迈向癌症芯片实验模型的不确定性量化

S. Bertoluzza[1], V. Bianchi[2], G. Bretti[3], L. Tamellini[1], 和 P. Zanotti[4]

[1]意大利国家研究委员会 – 应用数学与信息技术研究所，Via Ferrata 5, 27100, 帕维亚，意大利[2]帕维亚大学，Via Ferrata 5, 27100, 帕维亚，意大利[3]意大利国家研究委员会 – 计算应用研究所，Via dei Taurini 19, 00185, 罗马，意大利`米兰大学 – 数学系，Via Saldini 50, 20133, 米兰，意大利

arXiv:2602.06018v1 [cs.CE] 5 Feb 2026

摘要

本研究是利用数据驱动微分模型预测并控制癌症芯片实验动力学的初步探索。我们考虑一个概念化的一维装置，其中包含癌症细胞和一群白细胞。癌细胞与这群细胞之间的相互作用通过受Keller–Segel型方程启发的趋化模型进行描述，并采用混合间断Galerkin方法求解。我们的目标是利用（合成）数据调整控制方程的参数，并评估参数调整后仍存在的残余不确定性对动力学预测造成的不确定性。为此，我们采用参数化微分模型不确定性量化领域的技术。首先，我们利用 Sobol 和 Morris 指数进行全局敏感性分析，以评估参数不确定性对模型预测的影响，并固定影响可忽略的参数数值。随后，我们通过贝叶斯技术开展逆向不确定性量化分析，计算剩余模型参数的数据驱动概率分布。最后，我们进行正向不确定性量化分析，计算更新后的（残余）参数不确定性对模型关注量的影响。整个流程借助基于稀疏网格的代理模型加速，以近似不确定参数到关注量的映射。

关键词 芯片癌症、趋化模型（Keller–Segel 型方程）、可混合不连续伽辽金法、不确定性量化、稀疏网格、贝叶斯反演

AMS学科分类 41A10,62F15,65M60, 92-10.

1 引言

传统的二维细胞培养难以准确重现人体中复杂的细胞间相互作用，因此通常更倾向于使用动物模型。然而，由于动物与人体生理学存在显著差异，人们日益认识到动物实验无法完美代表人体系统，而且始终伴有伦理方面的顾虑。近年来，器官芯片技术（即通过工程设计重现整个生物体及其周围环境行为、且复杂程度可模块化调节的微流控装置） [1] 因此越来越多地被用作实验工具，以便在体外系统中重现体内系统。

在这项工作中，我们特别关注在芯片癌症技术（CoC）框架下，使用患者来源的人体细胞开展的实验室实验 [2], 一种免疫肿瘤学微流控芯片，旨在深入了解免疫功能行为的基本机制。基于生物学假设，并借助实验室实验（例如视频片段）所产生的数据，可以开发出能够复现实验的数学模型。此类模型经数据验证和校准后，有助于揭示复杂的、

细胞动力学的非线性和随机特性，并可助力开发生物数字孪生体。数字孪生体可被视为支持生物学实验的可持续、非侵入式计算工具，其经济成本低且易于复现。它们有助于更深入地理解复杂系统中的扰动如何促成疾病，并支持针对新型有效疗法的药物测试实验研究。然而，由于生物过程本身具有复杂性和异质性，许多方面仍未得到充分理解，因此可以通过不同方法构建多种数学模型：微观 [3], 宏观 [4], 动力学模型 [5] 及其流体动力学／扩散极限[6],离散—连续／混合模型 [7], 元胞自动机 [8]。

本文聚焦于用于 CoC 实验室实验的宏观数学模型。据我们所知， [4] 这是首个用于 CoC 的宏观模型，灵感源自 Keller-Segel 模型 [9], , 并基于包含趋化作用的反应扩散方程。该模型描述了细胞死亡过程、趋化剂的影响、交互作用，以及共存的两类细胞群体（即癌细胞和免疫细胞）之间的竞争，并以细胞密度表示这些细胞群体。由于相关生理过程错综复杂，用于描述 CoC实验的方程不可避免地十分复杂，这不仅给数值近似带来挑战，也使校准过程困难重重，而校准是获得最终生物数字孪生所必需的。

关于第一个问题，即模型解的数值近似，采用既准确又高效的数值算法至关重要：在本文中，我们采用了一种专用的混合可混合间断 Galerkin（HDG）求解器 [10]。

关于第二个问题，即模型校准，我们考虑采用贝叶斯方法 [11, 12]的不确定性量化（UQ）框架。我们不为感兴趣的参数计算单一数值，而是计算一个概率密度函数，为与现有数据更为一致的参数数值赋予更高的概率。该概率密度函数可视为校准后的残余不确定性，并可通过前向不确定性量化技术传播至模型的最终预测。在开展模型校准之前，我们先进行初步的全局敏感性分析 [13] ，以确定哪些参数确实会影响系统响应的变异性；校准过程中将根据数据对照这一响应，因此这些参数才适合校准。若尝试校准对系统影响微乎其微的参数，显然不会成功。我们通过计算 Sobol 指数和 Morris指数（并辅以图形化的单因素分析）开展敏感性分析，从而使分析结果更为可靠。因此，我们的方法可概括为以下三个步骤的工作流

$$\text{sensitivity analysis} \longrightarrow \text{Bayesian inversion} \longrightarrow \text{forward UQ.} \tag{1}$$

每个步骤都需要针对不确定参数的不同组合多次调用HDG求解器。我们通过采用模型关注量的HDG近似的合适代理模型，加快工作流。事实上，经过离线训练后，代理模型仍能准确近似关注量，而且评估所需的计算量要小得多。在文献中已有的多种方案中 [14], , 我们采用稀疏网格构建代理模型；这种方法已广泛应用于不确定性量化 [15]。

工作流（1）此前的应用不胜枚举。不过，据我们所知，这是整个工作流首次应用于CoC模型。与本文最相关的研究包括：

- [16, 17, 8] 进行局部敏感性分析，即在固定的名义设定附近，逐一研究参数值发生微小扰动所产生的影响；
- [18] 对Keller–Segel方程进行全局敏感性分析；
- [19, 20, 21] 采用基于深度学习的模型发现和高斯过程回归，学习一维Keller–Segel方程的函数形式及其相关系数；
- [22, 23] 将逆向-正向不确定性量化问题重新表述为一个联合的序贯状态-参数估计问题。

贡献与展望 本研究是一项概念验证，即朝着利用数据驱动的微分模型预测和控制 CoC实验动力学迈出的第一步。我们测试了

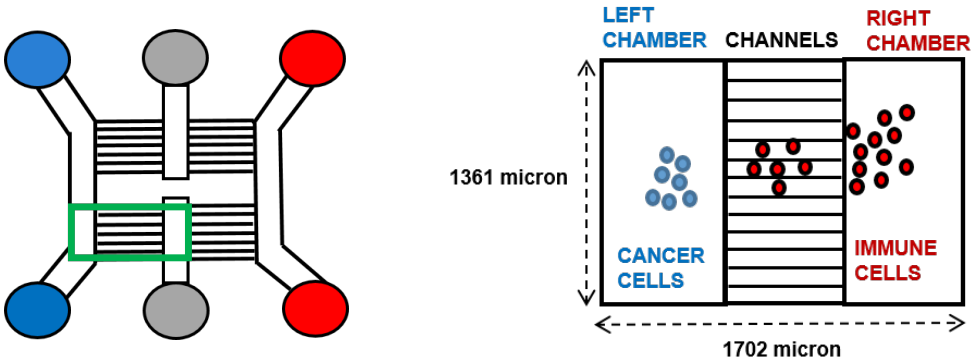


图1：CoC 器件的二维平面示意图（左），蓝色和红色孔中分别预先装有癌细胞和免疫细胞。绿色区域（左）及其放大图（右）为实验室实验期间的监测区域。

工作流 (1) 在文献[4]提出的简化一维模型中的应用。对于贝叶斯反演，我们使用合成数据：通过针对给定的参数组合运行混合间断伽辽金求解器获得数据，然后在其中加入高斯噪声。本文的结果为未来的工作奠定了基础，包括采用更真实的微分模型和二维几何结构，从而能够使用来自 CoC实验的真实数据。

组织结构 本文结构如下：第2节简要概述所关注的 CoC实验、相应的微分模型、混合间断伽辽金求解器以及稀疏网格替代模型。第3节介绍不确定性量化问题的建模，其结果将在第4节讨论。最后，第5节总结本文并展望未来研究方向。

2 微分模型与数值近似

本节介绍一种用于 CoC实验的一维模型，该模型是 [4, 16] 中所提出模型的简化版本。随后，我们将讨论用于近似模型解和/或其关注量的数值技术。

2.1 CoC实验

CoC器件的发展前景尤为广阔，因为它们能够模拟复杂的癌症微环境，而免疫细胞是其中的关键组成部分，在癌症生长和维持中发挥着重要作用。这种免疫功能行为备受关注，因为免疫细胞既可能抑制癌症进展，也可能促进癌症进展 [24]。在这一复杂情境中，免疫细胞和癌细胞的迁移能力对炎症过程至关重要。

[4, 16], 中引入并在下文第2节 2.2中进一步简化的模型，受到 [2] 的启发，因为后者对于研究癌细胞与免疫系统细胞之间的交互作用具有典型意义。在实验室实验中，癌细胞经化疗药物处理后逐渐死亡，且处于准静止状态，并分泌一种化学信号，该信号可作为野生型免疫细胞的趋化剂。具体而言，免疫细胞群由外周血单个核细胞组成，其中包括不同种类的细胞：单核细胞、树突状细胞、T淋巴细胞和B淋巴细胞。

CoC 实验所用的微流控器件示意图如图1所示：左侧展示了整个芯片几何结构，右侧则展示了实验室实验视频影像中监测的区域。

更具体地说，CoC 器件由两个封闭端中央腔室和两个细胞培养腔室组成（见图1左图），使细胞和化学信号能够在芯片所有相互连通的区域内循环。细胞腔室通过四组微米级通道与中央腔室相连。癌细胞被引入左侧细胞培养腔室的上、下两个储液室，而经荧光示踪剂标记的免疫细胞则被放入右侧腔室的两个储液室中。封闭端中央通道内仅含有

培养基，这些通道被设计为缓冲腔室，用于调节实验条件，并防止游离免疫细胞在初始加载时直接流入癌细胞腔室。

微流控器件的监测区域由两个主要培养腔室组成（见图1右图）：左侧腔室容纳贴壁癌细胞，右侧腔室容纳游离免疫细胞，两者通过微通道连接，从而实现化学和物理接触。免疫细胞在癌细胞产生的化学信号驱动下开始迁移。穿过微通道后，它们到达癌细胞并与之相互作用。

2.2 微分模型

我们考虑 [4, 16] 中针对受Keller-Segel方程 [9]启发的癌症芯片（CoC）器件动力学模型的一个简化版本。我们的模型由两个弱耦合的扩散-平流-反应偏微分方程组成。该模型描述了癌细胞分泌的趋化剂以及受趋化剂吸引的免疫细胞在时空中的演化。由于癌细胞接受化疗药物处理，它们不会增殖，趋化物产生也会随时间减少。

我们通过考虑一维空间区域来简化 CoC 器件的复杂几何结构。在这项初步研究中，采用降维处理是经过审慎考虑的选择，旨在建立不确定性量化框架的同时，将计算复杂度降至最低。尽管进行了这种简化，该模型仍能够捕捉实验所涉及的关键病理生理学方面，并初步揭示某些参数对动力学的影响。在第2.1节所述的 CoC实验背景下，我们将 CoC 监测区域的一维水平截面作为计算域，参见图1右图。我们假设该截面的长度为 $L = 1000\mu$m，从而使计算域能够重现横轴上两个腔室的长度（即左侧腔室与微通道的总长度）；这两个腔室容纳着两种不同的细胞群。

记 $[0, L]$ 为空间区域，我们假设癌细胞集中在左侧的点 $x = 250$附近，并为 $t > 0$产生趋化剂。免疫细胞则位于右侧的 $x = 750$附近。这一配置模拟了上述实验的初始状态。我们考虑时间区间 $[0, T]$，其中 $T = 1800s$。该区间远短于标准 CoC实验中的时间区间，但足以观察简化模型中相关的动力学。

