arXiv:2302.10986v1 [physics.geo-ph] 9 Feb 2023

FluidFlower 国际基准研究：过程、建模结果与实验数据比较

Bernd Flemisch[1*]、Jan M. Nordbotten[2,3],Martin Fernø[4],Ruben Juanes[5],Holger Class[1],Mojdeh Delshad[6], Florian Doster[7],Jonathan Ennis-King[8],Jacques Franc[9], Sebastian Geiger[7,10],Dennis Gl¨aser[1],Christopher Green[8], James Gunning[8],Hadi Hajibeygi[10],Samuel J. Jackson[8], Mohamad Jammoul[6],Satish Karra[12],Jiawei Li[9],Stephan K. Matth¨ai[13],Terry Miller[12],Qi Shao[13],Catherine Spurin[9], Philip Stauffer[12],Hamdi Tchelepi[9],Xiaoming Tian[10],Hari Viswanathan[12],Denis Voskov[10],Yuhang Wang[10],Michiel Wapperom[10],Mary F. Wheeler[6],Andrew Wilkins[14], AbdAllah A. Youssef[13]和 Ziliang Zhang[10]

[1*]水力学与水文系统建模系，斯图加特大学，德国斯图加特。[2]耦合地下动态建模中心，数学系，卑尔根大学，挪威卑尔根。[3]可持续地下资源中心，挪威研究中心，挪威卑尔根。[4]物理与技术系，卑尔根大学，挪威卑尔根。[5]土木与环境工程系，麻省理工学院，美国马萨诸塞州剑桥。[6]地下建模中心，奥登计算工程与科学研究所，德克萨斯大学奥斯汀分校，美国德克萨斯州奥斯汀。

[7]英国爱丁堡赫瑞瓦特大学地能工程研究所。[8]澳大利亚克莱顿北区澳大利亚联邦科学与工业研究组织能源部。[9]美国加利福尼亚州斯坦福斯坦福大学能源科学与工程系。[10]荷兰代尔夫特代尔夫特理工大学地球科学与工程系。[12]美国新墨西哥州洛斯阿拉莫斯洛斯阿拉莫斯国家实验室计算地球科学部。[13]澳大利亚帕克维尔墨尔本大学基础设施工程系及彼得·库克碳捕集与封存研究中心。[14]澳大利亚肯莫尔昆士兰先进技术中心澳大利亚联邦科学与工业研究组织矿产资源部。

*通讯作者。电子邮件：bernd@iws.uni-stuttgart.de；其他作者：
jan.nordbotten@uib.no；martin.ferno@uib.no；juanes@mit.edu；
holger.class@iws.uni- stuttgart.de；delshad@mail.utexas.edu；
f.doster@hw.ac.uk；jonathan.ennis-king@csiro.au；jfranc
@stanford.edu；s.geiger@tudelft.nl；dennis.glaeser
@iws.uni-stuttgart.de；chris.green@csiro.au；james.gunning@csiro.au；
h.hajibeygi@tudelft.nl；samuel.jackson@csiro.au；ja
mmoul@utexas.edu；satkarra@lanl.gov；jiaweili@stanford.edu；
stephan.matthai@unimelb.edu.au；tamiller@lanl.gov；
shao.q@unimelb.edu.au；cspurin@stanford.edu；stauffer@lanl.gov；
tchelepi@stanford.edu；x.tian-1@tudelft.nl；viswana@lanl.gov；
d.v.voskov@tudelft.nl；y.wang-25@tudelft.nl；
m.b.wapperom@tudelft.nl；mfw@ices.utexas.edu；
andrew.wilkins@csiro.au；abdallahy@student.unimelb.edu.au；
z.zhang-15@tudelft.nl；

摘要

成功部署地质碳封存（GCS）需要广泛使用储层模拟器，对封存场址进行筛选、排序和优化。然而，地质碳封存的时间尺度如此之长，目前尚无足够的长期数据可用于验证这些模拟器。因此，目前还没有可靠的依据来评估大规模地质碳封存作业动态特征的预测质量。

为弥补这一知识空白，我们开展了一项大型地质碳封存确认基准研究。为实现合理的时间尺度，我们建造了一个实验室规模的地质储存层（“FluidFlower”），作为实验和计算工作的基础。在 FluidFlower 中开展了一项由重复地质碳封存作业组成的验证实验，由此获得了我们定义为该系统真实物理动态特征的数据。来自世界各地的九个研究团队，基于对 FluidFlower 砂体进行的详细物理特征和岩石物理特征表征，分别或协作提交了预测。本文的主要贡献是报告并讨论此次验证基准研究的结果，同时介绍基准测试流程和参与的计算模型。我们通过多种具有指示意义的定性和定量指标，将各参与组提交的预测相互比较，并与实验数据进行对照。借此，我们详细评估了储层模拟器及其使用者捕捉地质碳封存作业注入动态和注入后动态特征的能力。

关键词：地质碳封存，验证基准，验证实验，代码互比

1 引言

地质碳封存（一氧化碳封存）有望弥合$_2$ 传统碳基能源的排放与政府间气候变化专门委员会报告所述所需减排量之间的差距 [1–4]。此外，地质碳封存还可与生物燃料结合，在负排放策略中发挥作用 [5]，并用于生产所谓的“蓝氢” [6]。为了安全且经济高效地实现这一潜力，大规模部署地质碳封存高度依赖建模和数值模拟研究，以评估潜在地质构造（主要是地下含水层）的适宜性。现有的储存潜力评估在很大程度上依赖此类建模研究 [7–12]。最终，要在地质碳封存这类对社会影响巨大的领域中生成基于模拟的数据和知识，就需要与政治决策者沟通。透明的模拟工作流程、数据的可复现性，以及对模拟结果信心的提升（例如通过全面的基准测试），都是开展沟通或让利益相关者参与建模过程的关键因素 [13]。

另一方面，目前全球仅有几十项大型碳封存作业处于运行状态 [14]，，其中没有任何一项是在经历了数十年的注入阶段后进入注入后阶段的。因此，建模与模拟领域缺乏用于评估其预测能力的可靠数据集，而对于我们能否准确捕捉与地质碳封存相关的主要物理过程，也存在很大的不确定性。试点研究提供了一定程度的信息 [15, 16]，，但地下环境的基本特性意味着，所收集的数据始终相对稀疏，尤其是在空间维度上。作为部分补救措施，已经开展了若干代码比较研究 [17–19]。然而，这些研究均未在存在物理真实值的情况下进行。

本研究旨在首次评估地质碳封存建模与模拟领域的预测能力。为实现这一目标，我们利用了卑尔根大学新建成的“FluidFlower”实验设施。在这套实验装置中，构建了一个具有挪威大陆架典型特征的地质模型。我们完成了初步的地质和岩石物理特征表征，并进行了单相示踪剂测试。在此基础上，我们开展了一项双盲研究：一方面，在卑尔根大学反复进行实验室尺度地质碳封存实验并测量结果，以下将相应团队称为 ExpUB。另一方面，我们邀请了全球从事地质碳封存研究的学术研究团队参加一项预测研究，该研究由斯图加特大学协调，以下称为 CoordUS。由于疫情减少了学术出行，我们得以充分确保参与组之间没有任何线下接触；同时，所有数字通信均受到限制、监管并存档，以确保双盲研究的完整性。如下文所述，我们还要求参与预测研究的参与者

提供独立预测，随后受邀根据群体互动更新预测。

在本文中，我们报告比较研究的最终结果，重点关注：1）来自不同参与组的预测之间的相关程度；2）这些预测与在 Flu-idFlower 中开展的实验室尺度地质碳封存测量结果之间的相关程度。综合来看，这既衡量了预测之间的可重复性（从实际操作角度来看），也反映了预测能力。

本文结构如下。第 2 节介绍一些必要的基本术语，描述验证实验，并说明基准测试过程。第 3 节介绍参与组及其相应模型。第 4 节通过定性和定量评估展示并讨论建模结果。第5 节比较建模结果与实验数据。第 6 节给出结论并展望未来。

2 基准测试方法

我们首先介绍一些基于 [20, 21]的基本概念和术语。验证这一术语描述的是“确定计算模型准确表示底层数学模型及其解的过程”，而确认指的是“从模型预期用途的角度，确定模型在多大程度上准确代表现实世界的过程”。此外，校准是调整计算模型中参数以提高其与数据一致性的过程。

一个验证实验，例如下面第2.1 节所介绍的实验，旨在“通过定量方式确定以计算机软件表示的数学模型模拟一个特征明确的物理过程的能力，并为此进行设计、执行和分析”。正如下面第2.2节将进一步详细说明的那样，我们执行一个验证基准 [20, 22], ，其中实验提供测量数据，用于与模拟结果进行比较。

2.1 验证实验

下文将简要介绍使用 FluidFlower 实验装置开展的实验。详情请参阅原始基准描述 [23] 和实验论文 [24]。图 1 展示了实验装置的几何结构。装置中填充了六种不同类型的沙粒，构建出若干渗透率各异的层，其中包括三个断层状结构。起初，孔隙空间完全被水饱和，水的顶部

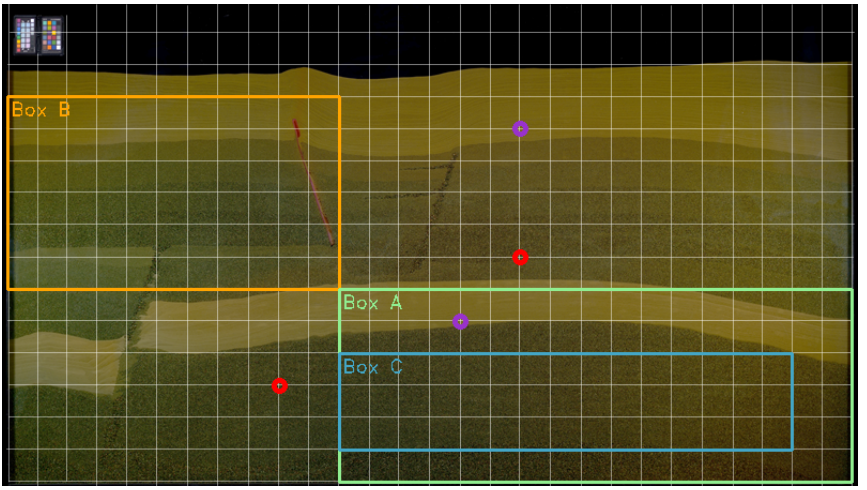

图 1 叠加激光网格的基准几何结构照片 [23, 图 8]。最亮的相为细砂屏障。红色圆圈表示注入点，紫色圆圈表示压力传感器。A-C 框对应于用于评估不同系统响应量的区域。

实验台的压力和温度均处于大气条件下[1]。通过两个注入端口注入气态一氧化碳$_2$，总时长为五小时。自注入开始后，连续五天监测一氧化碳$_2$ 在实验装置中的分布。2021年11月至2022年1月期间，共进行了五次实验运行。实验团队 ExpUB 尽力确保各次运行的操作条件一致。

[23] 中对实验装置的描述涉及外部几何结构、层理、相的性质、断层、流体性质、操作条件和井测试数据。具体而言，层理由高分辨率照片进行描述，参与组需据此确定不同砂层的位置。此外，描述还补充了沉积过程和注入前冲洗程序的详细信息。关于各相，提供了粒径分布信息，以及绝对渗透率、孔隙度、相对渗透率端点和毛细管进入压力的测量结果。井测试数据旨在用于校准数值模型。具体而言，提供的压力[2] 和示踪剂流动数据可用于估算不同相的渗透率分布。

该描述还定义了系统响应量（SRQ），基准研究参与者应报告这些量。各项 SRQ 将在第4节中详细介绍。

[1]显然，在这种低压下，一氧化碳$_2$ 与水之间的密度差远大于典型储层深度处的密度差。我们将在第6节讨论本研究对现实地质碳封存情景的适用性。[2]注入压力是在一个传感器处测得的，该传感器通过一段小直径管道与注入点相隔。将这段管道中的压降考虑在内，会影响校准结果。

2.2 基准测试流程

表1 展示了基准测试流程的时间顺序。在共同的准备阶段完成基准描述后 [23]，，为期三个月的所谓盲测阶段随即开始，期间不同参与组之间或与 ExpUB 之间均不允许直接交流。所有即将

2021年4月30日	封闭式参与征集开启。
2021年6月15日	征集截止。
2021年7月15日	向参与者提供初步基准测试描述。
2021年7月16日至8月19日	准备阶段，所有参与者均可参与讨论以及 ExpUB。
2021年8月20日	初步基准描述的反馈截止日期。
2021年9月16日	启动 Zoom 会议，第二轮基准描述已发布。
2021年9月17日至10月8日	公开讨论，以敲定描述。
2021年10月8日	向参与者传阅最终基准描述。
2021年10月9日–2022年1月11日	盲测阶段，不同参与者之间不得直接交流-与会者或与 ExpUB。
2022年1月9日	提交盲测基准数据的截止日期。
2022年1月12日	“完全盲测”模拟预测的线上研讨会和比较预测。
2022年1月12日至4月25日	同步阶段，所有参与者之间的通信已启用，但不包括 ExpUB。
2022年4月22日	提交最终基准数据的截止日期。
2022年4月26日至28日	在挪威举行研讨会，展示最终模拟预测、实验结果，模型校准研究，以及结果综合。

表 1 FluidFlower 基准测试流程的时间顺序。

建模人员提出的问题会提交给 CoordUS，并可能以匿名方式转交给 ExpUB。CoordUS 与 ExpUB 就答案达成一致后，该答案会发送给所有参与组，或仅回复提问者。盲测阶段结束时，每个参与组都会向 CoordUS 提交初始预测。随后，所有参与组首次召开会议，公布并讨论结果，ExpUB 仍不参与其中。此次会议开启了为期三个月的所谓同步阶段，让各预测组能够相互借鉴彼此的工作，并将这些知识融入自己的预测中。同步阶段尤其包括另外两次参与者共同会议。阶段结束时，各组记录最终预测，之后在挪威卑尔根附近举行现场研讨会，首次对比预测与实验结果。

为确保结果的完整性，在基准测试流程的不同阶段均遵循专门的通信规则。为了便于参与者远程交流，并保存这些交流内容以评估基准测试流程，建立了一个 Discord 服务器

已设立[3]。除最初向所有相关人员开放的通用频道外，还为每个参与组单独设立了一个私密频道，可用于与基准研究组织者沟通。

参与者将所有结果数据上传至“FluidFlower” GitHub 组织中的 Git代码库[4]。每个参与组均获得一个专用代码库的写入权限，该代码库以其所在机构命名。在盲测阶段，只有参与者本人可以访问各自的代码库。进入同步阶段后，所有参与者均获准读取所有参与者存储库。4月的研讨会结束后，这些代码库进一步开放，也收录了物理实验的结果。提交本文时，相关代码库已转为公开状态。

3 参与小组与模型

FluidFlower 基准研究共有九个小组参与，每组由两至五人组成。下文以相应机构的所在地或名称指代各组，分别为奥斯汀、CSIRO、Delft-DARSim、Delft-DARTS、赫瑞-瓦特、洛斯阿拉莫斯国家实验室、墨尔本、斯坦福和斯图加特。表2列出了各参与组相关的建模选择。

就构成数学模型主体的偏微分方程而言，几乎所有参与者都采用组分质量守恒方程。除奥斯汀和墨尔本这两个例外外，所有参与者都采用单元中心有限体积进行空间离散化。除墨尔本以外，所有团队都采用标准的隐式欧拉时间离散，并以全耦合全隐式的方式求解由此得到的离散方程。各团队的做法在本构关系方面开始出现较大差异。大多数参与者采用 Brooks–Corey关系来描述毛细压力和相对渗透率，但也采用线性关系等其他方法。此外，还考虑了各种用于确定相组成和相密度的状态方程。除了这些主要选择外，各参与者的计算模型在所采用的空间参数方面也存在差异，例如假设的固有渗透率、孔隙度、残余饱和度等。这些参数可能取决于所考虑的砂类型，即空间位置。每个参与团队所用的参数都汇总在各自 GitHub 仓库顶层目录下的文件`spatialparameters.csv` 中。

[3] https://discord.gg/8 47 45 3QfZST
https://github.com/fluidflower

表2 参与组的建模选择。所用缩写：组分质量守恒方程“CMB”、相质量平衡“PMB”、Brooks–Corey“BC”、混合有限元“MFEM”、单元中心有限体积“CC-FV”、采用嵌入式不连续性的同位有限元有限体积方法“DFEFVM”。

4 建模结果

以下将介绍并讨论基准描述要求以 SRQ 形式提交的建模结果。这些结果分为三类：第4.1节中的密集数据空间分布图、第 4.2节中的密集数据时间序列，以及第 4.3节中的稀疏数据。

