arXiv:2301.11195v1 [physics.flu-dyn] 26 Jan 2023

利用侵入式序贯数据同化进行城市环境的增广状态估计

L. Villanueva[a]，M. Mart´ınez Valero[b]，A. ˇSarki´c Glumac[c]，M. Meldi[b]

[a]*Institut Pprime，CNRS - ISAE-ENSMA -* 普瓦捷大学，*11 Bd. Marie et Pierre Curie, Site du Futuroscope, TSA 41123, 86073 Poitiers Cedex 9, France*[b]里尔大学，*CNRS，ONERA，Arts et Métiers ParisTech*，里尔中央理工学院，*UMR 9014- LMFL-Laboratoire de Mécanique des fluides de Lille - Kampé de Feriet, F-59000 Lille, France*

[c]卢森堡大学，安全、可靠性与信任跨学科中心（*SnT*），*6 Avenue de la Fonte,* 阿尔泽特河畔埃施，*4364 Luxembourg*

摘要

采用异质实验样本和RANS计算流体力学相结合的方法，对高层建筑周围的流动开展了数据驱动研究。该耦合采用基于集合卡尔曼滤波（EnKF）的技术实现，其中包括局部化和膨胀等先进操作。通过EnKF获得的增广状态估计还被用于优化 $\mathcal{K}-\varepsilon$ 湍流模型的五个自由全局模型常数，从而改善模型的预测特征，该湍流模型用于封闭方程组。优化值与经典值相距甚远，后者作为通用建议值规定并被编入相关代码；同时，优化值也不同于文献中报道的其他数据驱动分析结果。采用这一新的优化参数化描述所获得的结果表明，速度场和压力场均得到整体改善。此外，在远离传感器位置的下游区域，还观察到流动组织结构在拓扑方面有所改善。

关键词： 城市环境，数据同化，EnKF，CONES

1. 引言

在流体力学领域尚待解决的挑战中，准确预测和控制钝体周围的湍流，是当前自动化、数据驱动型城市环境发展中的一个紧迫课题。钝体是指浸没在流体中的固体（例如车辆或建筑物），

其形状与周围流场的相互作用会导致大尺度、高能量尾流的形成。在考虑高雷诺数时，整体气动相互作用呈现出多种复杂并存的现象，例如剪切层、流动分离、再附着和回流区。由于充分发展湍流中存在范围广泛的活跃动力学尺度，预测这些特征是一项复杂任务。对于城市环境等现实应用中遇到的雷诺数，通过直接数值模拟完整呈现湍流所需的计算资源高得难以承受。降阶计算流体力学（CFD），例如雷诺平均纳维–斯托克斯（RANS）[1, 2]，能够以可承受的资源描述完整的城市区域，但此类预测的准确度会受到用于封闭动力学方程的湍流模型特征的显著影响。这些模型由多个系数驱动，系数通常通过传统经验方法确定，因此往往无法准确呈现湍流所引发的不同物理现象之间的相互作用。

实验方法依赖于通过多种技术获取的测量数据，例如压力传感器和热线，可通过压力和速度测量，近乎精确地表征流动特征。然而，实验数据在空间和时间上可能是局部的，而要获得流动的全体积表征，成本高得难以承受。此外，实验可能会受到传感器布置困难的影响，导致无法在流动的敏感区域进行采样。

过去二十年来的研究一直致力于在数值模拟与实验之间建立稳固的联系，以发挥这两种方法各自的固有优势。在文献中提出的诸多方案中，数据同化 [3] 天然适用于结合实验数据与数值数据，从而更准确地预测流动。序贯数据同化利用概率和统计学工具，在提供不同信息来源及其相应置信度后，以最小的不确定性为目标估计物理状态（状态估计）。这类方法包括卡尔曼滤波器（KF） [4] 及其集合版本——集合卡尔曼滤波（EnKF）[5, 6]；后者实际上是现有最强大的数据同化工具之一。EnKF 不仅能够精确估计状态，还可以利用这一物理状态训练底层模型（例如计算流体力学模型），使其在实际运行条件下表现更佳。