用 u 和 φ 分别表示免疫细胞和趋化剂的浓度，单位为摩尔/单位长度。在上述简化和假设下，[4, 16] 中的模型可约化为以下初边值问题。

$$
\begin{aligned}
\partial_t u - \partial_x\big(\nu\partial_x u - \chi u\partial_x\varphi\big) &= f_u \quad \text{in } (0,L)\times(0,T)\\
\partial_t \varphi - \partial_x\big(\mu\partial_x\varphi\big) + a\varphi &= f_\varphi \quad \text{in } (0,L)\times(0,T)\\
u(\cdot,0) &= u_0 \qquad \text{in } (0,L)\\
\varphi(\cdot,0) &= \varphi_0 \qquad \text{in } (0,L)\\
\partial_x u(0,\cdot) = \partial_x u(L,\cdot) &= 0 \qquad \text{in } (0,T)\\
\partial_x \varphi(0,\cdot) = \partial_x \varphi(L,\cdot) &= 0 \qquad \text{in } (0,T)
\end{aligned}
\tag{2}
$$

与原始 Keller-Segel 模型不同，u和 φ的偏微分方程仅以单向方式耦合。耦合通过趋化（对流）项 $\partial_x(\chi u\partial_x\varphi)$实现，该项描述免疫细胞向趋化剂集中度较高的区域迁移。常数 χ 表示趋化敏感性系数。二阶项$\partial_x(\nu\partial_x u)$和 $\partial_x(\mu\partial_x\varphi)$描述遵循菲克定律的扩散，而 $a\varphi$ 表示趋化剂的自然降解。我们将两个强迫项设定如下：

$$
f_u(x,t) = 0 \qquad \text{and} \qquad f_\varphi(x,t) = \frac{k_\varphi \exp(-\rho t)}{\sqrt{2\pi}\sigma_\varphi} \exp\left(-\frac{(x - c_\varphi)^2}{2\sigma_\varphi^2}\right). \tag{3}
$$

第一个为零，因为不存在免疫细胞源。第二个表示癌细胞产生趋化剂。空间变量上的高斯依赖用于模拟癌细胞的位置（见图2），并假设癌细胞保持静止。随时间呈指数衰减

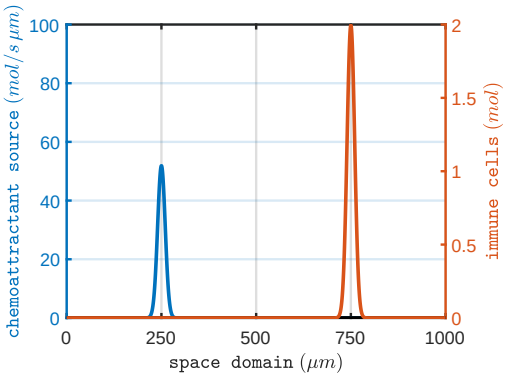


图2：参考参数组合下，一维空间域中趋化因子源和免疫细胞分布的初始分布示意图。

参数	描述		范围	数值	Unit	Ref
μ	趋化因子扩散率		$[700-1100]$	-	μm^2/秒	[25]
ν	免疫细胞扩散率		$[100-300]$	-	μm^2/秒	[25]
a	φ的消耗速率		-	10^{-4}	$1/s$	[26]
χ	趋化敏感性系数		-	5×10^{-2}	$\mu m^3/(s\,mol)$	(DC)
k_{u_0}	u_0的放大系数		-	50	mol/s	(DC)
c_{u_0}	免疫细胞初始条件中心		-	750	μm	(DC)
σ_{u_0}	免疫细胞初始条件的扩散范围		-	10	μm	(DC)
k_φ	f的放大系数	φ	$[600-2000]$	-	mol/s	(DC)
ρ	药物导致的死亡癌细胞衰减率		$[10-40]\times10^{-4}$	-	1/s	(DC)
c_φ	化学诱引物源项 f 的中心	φ	$[150-350]$	-	μm	(DC)
σ_φ	趋化因子源项的扩散 f	φ	$[5-15]$	-	μm	(DC)

表1：方程 (2)-(4) 中所涉及参数的汇总。

表示 表示药物施用导致癌细胞减弱。初始数据定义为
as

$$u_0(x)=\frac{k_{u_0}}{\sqrt{2\pi}\sigma_{u_0}}\exp\left(-\frac{(x-c_{u_0})^2}{2\sigma_{u_0}^2}\right)\qquad\text{and}\qquad\varphi_0(x)=0. \tag{4}$$

第一个模型用高斯分布描述免疫细胞的初始分布，见图2。第二个表示初始时刻不存在趋化剂。最后，齐次诺伊曼边界条件规定系统是封闭的。

表1汇总了方程 (2)-(4) 中涉及的参数，并说明了它们各自的作用和物理单位。对于已有相关数据的参数，其具体参数值或变化范围取自文献。对于其他与本节提出的模型特定属性相关的参数，其数值或变化范围则通过设计选择（DC）确定，即依据我们以往对模型动力学的经验确定。

在接下来的不确定性量化分析中，我们主要关注两个感兴趣量（QoI），即方程 (2) 的解 $(u,\ \varphi)$ 的两个泛函。首先，$M\colon\ [0,T]\to\mathbb{R}$ 是免疫细胞随时间变化的（空间）质心，其定义为

$$M(t):=\frac{1}{U}\int_0^L xu(x,t)\,dx,\quad\text{where}\quad U:=\int_0^L u(x,t)\,dx=\int_0^L u_0(x)\,dx. \tag{5}$$

其中第二个等式成立是由于式（3）和边界条件。这就是我们的一个感兴趣量范例，它可以从实验数据中推断得到，例如通过对延时视频记录进行跟踪分析。第二个感兴趣量是趋化剂在空间上的总量 $I\colon\ [0,T]\to\mathbb{R}$ 随时间变化，所得结果为

$$I(t):=\int_0^L\varphi(x,t)\,dx. \tag{6}$$

这体现了一种虽有必要了解、却很难通过直接实验测量推断得出的量。实际上，人们可能只关心某个子区域内 chamoattraftant 的含量

。我们研究整个空间区域，以便为 I 建立显式公式，并用它来检验混合间断伽辽金求解器和代理模型的精度，参见下文的式（27）。

2.3 混合间断伽辽金求解器

我们采用 [10] 中开发的 HDG 求解器的 MATLAB 实现，对初边值问题 (2) 的解进行近似。关于 HDG 求解器的一般介绍，请参阅 [27]。以下特性说明了在此问题中使用 HDG 求解器的原因。首先，近似解在单元层面满足与 (2) 中前两个方程所依据的守恒律相对应的关系。具体而言，免疫细胞的总量 U 随时间保持不变，见 (5)，其近似解也保证随时间保持不变。其次，针对众所周知颇具挑战性的对流主导区间（即第一条方程 (2) 中的动力学主要由输运项 $\partial_x(\chi u \partial_x \varphi)$ 驱动时），稳定化技术已相当成熟。最后，HDG 范式便于施加不同子区域之间的界面条件。对于第 2.2 节的一维模型，这一特性并不重要；但对于未来扩展至真实 CoC 器件的几何结构而言，它至关重要，因为其中必须耦合一维（微通道）和二维（腔室）组件，参见图1。在 HDG 离散框架中，可采用 [28] 提出的方法轻松实现此类耦合。

混合间断伽辽金求解器源于对弱形式进行离散化，该弱形式通过在（2）中引入免疫细胞通量 j 和趋化因子密度通量 ψ 获得。注意，方程（2）可改写为关于通量的形式：

$$\begin{aligned} \partial_t u + \partial_x j = 0 \quad &\text{and} \quad \frac{1}{\nu} j + \partial_x u + \frac{\chi}{\nu} u \partial_x \varphi = 0 \quad \text{in } (0,L)\times(0,T), \\ \partial_t \varphi + \partial_x \psi + a\varphi = f_\varphi \quad &\text{and} \quad \frac{1}{\mu}\psi + \partial_x \varphi = 0 \quad \text{in } (0,L)\times(0,T). \end{aligned} \tag{7}$$

类似地，边界条件变为

$$j(0,\cdot) = j(L,\cdot) = \psi(0,\cdot) = \psi(L,\cdot) = 0 \quad \text{in } (0,T). \tag{8}$$

对于空间离散化，我们使用空间区域的一个划分 $0 = x_0 < x_1 \cdots < x_{N_\mathsf{s}} = L$，该划分由 $N_\mathsf{s}+1$ 个节点组成。我们假设该划分是均匀的，即

$$x_k - x_{k-1} = \frac{1}{N_\mathsf{s}} =: h$$

对于 $k = 1$，……，$N_\mathsf{s}+1$。我们为每个节点 x_k、$k = 0$、……、$N_\mathsf{s}+1$ 关联近似值

$$\widehat{u}_h^k,\ \widehat{\varphi}_h^k : [0,T] \to \mathbb{R} \tag{9}$$

分别对应点值 $u(x_k)$ 和 $\varphi(x_k)$。此类未知量起拉格朗日乘子的作用。我们为每个区间 $I_k := (x_{k-1},\ x_k)$、$k = 1$、……、$N_\mathsf{s}+1$ 关联近似值

$$u_h^k,\ \varphi_h^k,\ \psi_h^k : [0,T] \to \mathbb{P}_1(I_k) \qquad \text{and} \qquad j_h^k : [0,T] \to \mathbb{P}_0(I_k) \tag{10}$$

浓度 u、φ 以及通量 j、ψ。这里，$\mathbb{P}_\ell(I_k)$、$\ell \in \mathbb{N}_0$表示 I_k上次数不超过 ℓ 的多项式空间。我们还在节点处引入了 j 和 ψ 的数值通量

$$\begin{aligned} \widehat{j}_+^k &:= -j_h^{k+1}(x_k) + \frac{\tau_u}{h}(u_h^{k+1}(x_k) - \widehat{u}_h^k), \quad \text{and} \quad \widehat{j}_-^k := j_h^k(x_k) + \frac{\tau_u}{h}(u_h^k(x_k) - \widehat{u}_h^k), \\ \widehat{\psi}_+^k &:= -\psi_h^{k+1}(x_k) + \tau_\varphi(\varphi_h^{k+1}(x_k) - \widehat{\varphi}_h^k), \quad \text{and} \quad \widehat{\psi}_-^k := \psi_h^k(x_k) + \tau_\varphi(\varphi_h^k(x_k) - \widehat{\varphi}_h^k), \end{aligned} \tag{11}$$

其中 $\tau_u, \tau_\varphi > 0$ 为惩罚参数。对于 $k = 0$（相应地 $k = L$），我们利用诺伊曼边界条件，并通过设定来覆盖上述表达式

$$j_-^0 = \psi_-^0 = 0, \qquad (\text{resp. } j_+^{N_\mathsf{s}} = \psi_+^{N_\mathsf{s}} = 0). \tag{12}$$

N \ w	1	2	3	4	5
4	9	41	137	401	1105
6	13	85	389	1457	4865

表 2：根据 w 和 N 构建稀疏网格所需的插值点数量。

(9)-(10) 中的未知量通过如下对 (7) 的半离散得到

$$\begin{aligned}
&\forall w \in \mathbb{P}_1(I_k) && \int_{I_k} \left(\partial_t u_h^k\, w - j_h^k\, \partial_x w\right) + \widehat{j}_-^k\, w(x_k) + \widehat{j}_+^{k-1}\, w(x_{k-1}) = 0,\\
&\forall w \in \mathbb{P}_1(I_k) && \int_{I_k} \left(\partial_t \varphi_h^k\, w - \psi_h^k\, \partial_x w + a\varphi_h^k\, w\right) + \widehat{\psi}_-^k\, w(x_k) + \widehat{\psi}_+^{k-1}\, w(x_{k-1}) = \int_{I_k} f_\varphi w,\\
&\forall q \in \mathbb{P}_0(I_k) && \int_{I_k} \left(\frac{1}{\nu} j_h^k\, q + \frac{\chi}{\nu}\frac{1}{\mu}\psi_h^k\, u_h^k\, q\right) + \widehat{u}_h^k\, q(x_k) - \widehat{u}_h^{k-1}\, q(x_{k-1}) = 0,\\
&\forall w \in \mathbb{P}_1(I_k) && \int_{I_k} \left(\frac{1}{\mu}\psi_h^k\, w - \varphi_h^k\, \partial_x w\right) + \widehat{\varphi}_h^k w(x_k) - \widehat{\varphi}_h^{k-1} w(x_{k-1}) = 0,
\end{aligned} \tag{13}$$