4.1 密集数据空间分布图

要求参与者提供注入开始后 24、48、72、96 和 120 h 时的空间相分布快照，具体包括气态一氧化碳$_2$ 的饱和度，以及液相中一氧化碳$_2$ 的浓度。每个参与组均可自行定义用于模拟的计算网格，但结果应以均匀网格报告，网格单元大小为 1 cm × 1 cm。

4.1.1 饱和度

图2 至6 展示了所有参与组在选定的每日时间步所报告的饱和度值。首先观察图2，可以发现，大多数参与者报告的下部细砂屏障下方一氧化碳$_2$ 羽流形状非常相似，时间为24小时后。此外，几乎没有气态一氧化碳$_2$ 。

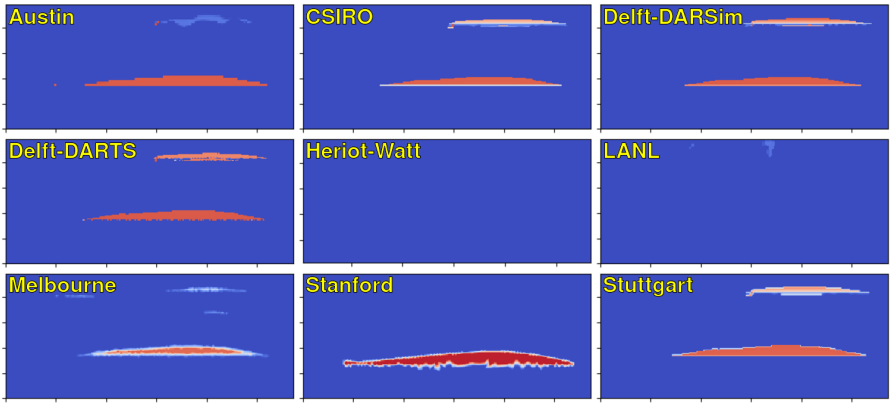

图224小时后气态一氧化碳$_2$ 的空间分布。色彩图的最小值为0，表示一氧化碳$_2$ 饱和度为0，以蓝色标示；最大值为1，以红色标示。

一天后在 B 箱内报告的结果。对于区域右侧部分的上部屏障，各结果的一致性明显较低。这可以解释为：下部和上部注入的一氧化碳$_2$ 量相差两倍以上，因此，上部区域的溶解行为差异会更早显现。赫瑞-瓦特 和 洛斯阿拉莫斯国家实验室的两位参与者报告称，模拟第一天结束后，区域内没有或几乎没有气态一氧化碳$_2$ 。

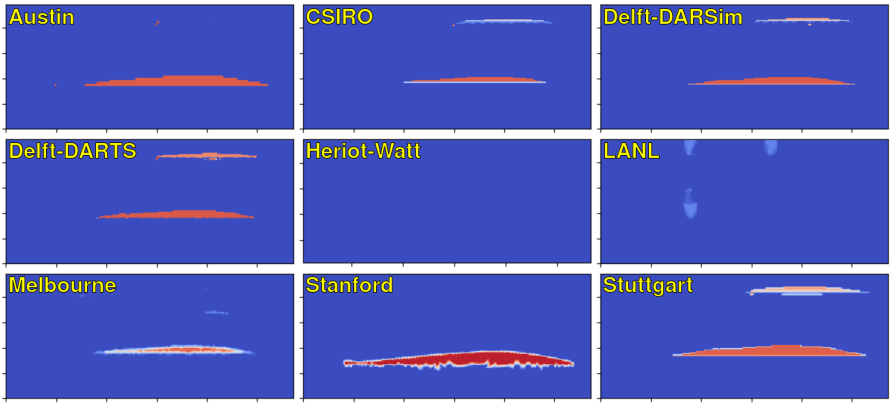

图 3 48 小时后气态一氧化碳$_2$ 的空间分布。色彩图的最小值为 0 一氧化碳$_2$ 饱和度，以蓝色表示；最大值为 1，以红色表示。

在赫瑞-瓦特的结果中，这是由于选择了用于毛细压力的范·根努赫滕关系，详见下文第4.2节。所报告的结果是在计算时间限制内能够解析的最小毛细带，预计会高估溶解量。对于洛斯阿拉莫斯国家实验室而言，情况则不同，其中一氧化碳$_2$会离开系统，因为几乎不会发生捕集，见下文。

考察图2至6在不同时间步下的饱和度分布，可以揭示不同溶解行为的影响。在此过程中，

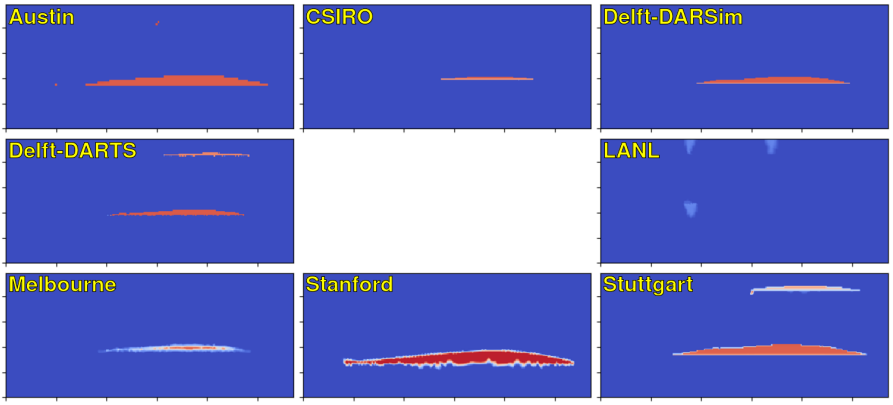

图 4 72 小时后气态一氧化碳$_2$ 的空间分布。色彩图的最小值为 0 一氧化碳$_2$ 饱和度，以蓝色表示；最大值为 1，以红色表示。

ular，CSIRO，Delft-DARSim，Delft-DARTS 和墨尔本 报告称，一氧化碳$_2$ 气体羽流随时间消散，而羽流形状对于奥斯汀、斯坦福 和斯图加特则相当稳定。从 72 小时开始，赫瑞-瓦特 未报告任何空间地图数据。

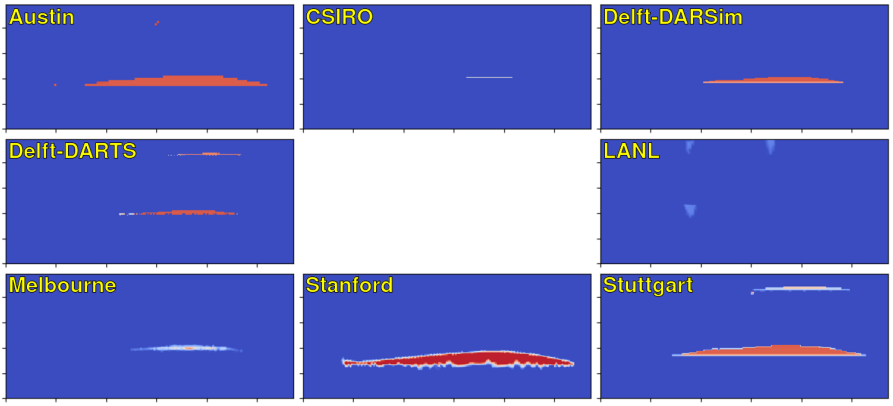

图 5 96 小时后气态一氧化碳$_2$ 的空间分布。色彩图的最小值为 0，以蓝色表示一氧化碳$_2$ 饱和度；最大值为 1，以红色表示。

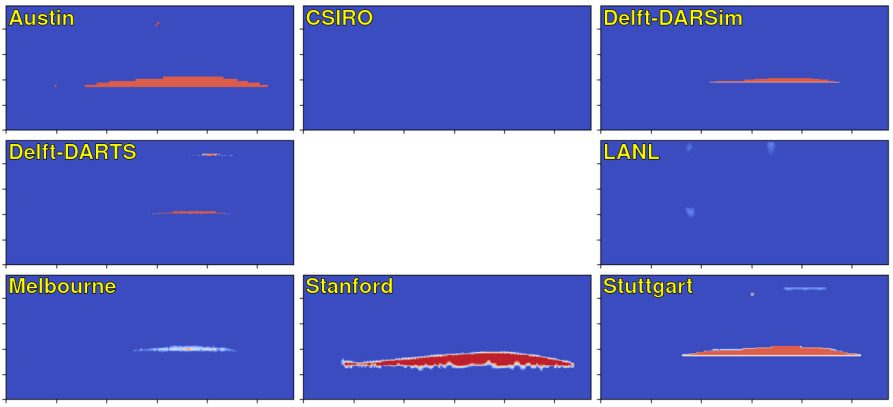

图 6 120小时后气态一氧化碳$_2$ 的空间分布。色彩图的最小值为 0，以蓝色表示一氧化碳$_2$ 饱和度；最大值为 1，以红色表示。

4.1.2 浓度

与饱和度类似，图7 至11 展示了所有参与组在选定的每日时间步上报告的浓度值。乍看之下，结果的变化似乎比饱和度更大，但大多数组报告的定性行为相似。一氧化碳$_2$溶解到液相中，并且由于纯水与富含一氧化碳的$_2$水之间存在密度差，后者向下移动并形成指进。这种运动会受到细砂屏障或区域底部的阻碍。洛斯阿拉莫斯国家实验室LANL的模拟所显示的结果明显偏离了这种较为一致的定性行为：一氧化碳$_2$ 相对笔直地向上移动，基本未受到细砂屏障的阻碍。可能的原因有很多，包括对相几何结构的不同解读、构建的计算网格不同、空间参数的变化幅度过小，以及

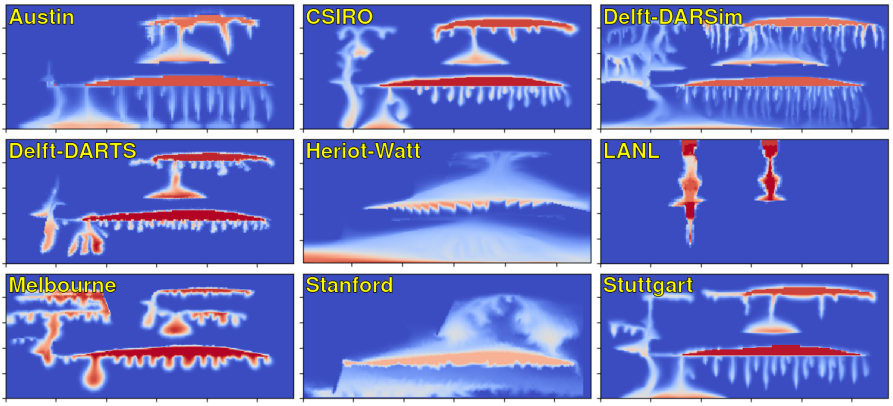

图 7 24小时后液相中 CO_2 浓度的空间分布。色彩图的最小值为 0 千克 m^{-3}，以蓝色表示；最大值为 1.8 千克 m^{-3}，以红色表示。

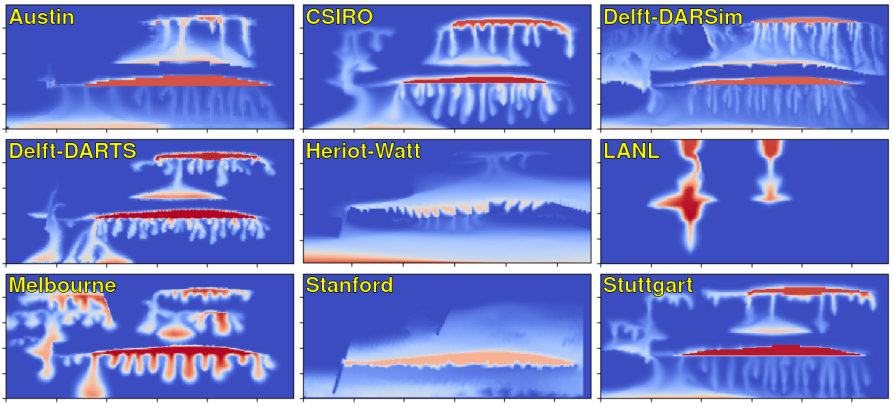

图 8 48小时后液相中 CO_2 浓度的空间分布。色彩图的最小值为 0 千克 m^{-3}，以蓝色表示；最大值为 1.8 千克 m^{-3}，以红色表示。

本构关系不足。由于使用 PFLOTRAN 和 FEHM 运行两套代码得出了相似结果，因此在研究过程中无法确定确切原因。

其余参与组之间可观察到的主要定量差异，源于溶解发生速度的不同。尤其是，赫瑞-瓦特 和 斯坦福 的溶解速度似乎远快于其他参与组。此外，对于有多少一氧化碳$_2$ 经相应断层带进入区域左上部分（即 B 箱），各方也存在相当大的分歧。

另一个值得关注的指标是演化指进的数量及其在水平方向上的厚度。这里的差异很大程度上

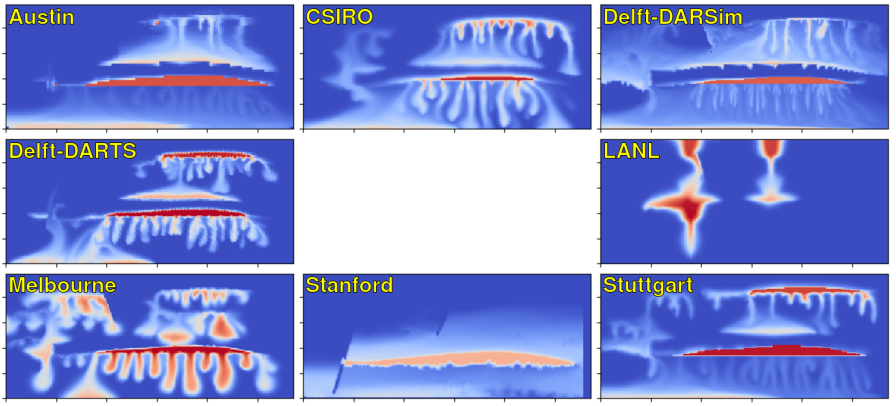

图 9 72 小时后液相中一氧化碳$_2$浓度的空间分布。色彩图的最小值为 0 千克 m^{-3}（以蓝色表示），最大值为 1.8 千克 m^{-3}（以红色表示）。

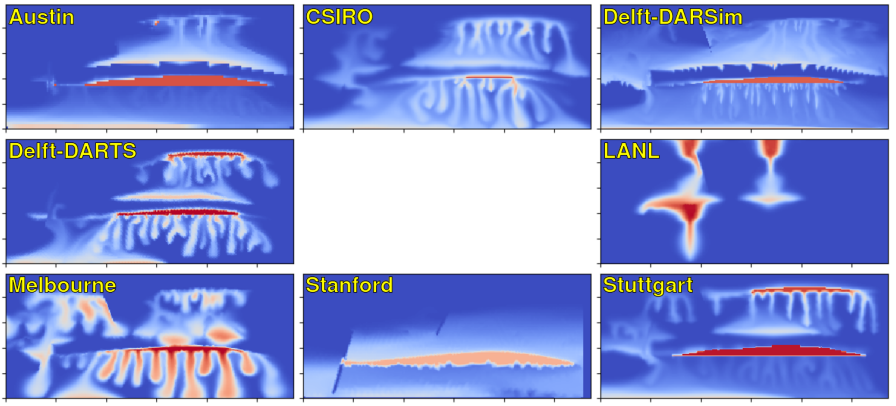

图 10 96 小时后液相中一氧化碳$_2$浓度的空间分布。色彩图的最小值为 0 千克 m^{-3}（以蓝色表示），最大值为 1.8 千克 m^{-3}（以红色表示）。

可归因于网格分辨率的相应差异。例如，参与的团队CSIRO、Delft-DARSim和Delft-DARTS 采用相对较高的分辨率（参见表2），显示出的指进数量明显更多、厚度也更薄；而奥斯汀和墨尔本的分辨率则相对较低。

4.1.3 定量比较

作为一种定量指标，我们采用 Wasserstein度量来分析两个快照之间的差异。该度量适用于质量相等的分布，用于衡量“重新配置一个分布中的质量以恢复另一个分布所需的最小工作量” [51] 为了将 Wasserstein度量应用于所报告的结果（这些结果通常具有