本文通过整合实

验数据，对矩形建筑周围的流动进行雷诺平均纳维–斯托克斯模拟增强。增强过程通过优化若干决定湍流封闭行为的全局自由常数来实现。为此，采用了在数量有限的传感器处采样的时间平均压力和时间平均速度。尽管文献中已有许多利用数据同化 [7, 8] 或基于机器学习的工具 [9, 10], 优化 RANS 建模的数据驱动分析，但其中大多数方法都以高保真数值数据为参考。原因在于，使用实验结果预计会遇到许多额外困难，其中包括可能导致模型发散的过拟合。研究将表明，基于集合卡尔曼滤波的方法得益于滤波器的平滑特性，适用于在降阶计算流体力学框架内稳健地整合实验数据。

第2节介绍了本研究所采用的数值策略和算法，包括对所用数值求解器的描述，以及对数据驱动策略的介绍；这些策略集成在一个专用的 C++ 库中。第3节介绍了数据同化分析的设置，并详细说明了将要比较的不同技术。第4节将所得结果与高保真模拟的数据进行比较。最后，第5节总结了主要结论，并探讨了未来的研究方向。

2. 数值方法

2.1. 数值代码：*OpenFOAM*

本研究中的数值模拟采用一个名为 $OpenFOAM$ 的 C++ 开源库进行。该库包含若干基于有限体积（FV）离散化的求解器 [11], ，以及用于预处理、后处理和数据处理的多种工具。由于该库采用免费许可证，且提供了大量可用于开展扩展多物理场分析的模块，因此该代码已广泛应用于流体力学研究文献中 [12, 13]。

对于本研究，采用有限体积（FV）数值离散方法求解不可压缩流和牛顿流体的RANS纳维-斯托克斯方程 [1]:

$$\overline{u}_j \frac{\partial \overline{u}_i}{\partial x_j} = -\frac{\partial \overline{p}}{\partial x_i} + \frac{\partial \overline{\tau}_{ij}}{\partial x_j} - \frac{\partial \tau_{ij}^T}{\partial x_j} \qquad i = 1, 2, 3 \tag{1}$$

$$\nabla^2 \overline{p} = -\frac{\partial \overline{u}_j}{\partial x_i} \frac{\partial \overline{u}_j}{\partial x_i} - \frac{\partial}{\partial x_i} \left(\frac{\partial \tau_{ij}^T}{\partial x_j} \right) \tag{2}$$

其中，式 1 为动量方程，式 2 为泊松方程。所用变量包括速度$\mathbf{u} = [u_1, u_2, u_3]$、归一化压力$p$、黏性应力张量 τ_{ij}（采用牛顿流体假设进行建模）和雷诺应力张量 τ_{ij}^T。上划线表示通过平均运算得到式 1–2。在雷诺平均纳维–斯托克斯框架内，必须对 τ_{ij}^T进行湍流封闭。 $\mathcal{K} - \varepsilon$模型 [14, 2] 采用涡黏性假设，在 τ_{ij}^T与平均速度$\mathbf{u}$的梯度之间建立联系：

$$-\tau_{ij}^T = 2\nu_T \overline{S}_{ij} - \frac{2}{3}\mathcal{K}\delta_{ij} \tag{3}$$

其中， ν_T 为湍流黏度， $\mathcal{K}$ 为湍流动能， $\overline{S}_{ij}$为平均应变率：

$$\overline{S}_{ij} = \frac{1}{2} \left(\frac{\partial \overline{u}_i}{\partial x_j} + \frac{\partial \overline{u}_j}{\partial x_i} \right) \tag{4}$$