对于 $k = 1, \ldots, N_{\mathrm{s}}$，以及（8）的以下半离散形式

$$\widehat{j}_-^k + \widehat{j}_+^k = 0, \qquad \widehat{\psi}_-^k + \widehat{\psi}_+^k = 0, \tag{14}$$

对于 $k = 0, \ldots, N_{\mathrm{s}}$。注意，根据（12），对于 $k = 0$ 和 $k = N_{\mathrm{s}}$，上述表达式编码了齐次诺伊曼边界条件。

我们在（13）—（14）中采用后向欧拉格式，使用 N_{t} 个均匀时间步进行全离散。初始时刻的近似值通过对初始值（4）进行 L^2-正交投影和点值计算获得。对于后续时间步的近似值，我们对（13）—（14）经静态凝聚后得到的非线性代数方程组，采用牛顿迭代求解，并结合 MATLAB 默认的（反斜杠）线性求解器。

2.4 稀疏网格代理模型

不确定性量化工作流程 (1) 基于对（5）和（6）中的感兴趣量 M 和 I 的计算，这些计算针对表1所列不确定参数的不同组合进行；详见第3节。由于此类参数组合数量相当庞大，因此，与其针对每种组合运行混合间断伽辽金求解器，不如使用代理模型近似感兴趣量，这样更为方便。事实上，代理模型能够以运行混合间断伽辽金求解器所需成本的一小部分，准确近似感兴趣量；不过，这种加速需要预先运行求解器，为精心选取的一组参数组合构建代理模型（离线训练阶段）。更准确地说，我们针对每个时间步计算的感兴趣量构建代理模型，即 $M(t_n)$ 和 $I(t_n)$，其中 $n = 0, \ldots\ldots, N_{\mathrm{t}}$。

文献中提供了许多构建代理模型的方法，例如综述专著 [14]。我们采用成熟的稀疏网格代理建模技术 [15], ，并使用 Sparse Grids Matlab Kit [29] 提供的实现工具。该技术通过对混合间断伽辽金求解器的输出进行插值来构建代理模型；所用参数组合（即参数域中的插值点，见第3节）按照特定模式选取，而该模式取决于不确定参数的概率密度函数（pdf），例如见图3。请注意，即使概率密度函数为均匀分布，稀疏网格节点也并非等距分布，以避免插值中的龙格现象。我们将在后续章节进一步阐述参数概率密度函数的选择。

参数域中用于插值的点数（因而也决定了代理模型的精度）由一个离散超参数 $w \in \mathbb{N}$ 控制，该参数称为稀疏网格层级，参见 [29, 式（15）–（20）]。表 2 列出了在后续讨论中所考虑的 w 值以及参数数量 N 所需的插值点数。请注意，针对特定插值点进行的每次混合间断伽辽金求解器计算，都完全独立于其他计算。因此，代理模型的构建可以并行化。

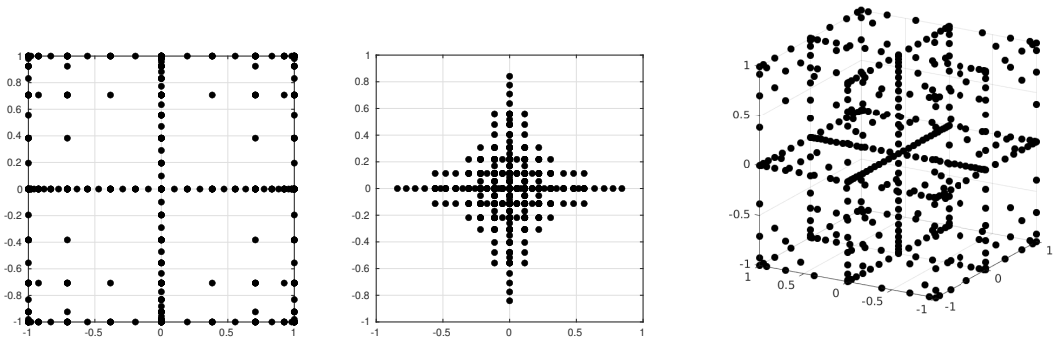


图3： $[-1,1]^N$ 上 $w = 5$的稀疏网格图。左： $N = 2$，参数服从相互独立的均匀概率密度。中： $N = 2$，参数服从相互独立的正态概率密度函数，均值为零，标准差为 $\sigma = 0.08$。右： $N = 3$，参数服从相互独立的均匀概率密度。

3 不确定性量化工作流程

本节将详细讨论不确定性量化工作流程 (1)，并特别强调每个步骤的目标以及其中将使用的工具。

3.1 总体框架

回顾一下，方程(2)-(4)中涉及的参数汇总于表1。对于其中一些参数，由于缺乏相关知识，考虑其变化范围而非特定数值是必要的做法；对于其他参数，这则是出于设计目的等考虑而作出的选择。我们将所有此类不确定参数收集为一个维度为 $N = 6$的向量

$$\mathbf{y} = \big(\mu,\, \nu,\, k_\varphi,\, \rho,\, c_\varphi, \sigma_\varphi\big) \in \Gamma =: \prod_{n=1}^{N} \Gamma_n \subset \mathbb{R}^N$$

其中 $\Gamma_n := [a_n, b_n]$、 $n = 1$， ....., N，是表1中列出的范围。

关键的是，我们可以将(5)-(6)中的感兴趣量视为不仅依赖于时间、也依赖于$\mathbf{y}$的函数，即我们有如下映射

$$\Gamma \times [0,T] \ni (\mathbf{y}, t) \mapsto M(\mathbf{y}, t) \qquad \text{and} \qquad \Gamma \times [0,T] \ni (\mathbf{y}, t) \mapsto I(\mathbf{y}, t).$$

接下来，我们的目标是在利用免疫细胞质心 M 的相关数据（来自 (5)）调优参数$\mathbf{y}$ 后，应用成熟的不确定性量化框架 (1)，量化由$\mathbf{y}$的不确定性所导致的 (6) 中趋化剂总量 I 的变异性。这个过程展示了如何借助与另一感兴趣量相关的数据，量化实验中难以测量的感兴趣量的变异性。

我们首先假设$\mathbf{y}$ 是一个概率密度函数已知的随机向量，出于稍后会更加明了的原因，我们将其称为先验概率密度函数，并记作 ρ_{prior}。我们还假设$\mathbf{y}$ 的 N个分量统计独立。因此，我们有

$$\rho_{\text{prior}}(\mathbf{y}) = \prod_{n=1}^{N} \rho_{\text{prior},n}(y_n) \qquad \text{with} \qquad \rho_{\text{prior},n}(y_n) = \frac{1}{b_n - a_n} \tag{15}$$

对于$\mathbf{y} = (y_1, \ldots, y_N) \in \Gamma$。先验边缘概率密度函数 $\rho_{\text{prior},n}$, $n = 1, \ldots, N$的表达式，是基于以下假设得出的：$\mathbf{y}$ 的每个分量在相应范围内均匀分布。换句话说，我们假设对于不确定参数可能取值，不存在任何先验偏好。

由于第2.2节中的模型高度概念化，我们使用合成数据而非实验数据来调整参数。为生成合成数据，我们考虑 $K \geq 1$ 个不同时刻 $0 < t_1 < \ldots < t_K \leq T$，并选择一个特定的可容许参数组合$\mathbf{y}_{\text{true}} \in \Gamma$。我们使用混合间断伽辽金求解器计算$\mathbf{y} = \mathbf{y}_{\text{true}}$，在指定时刻评估感兴趣量 M ，并

用随机噪声扰动所得结果。为模拟测量误差，我们假设每个时刻关联的噪声都是服从高斯分布、均值为零且标准差为 $\sigma_\varepsilon > 0$的随机变量。此外，不同时刻关联的噪声在统计上相互独立。记 $M_{\mathbf{HDG}}$ 为 M 的HDG近似，则合成数据可表示为

$$M_k^* := M_{\mathrm{HDG}}(\mathbf{y}_{\mathrm{true}}, t_k) + \varepsilon_k \qquad \text{with} \qquad \varepsilon_k \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad k = 1, \ldots, K. \tag{16}$$

注释 1 在更贴近现实、包含实验数据的情形中，可以考虑类似的设置，但需要使用更复杂的误差模型 ε_k。至少，这些误差应具有非零均值，以考虑建模误差和离散误差；此外，假设这些误差在整个 Γ中并非常量也是合理的，因此除了测量误差之外，可能还需要增大标准差 σ_ε ，以适应这种变化。

做好这些准备后，我们便可以说明不确定性量化工作流程（1）中各个步骤的作用和重要性。这里先进行一般性讨论，并将各步骤所用技术工具的详细介绍留到接下来的三个小节。

- 敏感性分析。我们首先验证 M 是否会对每个不确定参数作出响应。如果 M 对某些参数不敏感，就将这些参数固定为参考值，因为通过将模拟与数据进行比较来校准这些参数（即降低其方差）几乎是不可能的。随后，我们检查 I 是否对任何 M 不敏感的参数敏感。实际上，这会降低我们依据现有数据减少 I 不确定性的能力（在最坏的情况下，如果 I 仅取决于 M 不敏感的参数，那么接下来的校准步骤基本上就毫无用处）。

- 贝叶斯反演。我们为（剩余的）不确定参数定义并采样一个更新后的数据知情概率密度函数，以下简称后验概率密度函数（相对于前文介绍的先验概率密度函数，后者“不了解数据”），记为 ρ_{post}。后验概率密度函数旨在为参数组合$\mathbf{y}$ 赋予更高的概率，使得

  $$\big\{M(\mathbf{y}, t_1), \ldots, M(\mathbf{y}, t_K)\big\} \text{ is “closer” to } \big\{M_1^*, \ldots, M_K^*\big\}.$$

  这一步比经典的参数校准更为复杂，因为除了通过最小化过程确定最“可能”的数值外，我们还提供了概率密度函数的完整样本。

- 前向不确定性量化。对于 $t \in (0, T]$，我们在上一步生成的后验概率密度函数样本上评估 $I(\cdot, t)$。我们利用这些评估结果来估计 $\pi_I(\cdot, t)$ 的概率密度函数，即 I 的概率密度函数。换言之，我们将不确定性从后验 ρ_{post} 传播至 $\pi(\cdot, t)$。

3.2 敏感性分析

我们首先通过观察响应曲线，对感兴趣量的敏感性进行定性评估，即在其他参数固定为相应范围的中点时，仅改变一个参数，绘制两个感兴趣量相对于参数的曲线，见表1（单因素敏感性分析）。随后，我们采用两种不同且成熟的定量方法，即计算 Sobol 指数和 Morris指数 [13]。由于这两种方法的理念截然不同，若所得结果具有一致性，则可为判定某个参数是“相关”还是“不相关”提供明确证据，并在后一种情况下将其固定为常数。下面我们简要回顾这两个指数针对一般感兴趣量的定义，并将其记为$\mathbf{y} \mapsto$QoI(y)。此外，我们还将对这两种方法进行比较。

Sobol 指数。 Sobol 指数通常被称为全局敏感性指数，因为它们旨在量化在整个参数空间 Γ中，感兴趣量（QoI）(y) 对$\mathbf{y}$ 的敏感性。感兴趣量$\mathbb{V}$[QoI] 的方差被作为感兴趣量总体变异性的参考值，并被分解为 V_J 相对于任意参数子集之间交互作用的贡献，这些子集由指标集 $J \subseteq \{1, \ldots, N\}$标识。归一化后，可得恒等式

$$\frac{1}{\mathbb{V}[\mathrm{QoI}]} \sum_{J \subseteq \{1, \ldots, N\}} V_J = 1.$$

与第 n 个参数 $n = 1, \ldots, N$相关的主 Sobol 指数 S_n，用于衡量由单元素集合 $J = \{n\}$产生的方差占比，即

$$S_n := \frac{V_{\{n\}}}{\mathbb{V}[\mathrm{QoI}]}.$$

与第 n 个参数相关的总 Sobol 指数 S_n^T，衡量的是该参数与其他任意参数子集相互作用所产生的方差占比

$$S_n^T := \frac{1}{\mathbb{V}[\mathrm{QoI}]} \sum_{n \in J \subseteq \{1,\ldots,N\}} V_J \geq S_n.$$