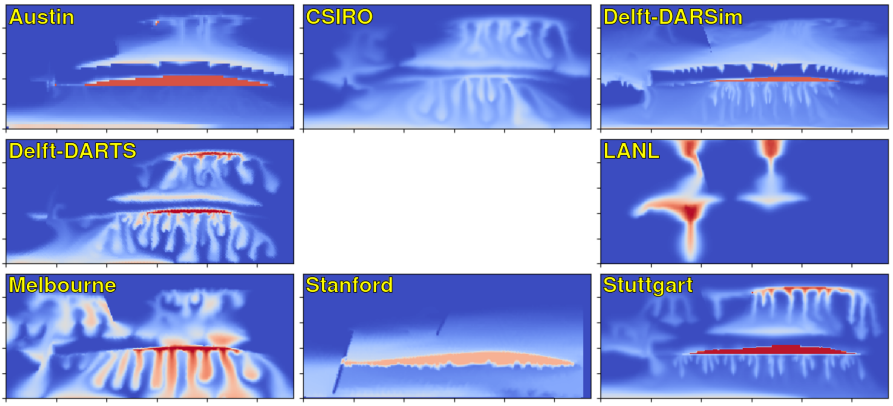

图 11 120小时后液相中一氧化碳$_2$ 浓度的空间分布。色彩图的最小值为 0 千克 m^{-3}，以蓝色表示；最大值为 1.8 千克 m^{-3}，以红色表示。

略有不同的质量（详见第 4.2.1 节中的讨论），我们首先粗略估算每个单元格中的一氧化碳$_2$ 质量密度，方法是通过以下公式结合所报告的浓度和饱和度

$$\widetilde{m} = \varrho_{\mathrm{g}} s + c(1-s).$$

上方，s 和 c 分别表示饱和度和浓度值，而气态 ϱ_{g} 一氧化碳$_2$ 的密度设为 2千克 m^{-3}，以反映实验条件。由此得到的数值可以通过相应的灰度图进行可视化，这些灰度图已上传至参与者的数据存储库。要使这些数值具有可比性，最后一步是对其进行归一化，使其能够正式视为实验域上的二维概率分布。给定归一化后的数值，可以应用 Python 库 POT [52] 来计算 Wasserstein 距离。这些数值列于附录A 中，对应每个要求的单独时间步。包含不同时间步结果之间距离的完整数据已提供于 FluidFlower 通用 GitHub 仓库中。对于报告结果中质量大致相等的组，这种方法可以提供合理的近似；但是，它不适用于洛斯阿拉莫斯国家实验室的结果，因为其模拟表明，注入质量中有相当一部分离开了区域。因此，洛斯阿拉莫斯国家实验室的结果被排除在 Wasserstein 距离计算之外。

我们在图12中以第一个和最后一个时间步为例，展示计算得到的 Wasserstein 距离。这些数值通过乘以系统中一氧化碳$_2$ 的实际质量进行了量纲化，单位为克·厘米。因此，100 克·厘米相当于 1 克质量（例如，系统中约 20% 的一氧化碳$_2$ ）移动了 1 米（例如，约占整个模拟域的三分之一）。因此，我们认为，100 克·厘米及以下的数值对应的结果相对接近

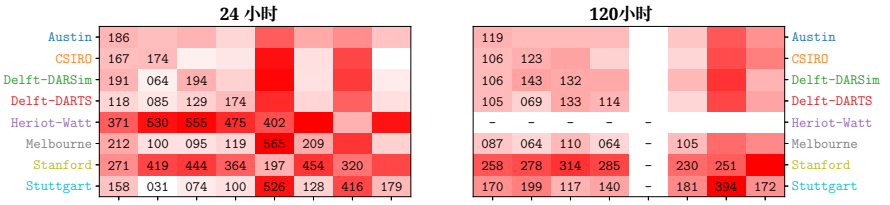

图 12 首个和最后一个时间步的 Wasserstein 距离，单位为克·厘米。颜色范围从低值的白色到高值的红色。对角线上的值是相应行/列的均值。由于数值具有对称性，因此不显示对角线以上的值。由于 Heriot-Watt 在 120小时没有报告空间分布图，因此相应字段留空。

；而远超 100 克·厘米的结果则表明存在显著差异。因此，图12 量化了上文各小节讨论的定性结果。尤其是，赫瑞-瓦特和斯坦福的空间图在所有时间步中与其他组的距离最大。由于溶解行为不同，它们的均值距离是其他组的两到三倍。总体而言，均值距离从第一个时间步到最后一个时间步大多逐渐减小，因为一氧化碳$_2$ 继续溶解于水中，其质量也更广泛地分布在区域内。

4.2 密集数据时间序列

参与团队被要求在为期五天的时间段内，每隔十分钟报告一次若干标量响应量（SRQ）：区域内 CO_2 的总质量、两个位置的压力、A 箱和 B 箱中的相组成，以及 C 箱中的对流。

4.2.1 CO_2的总质量

图13 展示了不同参与组报告的计算域内一氧化碳总质量随时间的演化情况。$_2$ 基准描述规定，注入速率以标准立方厘米每分钟（SCCM）为单位 [23]。虽然并未明确说明所采用的标准状态，但ExpUB 使用的是美国国家标准与技术研究院（NIST）对标准状态的定义，即 293。15 K 和 1。013 bar。假设没有一氧化碳$_2$离开区域，则最终总质量约为 8。5 g。大多数建模组采用了对标准状态的相应解释，但有三个组报告的数值较高，约为 9。4 g。对于洛斯阿拉莫斯国家实验室，有一组报告的数值明显偏低，这是因为一氧化碳$_2$ 会离开区域，详情见第4.1节。在大多数结果中，注入停止后，一氧化碳$_2$ 总量保持不变，表明没有质量离开该

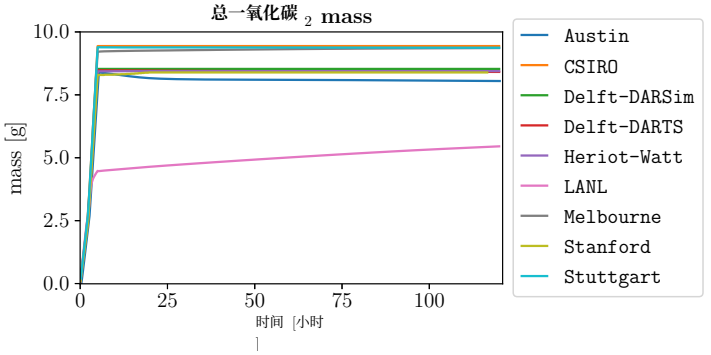

图14 计算域内一氧化碳$_2$ 总量的时间演化。

系统。不过，有些组报告称总量仍进一步增加或减少，这部分可由数值效应解释 [53]，也可能是因为气态一氧化碳$_2$ 离开了计算域。

4.2.2 压力

下一个报告的 SRQ 是在区域内两个传感器处测得的压力时间演化。图 14 展示了报告的结果。大多数

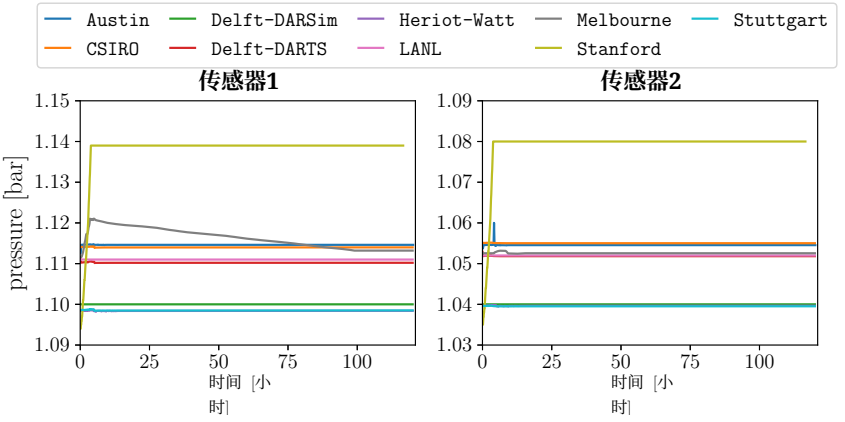

图14 计算域内两个位置的压力时间演化：传感器1（左）和传感器2（右）。

结果表明，一氧化碳$_2$ 注入对观测到的压力值影响很小。各传感器处的压力基本保持恒定，处于规定的初始条件和可能的边界条件下；这些条件对应于假定的环境大气压以及地下水位的影响。不过，斯坦福斯坦福和墨尔本两组报告称，注入过程产生了显著影响。为进一步详细考察这一现象，图15 展示了模拟开始后前十小时的放大图。结果

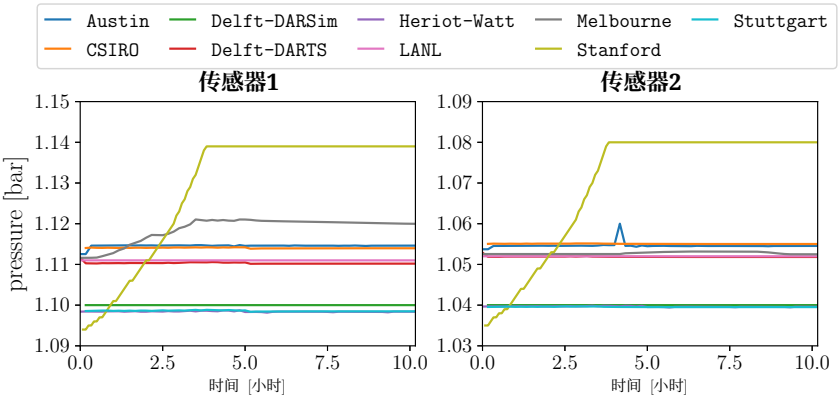

图15计算域内两个位置的压力时间演化：传感器1（左）和传感器2（右）。前六小时的放大图。

来自墨尔本的结果显示，只有第一个传感器的压力显著上升，并在停止注入后缓慢衰减至恒定水平。两个传感器的压力增幅不同，可用它们与注入井的距离不同来解释。相比之下，斯坦福报告称两个传感器的压力增幅相同。这可以解释为，各处均采用无流边界条件，且假设流体不可压缩。值得注意的是，两组均报告称，压力在约 3.5 小时后不再上升，而一氧化碳$_2$ 注入则在 5 小时后停止。

4.2.3 相组成

下面我们讨论所报告的一氧化碳在$_2$ A 箱和 B 箱的两种流体相中的分布。具体而言，参与者报告了流动和非流动气态一氧化碳的量的演化$_2$，一氧化碳$_2$ 在液相中溶解的一氧化碳的量，以及$_2$ 封盖相中所含一氧化碳的量。我们首先关注 A 箱及相应的图16。可以立即看出，各参与组之间结果的差异远大于此前的标量响应量。所有结果的共同点是，流动气态一氧化碳$_2$ 在约五小时处达到峰值（与注入停止时间一致），随后以不同速率溶解。八项结果可分为三个表现出相似速率的簇。最大的簇包括参与者CSIRO、Delft-DARSim、Delft-DARTS和墨尔本。在这些结果中，溶解贯穿整个模拟时段，其速率相较另外两个簇处于中等水平。参与者奥斯汀和斯图加特均显示出初始衰减后非常缓慢的溶解行为。相比之下，赫瑞-瓦特和斯坦福预测了一氧化碳最快的溶解，在不到一天后便没有流动气态一氧化碳$_2$ 残留。然而，斯坦福报告了大量气态一氧化碳$_2$ 转变为非流动状态，这是由于其采用的非流动气体识别方法导致了高估。另一个异常值是洛斯阿拉莫斯国家实验室，其结果显示没有一氧化碳$_2$ 进入 A 箱。这些观察结果

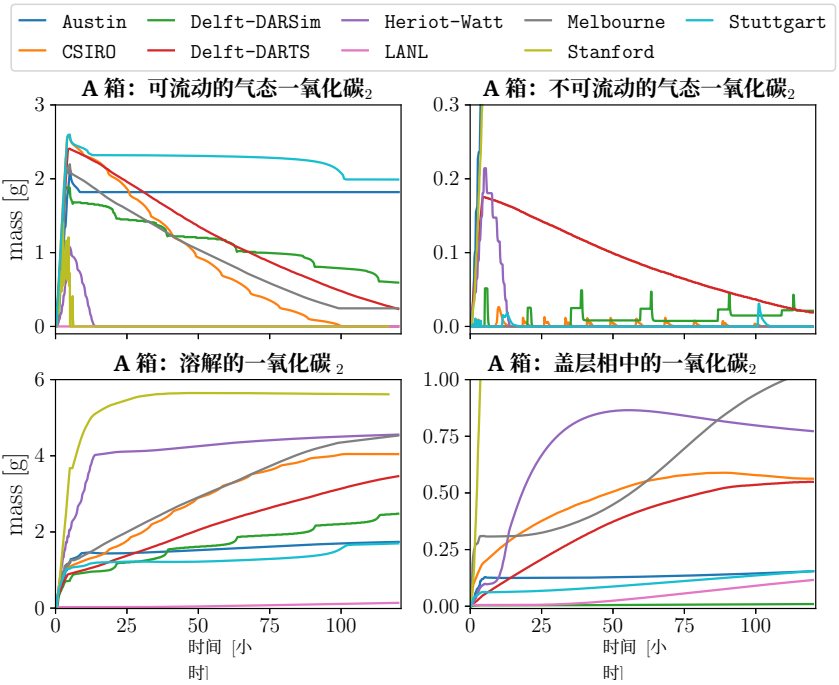

图 16 A 箱中一氧化碳$_2$ 相分布的时间演化。

与第4.1节关于空间图的结果和讨论一致。此外，一个显著特征是若干曲线呈阶梯状变化，尤其是CSIRO、Delft-DARSim和斯图加特所报告的结果。该数值效应是由于水气接触面与单元面重合时，溶解过程中出现网格依赖的突发。该现象最初也被赫瑞-瓦特观察到；为避免这一效应，他们决定对较粗的沙粒采用 van Genuchten 毛细压力—饱和度关系，另见表2。

转到 B 箱和图 17，结果呈现出更大的差异。这可归因于箱体的位置，以及量化有多少一氧化碳$_2$ 到达左下方的断层带，随后进入区域的左上方所带来的挑战。尽管所有参与者都预测，流动性气态一氧化碳$_2$ 最迟在两天后消失，但峰值量在 0 到 0.6 克之间存在显著差异。这些不同的峰值量，加上不同的溶解速率，解释了图 $_2$ 中所示溶解的一氧化碳$_2$ 的高度差异17。值得肯定的是，几乎所有预测 B 箱中存在大量一氧化碳$_2$ 的模型都报告了非常相近的出现时间。

4.2.4 对流

作为对流度量，参与者被要求报告 C 箱内浓度随时间的总变化，定义见 $M(t)$

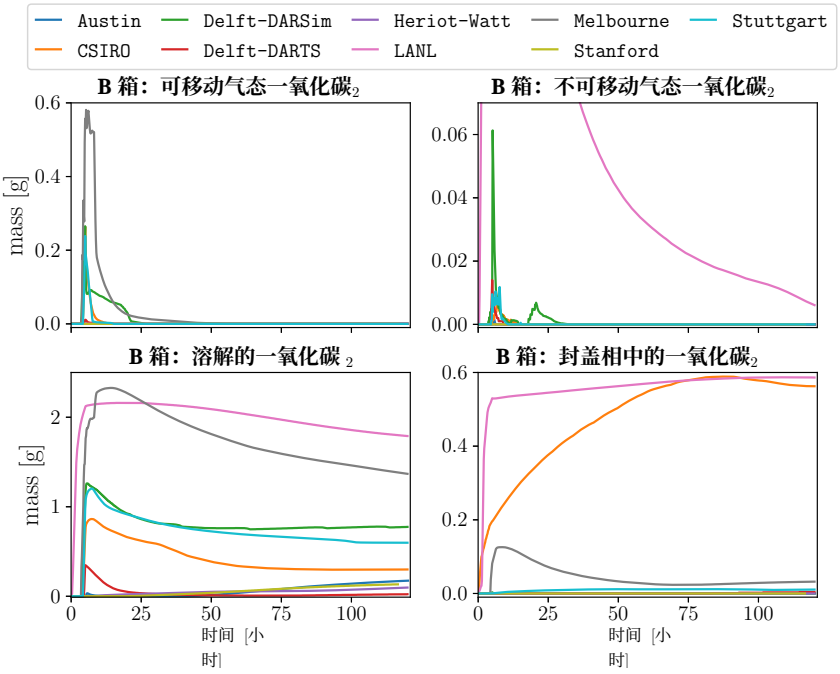

图 17 B 箱中一氧化碳$_2$ 相分布的时间演化。

见 [23, 第 2.8 节3]。结果如图 18 所示。相对较大的

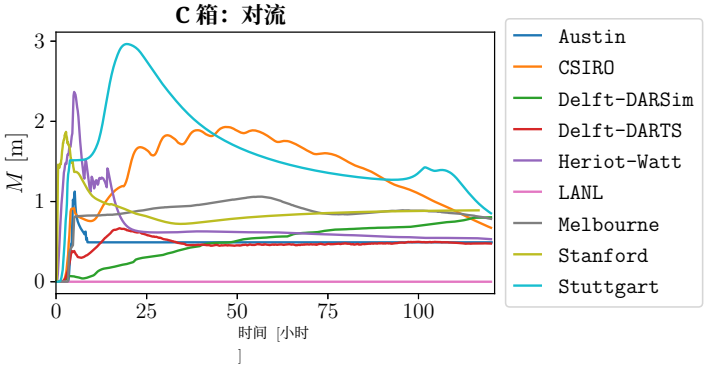

图 18 作为 C 箱对流度量的 $M(t)$的时间演化。

可以观察到其分布范围较广，峰值介于 0 至 3 米之间。此外，动态行为差异很大，从单调增加到相当强烈的振荡都有。尽管如此，大多数参与者报告称，随着时间推移，数值会趋于稳定，最终稳定在 0.5 至 1 米之间。