在 $\mathcal{K} - \varepsilon$ 模型中， ν_T表示为 $\mathcal{K}$ 和能量耗散率 ε的代数函数：

$$\nu_T = C_\mu \frac{\mathcal{K}^2}{\varepsilon} \tag{5}$$

其中， C_μ 为待校准的模型常数。为使问题闭合，必须加入关于 $\mathcal{K}$ 和 ε 的两个模型方程：

$$\frac{\partial \mathcal{K}}{\partial t} + \overline{u}_j \frac{\partial \mathcal{K}}{\partial x_j} = \frac{\partial}{\partial x_j} \left((\nu + \frac{\nu_T}{\sigma_\mathcal{K}}) \frac{\partial \mathcal{K}}{\partial x_j} \right) + \mathcal{P} - \varepsilon \tag{6}$$

$$\frac{\partial \varepsilon}{\partial t} + \overline{u}_j \frac{\partial \varepsilon}{\partial x_j} = \frac{\partial}{\partial x_j} \left((\nu + \frac{\nu_T}{\sigma_\varepsilon}) \frac{\partial \varepsilon}{\partial x_j} \right) + C_{\varepsilon 1} \frac{\varepsilon}{\mathcal{K}} \mathcal{P} - C_{\varepsilon 2} \frac{\varepsilon^2}{\mathcal{K}} \tag{7}$$

其中产生项为 $\mathcal{P} = \nu_T S^2$、 $S = \sqrt{2 S_{ij} S_{ij}}$。模型在

确定五个常数 C_μ、 $C_{\varepsilon 1}$、 $C_{\varepsilon 2}$、 $\sigma_\mathcal{K}$ 和 σ_ε 后即完整。Launder 和 Sharma [14] 提供了相关数值，这些数值是通过分析学术测试案例确定的，例如均匀各向同性湍流的自由衰减以及湍流平面通道。然而，这些系数并非常数，而是流动局部动力学及其与流动全局特征相互作用的函数（参见文献 [15, 16, 9] 中的讨论）。

2.2. 观测：在风洞中获取的实验数据

实验在德国波鸿鲁尔大学的大气边界层风洞中进行。风洞的横截面为 $1.6 \times 1.8m$，测试段长度为 9.4 m。图 1 a) 展示了安装在风洞旋转台上的木制建筑模型。

在风洞入口处设置尖塔并布置粗糙元，以生成边界层流动。如图 1 b) 所示，平均风速廓线符合指数为 0.2 的幂律。这代表地形类别 II [17]，可模拟孤立高层建筑周围流动的真实情况。

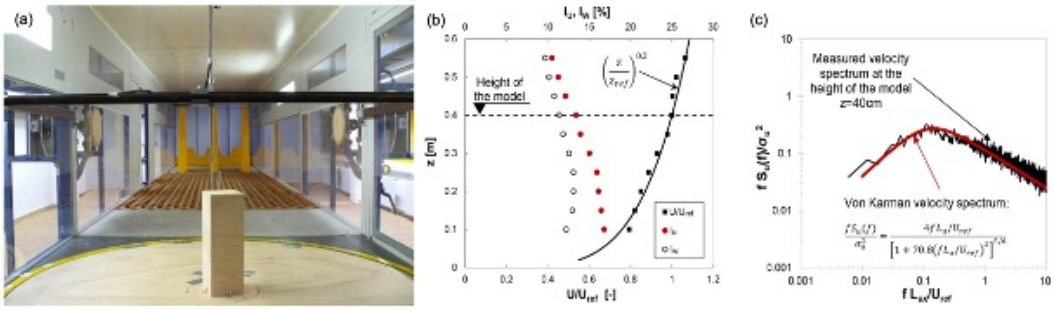


图 1：用于获取实验数据的风洞测试段

由于使用了不同的传感器来捕捉屋顶上方速度场的特征并测量建筑表面的压力，采样数据具有异质性。屋顶上方的速度主要在三个不同高度（$z/D = 0.075$、0.3 和 0.45）处进行测量，这些高度位于图 2 a) 中以红色标记的点上方。此外，在图 2 中标记为 P36 的屋顶中心上方，还考虑了十个高度，间距为 $z/D = 0.075$。测量使用热线风速仪进行，该仪器由两根交叉热线组成，可同时测量流向和垂向速度分量。所有速度数据均以 2000 Hz 的采样频率采集。