Sparse Grids Matlab Kit 提供了一个专用例程，可通过对稀疏网格代理模型进行后处理，计算主效应和总 Sobol 指数 [29, 第 4.5] 节。

Morris 指数。这种方法粗略而言，通过对参数空间中的感兴趣量 y_n 关于某个参数的偏导数 ∂_{y_n}（用有限差分近似）取平均，来衡量感兴趣量对该参数的敏感性，因此可视为一种全局-局部敏感性度量。更确切地说，设 $\{\mathbf{y}_p\}_{p=1}^P \subseteq \Gamma$ 为从所有可容许参数组合构成的集合中抽取的 $P \geq 1$ 个随机样本。对于 h_1，h_2，……，$h_N > 0$，以及所有 $p = 1$，……，P，考虑以下有限序列

$$\mathbf{y}_{p,0} := \mathbf{y}_p \;\mapsto\; \mathbf{y}_{p,1} := \mathbf{y}_{p,0} + h_1 \mathbf{e}_1 \;\mapsto\; \ldots \;\mapsto\; \mathbf{y}_{p,N} := \mathbf{y}_{p,N-1} + h_N \mathbf{e}_N$$

其中 $\mathbf{e}_1$，……，$\mathbf{e}_N$ 是 $\mathbb{R}^N$ 的标准单位基向量。假设序列中的每个点仍是一个可容许参数组合，即仍处于 Γ 中，则初等因子为

$$EE_{n,p} := \frac{\mathrm{QoI}(\mathbf{y}_{p,n}) - \mathrm{QoI}(\mathbf{y}_{p,n-1})}{h_n}, \qquad n = 1, \ldots, N. \tag{17}$$

然后，我们通过对相应的初等因子取平均，定义与第 n 个参数相关联的 Morris 指数

$$m_n := \frac{1}{P} \sum_{p=1}^{P} \left| EE_{n,p} \right|. \tag{18}$$

请注意，绝对值可以避免抵消效应。与 Sobol 指数类似，我们通过对稀疏网格代理模型进行后处理来计算 Morris 指数。

比较。计算出 Sobol 指数和 Morris 指数后，我们检查二者是否一致，以增强敏感性分析结果的可靠性。为此，值得注意的是，Sobol 指数是无量纲的归一化量，而 Morris 指数是有量纲数值。为进行公平比较，我们考虑以下归一化 Morris 指数

$$\overline{m}_n := \frac{m_n}{\sum_{i=1}^N m_i}.$$

3.3 贝叶斯反演

我们采用贝叶斯技术 [11, 12] 推导后验概率密度函数的表达式，并对其进行采样。根据条件概率的贝叶斯定理

$$\rho_{\mathrm{post}}(\mathbf{y}) = \frac{1}{C} \mathcal{L}(M_1^*, \ldots, M_K^*; \mathbf{y}) \rho_{\mathrm{prior}}(\mathbf{y}), \qquad \text{for } \mathbf{y} \in \Gamma \tag{19}$$

其中，C 是归一化常数，$\mathcal{L}$ 表示似然函数。后者定义为

$$\mathcal{L}(M_1^*, \ldots, M_K^*; \mathbf{y}) := \prod_{k=1}^{K} \rho_{\varepsilon_k}(M_k^* - M(\mathbf{y}, t_k)) \tag{20}$$

其中，ρ_{ε_k}是式（16）中的噪声 ε_k的概率密度函数——在我们的例子中，对于 $k = 1,\ldots, K$，它是高斯分布。粗略地说，在$\mathbf{y}$ 处计算似然，可以衡量数据（$M_1^*, \ldots, M_K^*$）由$\mathbf{y}_{\mathrm{true}} =\mathbf{y}$生成的可能性有多大。值得注意的是：

- 在$\mathbf{y}$ 处计算似然，需要运行混合间断伽辽金求解器（或者更方便地，计算稀疏网格代理模型），以获得 $M(\mathbf{y}\ t_k)$的数值；
- 能够计算乘积 $\mathcal{L}(M_1^*,\ldots,M_K^*;\mathbf{y})\rho_{\text{先验}}(\mathbf{y})$，也就是说，能够计算相差一个乘法常数 C 的后验分布，便足以继续我们的讨论。

具体而言，只需计算后验分布至相差 C，便足以生成服从 ρ_{post} 的样本，这将在后续的前向不确定性量化分析（不确定性量化工作流的第三步）中用到。本文同时考虑较为简单的高斯近似和更为先进的马尔可夫链蒙特卡罗算法。对于这两种技术，引入最大后验估计（MAP）$_{\text{MAP}} \in \Gamma\mathbf{y}$会很有帮助。后者定义为最有可能生成数据的参数组合（$M_1^*,\ldots,M_K^*$）：

$$\mathbf{y}_{\text{MAP}} := \arg\max_{\mathbf{y}\in\Gamma} \rho_{\text{post}}(\mathbf{y}) = \arg\min_{\mathbf{y}\in\Gamma} \left(-\log \rho_{\text{post}}(\mathbf{y})\right).$$

第二个等价定义源于对数函数的单调性；为保证数值近似的稳定性，文献中通常采用这一形式。我们将该公式与（15）、（19）和（20）结合起来。经过简单推导可得

$$\mathbf{y}_{\text{MAP}} = \arg\min_{\mathbf{y}\in\Gamma} \sum_{k=1}^{K} \left(M_k^* - M(\mathbf{y}, t_k)\right)^2, \tag{21}$$

这揭示了贝叶斯反演与参数经典最小二乘校准之间的联系。[1]

一旦可以获得$\mathbf{y}_{\text{MAP}}$，便可通过其样本估计值推导出测量噪声的标准差 σ_ε：

$$\widetilde{\sigma}_\varepsilon^2 := \frac{1}{K} \sum_{k=1}^{K} \left(M_k^* - M(\mathbf{y}_{\text{MAP}}, t_k)\right)^2. \tag{22}$$

请注意，这一估计值会自动同时纳入模型误差（如果数据实际来自真实实验）以及对 M 进行近似所产生的误差。因此，如果数据足够多且已确定一个清晰的极小值，即使像我们的情况这样明确知道 σ_ε^2，也建议使用 $\widetilde{\sigma}_\varepsilon$ 来评估 ρ_{post}，并用于下文的式（23）。完成这些准备后，我们便可以讨论上述两种对后验概率密度函数进行采样的技术。

高斯近似（GA）。该技术基于这样的假设：ρ_{post} 接近高斯分布。例如，在以下情况下，我们可以预期这一性质成立：

$$K \gg 1 \qquad \text{and} \qquad \sigma_\varepsilon \ll 1$$

式（16）中，这意味着有多个噪声较小的数据可用，参见 [11]。我们用均值为$\mathbf{y}_{\text{MAP}}$、协方差矩阵为.....的多元高斯分布来近似后验概率密度函数

$$\Sigma_{\text{post}} := \widetilde{\sigma}_\varepsilon^2 \left(J_M^T J_M\right)^{-1} \tag{23}$$

其中，$\widetilde{\sigma}_\varepsilon^2$ 来自公式（22），J_M 表示 $K \times N$ 雅可比矩阵

$$[J_M]_{k,n} := \partial_n M(\mathbf{y}_{\text{MAP}}, t_k). \tag{24}$$

此处 ∂_n 表示 M 关于 $n-th$ 参数的导数，我们通过对稀疏网格代理模型应用简单的有限差分来近似这些导数。式（23）中的 Σ_{post} 可推导为式（21）中最小化的最小二乘泛函之海森矩阵逆矩阵的近似，因此提供了一种衡量最大后验估计“置信度”的方法（如果待最小化泛函的极小值处是一个“狭窄的谷”，即我们“非常确信”

[1]通过假设不确定参数服从高斯先验，也可以在贝叶斯反演框架中恢复用于校准的蒂霍诺夫正则化项。

最大后验估计，那么海森矩阵将具有非常大的特征值，相应地，利用该公式计算出的后验将具有非常小的方差）。

计算出协方差矩阵后，即可通过采用多元高斯分布的标准样本生成方法，轻松生成后续前向不确定性量化分析所需的、服从 ρ_{post} 分布的样本。

马尔可夫链蒙特卡罗 (MCMC)。这一类算法不假设 ρ_{post} 服从某种特定分布。MCMC 算法是迭代算法，会生成一个链，其中包含 B 样本$\mathcal{C}_B = \{\mathbf{y}_i\}_{i=1}^B$，这些样本渐近地服从 ρ_{post} ，因此可直接用于后续的前向不确定性量化分析。我们将$\mathbf{y}_{\text{MAP}}$ 作为链的起始点；后续样本则根据所用的具体 MCMC 算法生成，生成过程基于通过公式 (19) 对 log(ρ_{post}) 的计算，忽略乘法常数 C。

在文献中已有的各种 MCMC 算法中，我们采用切片采样马尔可夫链蒙特卡罗算法 [30], 。简而言之，该算法通过迭代计算 log(ρ_{post}) 上水平集的近似逆像，并从这些逆像中生成样本。尽管计算近似逆像较为复杂，我们仍倾向于采用该算法，而非更常用的 Metropolis– Hastings 算法 [12] ，因为它无需设计所谓的提议分布。提议分布可视为一个关键的“超参数”，该参数的选择决定了此方法的计算效率。

无论采用何种具体的马尔可夫链蒙特卡罗算法，我们都要指出，链 $\mathcal{C}_B$ 的样本自相关可能随滞后增加而缓慢衰减，从而降低链的质量（即 $\mathcal{C}_B$ 中的样本分布是否能良好近似ρ_{post}）。每次保留一个样本后，丢弃 $t-1$ 个样本（稀疏抽样）可以缓解这一问题。此外，通常会在稀疏抽样 阶段之前丢弃链中的前 b个样本（预热期），以减轻上述问题：链仅渐近地 服从 ρ_{post}，从而使最终的链仅包含（$B-b$）个/t 样本。

集中度量。 每个参数的边缘后验概率密度函数预计比先验概率密度函数更集中。概率密度函数的集中度可通过均值为 μ 、标准差为 σ 的变异系数（CV）来量化，其定义为

$$\text{CV} := \frac{\sigma}{|\mu|}. \tag{25}$$

我们预计后验概率密度函数的变异系数小于先验概率密度函数的变异系数。此外，当先验均值与后验均值接近时，变异系数的差异主要取决于相应标准差 σ_{prior} 和 σ_{post}的比值。因此，不确定性降低可通过集中因子（CF）来衡量，其定义为

$$\text{CF} := \frac{\sigma_{\text{post}}}{\sigma_{\text{prior}}}. \tag{26}$$

3.4 前向不确定性量化

有了 ρ_{post} 的样本，我们终于可以对 I 进行正向不确定性量化分析，即将 ρ_{post} 中编码的残余不确定性从参数传播到感兴趣量。为此，我们首先对 ρ_{post} 和 $t \in [0,T]$ 中的$\mathbf{y}$ t进行评估。再次说明，使用代理模型而非混合间断伽辽金求解器来近似 I 更为方便。然后，我们计算 I 的统计量，如均值和标准差，并通过经典的核密度估计器近似 I 的概率密度函数。我们将其与传播参数的先验不确定性（15）所得结果进行比较，以量化不确定性的降低。为此，我们计算 $t \in [0,T]$ 的集中度量（25）和（26）。

4 数值结果

本节首先验证第2.3节介绍的混合间断伽辽金求解器，以及针对 M 和 I构建合适稀疏网格代理模型的方法。这一初步步骤有助于在后续计算中平衡精度

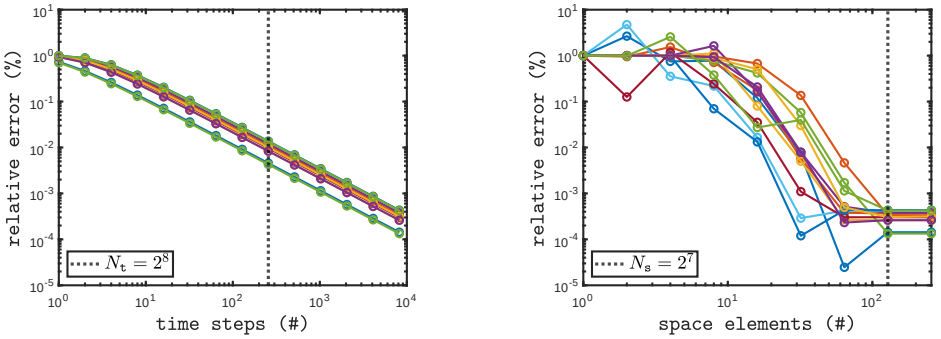


图4：12 个随机样本的 err(y N_t, N_s) 曲线图$\mathbf{y} \in \Gamma$（每个样本对应一条线）。左：误差随N_t 的变化， $N_s = 2^8$ 固定。右：误差随 N_s 的变化， $N_t = 2^{13}$固定。竖直虚线表示后续分析中在（28）中固定的 N_t 和 N_s 的数值。