4.3 稀疏数据

本节介绍所报告的所谓稀疏数据。每项稀疏数据都必须用六个数字报告，分别表示通过实验得出的量的均值预测（以 P10、P50 和 P90 值表示），以及该量在实验集合中的标准差预测（同样以 P10、P50 和 P90 值表示）。由于大多数组别没有报告期望标准差的 P10 和 P90 值，因此在以下比较中，我们只考虑 P50 值。生成预测和不确定性时，可以选择任何偏好的方法，从集合模拟和正式的不确定性量化方法，到基于经验的人类直觉，均可采用。我们先考察两个传感器处的最大压力，接着关注 A 箱中可移动气态一氧化碳$_2$ 达到最大值的时间，以及 C 箱中对流混合开始的时间，最后考察对三天后 A 箱和 B 箱内相态分布的预测。

4.3.1 两个传感器处的最大压力

研究要求参与者给出传感器1和2处的预期最大压力，以此评估覆盖层受到机械扰动的风险。报告的数值见图 19。从纵轴的刻度可以看出，所有参与组报告的压力值都非常接近。大多数组报告的期望均值的 P10、P50 和 P90 值也彼此十分接近，其中某一组的最大差异约为 10 毫巴。对于 奥斯汀 和 墨尔本，只有两个组预计实验集合中的标准差会较大。对于两个传感器，所有组报告的压力值中，最高 P90 与最低 P10 之间的差值均小于 40 毫巴。这表明参与者没有考虑实验装置所在地大气压力的典型变化；冬季月份期间，该变化超过 50 毫巴。尽管实验运行的具体日期没有明确告知参与者，但有关通常压力变化的信息是公开可查的[5]。

4.3.2 A 箱中可移动气态 CO_2 达到最大值的时间与 C 箱对流混合开始

我们现在关注 A 箱中可移动气态一氧化碳达到最大值的时间$_2$ ，将其作为泄漏风险开始降低时间的替代指标。相应的报告值显示在图 20 的上图中。现在，大多数参与组报告的期望均值和标准差的 P10 值与 P90 值之间均存在显著差异。然而，有几个参与组对期望均值非常确定，报告的范围也非常窄。各参与组之间的差异明显更大

[5]https://weatherspark.com/h/s/148035/2021/3/Historical-Weather-Winter-2021-at-Bergen-Flesland-Norway 图表压力

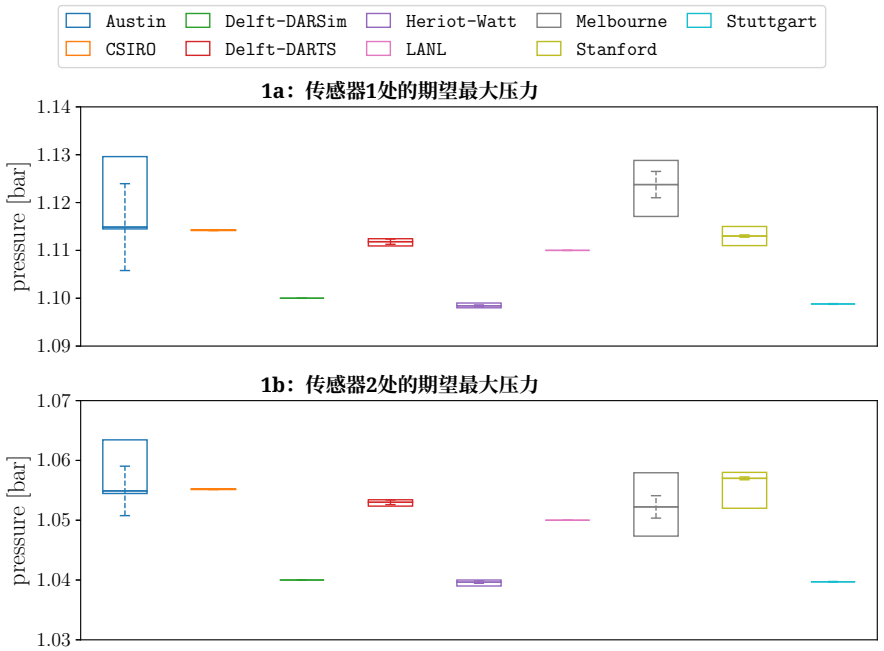

图 19 传感器1和2处最大压力的稀疏数据报告。箱体的底部、中部和顶部水平线分别表示期望均值的 P10、P50 和 P90 报告值。虚线从均值处延伸，长度为 ± 期望标准差的 P50 报告值。

，相比之下，前文讨论的压力的差异较小。这可以很好地解释为建模结果的差异更大，正如第4.2.2节和第 4.2.3节所讨论的那样。

作为衡量捕捉对流混合开始能力的替代指标，我们关注该量首次超过 C 箱宽度 110% 时的时间 $M(t$，其定义见 [23, 第 2.8 节。3]如图20下图所示。首先需要指出，有三个组完全没有报告任何数值。其余六个组中，有四个报告的数值非常相近，约为 4 小时，且 P10 与 P90 之间的范围较窄。至于 CSIRO，有一个组报告的期望值明显更大，P10 与 P90 之间的变化幅度也更大。为了进一步考察这一点，我们将其与相应的时间演化进行比较，如图 $M(t$所示，见图18。C 箱宽度的 110% 等于 1。65 米。我们可以看到，在整个模拟时段内，有若干结果始终未达到这一数值。这也解释了为何有三个组没有报告稀疏数据的任何数值。进一步放大模拟时间的前十小时，如图21 所示，可以明确看出报告的时间序列数值与稀疏数据之间的关系。从表示所报告期望均值的垂直线可以看出， $M(t)$ 的测量值通常远低于 110%。因此显而易见，若干参与组并非仅依据所报告的模拟结果来估算此处考虑的可测量量。

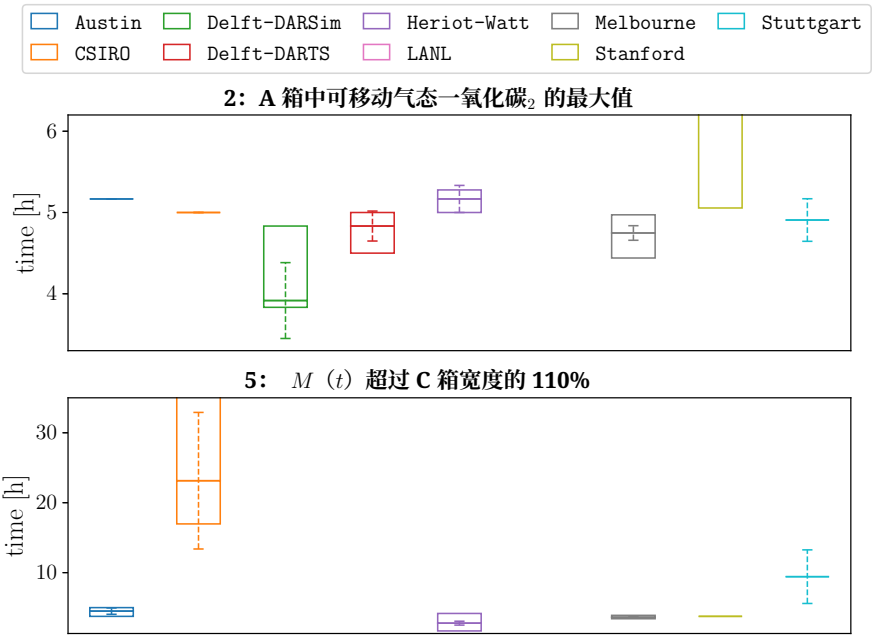

图 20 A 箱中可移动气态一氧化碳$_2$ 达到最大值的时间（上图），以及积分 $M(t)$首次超过 C 箱宽度 110%的时间所对应的稀疏数据（下图）。箱体底部、中部和顶部的水平线分别表示所报告的期望均值的 P10、P50 和 P90 值。虚线从均值延伸，表示 ± 所报告的期望标准差的 P50。

4.3.3 A 箱和 B 箱三天后的相态分布

接下来，我们考察注入开始 72 小时后 A 箱内相分布的稀疏数据，以此作为衡量准确预测近井相态分配能力的替代指标。由相应的图22可以立即看出，期望均值的 P10 与 P90 数值之间的报告范围明显大于前述指标，同时期望标准差也有所增加。关于可移动气态一氧化碳$_2$, 的量，均值的期望 P50 在 0.5 至 2 克之间；而对于溶解的一氧化碳$_2$, ，数值大多在 1 至 4 克之间。

注入开始后 72小时，B 箱中预期的相分布如图所示23，可将其视为衡量处理不确定地质特征能力的替代指标。可以观察到，大多数情况下预计不会出现可移动的气态一氧化碳$_2$ ，但相关不确定性被认为相当高。对于斯坦福，较大的变化源于其不确定性量化中纳入了一次以不混溶流体相为极端情况的模拟。再看左下图，各参与组预测的溶解一氧化碳$_2$ 含量差异显著。

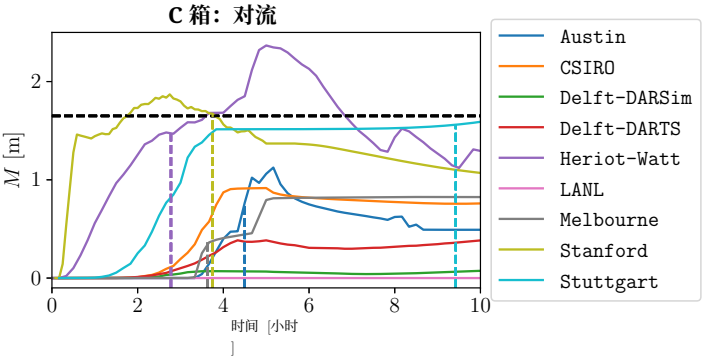

图 21 放大显示 $M(t)$ 时间演化的前十小时。黑色水平虚线表示 C 箱的 110%，竖直虚线对应报告的期望均值。

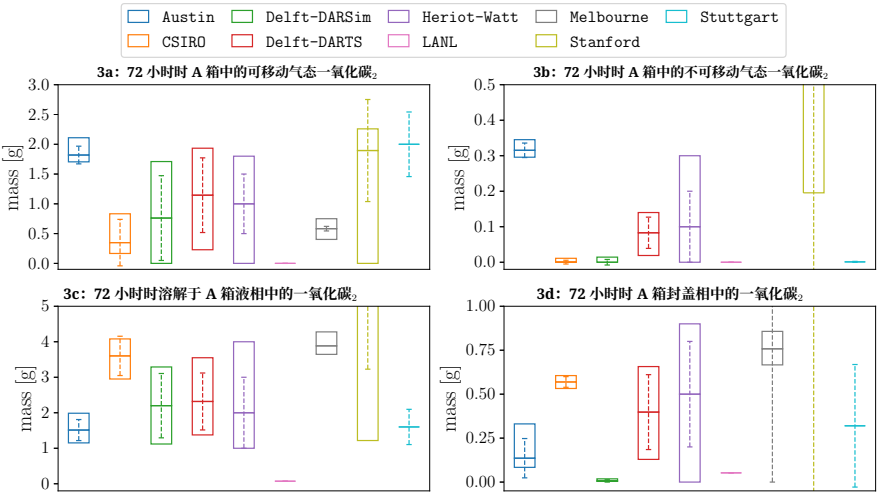

图 22 注入开始 72 小时后 A 箱内相分布的稀疏数据。方框底部、中部和顶部的水平线分别表示报告的期望均值的 P10、P50 和 P90 值。竖直虚线从均值处延伸，长度为 ± 报告的期望标准差的 P50。

4.3.4 A 箱内顶封盖相中的 CO_2 总质量

作为最后一项可测量指标，我们考察最终时刻 A 箱顶封盖相中的一氧化碳₂总质量，以评估捕获迁移进入低渗透性盖层的能力。图24 展示了相应的报告结果。这里同样可以观察到较大差异，不仅预期均值存在差异，期望标准差也存在较大差异。

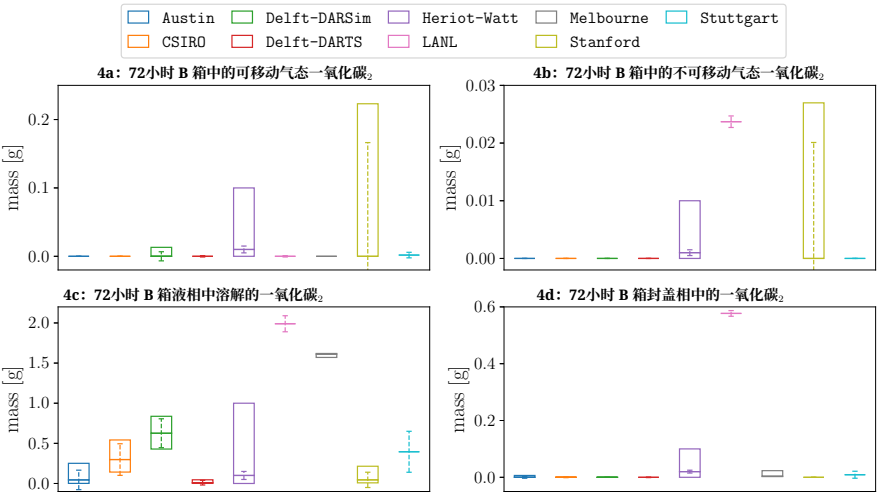

图 23 注入开始后 72 小时，B 箱中相分布的稀疏数据报告值。箱体底部、中部和顶部的水平线分别表示期望均值的 P10、P50 和 P90 报告值。虚线垂直线从均值延伸，长度为 ± 期望标准差的 P50 报告值。

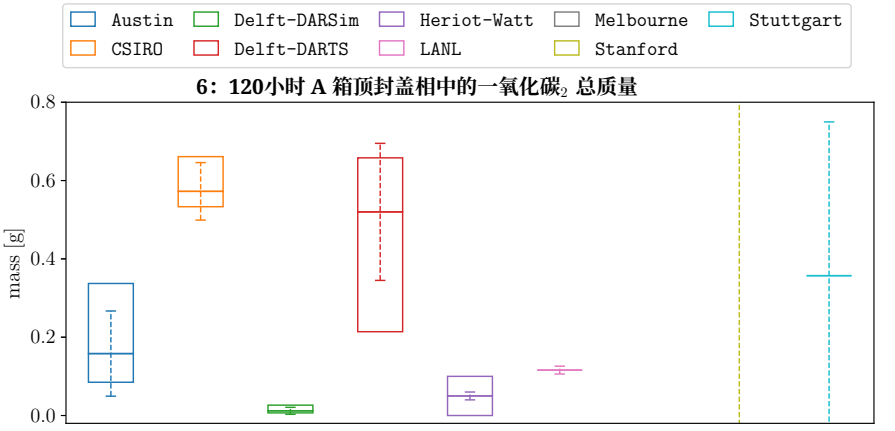

图 24 最终时刻，A 箱顶封盖相中一氧化碳的总质量稀疏数据报告值。箱体底部、中部和顶部的水平线分别表示期望均值的 P10、P50 和 P90 报告值。虚线垂直线从均值延伸，长度为 ± 期望标准差的 P50 报告值。

5 与实验数据比较

下面，我们将把上一节所述的建模结果与实际观测到的实验数据进行比较。相关实验方法和原始数据集介绍于 [24]，而图像

分析方法将在 [54]中讨论。我们首先关注密集数据空间分布图和时间序列，随后研究稀疏数据可测量量。

5.1 密集数据空间分布图

下文将比较以分割数据形式呈现的每日空间图。对于这些实验，这些数据是通过使用新开发的工具箱 DarSIA 分析相应图像生成的 [54]。图25展示了五次实验运行在24小时后的快照。从视觉上看，