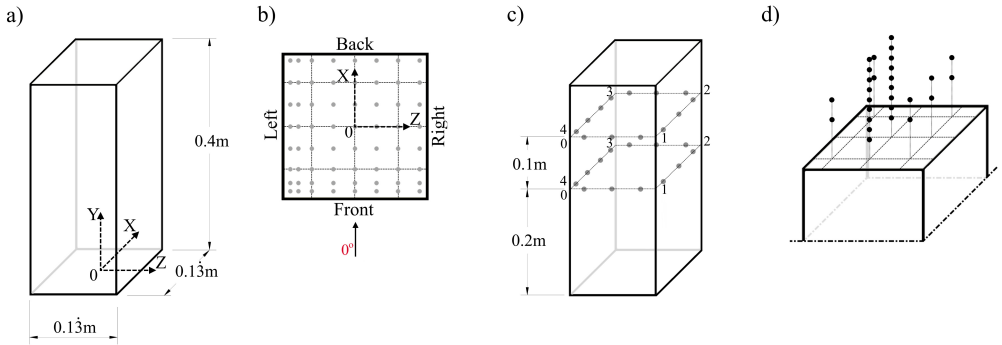


图 2：高层建筑的几何形状，包括 (a) 主要尺寸和坐标系；(b) 标有测压孔位置的俯视图；(c) 标有测压孔位置的立面；(d) 屋顶上测得的速度观测值。

除速度测量外，还在不同位置对表面压力进行采样。模型屋顶上的表面压力使用 64 个测压孔进行测量，其分布如图 2 b) 和 c) 所示，并以浅灰色圆圈标出。表面压力采用多通道同步扫描测量系统以 1000 Hz 的采样频率采集。管路效应通过数值方法进行补偿 [18]。关于风洞实验以及高层建筑周围流动分析的更多细节，尤其是屋顶上方区域的分析，见 [19]。在本分析中，实验数据用于提高定常雷诺平均纳维–斯托克斯模型的预测能力。因此，可获得的速度分量和压力时间序列均进行了时间平均。

2.3. 测试案例

所考虑的高层建筑案例是对风洞测试 [19] 的数值表示。模型的横截面为正方形，边长为 $B = 133.33mm$，建筑物高度为 $H = 400mm$，对应于全尺寸下高 120 m 的建筑物。建筑物为平屋顶，并研究了 0° 风向，因此渐近速度与流向 x 对齐。横向为 y，法向为 z。

计算域的尺寸按照 [20]提供的最佳实践指南选取。上游计算域长度为 $5H$。

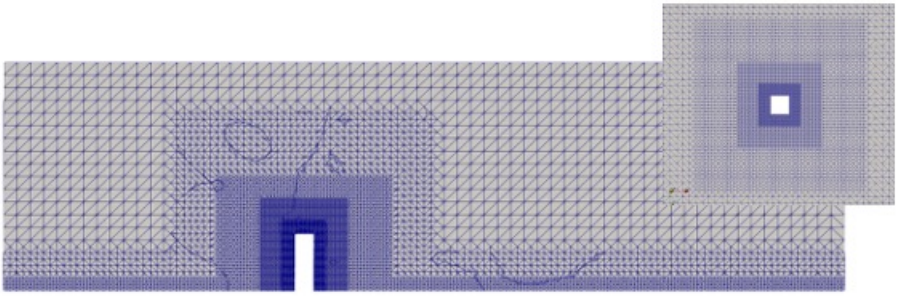


图3：用于雷诺平均纳维–斯托克斯计算的网格视图。图中显示了中央垂直平面和一个水平平面。

计算域最终尺寸为：长度（x） × 宽度（y） × 高度（z）
$15.5H \times 4.5H \times 4H = 6.2 \times 1.8 \times 1.6m^3$。对于 z 方向，高度选取为与风洞高度一致。