与成本。随后，我们按照第3节的指导，执行不确定性量化工作流程（1）的三个步骤，并展示结果。

所有测试均使用 Matlab 实现，并在一台配备英特尔 Core i7-1355U 处理器和 16 GB 随机存取存储器的标准笔记本电脑上运行。混合间断伽辽金求解器为内部开发，基于 [10]。在所有实验中，我们在时间离散中使用 N_t 个等长步长，在空间离散中使用 N_s 个等长单元。稀疏网格代理模型通过 Sparse Grids Matlab Kit 23-5 [29] 版构建和评估。

4.1 混合间断伽辽金求解器验证

为了评估混合间断伽辽金求解器的精度，我们注意到，闭式公式

$$I(\mathbf{y}, t) = \frac{\kappa_\varphi(e^{-at} - e^{-\rho t})}{\rho - a} \left(\int_0^L \exp\left(-\frac{(x - c_\varphi)^2}{2\sigma_\varphi^2} \right) dx \right)^{-1} \tag{27}$$

可通过对（2）的第二个方程进行空间积分，利用 φ 的边界条件，并求解所得常微分方程推导得出。我们将（11）中的惩罚参数设为 $\tau_u = \tau_\varphi = 1$。然后，考虑相对误差

$$\mathtt{err}(\mathbf{y}, N_t, N_s) := \max_{n=1,\dots N_t} \frac{|I(\mathbf{y}, t_n) - I_{\mathrm{HDG}}(\mathbf{y}, t_n)|}{I(\mathbf{y}, t_n)}$$

其中， I_{HDG} 是求解器返回的 I 近似值，而$(t_n)_{n=0}^{N_t}$ 是时间离散中的点。为考察误差相对于$\mathbf{y}$ 的变化情况，我们在集合 Γ中生成12个随机样本。对于每个样本，我们分别考察误差随时间离散和空间离散的衰减情况。对于前一种情况，我们固定 $N_s = 2^8$，并求解 $N_t = 2^0, 2_1, \dots, 2_{13}$。如预期的那样，所有样本均呈现一阶收敛，这与后向欧拉格式的性质一致，参见图4（左图）。对于另一种情况，我们固定 $N_t = 2^{13}$，并求解 $N_s = 2^0, 2_1, \dots 2_8$。起初，由于空间分辨率不足以应对数据的变化，我们观察到误差停滞。随后出现超线性收敛（样本间的变化比前一种实验更大），之后误差再次停滞，这是由于时间分辨率不足所致，参见图4（右图）。

鉴于上述结果，通过固定以下设置，可以在后续分析中获得精度与成本之间的合理平衡

$$N_t = 2^8 \qquad \text{and} \qquad N_s = 2^7. \tag{28}$$

我们通过在 Γ 中生成 60 个随机样本来验证这一选择。相对于所有样本，err 的最大值和均方根分别为

$$\mathtt{max} = 1.40\% \qquad \text{and} \qquad \mathtt{rms} = 0.98\%.$$

不进行并行化时，对所有样本运行求解器大约需要 $130s$。这些结果以及后续结果均采用惩罚参数 $\tau_u = 1 = \tau_\varphi$ 计算数值通量（11）。在牛顿求解器中，我们将容差设为 10^{-5}，最多执行 10 次迭代。与上述误差相比，这些设置下牛顿求解器的误差似乎可以忽略不计。

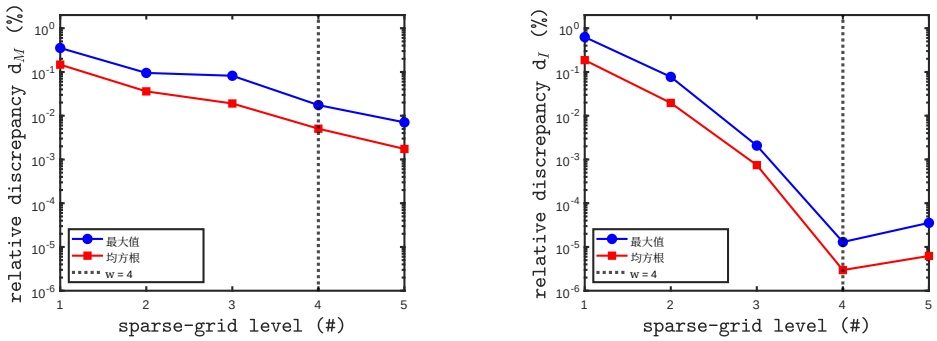


图 5：感兴趣量 M （左）和 I （右）的相对偏差的最大值和均方根随稀疏网格层级的变化 w。垂直点线表示后续分析中在（29）中固定的 w 数值。

4.2 替代模型验证

我们采用 HDG 求解器，并使用式（28）中的 N_{t} 和 N_{s}，为感兴趣量M 和 I构建稀疏网格代理模型。更具体地说，我们针对 HDG 求解器的每个时间步，为每个感兴趣量构建一个代理模型。为此，我们使用 Sparse Grids Matlab Kit，并采用以下标准设置：

- 每个不确定参数的克伦肖–柯蒂斯配置节点，参见 [29, 第 3]节；
- “层级到节点”函数，类型为“加倍”，参见 [29, 式（5）]；
- 指标集，类型为“总次数”，参见 [29, 式（15）、（17）]。

这种超参数组合产生了所谓的 Smolyak 稀疏网格，适用于具有均匀先验分布的不确定参数，正如式（15）所规定的那样。

如第 2.4 节所述，稀疏网格中的节点数（因此代理模型的精度）由稀疏网格层级 w 控制，见表 2。为合理选择 w，我们在 Γ 中生成 60 个随机样本。对于每个样本，我们计算由混合间断伽辽金求解器计算得到的关注量 M_{HDG} 和 I_{HDG} 与其由层级为 w 的稀疏网格代理模型得到的近似值 $M_{\mathrm{S}(w)}$ 和 $I_{\mathrm{S}(w)}$ 之间的相对偏差

$$\mathtt{d}_M(\mathbf{y}, w) := \max_{k=1,\dots,N_{\mathrm{t}}} \frac{|M_{\mathrm{HDG}}(\mathbf{y}, t_k) - M_{\mathrm{S}(w)}(\mathbf{y}, t_k)|}{|M_{\mathrm{HDG}}(\mathbf{y}, t_k)|}$$

and

$$\mathtt{d}_I(\mathbf{y}, w) := \max_{k=1,\dots,N_{\mathrm{t}}} \frac{|I_{\mathrm{HDG}}(\mathbf{y}, t_k) - I_{\mathrm{S}(w)}(\mathbf{y}, t_k)|}{|I_{\mathrm{HDG}}(\mathbf{y}, t_k)|}.$$

两者的最大值 以及相对于所有样本的 $\mathtt{d}_I$ 均方根，都迅速衰减至零 ，见图 5（侧面板）。对于 $\mathtt{d}_M$ （左侧面板），收敛趋势不太明显，但仍然清晰。

由于 t 根据上述结果，为后续开发兼顾精度与成本的方案提供了便利的平衡 - ments is obtai 的方法是固定

$$w = 4. \tag{29}$$

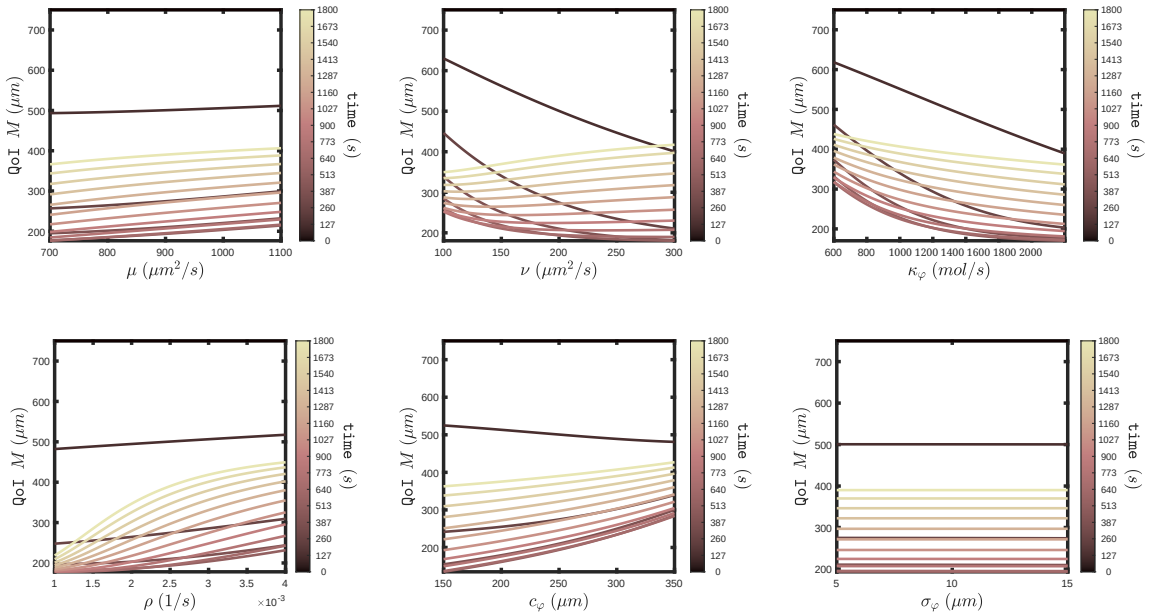
在此情况下，相对于样本的最大值和均方根分别为

$$\mathtt{max} = 1.75\% \qquad \text{and} \qquad \mathtt{rms} = 0.51\%$$

对于 $\mathtt{d}_M$ 和

$$\mathtt{max} < 0,01\% \qquad \text{and} \qquad \mathtt{rms} < 0.01\%$$

对于 $\mathtt{d}_I$。评估代理模型的计算成本并非一目了然，因为专用例程在一次调用中处理多个样本时比逐个样本调用更高效，原因是无论所需评估次数多少，都必须执行一定数量的运算。例如，就上述 60 个样本而言，一次性评估全部 60 个样本约需 $0.15s$，而逐个样本进行 60 次评估（不进行并行化）约需 $0.60s$ 。



Fi图6： M 响应曲线随时间的演变（相对于各参数）。

4.3 不确定性量化工作流步骤1：敏感性分析

我们使用第4.2节中的代理模型，按照第3.2节所述，对 M 和 I进行敏感性分析。我们首先绘制感兴趣量在不同时刻的若干响应曲线。如第3.2节已指出的，这些曲线是通过让一个参数在其范围内变化、而其余参数固定在各自范围的中点得到的，见表1（单因素敏感性分析）。根据图6，参数 μ 和σ_φ 对 M的作用可忽略不计，而其余参数似乎具有显著影响。图 7表明， I对 $\mu, \nu, c_\varphi, \sigma_\varphi$不敏感，而对其他参数则明显有响应。

我们通过计算 Sobol 和 Morris 指数，进一步验证了从响应曲线中得到的定性直觉。对于后者，我们使用 $P = 10000$ 个样本点，增量为$h_n = |\Gamma_n|/P$, $n = 1$, ..., N，其中 $|\Gamma_n|$表示第 n个参数的范围大小，参见式（17）和（18）。图8-9展示了与各参数相关的指数随时间的演变。初步观察发现，总体而言，Sobol 和 Morris 指数非常一致。这使结果更易于解读，也更加可靠。特别是，结果证实

