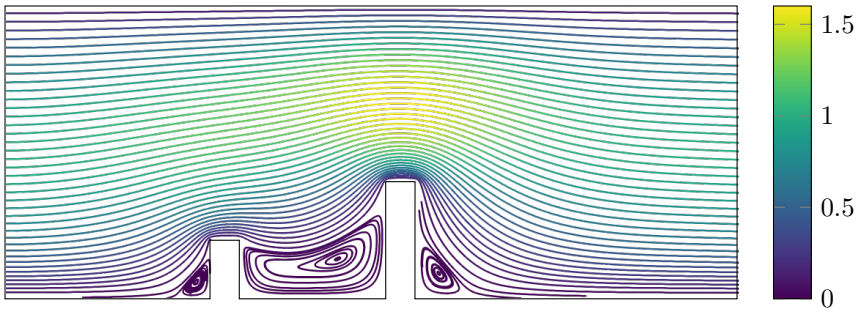


图 A.10：速度场的流线 v

[44] H. Yang、M. Fan、A. Liu 和 L. Dong。基于理论定律的球体阻力系数和沉降速度通用公式。《国际矿业科学与技术期刊》，25(2)：219–223，2015年。

附录 A. 污染物输运示例的速度场

污染物输运问题的速度场 v 通过求解斯托克斯方程并采用以下边界条件获得

$$
\begin{aligned}
-\nu \Delta v + \nabla \beta &= 0, && \text{in } \Omega, && \text{(A.1a)}\\
\nabla \cdot v &= 0, && \text{in } \Omega, && \text{(A.1b)}\\
v &= 0, && \text{on } \Gamma_{\text{banks}} \cup \Gamma_{\text{dock}}, && \text{(A.1c)}\\
v &= v_D, && \text{on } \Gamma_{\text{west}}, && \text{(A.1d)}\\
n \cdot (\nu \nabla v - \beta I) &= 0, && \text{on } \Gamma_{\text{east}}, && \text{(A.1e)}
\end{aligned}
$$

其中，$\nu \approx 0.001$ 约为水的动力黏度，β 为压力。狄利克雷条件 v_D 由充分发展抛物线流构成

$$v_D(z) = v_{\max}(2 - z_2)z_2, \tag{A.2}$$

其中，此处最大速度 $v_{\max}$ 取为 1。我们使用开源有限元代码 FEniCSx [1]求解该问题的弱形式。流动的流线如图 A.10 所示。

我们观察到存在三个回流区：第一个码头前方一个、两个码头之间一个较大的，以及第二个码头后方另一个较小的回流区。对于此应用而言，选择斯托克斯模型来表示速度场只是一个粗略的近似。不过，该模型的唯一目的是获得污染物输运模型所需的速度场。例如，如果速度场由纳维–斯托克斯方程的解给出，那么污染物输运模型 (23) 将不同于本文研究的模型。不过，我们仍会采用相同的方法来获得最优验证场景。

附录B. 污染物输运示例的梯度

计算第 4.2 节所述污染物输运模型中感兴趣量和可观测量（关于参数）的梯度，需要进行一些较为复杂的计算。为此，我们计算扩散-对流方程 (23) 的伴随解，并利用它求得导数 [20]。如前所述，水的速度和压力对应的状态变量（v，β）可以独立于对流扩散方程求解，且不随参数变化。

¯我们明确引入提升函数 ϕ_D ，以处理狄利克雷边界条件 $\phi = \phi_D$，因为我们希望计算感兴趣量和可观测量关于参数的导数，而这些参数¯ ¯定义了 ϕ_D，因此 $\phi = \phi + \phi_D$。对流扩散方程的弱形式为:

$$\begin{aligned}&\text{Find } \bar{\phi} \in V = \{\psi \in H^1(\Omega);\ \psi|_{\Gamma_{\text{west}}} = 0\} \text{ such that}\\ &\underbrace{\int_\Omega \exp(k)\nabla\bar{\phi}\cdot\nabla\psi + \nabla\cdot(v\bar{\phi})\psi\,\mathrm{d}z}_{=a(\bar{\phi},v)} + \underbrace{\int_\Omega \exp(k)\nabla\bar{\phi}_D\cdot\nabla\psi + \nabla\cdot(v\bar{\phi}_D)\psi\,\mathrm{d}z}_{=a(\bar{\phi}_D,\psi):=l(\psi)} = 0, \quad \forall\ \psi \in V.\end{aligned} \tag{B.1}$$