如图3所示，高层建筑表面附近采用结构化网格。墙面相邻网格单元中心点到建筑物的距离使其平均值为 $y^+ = 141$ 、最小值为 $y^+ = 40$ ，从而确保该点位于对数层内。用于离散计算域的网格单元总数为 513 266 个。

通过将结果与更细网格进行比较，开展了网格依赖性研究。更细网格由 430 万个网格单元组成，其建筑模型附近的空间分辨率是粗网格算例的 2 倍。比较了粗、细网格模拟预测的平均压力在屋顶测压孔位置处的数值。比较结果显示，建筑物表面上 86% 的点相对差异低于 10%。

模拟中的入口边界条件，即平均速度$\mathbf{u}$、湍流动能 $\mathcal{K}$ 和湍流耗散率 ε，是根据平均风速 U 和纵向湍流强度 I_u 的来流垂直剖面确定的。湍流动能 $\mathcal{K}$ 由 U 计算得出

和 I_u:

$$\mathcal{K}(z) = a(I_u(z)u(z))^2 \tag{8}$$

$$\varepsilon(z) = \frac{u_{ABL}^{*3}}{\kappa(z+z_0)} \tag{9}$$

其中 $a \in [0.5, 1.5]$ [21]。本研究采用 $a = 1$ ，这是 [20] 的建议。湍流耗散率 ϵ 由式9给出，其中冯·卡门常数为 $\kappa = 0.42$。压力—速度耦合采用SIMPLE算法。数值格式采用经典选择。对流项采用一阶迎风格式，黏性项采用二阶中心格式。压力从单元中心插值到面中心时，采用OpenFOAM 求解器自带的二阶线性格式。

2.4. 数据同化：集合卡尔曼滤波

数据同化（DA） [3, 6]是一类工具，可用于整合多个信息来源，以获得增强预测并提高准确度。经典应用通常依赖于：

- 一种模型，它能够对所研究的物理现象进行（准）连续表征。基于物理的模型，例如计算流体力学求解器，就是流体力学应用中的一种模型

- 一些观测，通常比模型更准确，但在时空上具有局部性。在流体力学中，这些数据可能来自高保真数值模拟或实验

通过操纵信息源获得的增强预测 也可用于主动推断模型的优化参数描述，目的是获得一种无需依赖观测即可提供准确预测的预测工具。数据同化传统上应用于环境科学和气象科学，但近年来在流体力学中的应用迅速增加 [22, 23, 7, 24, 25,26, 27, 8, 28, 29]。方法种类繁多，但可分为两大类 [6, 30]:

变分方法： 这类方法旨在针对所研究的情形最小化代价函数。通常通过对模型进行参数优化来

达到这一最小值，对应的流动状态也较为准确。

统计（序贯）方法： 其目标是最小化解的方差（即提高对预测的置信度）的方法。

变分方法（如四维变分）已广泛应用于流体力学，尤其是稳态模拟中 [31, 22, 32, 27] 。尽管序贯工具据称更适合预测非平稳现象，但文献中也报道了其在稳态流动中的应用 [24, 8, 33]。在本文中，我们将重点讨论源自卡尔曼滤波器的工具，卡尔曼滤波器是一种广为人知的序贯方法。