- M 对 μ 和σ_φ（几乎）不敏感，并且
- I 对 μ、 ν、 c_φ 和σ_φ不敏感。

所有其他参数都会影响感兴趣量，其中 ρ 在较晚时刻的作用更为显著，而其他参数的重要性则有所降低。

4.4 UQ工作流步骤2：逆UQ分析

根据上述敏感性分析，参数 μ 和σ_φ 与两个感兴趣量均无关。因此，我们将它们固定为一个标称值，即其变化范围的中点，参见表1。值得一提的是，与 I 相关的参数也与 M相关。因此，可以预期贝叶斯反演能够恢复这些参数，从而大幅降低对 I进行预测时的不确定性，见下文第3.4节。请注意，我们在校准中还纳入了ν 和c_φ，尽管它们与 I无关。校准这些参数并不能降低 I的不确定性，但对于最大后验估计的精度是必要的，因为 ν 和c_φ 都与 M相关，参见第3.3节。

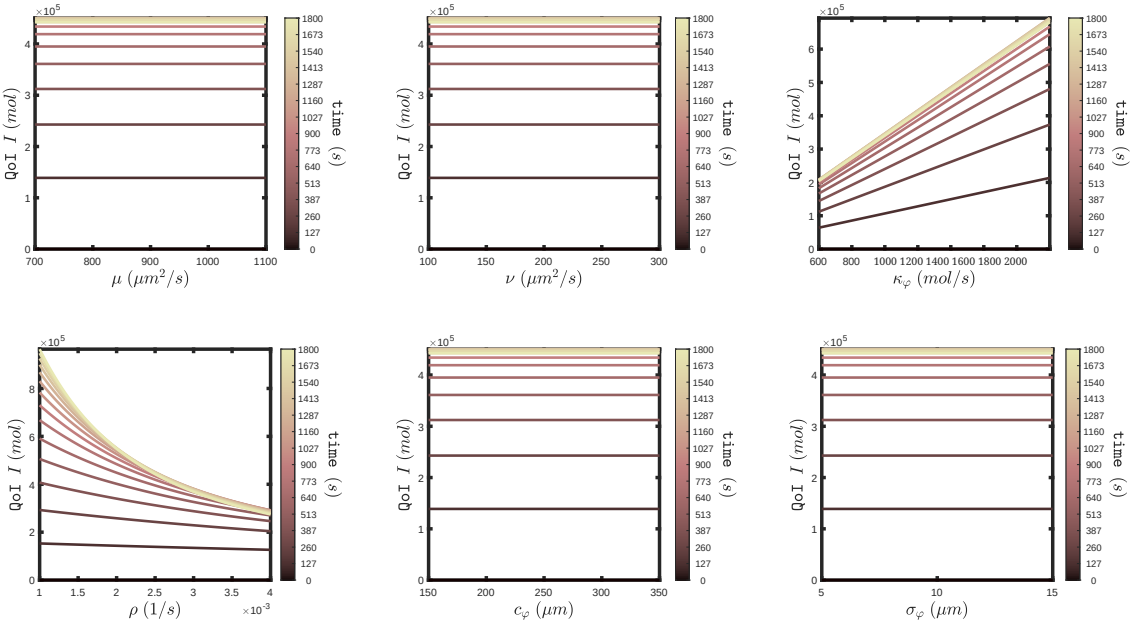


图 7： I 响应曲线随参数的时间演变

rs.

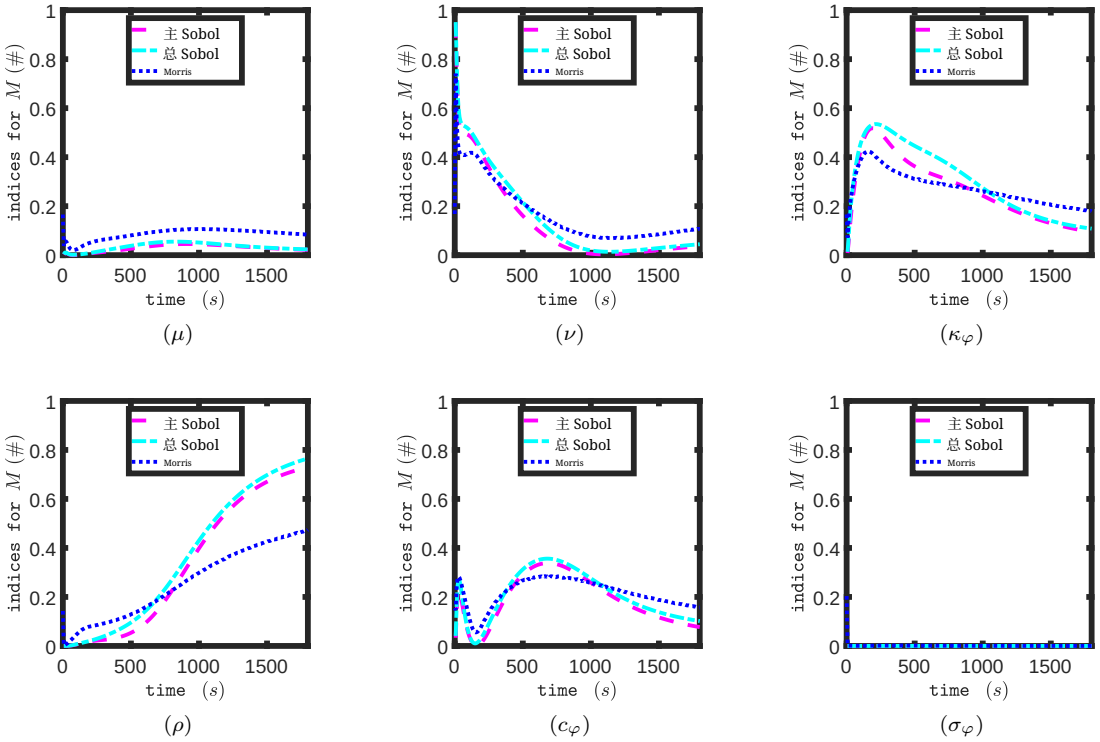


图8： M敏感性指数随时间的演变。

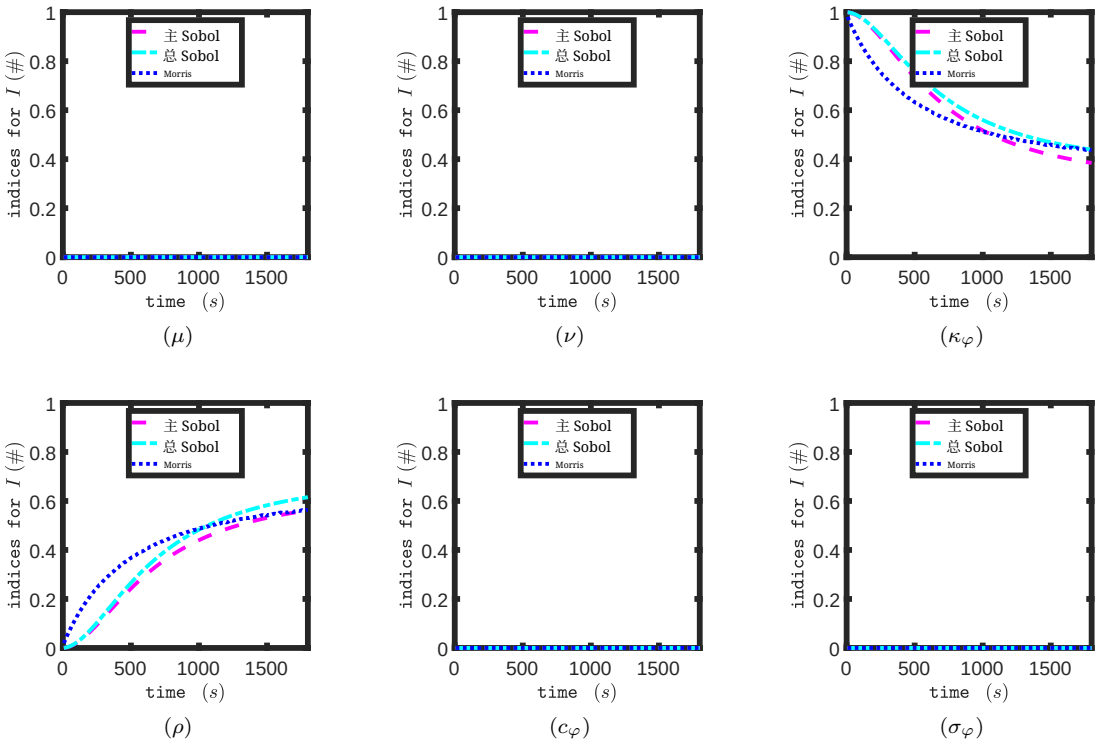


图9： M的敏感性指数随时间的变化。

为方便起见，我们调整第2节中引入的记号，以适应新的设置。现在将不确定参数汇集为一个维度为 $N = 4$的向量

$$\mathbf{y} = (\nu,\, k_\varphi,\, \rho,\, c_\varphi) \in \Gamma =: \prod_{n=1}^{N} \Gamma_n \subset \mathbb{R}^N$$

其中， $\Gamma_n := [a_n, b_n]$, $n = 1$, ..., N为表1中列出的范围。

减少不确定参数的数量后，我们按照与第4.2节相同的标准设置，为感兴趣量 M 和 I构建新的稀疏网格替代模型。我们重复精度测试，得到相似且略小的相对差异数值（此处不予展示）。因此，我们使用与之前相同的稀疏网格层级 $w = 4$ 。

我们通过公式（16）为感兴趣量 M 生成合成数据。对于参考参数组合$\mathbf{y}_{\text{true}}$，我们选择 Γ的重心，即

$$\mathbf{y}_{\text{true}} = (200, 1300, 25 \times 10^{-4}, 250).$$

我们将高斯噪声的标准差设为 $\sigma_\varepsilon = 2$。针对 $K = 30$个等间距时刻 $0 < t_1 < \cdots < t_K < T$生成数据。下文将报告并讨论在此设置下针对单个噪声实现获得的结果。对于不同的 σ_ε和 K ，和/或不同的噪声实现，我们观察到了一致的结果。在选定时刻，用于后续分析的数据的相对误差如图10所示。

我们通过最小化问题 (21) 求得最大后验估计 $\mathbf{y}_{\text{MAP}}$ ，并使用代理模型计算 M。我们采用 Matlab 内置的无导数例程 fminsearch 进行无约束多变量优化。由于可能存在局部极小值，我们从 Γ 中均匀分布的 100 个随机初始点分别运行该例程，最后选取所有运行结果中使目标值最小的优化结果。结果列于下方表3的第二列。

有了最大后验估计后，我们首先用均值为 $\mathbf{y}_{\text{MAP}}$ 、协方差矩阵由式（23）给出的多元高斯分布来近似后验概率密度函数 ρ_{post} 。不确定参数的（归一化）边缘概率分布见图11。

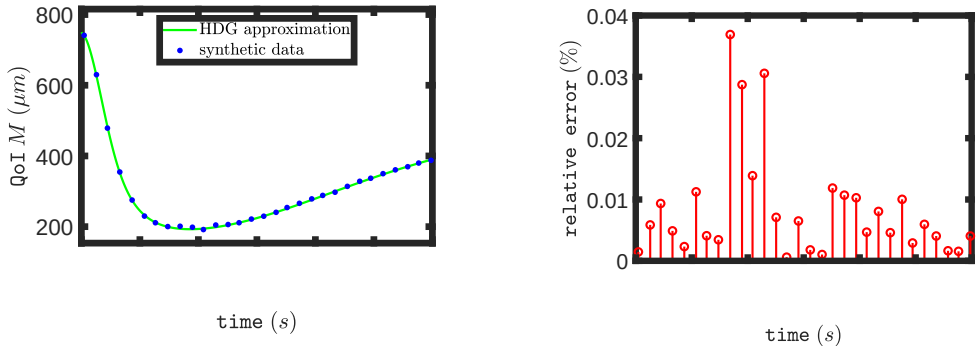


图10：用于逆UQ分析的合成数据。左： M 的HDG近似和合成数据。右：合成数据的相对误差。

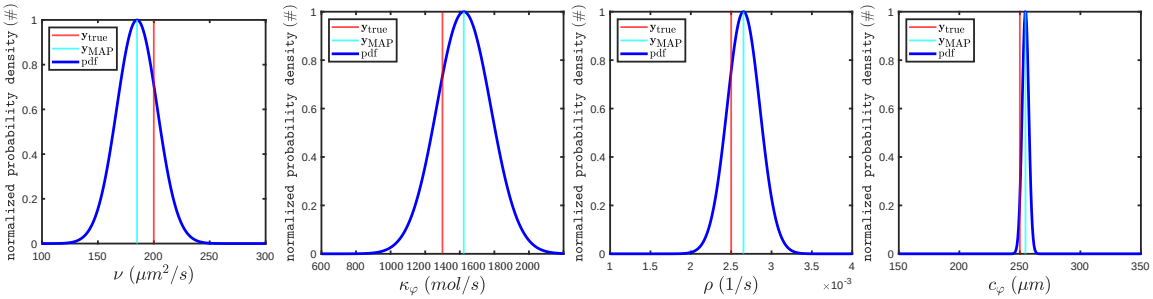


图11：不确定参数边缘后验分布的最大后验估计和归一化高斯近似，与参考参数值的对比。横轴表示参数的先验范围，因此该图定性地展示了 ρ_{post}的集中度。

接下来，我们通过切片采样马尔可夫链蒙特卡罗算法对后验 ρ_{post} 进行抽样，以验证高斯近似的精度。为此，我们首先使用替代模型计算 M，并通过式（22）估计噪声的标准差。然后，我们运行 Matlab 内置例程 slicesample，生成一条包含 6000 个样本、服从 ρ_{post} 的链。事实证明，为提高样本质量，即提高概率密度函数估计的精度，将预热期和抽稀超参数设置得足够高至关重要。根据我们的调参经验，我们将可选参数设置为

$$\texttt{burnin} = 2000 \qquad \text{and} \qquad \texttt{thin} = 100$$

在 slicesample 例程中。这意味着算法会生成近乎 6×10^5 个样本，才能返回所需的链。因此，再次强调，通过替代模型而非混合间断伽辽金求解器来近似 M 至关重要。fff