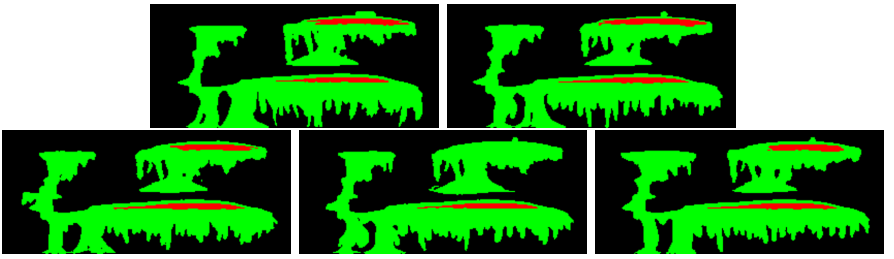

图 25 五次实验运行24小时后的分割数据。黑色、绿色和红色分别表示纯水、含有溶解的一氧化碳的水和气体。

五次实验运行的结果非常一致，只有细节上存在差异。一个小小的例外是第四次运行，其区域右上部似乎没有气体。不过，这是图像分析过程中的数值效应所致，而非物理真实性有所不同。我们将在下文进一步进行定量分析。

在此之前，先对建模结果进行直观比较。为此，参与者提供的 24 小时浓度图和饱和度图被转换为分割数据。为了便于更直接地比较，在建模结果上叠加了与实验数据相对应的等值线。结果如图 26 所示。可以看出，多个模型较好地捕捉到了两个气体羽流的位置，具体包括 奥斯汀、CSIRO、Delft-DARSim、Delft-DARTS、墨尔本和斯图加特，但整体而言，这些模型对气体羽流大小的估计偏大。正如第 4.1 节讨论的浓度分布存在很大差异所表明的，对于溶有一氧化碳的水所覆盖的区域，各模型之间的一致性明显较低$_2$。这一点在区域左上方的 B 箱尤为明显，只有 CSIRO 的建模结果在形状和延伸范围上与实验结果基本吻合，视觉效果令人满意。

为了更定量地理解这一点，可以按照 4.1.3 节中的类似分析，采用 Wasserstein度量进行分析。这需要计算所有由两个参与组、两次实验运行，或一名参与者与一次实验运行组成的配对之间的距离。与上文类似，可以计算各分割图与其他建模结果之间的均值距离，以及现在还包括与实验数据之间的均值距离，从而为每个分割图得到两个值

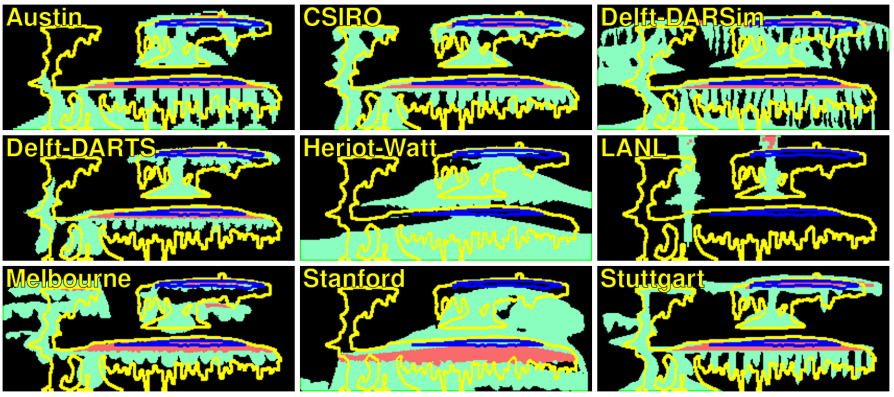

图 26 24小时后的分割数据比较。每个建模结果均叠加了实验运行2的等值线。预测结果分别以黑色、浅绿色和浅红色显示，分别表示纯水、含溶解的一氧化碳的水和气体。对于实验数据，黄色等值线表示含溶解的一氧化碳的水所在区域，而蓝线表示气体羽流。

。图 27绘出了所选时间步下所有分割图对应的这些值。可以看出，实验数据集之间的距离在 50 gr.cm 以内，这证实了实验重复性良好，而且不同实验条件的影响很小（主要归因于大气压力、实验装置内部的一些化学变化，以及实验期间沉降的沙粒量极少）。约有一半的建模结果在所有报告时间内与实验数据的距离约为 100 gr.cm，我们认为这是相对较好的匹配。在最终时间，最接近实验数据的模拟结果与实验均值的距离仅为 50 gr.cm，这在该时间点实验变异性的两倍范围内。这也与上文对分割图像的视觉观察相符。随着时间推移，大多数建模结果与实验数据及预测之间的距离都在减小；实验数据集与预测之间的距离也同样如此。这可以用富含一氧化碳$_2$的水在整个区域内扩散范围不断扩大，以及一氧化碳$_2$ 质量随之趋于平衡来解释。

5.2 密集数据时间序列

下面，我们将参与组报告的选定密集数据时间序列与相应的实验数据进行比较。如 [24, 54], 所述，根据实验照片推导饱和度和浓度值是一项极具挑战性的工作，需要建立多项假设。相应计算出的质量值存在显著不确定性。因此，必须非常谨慎地看待这些比较背后的物理真实性。

图 28展示了 BoxA 中相分布时间演化的比较。为了能够观察所研究时间范围开始阶段的更多细节

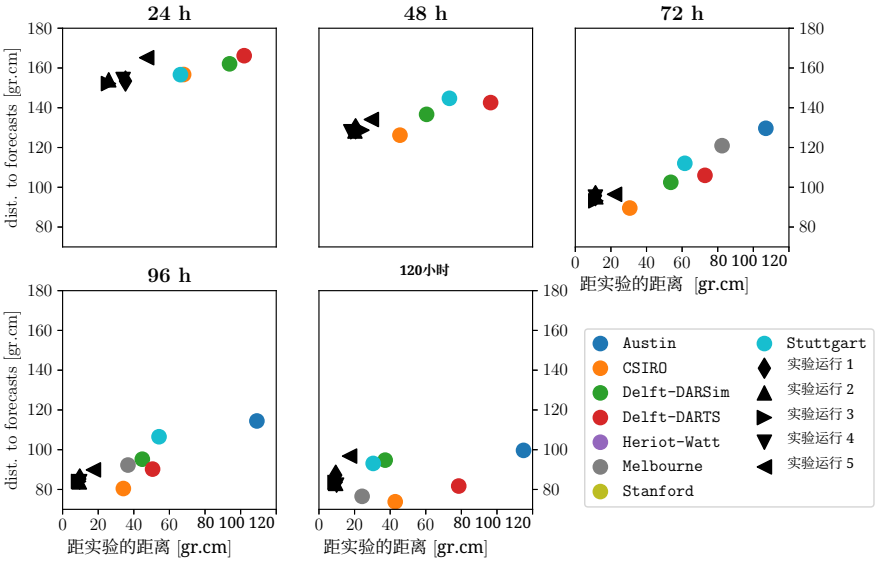

图 27 分割图与实验结果及预测的 Wasserstein 距离。均值与实验结果的距离图放大显示 0 至 120 gr.cm 的范围，均值与建模预测的距离图放大显示 70 至 180 gr.cm 的范围。因此，某些结果离群的组别并未出现在所有图中，而 Heriot-Watt 和 Stanford 始终位于图表范围之外（参见图12）。

，图中的 x 轴采用对数缩放。对于流动的一氧化碳气体$_2$，而言，实验均值的基本形态与建模结果的中位数相当相似。不过，预测的峰值明显低于实验值。在对流驱动阶段，数值不断升高时，建模结果的离散程度远小于随后溶解驱动阶段数值下降时的离散程度。因此，数值保持相对稳定的时间段要长得多。总体而言，数值上升和下降阶段均滞后于实验结果，但CSIRO 和 斯图加特 的结果与第一阶段吻合得非常好。

聚焦于溶解的一氧化碳$_2$ 的质量随时间的变化，可以看出，大多数建模结果在初期与实验数据吻合良好。注入停止后，预测结果的离散程度开始增大，此前讨论过的溶解行为差异也逐渐占据主导。在模拟的大部分时间里，大多数建模结果都低估了溶解的一氧化碳$_2$ 的量，但其数值往往比相应的实验数据持续更长时间地增加，而实验数据则较早达到饱和。观察第三幅图可见，各参与组所预测的盖层中一氧化碳$_2$ 质量变化差异很大，并且与实验数据存在显著差别。实验均值呈现非单调行为的一个可能原因将在 [24]中讨论。

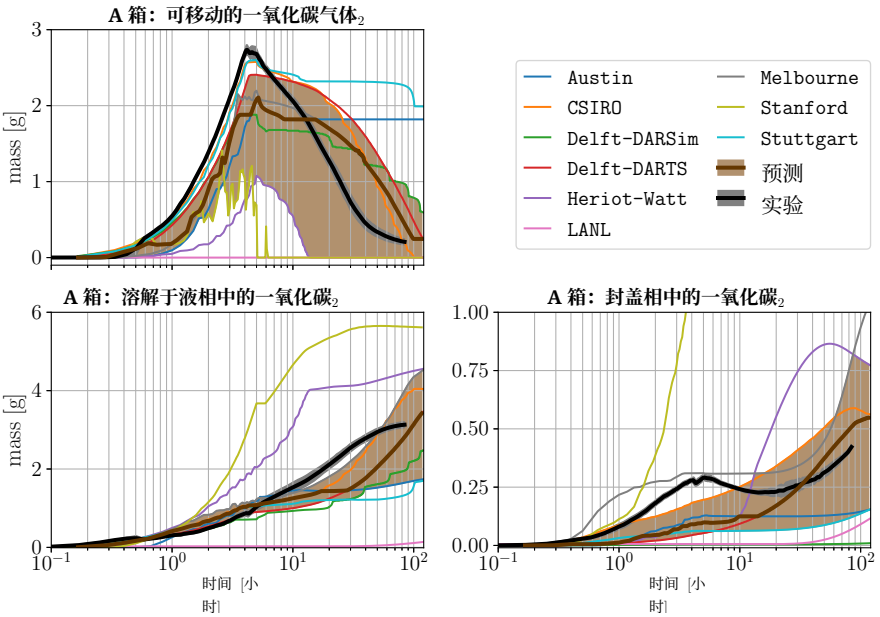

图 28 建模预测与实验观测之间的比较：Box A 中一氧化碳$_2$ 相分布的时间演化。棕色线表示所报告建模结果的中位数，相关的浅棕色区域表示对应第一四分位数与第三四分位数之间的范围。黑色线表示实验数据的均值，相关的灰色区域通过标准差表示相应的变化范围。

另外两条时间序列也提供了实验数据，相应的比较见图29。首先来看 B 箱中溶解的一氧化碳$_2$ 的量，图中较大的离散范围同样显示出建模结果存在显著差异。与 A 箱类似，有两个参与组准确捕捉到了初期由平流驱动的增加趋势。这里的差异也在注入停止后变得更加明显。由于富含一氧化碳$_2$ 的水从右侧进入 B 箱，实验数据中的一氧化碳$_2$含量随时间继续增加。大多数模型未能捕捉到这一效应。

最后，我们研究右图中图 $M(t)$ 所示对流测量值的时间演化 29。然而，建模结果与实验数据之间的差异过大，无法在此得出任何有意义的结论。这很可能与积分值的数值计算并不直接明了、强烈依赖于离散化，并且完全交由参与者处理这一事实有关。

5.3 稀疏数据

在收集稀疏数据结果的同时，还通过问卷调查来监测每位参与者对自己预测结果的信心

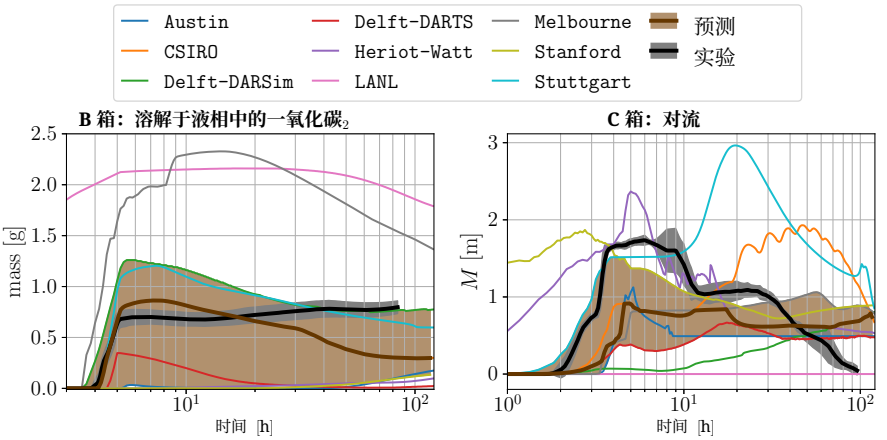

图 29 建模预测与实验观测对 B 箱中溶解的一氧化碳$_2$ 质量的时间演化（左）以及积分量 $M(t)$（右）的比较。棕色线表示报告的建模结果的中位数，相关的浅棕色区域表示对应的第一四分位数与第三四分位数之间的范围。黑色线表示实验数据的均值，相关的灰色区域表示以标准差衡量的相应变化。

以及其他各工作组的预测结果。由于对这一过程及其结果的描述和分析超出了本文的范围，因此另有一篇论文专门对此进行阐述 [55]。因此，在接下来与实验数据的比较中，我们仅对少数汇总指标作较为简要的介绍。

为概括第 4.3 节所呈现的各参与组的单次实验结果，我们仅考虑所报告的预期均值和标准差的 P50 值。均值将作为单独的数据点绘制，并同时呈现其均值以及预期标准差的中位数。对于实验数据，将绘制各单次实验的结果，并同时呈现其均值和标准差。

在图 30 中，我们首先考察两个传感器中可测量的预期和观测最大压力。正如大多数参与组所预测的那样，注入一氧化碳$_2$ 对两个传感器所观测到的压力几乎没有影响。报告的实测实验值对应于相应单次实验期间的最大大气压力，加上相应上覆水柱产生的静水压力贡献。各组单独报告的预期均值均在实验均值的 10 mbar 范围内，且预期均值的中位数与实验均值非常吻合。然而，正如第 4.3.1 节已经指出的那样，参与者预计实验结果几乎不会发生变化。由于大气压力的自然波动，观测到的变化最终明显大于预期变化。

图 31展示了可测量量 2 和 5 的比较，即 A 箱中可动气相达到最大值的时间，以及 $M(t)$

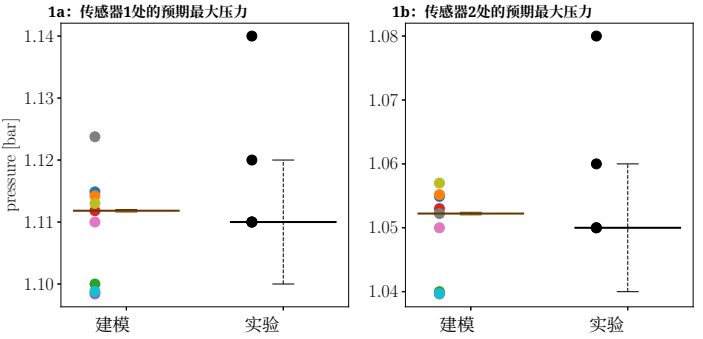

图 30 参与组报告的稀疏数据与可测量量 1 的实验数据比较。对于建模结果，彩色圆圈表示各次实验的预期均值，水平棕色线表示其中位数。一条棕色虚竖线从该值延伸，长度为 ± 所有报告的标准差 P50 值的中位数。某些异常值可能超出绘图范围。对于实验数据，黑色圆圈表示各次单次实验的结果，水平黑色线表示其均值。一条黑色虚竖线从均值延伸，长度为 ± 标准差。

分别超过 C 箱宽度的 110%。关于前者，可以

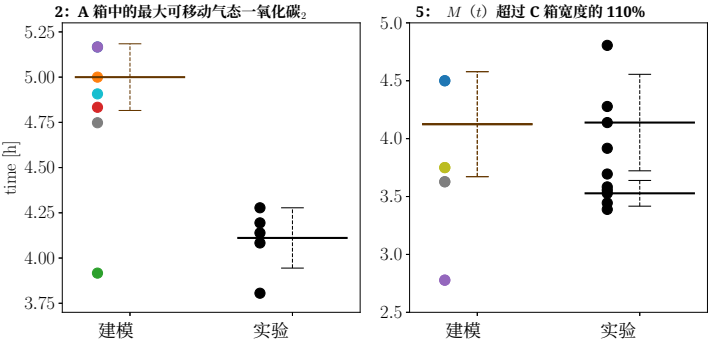

图 31 参与组报告的稀疏数据与可测量量 2（左）和 5（右）的实验数据比较。有关所绘制量的更多详细信息，请参见图 30 。

可以观察到，大多数参与组都高估了实验均值，且报告值范围与观测值范围基本不重叠。对于后者，情况有所不同，因为提供了两组实验数据，它们采用的图像分析参数不同，分别构成目标量的上限和下限。在这里，预期均值的中位数接近相应的实验均值上限。