2.4.1. 卡尔曼滤波器

卡尔曼滤波器（KF） [4] 是一种基于贝叶斯定理的序贯数据同化方法。它为随时间变化的离散数据提供了线性滤波的解。对物理量$\mathbf{x}$ 进行分析的经典表述依赖于通过离散模型$\mathbf{M}$产生的结果与某个观测$\mathbf{y}$的结合；在原始卡尔曼滤波器中，该离散模型是线性的。在卡尔曼滤波器框架内，模型和观测都会受到误差/不确定性的影响，这里分别称为 v 和 w。卡尔曼滤波器的一个核心假设是，这些不确定性可以用高斯分布准确描述，即 $v = \mathcal{N}(\mathbf{0}, \mathbf{Q})$和 $w = \mathcal{N}(\mathbf{0}, \mathbf{R})$。$\mathbf{Q}$和$\mathbf{R}$同样是时间的函数，分别表示模型和观测的方差。鉴于这些误差可以用高斯分布描述，该解完全由状态的前两个矩确定，即物理量$\mathbf{x}$ 和误差协方差矩阵$\mathbf{P} = \mathbb{E}((\mathbf{x} - \mathbb{E}(\mathbf{x}))(\mathbf{x} - \mathbb{E}(\mathbf{x}))^T)$。考虑物理系统在时刻 k 和 $k+1$之间的时间推进。对于后一个时刻，可获得观测$\mathbf{y}_{k+1}$ 。在这种情况下，数据同化过程包括两个阶段（见算法1）：

1. 一次预报步骤（上标 f），其中时刻 k 的物理状态和误差协方差矩阵通过模型推进至下一时刻：$\mathbf{x}^f_{k+1} = \mathbf{M}\mathbf{x}_k \mathbf{P}^f_{k+1} = \mathbf{M}\mathbf{P}_k\mathbf{M}^T + \mathbf{Q}_{k+1}$

2. 一次分析步骤，其中将观测与预报结合，以获得增强预测：

$$\mathbf{K}_{k+1} = \mathbf{P}^f_{k+1}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f_{k+1}\mathbf{H}^T + \mathbf{R}_{k+1}\right)^{-1}$$
$$\mathbf{x}^a_{k+1} = \mathbf{x}^f_{k+1} + \mathbf{K}_{k+1}(\mathbf{y}_{k+1} - \mathbf{H}\mathbf{x}_{k+1})$$
$$\mathbf{P}^a_{k+1} = (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H})\mathbf{P}^f_{k+1}$$

其中，$\mathbf{H}$ 是一个将模型结果映射到观测空间的线性算子，而$\mathbf{K}$ 是卡尔曼增益。该矩阵考虑了状态向量值与观测值之间的相关性，是提供物理系统最终状态估计的核心要素。

该算法在流体力学复杂应用中的主要缺点是：i）它是为线性模型$\mathbf{M}$设计的；ii）P 的规模与所研究问题的自由度数量直接相关。第一个问题可以通过对数据驱动策略进行专门改进来绕过，例如扩展卡尔曼滤波器 [6, 30], 中采用的改进；第二个问题则更为严重。事实上， P 必须像物理变量一样随时间推进。此外，在分析 阶段，还需要对$\mathbf{P}$ 进行大量操作，包括矩阵求逆。对于计算流体力学中使用的自由度数量（通常在 $10^6 - 10^8$ 范围内），这会导致对随机存取存储器和计算资源的需求高得难以承受。

算法 1： 卡尔曼滤波器算法

预报步骤

$\mathbf{x}^f_{k+1} = \mathbf{M}\mathbf{x}^a{}_k$

$\mathbf{P}^f_{k+1} = \mathbf{M}\mathbf{P}^a_k\mathbf{M}^T + \mathbf{Q}_{k+1}$

分析步骤

$\mathbf{K}_{k+1} = \mathbf{P}^f_{k+1}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f_{k+1}\mathbf{H}^T + \mathbf{R}_{k+1}\right)^{-1}$

$\mathbf{x}^a_{k+1} = \mathbf{x}^f_{k+1} + \mathbf{K}_{k+1}(\mathbf{y}_{k+1} - \mathbf{s}^f{}_{k+1})$

$\mathbf{P}^a_{k+1} = (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H})\mathbf{P}^f{}_{k+1}$