我们借助两类诊断图评估切片抽样生成的链的质量。图12中，我们绘制了链在各个坐标方向上的投影，即每个参数取值的变化。图13中，我们绘制了各投影链的自相关。两幅图均证实，该链能够准确地对后验概率密度函数进行采样。事实上，自相关相对较快地衰减至零，且样本在其变异范围内分布良好，没有明显的趋势。

验证切片抽样算法生成的链的质量后，我们将边缘后验分布与图11中的结果进行比较，见图14。两者吻合良好，仅标准差存在一些细微差异，参见表3。因此，可以确认高斯近似在这种情况下是合理的。为了更直观地展示后验概率密度函数，我们还给出了 MCMC 样本在参数空间所有坐标二维平面上的投影，见图12。根据 MCMC 样本计算参数的相关矩阵

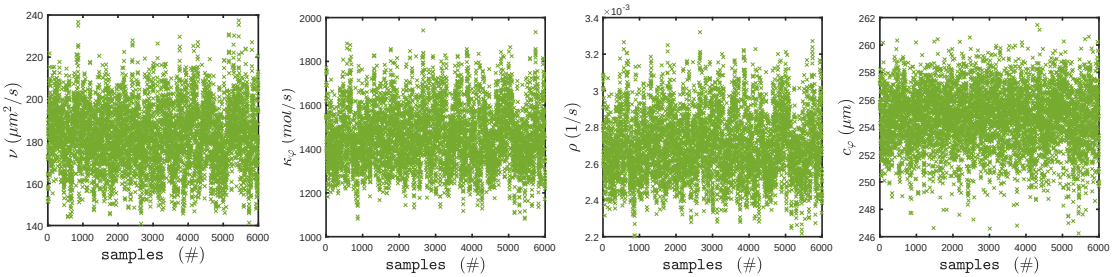


图12：切片采样MCMC算法在参数空间中各坐标轴上生成的链的投影。

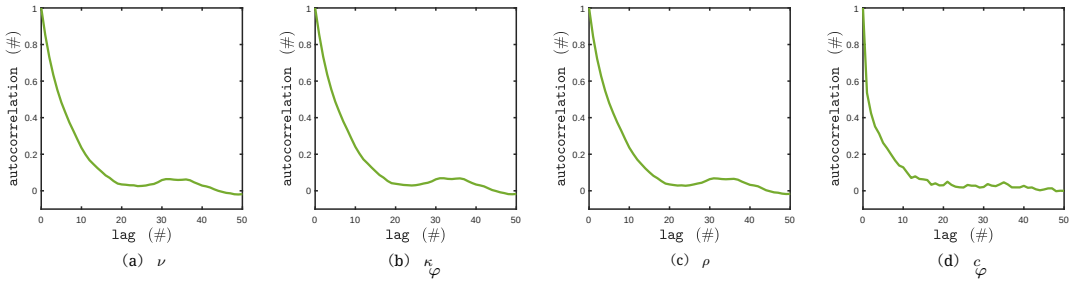


图13：各不确定参数的投影链的自相关。

得到下方的矩阵，这表明ν、κ_{φ}和 ρ之间几乎存在线性依赖：

$$R = \begin{pmatrix} 1.0000 & -0.9924 & -0.9896 & -0.6904 \\ -0.9924 & 1.0000 & 0.9953 & 0.6819 \\ -0.9896 & 0.9953 & 1.0000 & 0.6228 \\ -0.6904 & 0.6819 & 0.6228 & 1.0000 \end{pmatrix}$$

图14和图15显示，参数分布的不确定性显著降低。为更精确地描述这一现象，我们在表3和表4中报告了一些基本统计量和集中度指标。后者尤其涉及变异系数（25）和集中因子（26），用于量化贝叶斯反演步骤后参数变异性的降低。我们观察到，高斯近似对均值的预测（通过$\mathbf{y}_{\mathrm{MAP}}$）略优于切片采样MCMC算法；而后者的标准差则适度更小，这一点在图14中也有所体现。因此，集中度量也有适度改善。所有参数的不确定性均有所降低，其中c_{φ}的降低更为显著。

4.5 不确定性量化工作流第3步：前向不确定性量化分析

贝叶斯反演完成后，我们便可以将后验分布中包含的残余不确定性从参数传播到感兴趣量 I。为此，我们基于上一步（采用上述任一策略）生成的后验概率密度函数 ρ_{post}的样本，对感兴趣量 I 的代理模型进行评估。尽管我们可以重复使用第4.2节中针对 I 构建的六参数代理模型，但建议重新构建模型，就像我们在第4.4节针对 M 所做的那样，仅基于敏感性分析步骤后选出的四个相关参数。采用MCMC策略时，无法假设 ρ_{post} 服从某种特定分布，因此我们也使用第4.2节中用于构建新代理模型的相同*Clenshaw–Curtis* 点：实际上，如果 ρ_{post} 服从均匀概率密度，这些点便是最优的；但对于一般的 ρ_{post}，也可将其作为稳健的选择。相反，采用高斯近似策略时，（近似）后验分布是明确已知的。因此，我们使用所谓的加权*Leja* 点，这些点专用于高斯分布，预计能在参数空间中后验分布更为集中的区域，对 I 给出更准确的近似。

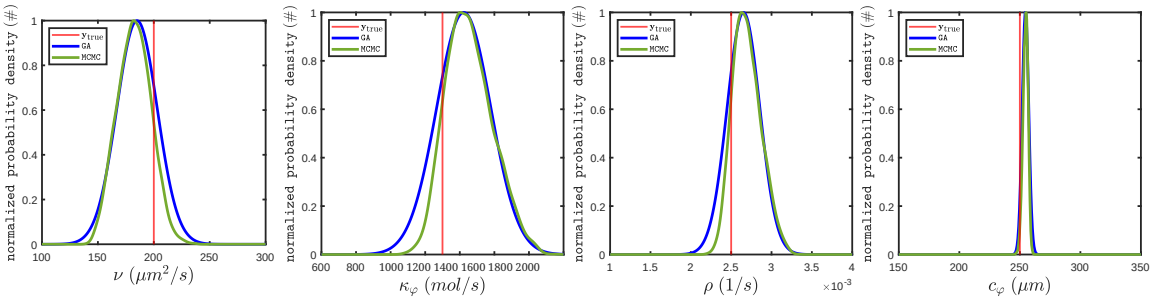


图14：高斯分布和切片采样MCMC对不确定参数边缘后验分布的近似，以及与“真实”参数值的对比。横轴显示参数的先验范围，因此该图定性地展示了ρ_{post}的集中度。

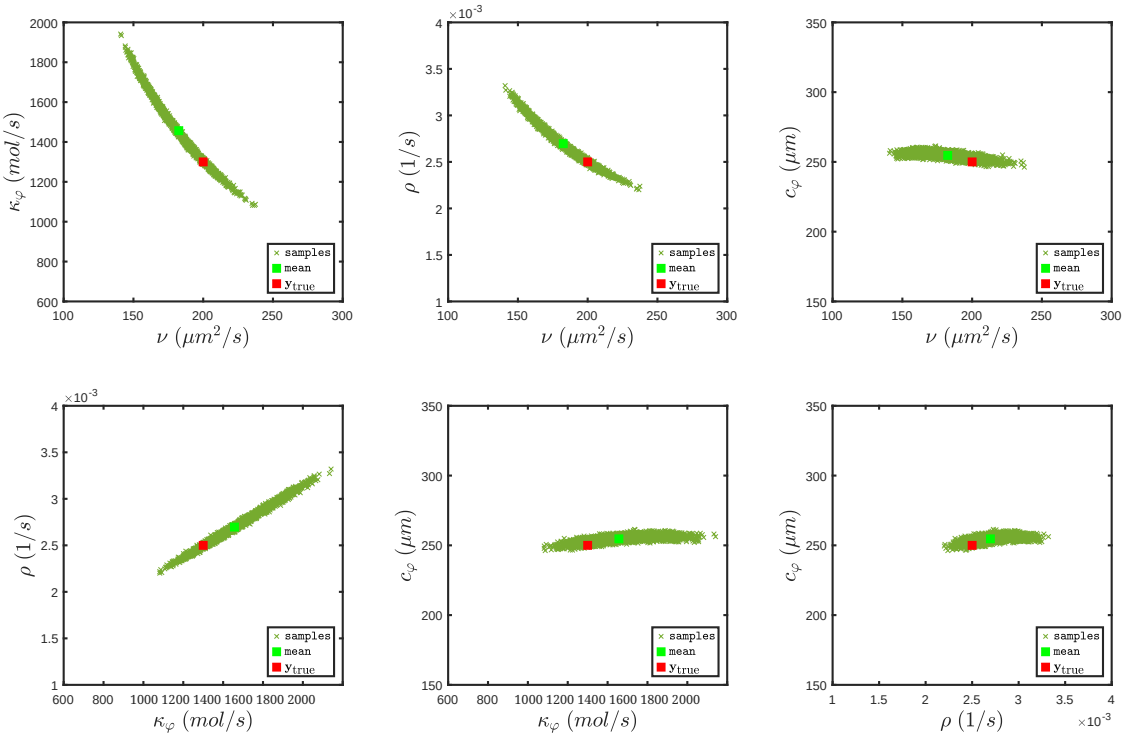


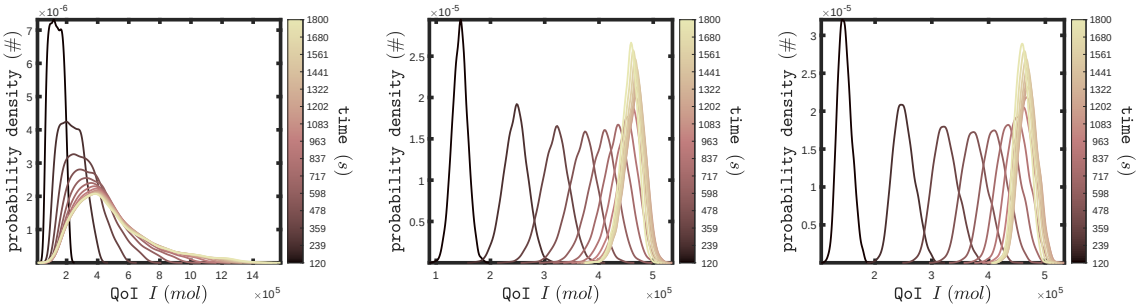
图15：切片采样MCMC算法生成的链在参数空间中各二维坐标平面上的投影。

	先验 mean	GA mean	MCMC mean	先验 标准差	GA 标准差	MCMC 标准差
ν	2.00×10^{2}	1.85×10^{2}	1.82×10^{2}	5.77×10^{1}	1.81×10^{1}	1.50×10^{1}
k_φ	1.30×10^{3}	1.42×10^{3}	1.46×10^{3}	4.04×10^{2}	1.57×10^{2}	1.35×10^{2}
ρ	2.50×10^{-3}	2.65×10^{-3}	2.70×10^{-3}	8.66×10^{-4}	2.03×10^{-4}	1.76×10^{-4}
c_φ	2.50×10^{2}	2.55×10^{2}	2.55×10^{2}	5.77×10^{1}	2.71×10^{0}	2.12×10^{0}