接下来，我们将比较稀疏数据可测量量 3a、3c、4a 和 4c，考察 CO 的相分布$_2$ 在 A 箱和 B 箱中分别经过 72 小时后的情况。图 32以 CO_2在气相或液相中的质量表示相应的量。从 3a 开始，可以观察到

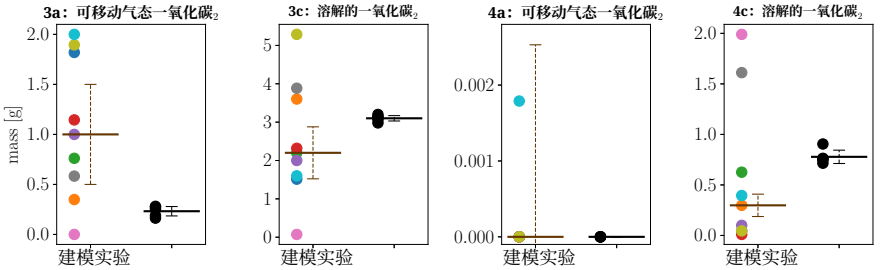

图32 参与组报告的稀疏数据与可测量量3a、3c、4a和4c的实验数据比较（从左至右）。有关图示量的更多详细信息，请参见图30 。

A 箱中流动气态 CO 的均值$_2$ 被大多数参与组高估，只有部分参与组报告的数值处于观测实验范围内。这与第 5.1节中讨论的视觉印象一致。对于 3c，A 箱中溶解于水中的 CO 的均值$_2$ 则被建模人员明显低估。转向 B 箱，所有实验运行结果都表明，经过 72 小时后已没有气态 CO_2 残留。这也得到了大多数参与者的预期，但他们仍然预计其平均标准差略大于零。尽管所报告的溶解的一氧化碳期望均值$_2$ 分布相当广泛，但其中位数却与观测实验均值非常接近。

对于最后一项可测量量6，我们考察模拟结束时 A 箱顶封盖相中的一氧化碳$_2$ 总质量，如图33所示。

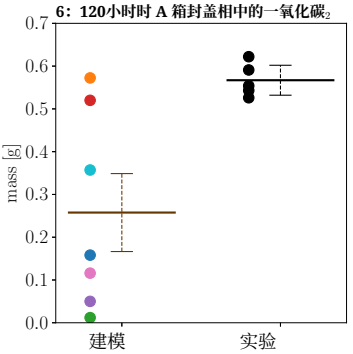

图33 参与组报告的稀疏数据与可测量量6的实验数据比较。有关图示量的更多详细信息，请参见图30 。

预期均值的中位数约为观测实验均值的 50%。相应地，大多数参与组低估了顶封盖相中的一氧化碳$_2$ 含量。然而，有两个参与组的结果与实验结果非常接近。

6 结论与展望

下文将总结本验证基准研究得出的若干结论，并介绍后续工作的挑战与机遇。

首先，我们可以非常有把握地指出，达西尺度平衡方程与标准的毛细压力和相对渗透率本构关系相结合，能够充分描述在所考虑的空间和时间尺度上观测到的相关物理过程。将模拟饱和度和浓度分布与相应的实验分割图进行比较，清楚地证明了这一点。尤其是，大多数参与组都很好地捕捉到了地层圈闭和残余圈闭机制。此外，对流混合过程因密度差异而发生，这一现象也得到了定性上的合理考虑。

从定量角度看，建模结果差异较大，尤其是在溶解行为及由此产生的指进方面。这可归因于不同的建模选择，例如对 CO_2 在水中溶解度上限的设定，以及毛细压力—饱和度关系、用于确定相组成的状态方程或相密度计算等本构关系的选择。还可以观察到，网格分辨率的差异明显影响对流混合行为。尽管如此，若干参与组的结果与实验结果十分接近，Wasserstein度量对此进行了量化。随着时间推移，越来越多的 CO_2 溶解，其质量在整个区域内趋于平衡，相应的距离也随之减小。

本研究要求报告预先定义的“稀疏数据”，这些数据是可用于表征储存能力和储存安全性不同方面的代理指标。报告内容既包括最可能的超越值（P50），也包括 P10-P90 区间。虽然 P50 值大多再现了所报告的密集数据，但 P10-P90 值为结果增加了一个维度。值得注意的是，对于大多数要求报告的指标，各组报告的 P10-P90 区间彼此不重叠。从逻辑上讲，如果两个 P10-P90 区间不重叠，那么一组认为另一组得出的实验结果落在其所报告区间内的概率至多为 10%（反之亦然）。这表明，尽管各组在研究过程中进行了大量互动，但它们并未认真考虑其他组的定量响应，而是对自己的结果抱有高度信心，甚至完全确信。与此相辅相成的是，互动很好地帮助几乎所有组就预期的定性行为（例如毛细屏障的影响）达成了共识。

本研究揭示的一个特别关键的物理过程（在稀疏数据和密集数据中均有体现）是对流混合在加速气态一氧化碳溶解方面所发挥的作用$_2$。这一点既通过 A 箱和 B 箱中的实际相组成进行了量化，也通过指标 M进行了量化；该指标可作为完全发育指状流的时间的代理指标（关于数值模拟中密度驱动指状流的各种起始时间，详见 [56]。起始

对于该系统而言，对流指进的演化尤其难以模拟，因为本研究采用的低阶数值方法（适用于捕捉非均质性并实现多相流的稳定离散化）在表示气水界面时往往过于弥散。这导致气相向水相的传质量被显著高估，因此需要在垂直方向采用精细网格。此外，该系统中密度驱动指进的特征波长约为 5 厘米（实验观测结果），这进一步要求水平方向的网格分辨率小于 1 厘米。综合来看，这可能是参与者报告的密度驱动指进形态及其重要性差异较大的原因，也促使我们进一步研究如何利用油藏模拟工具可靠、准确地捕捉这一过程。

虽然本研究是在实验室尺度开展的，但非均质多孔介质中多相多组分流的基本物理过程与油藏条件下的相同。因此，我们认为，本研究的发现和观察结果也能反映油田尺度模拟的情况（详细的尺度分析见本卷 Kovscek 等人的文章）。尽管如此，实际的油田尺度模拟在若干重要方面会与本研究有所不同，以下重点介绍其中几点：

- 非均质性。本研究采用均质相（在实验室条件下尽可能实现），重点关注较大尺度的结构非均质性。在场地尺度上，预计各地质结构内部也会存在显著的亚尺度非均质性。
- 地质表征质量。本研究采用准二维几何，并得到了较好的表征（实验开始时进行了高分辨率摄影和厚度测量）。在场地尺度上，地质表征依据地震勘测，而地震勘测无法提供同等精度。
- 维度。现实世界是三维的，这会影响模拟时间，进而间接影响可实现的网格细化程度。
- 对流混合。在油田尺度模拟中，为准确捕捉对流混合所需的空间和时间分辨率，在实际中难以实现。
- 压力和温度条件。在实验室条件下，一氧化碳$_2$ 以气相存在；而在场地尺度上，通常寻求压力和温度条件与超临界一氧化碳$_2$ 相适配的储层。这对黏度的影响较小，但会形成密度更高、可压缩性更低的一氧化碳$_2$ 相。

与深部储层条件相比，真正截然不同的是压力测量的重要性。在实验中，压力信号提供的信息相当有限，还可能导致对渗透率的解读出现差异；而在储层环境中，压力信号则很有价值。另一个重要考量是，地下环境比实验装置更难表征，因此预测不确定性将主要由地质表征的不确定性决定。这项代码比较阐明了

在特征相对明确的系统中，各种可能预测结果的范围。

从油藏模拟的角度来看，所有参与者都报告称，他们难以达到可接受的运行时间，因此不得不在本研究中使用相对粗糙的网格。我们推测，这是由于气相密度较低所致，其结果是，当一氧化碳$_2$ 溶解于水中时，所得混合物的体积显著小于混合前的体积。因此，本研究推动了可溶性气体-水体系高效非线性求解器的进一步开发。

致谢。 B. Flemisch 感谢德国研究基金会（DFG）通过资助 SFB 1313（项目编号327154368）支持本研究。S. Geiger 感谢 Energi Simulation 提供的部分资助。H. Hajibeygi 获得荷兰科学研究组织（NWO）Vidi人才计划项目“ADMIRE”（项目编号17509）的资助。

参考文献

[1]Bachu, S., Bonijoly, D., Bradshaw, J., Burruss, R., Holloway, S., Christensen, N.P., Mathiassen, O.M.: 一氧化碳$_2$ 储存容量估算：方法与不足。《国际温室气体控制杂志》**1**(4), 430–443 (2007)。https://doi.org/10.1016/S1750-5836(07)00086-2[2] Pacala, S., Socolow, R.:稳定楔：利用现有技术解决未来50年的气候问题。《科学》**305**(5686), 968– 972 (2004)。https://doi.org/10.1126/science.1100103[3]Halland, E.K., Riis, F., Magnus, C., Johansen, W.T., Tappel, I.M., Gjeldvik, I.T., Solbakk, T., Pham, V.T.H.: 一氧化碳$_2$ 挪威北海区域储存图集。《能源学报》**37**, 4919–4926 (2013)。https://doi.org/10.1016/j.egypro.2013.06.403[4] Metz, B., Davidson, O., De Coninck, H., Loos, M., Meyer, L.: 政府间气候变化专门委员会《二氧化碳捕集与封存特别报告》。剑桥大学出版社，剑桥（2005）[5] Johnson, N., Parker, N., Ogden, J.: 配备碳捕集与封存的生物燃料能实现多大程度的负排放？成本又是多少？运用空间显式建模，重新评估配备碳捕集与封存的生物燃料生产的经济潜力。《能源学报》**63**, 6770–6791 (2014)。https://doi.org/10.1016/j.egypro.2014.11.712[6] Noussan, M., Raimondi, P.P., Scita, R., Hafner, M.: 绿色氢和蓝氢在能源转型中的作用——技术与地缘政治视角。《可持续性》**13**(1)（2021）。https://doi.org/10.3390/su13010298

36 *FluidFlower* 国际基准研究[7] Juanes, R., MacMinn, C.W., Szulczewski, M.L.：咸水含水层中二氧化碳封存期间一氧化碳$_2$羽流的范围：盆地尺度毛细管圈闭的封存效率。《多孔介质中的传输》**82(1)**, 19–30（2010）。https://doi.org/10.1007/s11242-009-9420-3

[8] Lindeberg, E., Vuillaume, J.-F., Ghaderi, A.：一氧化碳$_2$乌齐拉地层的封存容量。《能源》论文集**1(1)**, 2777–2784（2009）。https://doi.org/10.1016/j.egypro.2009.02.049

[9] Kopp, A., Class, H., Helmig, R.：咸水含水层中的一氧化碳$_2$ 封存容量研究：第**1**部分。流动过程和储层特征的量纲分析。《国际温室气体控制杂志》**3(3)**, 263–276（2009）。https://doi.org/10.1016/j.ijggc.2008.10.002[10] Kopp, A., Class, H., Helmig, R.：咸水含水层中的一氧化碳$_2$ 封存容量研究——第2部分：封存容量系数的估算。《国际温室气体控制杂志》**3(3)**, 277–287（2009）。https://doi.org/10.1016/j.ijggc.2008.10.001

[11] Niemi, A., Bensabat, J., Shtivelman, V., Edlmann, K., Gouze, P., Luquot, L., Hingerl, F., Benson, S.M., Pezard, P.A., Rasmusson, K., Liang, T., Fagerlund, F., Gendler, M., Goldberg, I., Tatomir, A., Lange, T., Sauter, M., Freifeld, B.：Heletz 实验场地中一氧化碳$_2$ 注入和地质封存的概述、表征与数据分析。《国际温室气体控制期刊》**48**，3–23（2016）。https://doi.org/10.1016/j.ijggc.2015.12.030

[12] Sharma, S., Cook, P., Jenkins, C., Steeper, T., Lees, M., Ranasinghe, N.：《CO2CRC 奥特韦项目：借鉴经验，探索澳大利亚首个碳捕集与封存项目场址的新机遇》。《能源学报》**4**, 5447–5454（2011年）。https://doi.org/10.1016/j.egypro.2011.02.530[13] Scheer, D., Class, H., Flemisch, B.：《科学与政策之间的地下环境建模》。施普林格，查姆（2021年）[14]Steyn, M., Oglesby, J., Turan, G., Zapantis, A., Gebremedhin, R., Zapantis, A., Amer, N.A., Havercroft, I., Ivory-Moore, R., Steyn, M., Yang, X., Gebremedhin, R., Zahra, M.A., Pinto, E., Ras- sool, D., Williams, E., Consoli, C., Minervini, J.：《2022年全球碳捕集与封存现状》（2022年）。
h
t
t
ps://status22.globalccsinstitute.com/wp-content/uploads/2022/11/Global-Status-of-CCS-2022 Download.pdf[15] Preston, C., Monea, M., Jazrawi, W., Brown, K., Whittaker, S., White, D., Law, D., Chalaturnyk, R., Rostron, B.：IEA GHG Weyburn 二氧化碳$_2$监测与封存项目。《燃料加工技术》**86**（14），1547–1568（2005年）。https://doi.org/10.1016/j.fuproc.2005.01.019

[16] L¨uth, S., Henninges,J.,Ivandic, M., Juhlin, C., Kempka, T., Norden, B., Rippe, D., Schmidt-Hattenberger, C.：第6.2章——Ketzin试验场注入和封闭后阶段的地球物理监测。载于：Kasahara, J., Zhdanov, M.S., Mikada, H.（编），《主动地球物理监测》，第2版，第523–561页。爱思唯尔，阿姆斯特丹（2020）。https://doi.org/10.1016/B978-0-08-102684-7.00025-X

[17] Pruess, K., Garciia, J., Kovscek, T., Oldenburg, C.,Rutqvist, J., Steefel, C., Xu, T.：代码互比增强了人们对用于CO地质处置的数值模拟模型的信心$_2$。《能源》**29**(9–10), 1431–1444（2004）。https://doi.org/10.1016/j.energy.2004.03.077

[18] Class, H., Ebigbo, A., Helmig, R., Dahle, H.K., Nordbotten, J.M., Celia, M.A., Audigane, P., Darcis, M., Ennis-King, J., Fan, Y., Flemisch, B., Gasda, S.E., Jin, M., Krug, S., Labregere, D., Naderi Beni, A., Pawar, R.J., Sbai, A., Thomas, S.G., Trenty, L., Wei, L.：关于一氧化碳$_2$ 地质构造封存相关问题的基准研究。《计算地球科学》**13**(4), 409–434（2009）。https://doi.org/10.1007/s10596-009-9146-x[19]Nordbotten, J.M., Flemisch, B., Gasda, S.E., Nilsen, H.M., Fan, Y., Pickup, G.E., Wiese, B., Celia, M.A., Dahle, H.K., Eigestad, G.T., Pruess, K.：实际模拟一氧化碳$_2$ 封存过程中的不确定性。《国际温室气体控制杂志》**9**，234–242（2012年）。https://doi.org/10.1016/j.ijggc.2012.03.007[20]Oberkampf, W.L., Roy, C.J.：《科学计算中的验证与确认》。剑桥大学出版社，剑桥（2010）[21] 美国机械工程师学会：《计算固体力学中的验证与确认指南：美国国家标准》。ASME出版社，纽约（2006年）。https://www.asme.org/products/codes-standards/v-v-10-2006-guide-verification-validation[22] Oberkampf, W.L., Trucano, T.G.：验证与确认基准。《核工程与设计》**238**(3), 716–743（2008年）。https://doi.org/10.1016/j.nucengdes.2007.02.032[23]Nordbotten, J.M., Fernø, M., Flemisch, B., Juanes, R., Jørgensen, M.：最终基准描述：FluidFlower国际基准研究。（2022年）。https://doi.org/10.5281/zenodo.6807102[24] Fernø, M.A., Haugen, M., Eikehaug, K., Folkvord, O., Benali, B., Nord- botten, J.M.：米尺度一氧化碳$_2$ 注入实验与数据集。《多孔介质中的传输》，本期专辑（2023年）[25] Peng, D.-Y., Robinson, D.B.：一种新的双常数状态方程。