2.4.2.（随机）集合卡尔曼滤波

集合卡尔曼滤波（EnKF）[34, 6] 是一种基于卡尔曼滤波器的流行数据驱动策略，能够高效解决前面讨论的问题。其基本思想是不再随时间推进误差协方差矩阵$\mathbf{P}$ ，而是通过模型运行集合对其进行近似。这一策略能够充分考虑模型的非线性，并几乎消除了与$\mathbf{P}$操作相关的计算负担。现将详细介绍集合卡尔曼滤波的完整结构，其概要见算法 2。

集合卡尔曼滤波依赖于模型的 N_e 个实现，这些实现构成模型集合。在给定时刻 k，这些实现可组成一个状态矩阵 $\mathbf{X}_S$ ，其大小为 $[N, N_e]$，其中 N 是所研究物理问题的自由度数量。因此，沿用卡尔曼滤波器中引入的上标 f 和 a ，$\mathbf{X}_S$ 的每一列都对应第 i^{th} 个集合成员的状态 $\mathbf{x}_{i\ ,k+1}^{f}$ ，其中 $i \in [1, N_e]$。利用集合成员在统计上相互独立这一假设，可以近似得到误差协方差矩阵 $\mathbf{P}_e$ ：

$$\mathbf{P}_e^f = \mathbf{X}^f (\mathbf{X}^f)^T \tag{10}$$

其中，$\mathbf{X}^f$ 是异常矩阵，表示状态向量的所有值相对于其集合均值的偏差：

$$\mathbf{X}_{k+1} = \frac{\mathbf{x}_{i,k+1} - \overline{\mathbf{x}_{k+1}}}{\sqrt{N_e - 1}}\ , \qquad \overline{\mathbf{x}_{k+1}} = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_{i,k+1} \tag{11}$$

采样观测由 N_o 个元素组成，也会进行扩展以获得 N_e 组数值。为此，将基于测量误差协方差矩阵 $\mathbf{R}_{k+1}$ 的高斯噪声添加到观测向量中：

$$\mathbf{y}_{i,k+1} = \mathbf{y}_{k+1} + \mathbf{e}_{i,k+1},\ with\ \mathbf{e}_{i,k+1} \sim \mathcal{N}(0, \mathbf{R}_{k+1}) \tag{12}$$

最后，模型实现被投影到观测空间 N_e次，与经典卡尔曼滤波器类似：

$$\mathbf{s}_{i,k+1} = \mathbf{H}\mathbf{x}_{i,k+1} \tag{13}$$

所有这些要素共同使得能够确定卡尔曼

增益:

$$\mathbf{S}_{k+1} = \frac{\mathbf{s}_{i,k+1} - \overline{\mathbf{s}_{k+1}}}{\sqrt{N_e - 1}}, \qquad \overline{\mathbf{s}_{k+1}} = \frac{1}{N_e}\sum_{i=1}^{N_e} \mathbf{s}_{i,k+1} \tag{14}$$

$$\mathbf{E}_{k+1} = \frac{\mathbf{e}_{i,k+1} - \overline{\mathbf{e}_{k+1}}}{\sqrt{N_e - 1}}, \qquad \overline{\mathbf{e}_{k+1}} = \frac{1}{N_e}\sum_{i=1}^{N_e} \mathbf{e}_{i,k+1} \tag{15}$$

$$\mathbf{K}_{k+1} = \mathbf{X}^f_{k+1}(\mathbf{S}^f_{k+1})^T \left[\mathbf{S}^f_{k+1}(\mathbf{S}^f_{k+1})^T + \mathbf{E}_{k+1}(\mathbf{E}_{k+1})^T\right]^{-1} \tag{16}$$

当集合规模趋于无穷时，$\mathbf{E}_{k+1}(\mathbf{E}_{k+1})^T$ 趋于卡尔曼滤波器的矩阵$\mathbf{R}$。在实际应用中，集合规模有限，因此利用对角矩阵$\mathbf{R}_{k+1}$ 简化扰动乘积，从而简化计算并降低计算成本 [30, 35]。此外，对于每个分析阶段，都可以直接根据集合成员估计$\mathbf{P}_e$ ，无需存储内存或推进时间。