表3：不确定参数的先验和后验统计量。

	先验 CV	GA CV	MCMC CV	GA CF	MCMC CF
ν	2.89×10^{-1}	9.79×10^{-2}	8.21×10^{-2}	3.19×10^{0}	3.86×10^{0}
k_φ	3.11×10^{-1}	1.10×10^{-1}	9.28×10^{-2}	2.58×10^{0}	2.99×10^{0}
ρ	3.46×10^{-1}	7.65×10^{-2}	6.51×10^{-2}	4.27×10^{0}	4.93×10^{0}
c_φ	2.31×10^{-1}	1.07×10^{-2}	8.34×10^{-3}	2.13×10^{1}	2.72×10^{1}

表4：不确定参数的先验和后验浓度度量。



(a) 先验 (b) 遗传算法 (c) 马尔可夫链蒙特卡罗

图16：感兴趣量 I 的先验和后验概率密度函数随时间的演变。

图16展示了 I的先验与后验分布随时间的演化。尽管由高斯近似与马尔可夫链蒙特卡罗得到的后验分布之间并无实质性差异，但在比较后验概率密度函数与先验概率密度函数时，我们观察到显著的不确定性降低。（请注意坐标轴上的刻度不同。）我们再次分别借助式（25）和式（26）中的比较度量变异系数与CF来量化不确定性降低。值得注意的是，在更大的时刻，这种降低更为明显，见图17。与表4中一样，由马尔可夫链蒙特卡罗得到的后验分布近似似乎比高斯近似略为集中。

5 结论与展望

本文阐述了一套完整的工作流，其中涉及数据驱动微分模型，用以预测并控制CoC实验的动力学。将该工作流应用于带有合成数据的简化一维模型，可作为概念验证，展示所有相关步骤，并为应用于更现实的情形奠定基础。本文还证明了代理模型在加速数值模拟的同时不会对总体精度产生实质性影响这一关键作用。

未来的工作将致力于将所提出的工作流应用于能够更准确描述 CoC实验的模型。首先，这需要采用更复杂的方程（2）版本，其中包含更多未知场和耦合项；同时，还需要更符合实际的几何形状，并为设备的不同组件之间设置适当的界面条件，如图1所示。为此，我们计划扩展第2.3节讨论的混合间断伽辽金求解器，并纳入适用于对流主导情形的稳定化项。当不确定性影响（2）中的 χ 等参数时，这一点尤为重要。预计这种求解器的计算开销将远高于本研究中使用的求解器。因此，我们计划采用多保真代理模型，例如 [31]。我们还将探索切片采样马尔可夫链蒙特卡罗算法的替代方案，例如哈密顿蒙特卡罗采样，以降低贝叶斯反演的计算负担。作为本研究的进一步拓展，我们将应用所提出的框架，利用现有的 CoC实验实验数据校准模型参数。

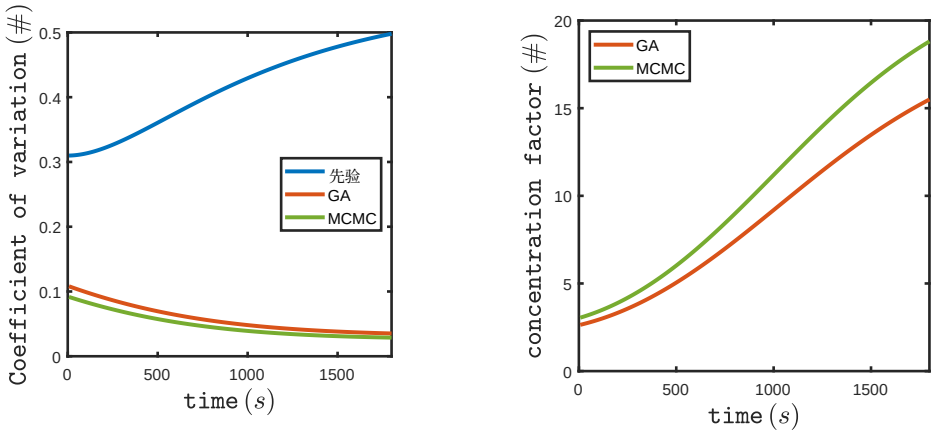


图17：感兴趣量 I 的先验和后验浓度测度随时间的演化。

6 致谢

本研究得到意大利研究部通过国家复苏与韧性计划（NRRP）补充行动的资助，资助项目为“D34Health——面向可持续医疗保健的数字驱动诊断、预后和治疗”，项目编号 #PNC0000001，CUP B53C22006100001。S.B.、G.B.、L.T. 和 P.Z. 均为国家科学计算组—国家高等数学研究所（GNCS-INdAM）的成员。L.T. 和 P.Z. 还部分获得了 PRIN 2022 PNRR项目“水流与污染物运移耦合模型的不确定性量化”（编号 P2022LXLYY）的资助，该项目由欧盟“下一代欧盟”计划资助。

参考文献

[1] N. Beißner, T. Lorenz, and S. Reichl. 器官芯片。载于 A. Dietzel（编），制药技术微系统：流体、颗粒、液滴和细胞操控，第299–339页，Cham，2016。施普林格国际出版公司。[2] E. Vacchelli, Y. Ma, E. E. Baracco, A. Sistigu, D. P. Enot, F. Pietrocola, H. Yang, S. Adjemian, K. Chaba, M. Semeraro, M. Signore, A. De Ninno, V. Lucarini, F. Peschiaroli, L. Businaro, A. Gerardino, G. Manic, T. Ulas, P. G¨unther, J. L. Schultze, O. Kepp, G. Stoll, C. Lefebvre, C. Mulot, F. Castoldi, S. Rusakiewicz, S. Ladoire, L. Apetoh, J. M. B.-S. Pedro, M. Lucattelli, C. Delarasse, V. Boige, M. Ducreux, S. Delaloge, C. Borg, F. Andr´e, G. Schiavoni, I. Vitale, P. Laurent-Puig, F. Mattei, L. Zitvogel, and G. Kroemer. 化疗诱导的抗肿瘤免疫需要甲酰肽受体1。《科学》，350(6263):972–978，2015。[3] J.-H. Lee, S.-K. Kim, I. A. Khawar, S.-Y. Jeong, S. Chung, and H.-J. Kuh. 胰腺肿瘤球体与星状细胞的微流控共培养：一种用于研究基质介导的细胞运动性和耐药性的新型三维模型。《实验与临床癌症研究杂志》，37，2018。[4] E. C. Braun, G. Bretti, and R. Natalini. 微流控芯片趋化多区域传输模型的质量守恒近似。《数学》，9(6)，2021。[5] H. G. Othmer and T. Hillen. 输运方程的扩散极限 II：趋化方程。*SIAM*应用数学杂志，62(4):1222–1250，2002。[6] R. Natalini and T. Paul. 关于Cucker-Smale模型的平均场极限。*Discrete and Contin-uous Dynamical Systems-B*，27:2873–2889，2022。[7] G. Bretti, A. De Ninno, R. Natalini, D. Peri, and N. Roselli. 受免疫活性癌症芯片实验启发的偏微分方程-常微分方程混合模型估计算法。*Axioms*，10，2021。

[8] M. Pompa, D. Torre, G. Bretti, and A. De Gaetano. 用于癌症芯片实验的二维随机基于智能体模型与偏微分方程扩散模型的敏感性分析。*Axioms*，12，2023。[9] E. F. Keller and L. A. Segel. 将黏菌聚集的启动视为一种不稳定性。*Journalof theoretical biology*，26(3):399–415，1970。[10] S. Bertoluzza, Bretti G., M. Pennacchio, and C. Prud'homme. Keller-Segel方程的HDG离散。准备中。[11] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler. 无限维贝叶斯反问题的计算框架（第一部分）：线性化情形及其在全球地震反演中的应用。*SIAM*科学计算杂志，35(6):A2494–A2523，2013。[12]N. Petra, J. Martin, G. Stadler, and O. Ghattas. 无限维贝叶斯反问题的计算框架（第二部分）：随机牛顿马尔可夫链蒙特卡罗及其在冰盖流反问题中的应用。*SIAM*科学计算杂志，36(4):A1525–A1555，2014。[13] A. Saltelli. 全局敏感性分析：入门。John Wiley & Sons, Ltd，2008。[14] R. Ghanem, D. Higdon, and H. Owhadi. 不确定性量化手册。Springer，2017。[15] I. Babuˇska, F. Nobile, and R. Tempone. 一种用于具有随机输入数据的椭圆型偏微分方程的随机配置法。*SIAM*评论，52(2):317–355，2010。[16] E. C. Braun, G. Bretti, and R. Natalini. 受癌症芯片（COC）实验启发的趋化模型参数估计技术。*International Journal of Non-Linear Mechanics*，140，2022。[17] G. Bretti, E. Campanile, M. Menci, and R. Natalini. 癌症芯片实验偏微分方程-常微分方程混合模型的情景研究。载于 A. D'Onofrio, A. Fasano, F. Papa, and C. Sinisgalli（编），数学生物物理学问题：纪念*Alberto Gandolfi*文集，第37–57页，Cham，2024。Springer Nature Switzerland。[18] E. Campanile, A. Colombi, and G. Bretti. 癌症芯片实验非局部积分微分模型的两步全局敏感性分析。*Mathematical Biosciences*，378，2024。[19] G.-J. Both, S. Choudhury, P. Sens, and R. Kusters. DeepMoD：利用深度学习从噪声数据中发现模型。*Journal of Computational Physics*，428，2021。[20] S. Lee, Y. M. Psarellis, C. I. Siettos, and I. G. Kevrekidis. 从基于智能体的蒙特卡罗模拟数据中学习黑箱和灰箱趋化偏微分方程/闭合关系。数学生物学杂志，87(1):15，2023年6月。[21] Y. M. Psarellis, S. Lee, T. Bhattacharjee, S. S. Datta, J. M. Bello-Rivas, and I. G. Kevrekidis. 通过坐标不变机器学习，从数据中发现细菌的趋化迁移。*BMC*生物信息学，25(1):337，2024。[22] A. Collin, T. Kritter, C. Poignard, and O. Saut. 肿瘤生长模型的联合状态-参数估计。*SIAM*应用数学杂志，81(2):355–377，2021。[23] A. Flavien, D. Ben Mansour, and S. Mazen. 由反应-平流-扩散型偏微分方程支配的联合状态-参数估计与反问题：在生物Keller–Segel方程和图案形成中的应用。计算与应用数学杂志，461，2025。[24] V. Kumar and S. Varghese. 用于研究肿瘤微环境中细胞间相互作用的离体肿瘤芯片平台。*Advanced Healthcare Materials*，8(1801198)，2019。[25] J. D. Murray. 数学生物学 *II*：空间模型与生物医学应用。Springer，纽约，NY，2003。[26] T. Curk, D. Marenduzzo, and J. Dobnikar. 对环境中和分泌的趋化物的趋化感知驱动大肠杆菌的集体行为。*PLoS ONE*，8(10)，2013。

[27] S. Du 和 F.-J. Sayas。《可混合化间断 Galerkin 方法理论：投影、估计与工具》。Springer，2019。[28] S. Bertoluzza、G. Guidoboni、R. Hild、D. Prada、C. Prud'homme、R. Sacco、L. Sala 和 M. Szopos。用于带积分边界条件椭圆问题的 HDG 方法：理论与应用。《科学计算杂志》，95(1)，2023。[29] C. Piazzola 和 L. Tamellini。算法 1040：Sparse Grids Matlab Kit——用于高维函数逼近和不确定性量化的稀疏网格 Matlab 实现。《*ACM* 数学软件汇刊》，50(1)，2024。[30] R. M. Neal。切片抽样。《统计年鉴》，31(3)：705–767，2003。[31] C. Piazzola、L. Tamellini、R. Pellegrini、R. Broglia、A. Serani 和 M. Diez。比较多指标随机配置法与多保真随机径向基函数在船舶阻力正向不确定性量化中的应用。《计算机辅助工程》，39：2209–2237，2023。