该问题的平衡关系为:

$$r(x,\theta,\phi) = a(\phi,\cdot) = a(\bar{\phi},\cdot) + a(\bar{\phi}_D,\cdot) \in V^*, \tag{B.2}$$

其中， V^* 是 V的对偶空间。根据 [20, 第 1 节，6], 导数的计算分为两步。首先，我们求解伴随问题，求得 λ 使得

$$r_\phi(p,\phi(p))^*\lambda = -h_\phi(p,\phi(p)), \tag{B.3}$$

然后根据下式计算梯度:

$$\nabla_p h(p,\theta) = r_p\big(p,\phi(p)\big)^*\lambda + h_p\big(p,\phi(p)\big). \tag{B.4}$$

下标 ϕ 和 p 表示进行求导时所针对的变量， *表示伴随变量。对于当前问题，平衡关系沿ˆ方向 ϕ 的偏导数为

$$r_\phi(p,\phi(p))\hat{\phi} = a(\hat{\phi},\cdot), \tag{B.5}$$

而其伴随算子为

$$\begin{aligned}\langle r_\phi(p,\phi(p))^*\psi,\phi\rangle_{V^*,V} &= \langle r_\phi(p,\phi(p))\phi,\psi\rangle_{V^*,V}\\ &= a(\phi,\psi)\\ &= \int_\Omega \exp(k)\nabla\phi\cdot\nabla\psi + \nabla\cdot(v\phi)\psi\,\mathrm{d}z.\end{aligned} \tag{B.6}$$

因此，伴随问题为:

$$\begin{aligned}&\text{Find } \lambda \in V \text{ such that}\\ &\int_\Omega \exp(k)\nabla\psi\cdot\nabla\lambda + \nabla\cdot(v\psi)\lambda\,\mathrm{d}z = -\frac{1}{|\omega|}\int_\omega \psi\,\mathrm{d}z\,, \qquad \forall\,\psi \in V,\end{aligned} \tag{B.7}$$

其中 ω 可以是 Ω_1、 Ω_2 或 Ω_{obs}。现在，我们需要计算平衡关系 $r(p,\phi(p))$ 以及感兴趣量/可观测量 $h(p,\phi(p),z)$ 关于参数 p 的偏导数。提醒一下，该问题的控制参数是狄利克雷条件的位置、宽度和高度，即分别为 $x = (z_0, L, c)$，而模型参数是扩散率项 $\theta = k$。此时将使用提升函数。应用链式法则，可得沿方向 ˆx 关于控制参数 x 的偏导数

$$r_x(p,\phi(p))\hat{x} = r_{\bar{\phi}_D}(p,\phi(p))\bar{\phi}_{D,x}(x,z)\hat{x} = a(\bar{\phi}_{D,x}(x,z)\cdot\hat{x},\cdot), \tag{B.8}$$