《工业与工程化学基础》**15**(1), 59–64 (1976)。
https://doi.org/10.1021/i160057a011

[26] Spycher, N., Pruess,K., Ennis-King, J.: 一氧化碳$_2$-H_2O 混合物在地质封存一氧化碳$_2$中的应用。I. 评估并计算 12 至 100°C、最高 600 bar 条件下的互溶性。《地球化学与宇宙化学学报》**67**(16), 3015–3031 (2003)。
https://doi.org/10.1016/S0016-7037(03)00273-4

[27] Spycher, N., Pruess, K.: 一氧化碳$_2$-H_2O 混合物在地质封存一氧化碳$_2$中的应用。II. 12–100°C、最高 600 bar 条件下在氯化物卤水中的分配。《地球化学与宇宙化学学报》**69**(13), 3309–3320 (2005)。https://doi.org/10.1016/j.gca.2005.01.015

[28] Ziabakhsh-Ganji, Z., Kooi, H.：一种用于气体混合物和卤水热力学平衡的状态方程，以实现对地下二氧化碳杂质影响的模拟$_2$ 储存。《International Journal of Green- house GasControl》**11**，21–34（2012年）。
https://doi.org/10.1016/j.ijggc.2012.07.025

[29] Duan, Z., Sun，R.：一种改进的模型，用于计算273至533 K、0至2000 bar条件下二氧化碳$_2$ 在纯水和氯化钠水溶液中的溶解度。《化学地质学》**193**（3），257–271（2003年）。https://doi.org/10.1016/S0009-2541(02)00263-2

[30] Span, R., Wagner, W.：一种新的二氧化碳状态方程，涵盖从三相点温度至1100 K、压力最高达800 MPa的流体区域。《物理与化学参考数据》**25**, 1509–1596（1996年）。https://doi.org/10.1063/1.555991

[31] Span, R., Wagner, W.：《技术应用状态方程》。I. 非极性流体和极性流体的同时优化函数形式。《国际热物理学杂志》**24**(1), 1–39（2003年）。
https://doi.org/10.1023/A:1022390430888[32] Weiss, R.F.：水和海水中的二氧化碳：非理想气体的溶解度。《海洋化学》**2**(3), 203–215（1974年）。
https://doi.org/10.1016/0304-4203(74)90015-2[33]Sandve, T.H., Gasda, S.E., Rasmussen, A., Rustad, A.B.：使用 OPM flow 模拟器进行场地尺度模拟中的对流溶解。载于：TCCS–11。CO_2 捕集、运输与封存。特隆赫姆，2021年6月22日至23日，第11届特隆赫姆国际碳捕集与封存会议论文集（2021年）。SINTEF 学术出版社[34] IAPWS：关于 IAPWS 1997 工业公式中水和蒸汽热力学性质的修订发布文件。技术报告，

IAPWS（2007年）。www.iapws.org/relguide/IF97-Rev.pdf[35] Garcia, J.E.：CO_2水溶液的密度。技术报告 LBNL-49023，LBNL（2001年）[36] Soave, G.：基于修正的 Redlich-Kwong状态方程的平衡常数。《化学工程科学》27(6)，1197–1203（1972年）。https: // doi.org/10.1016/0009-2509(72)80096-4[37] Carroll, J.J., Slupsky, J.D., Mather, A.E.：低压下二氧化碳在水中的溶解度。《物理与化学参考数据杂志》20(6)，1201–1209（1991年）。https://doi.org/10.1063/1.555900[38] Fenghour, A., Wakeham, W.A., Vesovic, V.：二氧化碳的黏度。《物理与化学参考数据杂志》27(1)，31–44（1998年）。https://doi.org/10.1063/1.556013[39] Wilkins, A., Green, C.P., Ennis-King, J.：用于多孔介质中耦合问题的开源多物理场模拟代码。《计算机与地球科学》154，104820（2021年）。https://doi.org/10.1016/j.cageo.2021. 104820[40] Wang, Y., Vuik, C., Hajibeygi, H.：深部咸水含水层中一氧化碳$_2$ 全周期注入和迁移期间流体动力学捕集相互作用的分析。《水资源进展》159，104073（2022年）。https://doi. org/10.1016/j.advwatres.2021.104073[41] Lyu, X., Khait, M., Voskov, D.：基于算子的线性化方法，用于模拟浮力和毛细作用影响下的多相流。《SPE期刊》26(4)，1858–1878（2021年）。https://doi.org/10.2118/205378-PA[42]Lie, K.-A.：《使用MATLAB/GNU Octave进行油藏模拟导论：MATLAB油藏模拟工具箱（MRST）用户指南》。剑桥大学出版社，剑桥（2019年）。https: // doi.org/10.1017/9781108591416[43] Zyvoloski, G.A., Robinson, B.A., Dash, Z.V., Trease, L.L.：FEHM应用的模型和方法概要——一种有限元热质传递代码。OSTI（1997年）。https://doi.org/10.2172/14903[44]Lichtner, P.C., Hammond, G.E., Lu, C., Karra, S., Bisht, G., Andre, B., Mills, R., Kumar, J.：PFLOTRAN用户手册：用于描述地表与地下过程的海量并行反应性流动与传输模型。OSTI（2015年）。https://doi.org/10.2172/1168703[45] Matth¨ai, S., Geiger, S., Roberts, S.：复杂系统平台：Csp3d3.0：用户指南。技术报告，苏黎世联邦理工学院（2001年）

[46] Zhou, Y., Jiang, Y., Tchelepi, H.A.：一种用于含多段井油藏模型的可扩展多阶段线性求解器。《计算地球科学》**17**(2), 197–216 (2013)。
https://doi.org/10.1007/s10596-012-9324-0

[47] Garipov, T.T., Tomin, P., Rin, R., Voskov, D.V., Tchelepi, H.A.：用于油藏模拟的统一热-组分-力学框架。《计算地球科学》**22**(4), 1039–1057 (2018)。https://doi.org/10.1007/s10596-018-9737-5

[48] Zhou, Y.：耦合油藏模型与多段井的并行通用油藏模拟。博士论文，斯坦福大学（2012年）

[49] Younis, R., Tchelepi, H.A., Aziz, K.：自适应局部化延续牛顿法——一种始终能够收敛的非线性求解器。《SPE期刊》**15**(02),526–544 (2010).
https://doi.org/10.2118/119147-PA

[50] Koch, T., Glaeser, D., Weishaupt, K., Ackermann, S., Beck, M., Becker, B., Burbulla, S., Class, H., Coltman, E., Emmert, S., Fetzer, T., Gru- eninger, C., Heck, K., Hommel, J., Kurz, T., Lipp, M., Mohammadi, F., Scherrer, S., Schneider, M., Seitz, G., Stadler, L., Utz, M., Weinhardt, F., Flemisch, B.：《DuMux 3——一款用于求解多孔介质中流动与传输问题、侧重模型耦合的开源模拟器》。《计算机与数学及其应用》**81**，423–443（2021年）。
https://doi.org/10.1016/j.camwa.2020.02.012[51] Panaretos, V.M., Zemel, Y.：Wasserstein 距离的统计学方面。《统计学与应用年度评论》**6**(1), 405–431（2019年）。
https://doi.org/10.1146/annurev-statistics-030718-104938[52]Flamary, R., Courty, N., Gramfort, A., Alaya, M.Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gau- theron, L., Gayraud, N.T.H., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz,A.,Seguy, V., Sutherland, D.J., Tavenard, R., Tong, A., Vayer, T.：POT：Python 最优传输。《机器学习研究杂志》**22**(78)，1–8（2021年）
[53] Youssef, A.A., Shao, Q., Matth¨ai, S.K.：近地表条件下二氧化碳$_2$, g-水流动与捕集的简化数值模拟方法。《多孔介质中的传输》，本期特刊（2023年）[54]Nordbotten, J.M., Benali, B., Both, J.W., Brattek˚as, B., Storvik, E., Fernø, M.：多孔介质的双尺度图像处理。《多孔介质中的传输》，本期特刊（2023年）[55] Nordbotten, J.M., Jørgensen, M., Fernø, M., Flemisch, B., Juanes, R.：

预测地质碳封存时的过度自信偏差。准备中（2023年）

[56] Elenius, M.T.,Johannsen,K.：混相驱替中多孔介质流动非线性失稳的时间尺度。《计算地球科学》**16**（4），901–911（2012年）。
https://doi.org/10.1007/s10596-012-9294-2

声明与说明

资金支持

S. Geiger 感谢 Energi Simulation 提供的部分资助。H. Hajibeygi 获荷兰国家科学基金会（NWO）Vidi人才计划项目“ADMIRE”（项目编号17509）资助。

利益冲突

作者声明不存在需要披露的相关财务或非财务利益。

作者贡献

B. Flemisch、J. Nordbotten、M. Fernø 和 R. Juanes 共同构思、设计并实施了这项基准研究。所有其他作者均为参与组成员，并相应地搭建、执行和评估了模拟，提供了所要求的结果以及底层模型的描述。B. Flemisch 和 J. Nordbotten 撰写了手稿初稿，所有其他作者均参与了内部审阅和编辑工作。B. Flemisch 编写并运行了生成手稿中所有图的脚本。所有作者均已阅读并批准最终手稿。

数据可用性

本文用于生成图的所有数据均收集在 GitHub “FluidFlower”组织的相应代码库中，可通过 github.com/fluidflower访问。具体而言，参与组提供的结果收集在代码库 github. com/fluidflower/$groupname$.git中，其中 $groupname$ 为以下名称之一：奥斯汀、 CSIRO、代尔夫特、 赫瑞-瓦特、 洛斯阿拉莫斯国家实验室、 墨尔本、 斯坦福和 斯图加特。用于与建模结果进行比较的实验数据汇总于github.com/fluidflower/experiment。生成所有图的脚本均存放在 github.com/fluidflower/general。

附录A Wasserstein 距离

下表列出了参与组针对各指定时间步所提供空间图之间的 Wasserstein 距离。计算中使用了 Python 库 POT [52] 。完整数据包括

不同时间步结果之间的距离可在 FluidFlower 通用 GitHub 仓库中找到。要获得图12中所示的数值，需将以米为单位的归一化表格值乘以 850 克·厘米$^{-1}$，从而得到所需的克·厘米单位。数值 8.5 表示注入的一氧化碳$_2$ 的质量，单位为克。

A.1 24 小时

距离 [米]	奥斯汀	CSIRO	DARSim	DARTS	LANL	墨尔本	斯坦福	斯图加特
奥斯汀	0。00e0	1.97e − 1	2。25e − 1	1.39e − 1	7。12e − 1	2.50e − 1	3.19e − 1	1.87e − 1
CSIRO	1.97e − 1	0。00e0	7。62e − 2	1.01e − 1	5。27e − 1	1.18e − 1	4.93e − 1	3.71e − 2
DARSim	2。25e − 1	7.62e − 2	0。00e0	1.52秒 − 1	5.00e − 1	1.13e − 1	5.23e − 1	8.81e − 2
DARTS	1.39e − 1	1.01e − 1	1.52秒 − 1	0。00e0	5.91e − 1	1.40e − 1	4.28e − 1	1.18e − 1
LANL	7。12e − 1	5。27e − 1	5.00e − 1	5.91e − 1	0。00e0	4.85e − 1	1.02e0	5.33e − 1
墨尔本	2.50e − 1	1.18e − 1	1.13e − 1	1.40e − 1	4.85e − 1	0。00e0	5。35e − 1	1.51e − 1
斯坦福	3.19e − 1	4.93e − 1	5.23e − 1	4.28e − 1	1。02e0	5。35e − 1	0。00e0	4.90e − 1
斯图加特	1.87e − 1	3.71e − 2	8。81e − 2	1.18e − 1	5.33e − 1	1.51e − 1	4.90e − 1	0。00e0

A.2 48小时

距离 [m]	奥斯汀	CSIRO	DARSim	DARTS	LANL	墨尔本	斯坦福	斯图加特
奥斯汀	0。00e0	1.58e − 1	1.94e − 1	1.46e − 1	6.47e − 1	2.02e − 1	3。39e − 1	2。32e − 1
CSIRO	1.58e − 1	0。00e0	8.94e − 2	6.62e − 2	4。99e − 1	9.40e − 2	4.52e − 1	9。45e − 2
DARSim	1.94e − 1	8.94e − 2	0。00e0	1.17e − 1	4.60e − 1	9.46e − 2	4.92e − 1	1.13e − 1
DARTS	1.46e − 1	6.62e − 2	1.17 秒 − 1	0。00e0	5。08e − 1	1.13e − 1	4。43e − 1	1.12e − 1
LANL	6.47e − 1	4。99e − 1	4.60e − 1	5。08e − 1	0。00e0	4。58e − 1	9.49e − 1	4.20e − 1
墨尔本	2.02e − 1	9.40e − 2	9.46e − 2	1.13e − 1	4。58e − 1	0。00e0	4.92e − 1	1.36e − 1
斯坦福	3.39e − 1	4。52e − 1	4.92e − 1	4.43e − 1	9.49e − 1	4.92e − 1	0。00e0	5.35e − 1
斯图加特	2。32e − 1	9。45e − 2	1。13e − 1	1.12e − 1	4.20e − 1	1.36e − 1	5。35e − 1	0。00e0

A.3 72小时

距离 [米]	奥斯汀	CSIRO	DARSim	DARTS	LANL	墨尔本	斯坦福	斯图加特
奥斯汀	0。00e0	1.15e − 1	1.59e − 1	1.43e − 1	6。33e − 1	1.41e − 1	3。22e − 1	2.40e − 1
CSIRO	1.15e − 1	0。00e0	1.32秒 − 1	8。68e − 2	5.64e − 1	7。48e − 2	3.60e − 1	1.70e − 1
DARSim	1.59e − 1	1。32e − 1	0。00e0	1。32e − 1	4.83e − 1	9.89e − 2	4.37e − 1	1.46e − 1
DARTS	1.43e − 1	8。68e − 2	1.32秒 − 1	0。00e0	5.06e − 1	8.39e − 2	4。14e − 1	1.21e − 1
LANL	6。33e − 1	5.64e − 1	4.83e − 1	5.06e − 1	0。00e0	5。28e − 1	9.19e − 1	3.97e − 1
墨尔本	1.41e − 1	7.48e − 2	9。89e − 2	8.39e − 2	5。28e − 1	0。00e0	3.92秒 − 1	1.64e − 1
斯坦福	3.22e − 1	3.60秒 − 1	4。37e − 1	4.14e − 1	9.19e − 1	3。92e − 1	0。00e0	5.26e − 1
斯图加特	2.40e − 1	1.70e − 1	1.46e − 1	1.21e − 1	3.97e − 1	1.64e − 1	5.26e − 1	0。00e0

A.4 96小时

距离 [m]	奥斯汀	CSIRO	DARSim	DARTS	LANL	墨尔本	斯坦福	斯图加特
奥斯汀	0。00e0	1.08e − 1	1.37秒 − 1	1。33e − 1	6。14e − 1	1.10e − 1	3。12e − 1	2.28e − 1
CSIRO	1.08e − 1	0。00e0	1.59e − 1	9.81e − 2	6.10e − 1	6。81e − 2	3。33e − 1	2.25e − 1
DARSim	1.37e − 1	1.59e − 1	0。00e0	1.45e − 1	4。86e − 1	1。15e − 1	4.00e − 1	1.50秒 − 1
DARTS	1。33e − 1	9.81e − 2	1.45e − 1	0。00e0	5。27e − 1	7.38e − 2	3.61e − 1	1.43e − 1
LANL	6。14e − 1	6.10e − 1	4.86e − 1	5。27e − 1	0。00e0	5。74e − 1	8.86e − 1	3。87e − 1
墨尔本	1.10e − 1	6。81e − 2	1.15e − 1	7.38e − 2	5。74e − 1	0。00e0	3。14e − 1	1.96秒 − 1
斯坦福	3。12亿 − 1	3.33e − 1	4.00e − 1	3.61e − 1	8.86e − 1	3.14e − 1	0。00e0	5。01e − 1
斯图加特	2.28e − 1	2.25e − 1	1.50e − 1	1.43e − 1	3。87e − 1	1.96秒 − 1	5。01e − 1	0。00e0