¯其中 $\phi_{D,x}$ 表示 ϕ_D 相对于 x的梯度。其伴随算子的计算方式为

$$\begin{aligned}\langle r_x(p,\phi(p))^*\psi,\hat{x}\rangle_{\mathcal{X}^*,\mathcal{X}} &= \langle r_x(p,\phi(p))\hat{x},\psi\rangle_{V^*,V} = a(\bar{\phi}_{D,x}(x,z)\cdot\hat{x},\psi)\\ &= \int_\Omega \exp(k)\nabla(\bar{\phi}_{D,x}(x,z)\cdot\hat{x})\cdot\nabla\psi + \nabla\cdot\big(v(\bar{\phi}_{D,x}(x,z)\cdot\hat{x})\big)\psi\,\mathrm{d}z\\ &= \int_\Omega \Big(\exp(k)\nabla\bar{\phi}_{D,x}(x,z)\cdot\nabla\psi + \nabla\cdot(v\bar{\phi}_{D,x}(x,z))\psi\Big)\cdot\hat{x}\,\mathrm{d}z.\end{aligned} \tag{B.9}$$

So 感兴趣量/可观测量 h 相对于控制参数的梯度由 y

$$\nabla_x h(p,\phi(p)) = \int_\Omega \exp(k)\nabla\bar{\phi}_{D,x}(x,z)\cdot\nabla\lambda + \nabla\cdot(v\bar{\phi}_{D,x}(x,z))\lambda\,\mathrm{d}z + \frac{1}{|\omega|}\int_\omega \bar{\phi}_{D,x}(x,z)\,\mathrm{d}z. \quad \text{(B.10)}$$

对模型参数 $\theta = k$ 的导数也可以类似地计算为

$$\nabla_\theta h(p,\phi(p)) = \int_\Omega \exp(k)\nabla\phi\cdot\nabla\lambda\,\mathrm{d}z. \quad \text{(B.11)}$$

伴随解 (B.7)、梯度 (B.10) 和导数 (B.11) 均使用模块 UFL [2]在有限元代码 FEniCSx [1]中计算。

附录 C. 最优设计问题的求解

优化问题 (16) 使用求解器 NOMAD [7] 求解，所用的是 Julia 编程语言中的 NOMAD 代码实现 [26]。主要原因是 NOMAD 是一种无导数求解器，这意味着只需要目标函数的值。问题 (16) 和 (17) 中目标函数导数的计算需要求观测泛函和感兴趣量泛函 h_{obs} 与 h_{qoi} 关于控制参数和模型参数 x 与 θ的海森矩阵。计算这些海森矩阵通常需要耗费过多的计算资源，因此我们倾向于使用无导数优化求解器。NOMAD 可以找到局部极小值，但也可以投入额外资源进行全局搜索。出于下文所述的原因，本文采用了这一选项。

如第 3.3 节所述，问题（16）和问题（17）的目标函数可能不存在唯一的全局极小值。此外，目标函数还可能存在多个局部极小值。因此，值得在受控环境$\mathcal{X}_{\text{lab}}$ 和约束集$\mathcal{Z}_{\text{lab}}$中进行全局搜索，以摆脱局部极小值的吸引域，并期望找到全局极小值。

为了深入了解问题（16）的目标函数的行为，并评估 NOMAD 的性能，我们进行了两项验证测试。这两项验证测试分别针对第 4.1 节和第 4.2 节所述的两个模型求解问题（16），但此时受控环境$\mathcal{X}_{\text{lab}}$ 等于控制空间$\mathcal{X}$。在这种情况下，可保证存在全局极小值，这恰好对应于预测场景。对于抛射体模型，控制参数空间为$\mathcal{X} = [0.005, 5]\times[0.005, 0.1]\times[0, 2]\times[10, 120]$。如前所述，预测场景为 $x_{\text{pred}} =$ (0.05,0.01,1,100)。我们使用拉丁超立方算法生成十个不同的初始点，并以此初始化优化算法。表 C.4 列出了每个初始点所达到的局部极小值。相对于 $\left\|M_{h_{\text{qoi}}}(x_{\text{pred}})\right\|$归一化后的目标函数值表明，所有局部极小值的质量相近。然而，考察误差（即计算得到的最优点与预测场景之间的 l_2 距离 x_{pred}）时，我们发现没有任何初始点能够得到预测场景。通过更仔细地分析最优点，可以解释这一结果。一方面，所有最优点都成功恢复了预测场景的质量 m、直径 ℓ和初始速度 v_0 值 x_{pred}。另一方面，得到的初始位置 u_0 值略有不同。这是因为感兴趣量（18）关于初始位置 u_0是仿射的。因此，无论初始位置 u_0 取何值（其他控制参数固定时），影响矩阵都相同，并由目标函数值体现。我们可以得出结论：该问题不存在唯一的全局极小值。不过，存在一个能够达到全局极小值的一维线性流形。

我们对污染物输运模型重复相同的验证过程。预测场景为 x_{pred} =(0.75,0.6, 2)，控制参数空间为 $\mathcal{X} = [0.1, 1.9]\times[0.005, 1]\times[0.1, 10]$。表 C.5 给出了使用 NOMAD 针对每个初始控制参数得到的解。编号为 3、9 和 10 的验证实验似乎表明，预测场景附近存在局部极小值。然而，它们的归一化目标函数值表明，最优点处与预测场景 x_{pred} 处的感兴趣量影响矩阵相似。

表 C.4：抛射体问题的十次验证实验结果

	初始点				最优点				误差	归一化的目标函数
	m	ℓ	u_0	v_0	m	ℓ	u_0	v_0		
1	2.535	0.02359	0.8801	76.87	0.05	0.009999	1.126	100	0.1263	1.6e-05
2	1.438	0.02451	0.4384	56.95	0.05	0.01	0.5074	100	0.4926	1e-06
3	0.7639	0.05092	0.1344	43.16	0.05	0.01	1.752	100	0.7515	1e-06
4	1.721	0.05431	1.291	112.5	0.05	0.01	0.8704	100	0.1296	1e-06
5	4.257	0.04152	0.2073	97.81	0.05	0.009999	0.41	100	0.59	4e-06
6	3.573	0.08897	0.6992	106.9	0.05	0.01	1.977	100	0.9768	1e-06
7	4.754	0.0655	1.663	33.4	0.05	0.01	1.703	100	0.7028	1e-06
8	2.149	0.01172	1.061	27.74	0.05	0.01	0.06525	100	0.9348	1e-06
9	0.3243	0.09373	1.96	18.72	0.05	0.01	1.23	100	0.23	1e-06
10	3.14	0.07903	1.491	75.87	0.05	0.01	1.606	100	0.6062	1.6e-05

表 C.5：污染物输运问题的十次验证实验结果

	初始点			最优点			误差	归一化的目标函数
	z_0	L	c	z_0	L	c		
1	0.9751	0.9485	5.141	0.7445	0.5915	1.994	0.01178	0.000178
2	1.057	0.8015	1.978	0.7001	0.5231	1.95	0.1047	0.00052
3	1.275	0.7156	6.788	0.9793	0.9626	2.327	0.5392	0.01186
4	0.6539	0.6264	0.9501	0.8002	0.6802	2.059	0.1115	0.002432
5	0.541	0.5098	7.535	0.7852	0.6554	2.041	0.07719	0.00158
6	1.496	0.4369	2.966	0.6955	0.5164	1.946	0.1136	0.00011
7	1.651	0.3441	9.502	0.8258	0.7204	2.092	0.1696	0.003887
8	0.3233	0.2538	8.832	0.7561	0.6098	2.007	0.0133	0.000274
9	1.832	0.1114	3.718	0.9329	0.8938	2.253	0.4287	0.009839
10	0.1281	0.04033	4.255	0.9048	0.8467	2.207	0.3571	0.008336

其他验证实验的归一化目标函数值较低，即使有些最优点与预测点并不十分接近。

这些验证实验表明，求解器 NOMAD 在求解最优设计问题（16）方面具有稳健性。因此，我们确信，先前针对抛射体和污染物输运问题获得的最优验证场景，是目标函数（16）在各自约束集 $\mathcal{X}_{\mathrm{lab}}$ 下的合理局部极小值。