A.5 120小时

距离 [m]	奥斯汀	CSIRO	DARSim	DARTS	LANL	墨尔本	斯坦福	斯图加特
奥斯汀	0。00e0	1.25e − 1	1.25e − 1	1.24e − 1	5.91e − 1	1.03e − 1	3.05e − 1	2.00e − 1
CSIRO	1.25e − 1	0.00e0	1.69e − 1	8.21秒 − 2	6.27e − 1	7.59e − 2	3.27e − 1	2.35e − 1
DARSim	1.25e − 1	1.69亿 − 1	0.00e0	1.57e − 1	4.85e − 1	1.30秒 − 1	3。69e − 1	1。39e − 1
DARTS	1.24e − 1	8.21秒 − 2	1.57秒 − 1	0.00e0	5。57e − 1	7。54e − 2	3.36e − 1	1.66e − 1
LANL	5.91e − 1	6.27e − 1	4.85e − 1	5。57e − 1	0.00e0	6。03e − 1	8.54秒 − 1	3。92e − 1
墨尔本	1.03e − 1	7.59e − 2	1.30e − 1	7。54e − 2	6。03e − 1	0.00e0	2。71e − 1	2。14e − 1
斯坦福	3.05e − 1	3。27e − 1	3。69e − 1	3.36e − 1	8。54e − 1	2.71e − 1	0.00e0	4。64e − 1
斯图加特	2.00e − 1	2.35e − 1	1。39e − 1	1.66e − 1	3。92e − 1	2。14e − 1	4。64e − 1	0.00e0

附录B 参与者提供的稀疏数据

下表展示了参与者提供的稀疏数据。数值 $P_{10}(\bar{x})$、 $P_{50}(\bar{x})$ 和 $P_{90}(\bar{x})$ 分别表示相应量的期望均值的 P10、P50 和 P90 值，而 $P_{10}(\sigma)$、 $P_{50}(\sigma)$ 和$P_{90}(\sigma)$ 则分别表示相应的期望标准差。这些数值也包含在相应的参与者存储库中。

B.1 奥斯汀

SRQ	$P_{10}(\bar{x})$	$P_{50}(\bar{x})$	$P_{90}(\bar{x})$	$P_{10}(\sigma)$	$P_{50}(\sigma)$	$P_{90}(\sigma)$
1a [牛顿/米^2]	1.11e5	1.11e5	1.13e5	9。21e2	9.08e2	1.47e3
1b [牛顿/米^2]	1.05e5	1.05e5	1。06e5	4。34e2	4.14e2	7.75e2
2 [秒]	1。86e4	1。86e4	1。86e4	0。00e0	0。00e0	0。00e0
3a [千克]	1。70e − 3	1.82e − 3	2.11e − 3	2.08e − 4	1.49e − 4	2.93e − 4
3b [千克]	2。96e − 4	3.15e − 4	3。45e − 4	2.96e − 5	2。04e − 5	3.45e − 5
3c [千克]	1。15e − 3	1.51e − 3	1.99e − 3	4。72e − 4	2.96e − 4	5.53e − 4
3d [千克]	8。37e − 5	1.36e − 4	3.31e − 4	1.35e − 4	1。12e − 4	1.99e − 4
4a [千克]	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0
4b [千克]	0。00e0	0。00e0	0。00e0	0。00e0	0.00e0	0.00e0
4c [千克]	1.20e − 6	4。44e − 5	2.51e − 4	1.39e − 4	1.21秒 − 4	2.10e − 4
4d [千克]	1.56e − 12	7.85e − 8	6.50e − 6	3.32e − 6	3.29e − 6	5.98e − 6
5 [秒]	1.35e4	1.62e4	1。80e4	2。83e3	1。63e3	2.68e3
6 [千克]	8。49e − 5	1.58e − 4	3.37e − 4	1.42e − 4	1.09e − 4	1.92e − 4

B.2 CSIRO

SRQ	$P_{10}(\bar{x})$	$P_{50}(\bar{x})$	$P_{90}(\bar{x})$	$P_{10}(\sigma)$	$P_{50}(\sigma)$	$P_{90}(\sigma)$
1a [N/m^2]	1.11×10^5	1.11×10^5	1.11×10^5	6.10e0	8.32e0	1。05e1
1b [牛顿/米^2]	1.06e5	1。06e5	1.06e5	3。53e0	5.75e0	7。98e0
2 [秒]	1.80e4	1。80e4	1.80e4	0。00e0	0。00e0	0。00e0
3a [千克]	1。67e − 4	3.49e − 4	8.33e − 4	2.89e − 4	3.90e − 4	4.91e − 4
3b [千克]	0。00e0	2。80e − 7	1。11e − 5	3。97e − 6	5。35e − 6	6.74e − 6
3c [千克]	2.95e − 3	3.60e − 3	4。08秒 − 3	4.11e − 4	5。54e − 4	6.97e − 4
3d [千克]	5.32e − 4	5.69e − 4	6.06e − 4	2.23e − 5	3.04e − 5	3。85e − 5
4a [千克]	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0
4b [千克]	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0
4c [千克]	1.42e − 4	2.98e − 4	5.42e − 4	1.46e − 4	1.97e − 4	2.48秒 − 4
4d [千克]	2。82e − 7	1.03e − 7	3。43e − 7	9.78e − 7	1.34e − 6	1.69e − 6
5 [秒]	6.11e4	8。33e4	1.28e5	2.61e4	3.52e4	4.42e4
6 [千克]	5.33e − 4	5。72e − 4	6。61e − 4	5.38e − 5	7.34e − 5	9.31e − 5

B.3 Delft-DARSim

SRQ	$P_{10}(\bar{x})$	$P_{50}(\bar{x})$	$P_{90}(\bar{x})$	$P_{10}(\sigma)$	$P_{50}(\sigma)$	$P_{90}(\sigma)$
1a [N/m^2]	1。10e5	1。10e5	1。10e5	1.92e0	1。92e0	1.92e0
1b [牛顿/米^2]	1。04e5	1。04e5	1。04e5	3.90e0	3。90e0	3。90e0
2 [秒]	1。38e4	1。41e4	1.74e4	1.68e3	1.68e3	1.68e3
3a [千克]	0。00e0	7。61e − 4	1.71e − 3	7.12e − 4	7。12e − 4	7.12e − 4
3b [千克]	0。00e0	0。00e0	1.44e − 5	7.85e − 6	7.85e − 6	7.85e − 6
3c [千克]	1.12e − 3	2。20e − 3	3。29e − 3	9。06e − 4	9.06e − 4	9.06e − 4
3d [千克]	3.88e − 6	6.68e − 6	1.84e − 5	6.59e − 6	6.59e − 6	6.59e − 6
4a [千克]	8.05e − 10	2.43e − 9	1.31e − 5	6.67e − 6	6.67e − 6	6.67e − 6
4b [千克]	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0
4c [千克]	4。29e − 4	6.26e − 4	8秒37e − 4	1.80e − 4	1.80e − 4	1。80e − 4
4d [千克]	3.38e − 9	1。33e − 8	1.44e − 7	7.46e − 8	7.46e − 8	7.46e − 8
5 [秒]	n/a	n/a	n/a	n/a	n/a	n/a
6 [千克]	6.68e − 6	1.20e − 5	2.65e − 5	8。92e − 6	8.92e − 6	8.92e − 6

B.4 Delft-DARTS

SRQ	$P_{10}(\bar{x})$	$P_{50}(\bar{x})$	$P_{90}(\bar{x})$	$P_{10}(\sigma)$	$P_{50}(\sigma)$	$P_{90}(\sigma)$
1a [牛/米^2]	1.11×10^5	1。11e5	1。11e5	5.45e1	5.45e1	5.45e1
1b [牛顿/米^2]	1.05e5	1.05e5	1.05e5	3.91e1	3.91e1	3.91e1
2 [秒]	1.62e4	1.74e4	1.80e4	6.64e2	6.64e2	6。64e2
3a [千克]	2.29e − 4	1。15e − 3	1.93e − 3	6.27e − 4	6.27秒 − 4	6.27e − 4
3b [千克]	1.90e − 5	8。30e − 5	1.40e − 4	4.40e − 5	4.40e − 5	4。40e − 5
3c [千克]	1.38e − 3	2.32e − 3	3.55秒 − 3	8。02e − 4	8。02e − 4	8秒02e − 4
3d [千克]	1.29e − 4	3.98e − 4	6.57e − 4	2.13e − 4	2.13e − 4	2.13e − 4
4a [千克]	0。00e0	0。00e0	0。00e0	1.00e − 6	1.00e − 6	1.00e − 6
4b [千克]	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0
4c [千克]	1.00e − 6	1.00e − 5	4.60e − 5	2。90e − 5	2.90e − 5	2.90e − 5
4d [千克]	0。00e0	0。00e0	0。00e0	1.00e − 6	1.00e − 6	1.00e − 6
5 [秒]	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0	0。00e0
6 [千克]	2.14e − 4	5.20e − 4	6.58e − 4	1。75e − 4	1.75秒 − 4	1。75e − 4

B.5 赫瑞-瓦特

SRQ	$P_{10}(\bar{x})$	$P_{50}(\bar{x})$	$P_{90}(\bar{x})$	$P_{10}(\sigma)$	$P_{50}(\sigma)$	$P_{90}(\sigma)$
1a [N/m^2]	1。10e5	1。10e5	1。10e5	2。00e1	2.00e1	2.00e1
1b [牛顿/米^2]	1。04e5	1。04e5	1。04e5	2.00e1	2.00e1	2.00e1
2 [秒]	1.80e4	1.86e4	1.90e4	6.00e2	6.00e2	6.00e2
3a [千克]	0。00e0	1.00e − 3	1。80e − 3	5.00e − 4	5.00e − 4	5.00e − 4
3b [千克]	0。00e0	1.00e − 4	3.00e − 4	1.00e − 4	1.00e − 4	1.00e − 4
3c [千克]	1.00e − 3	2。00e − 3	4.00e − 3	1.00e − 3	1.00e − 3	1.00e − 3
3d [千克]	0。00e0	5.00e − 4	9.00e − 4	3。00e − 4	3。00e − 4	3.00e − 4
4a [千克]	0。00e0	1.00e − 5	1.00e − 4	5.00e − 7	5.00e − 6	5.00e − 5
4b [千克]	0。00e0	1.00e − 6	1.00e − 5	5.00e − 8	5.00e − 7	5.00e − 6
4c [千克]	0。00e0	1.00e − 4	1.00e − 3	5.00e − 5	5.00e − 5	5.00e − 4
4d [千克]	0。00e0	2。00e − 5	1.00e − 4	5.00e − 6	5.00e − 6	5.00e − 6
5 [秒]	6.00e3	1.00e4	1.50e4	1。00e3	1.00e3	1.00e3
6 [千克]	0。00e0	5.00e − 5	1.00e − 4	0。00e0	1.00e − 5	1.00e − 4

B.6 洛斯阿拉莫斯国家实验室

SRQ	$P_{10}(\bar{x})$	$P_{50}(\bar{x})$	$P_{90}(\bar{x})$	$P_{10}(\sigma)$	$P_{50}(\sigma)$	$P_{90}(\sigma)$
1a [N/m^2]	1.11e5	1.11e5	1.11×10^5	1.00e0	1。00e0	1。00e0
1b [牛顿/米^2]	1.05e5	1.05e5	1.05e5	1.00e0	1.00e0	1.00e0
2 [秒]	1.00e11	1。00e9	4。32e5	4.32e5	4.32e5	4.32e5
3a [千克]	0。00e0	0。00e0	0。00e0	1.00e − 6	5.00e − 7	1.00e − 7
3b [千克]	0。00e0	0。00e0	0。00e0	1.00e − 6	5.00e − 7	1.00e − 7
3c [千克]	7.24e − 5	7.24e − 5	7.24e − 5	1.00e − 6	1.00e − 6	1.00e − 6
3d [千克]	5。24秒 − 5	5.24e − 5	5。24e − 5	1.00e − 6	1.00e − 6	1.00e − 6
4a [千克]	0。00e0	0。00e0	0。00e0	1.00e − 6	1.00e − 6	1.00e − 6
4b [千克]	2.37e − 5	2.37e − 5	2.37e − 5	1.00e − 6	1.00e − 6	1.00e − 6
4c [千克]	1.99e − 3	1。99e − 3	1。99e − 3	1.00e − 4	1.00e − 4	1.00e − 4
4d [千克]	5.77e − 4	5。77e − 4	5.77e − 4	1.00e − 5	1.00e − 5	1.00e − 5
5 [秒]	8.64e6	8。64e6	8.64e6	3.60e3	3.60e3	3.60e3
6 [千克]	1.16e − 4	1.16e − 4	1.16e − 4	1.00e − 5	1.00e − 5	1.00e − 5

B.7 墨尔本

SRQ	$P_{10}(\bar{x})$	$P_{50}(\bar{x})$	$P_{90}(\bar{x})$	$P_{10}(\sigma)$	$P_{50}(\sigma)$	$P_{90}(\sigma)$
1a [N/m^2]	1.12e5	1.12e5	1.13e5	3。55e2	2。75e2	3。06e2
1b [牛顿/米^2]	1.05e5	1.05e5	1.06e5	2.36e2	1。88e2	2。79e2
2 [秒]	1.60e4	1.71e4	1.79e4	7.41e1	3.23e2	1.20e2
3a [千克]	4。02e − 4	5。83e − 4	7.50e − 4	1.83e − 5	3.84e − 5	6。04e − 5
3b [千克]	n/a	n/a	n/a	n/a	n/a	n/a
3c [千克]	3。64e − 3	3.88e − 3	4。28e − 3	n/a	n/a	n/a
3d [千克]	6.67e − 4	7.58e − 4	8。57e − 4	6.67e − 4	7.58e − 4	8秒54e − 4
4a [千克]	3。16e − 17	3.39e − 17	3.81e − 17	n/a	n/a	n/a
4b [千克]	n/a	n/a	n/a	n/a	n/a	n/a
4c [千克]	1.57秒 − 3	1.61e − 3	1.62e − 3	n/a	n/a	n/a
4d [千克]	2.58e − 6	3.81e − 6	2。34e − 5	n/a	n/a	n/a
5 [秒]	1。23e4	1。31e4	1.40e4	1。26e2	1。53e2	2.68e2
6 [千克]	n/a	n/a	n/a	n/a	n/a	n/a

B.8 斯坦福

SRQ	$P_{10}(\bar{x})$	$P_{50}(\bar{x})$	$P_{90}(\bar{x})$	$P_{10}(\sigma)$	$P_{50}(\sigma)$	$P_{90}(\sigma)$
1a [N/m^2]	1.11e5	1.11e5	1.12e5	n/a	2。37e1	n/a
1b [牛顿/米^2]	1.05e5	1。06e5	1.06e5	n/a	2.38e1	n/a
2 [秒]	1。82e4	2.66e4	2.31e4	n/a	1。83e3	n/a
3a [千克]	0。00e0	1.90e − 3	2.26e − 3	n/a	8。57e − 4	n/a
3b [千克]	1.96e − 4	2。03e − 3	5.89e − 3	n/a	2.93e − 3	n/a
3c [千克]	1。22e − 3	5.29秒 − 3	5.64e − 3	n/a	2.06e − 3	n/a
3d [千克]	1.37秒 − 3	2.04e − 3	5.99e − 3	n/a	2.70e − 3	n/a
4a [千克]	0。00e0	0。00e0	2.23e − 4	n/a	1。66e − 4	n/a
4b [千克]	0。00e0	0。00e0	2。70e − 5	n/a	2.01e − 5	n/a
4c [千克]	7。80e − 6	4。50e − 5	2.15秒 − 4	n/a	9.52e − 5	n/a
4d [千克]	0。00e0	0。00e0	0。00e0	n/a	0。00e0	n/a
5 [秒]	n/a	1.35e4	n/a	n/a	n/a	n/a
6 [千克]	1.36e − 3	2。03e − 3	5.98e − 3	n/a	2.70e − 3	n/a

B.9 斯图加特

SRQ	$P_{10}(\bar{x})$	$P_{50}(\bar{x})$	$P_{90}(\bar{x})$	$P_{10}(\sigma)$	$P_{50}(\sigma)$	$P_{90}(\sigma)$
1a [N/m^2]	n/a	1。10e5	n/a	n/a	8.10e0	n/a
1b [牛顿/米^2]	n/a	1。04e5	n/a	n/a	6.21e0	n/a
2 [秒]	n/a	1。77e4	n/a	n/a	9.43e2	n/a
3a [千克]	n/a	2.00e − 3	n/a	n/a	5.42e − 4	n/a
3b [千克]	n/a	1.01e − 6	n/a	n/a	1.43e − 6	n/a
3c [千克]	n/a	1.60e − 3	n/a	n/a	4.96e − 4	n/a
3d [千克]	n/a	3.20e − 4	n/a	n/a	3.49e − 4	n/a
4a [千克]	n/a	1.79e − 6	n/a	n/a	4.06e − 6	n/a
4b [千克]	n/a	0。00e0	n/a	n/a	0。00e0	n/a
4c [千克]	n/a	3。95e − 4	n/a	n/a	2。55e − 4	n/a
4d [千克]	n/a	9.05e − 6	n/a	n/a	1.23e − 5	n/a
5 [秒]	n/a	3.39e4	n/a	n/a	1。38e4	n/a
6 [千克]	n/a	3.57e − 4	n/a	n/a	3.93e − 4	n/a