最后，状态向量的更新方式与经典卡尔曼滤波器相同，唯一的区别在于必须执行 N_e 次更新:

$$\mathbf{x}^a_{i,k+1} = \mathbf{x}^f_{i,k+1} + \mathbf{K}_{k+1}(\mathbf{y}_{i,k+1} - \mathbf{s}^f_{i,k+1}) \tag{17}$$

集合卡尔曼滤波也可用于优化模型的参数化描述。其基本思想是在分析阶段结束时更新参数，使模型能够更准确地预测所研究的物理现象，从而缩小模型预测状态与最终状态估计之间的差异。文献中提出了多种策略 [6] ，其中一种适用于参数集相对较小（称为 θ）且易于实现的策略，就是所谓的扩展状态方法。在该策略中，集合卡尔曼滤波的步骤应用于一个状态$\mathbf{x}^\star$ ，其定义如下:

$$\mathbf{x}^\star = \begin{bmatrix}\mathbf{x} \\ \theta\end{bmatrix} \tag{18}$$

也就是说，集合卡尔曼滤波所使用的状态同时包含物理状态和模型的参数化描述。对于这个非常简单的算法，全局状态的大小现在等于 $N^\star = N + N_\theta$，其中 N_θ 是待优化参数的数量

。如果 $N_\theta << N$，这种修改带来的计算成本增加可以忽略不计，而且我们可以在分析阶段结束时，同时获得更新后的状态估计和优化后的模型参数化描述。

算法 2： 集合卡尔曼滤波算法

对于 N_e 集合中的成员数量 ，$i = 1, ..., N_e$**：**

1 状态向量的时间推进：

$$\boldsymbol{x}^f_{i,k+1} = \mathcal{M}\boldsymbol{x}^a_{i,k}$$

2 通过引入误差，根据观测数据创建观测矩阵：

$$\boldsymbol{y}_{i,\ k+1} = \boldsymbol{y}_{k+1} + \boldsymbol{e}_i,\ \text{其中}\ \ \boldsymbol{e}_i \sim \mathcal{N}(0,\ \ \boldsymbol{R})$$

3 预测观测值的计算：

$$\boldsymbol{s}^f_{i,k+1} = \mathcal{H}\boldsymbol{x}^f_{i,k+1}$$

4 集合均值的计算：

$$\overline{\boldsymbol{x}^f_{k+1}} = \frac{1}{N_e}\sum_{i=1}^{N_e}\boldsymbol{x}^f_{i,k+1},\ \overline{\boldsymbol{s}^f_{k+1}} = \frac{1}{N_e}\sum_{i=1}^{N_e}\boldsymbol{s}^f_{i,k+1},$$
$$\overline{\boldsymbol{e}_{k+1}} = \frac{1}{N_e}\sum_{i=1}^{N_e}\boldsymbol{e}_{i,k+1}$$

5 异常矩阵的计算：

$$\boldsymbol{X}_{k+1} = \frac{\boldsymbol{x}_{i,k+1} - \overline{\boldsymbol{x}_{k+1}}}{\sqrt{m-1}},\ \boldsymbol{S}_{k+1} = \frac{\boldsymbol{s}_{i,k+1} - \overline{\boldsymbol{s}_{k+1}}}{\sqrt{m-1}},\ \boldsymbol{E}_{k+1} = \frac{\boldsymbol{e}_{i,k+1} - \overline{\boldsymbol{e}_{k+1}}}{\sqrt{m-1}}$$

6 卡尔曼增益的计算：

$$\boldsymbol{K}_{k+1} = \boldsymbol{X}^f_{k+1}(\boldsymbol{S}^f_{k+1})^T\left[\boldsymbol{S}^f_{k+1}(\boldsymbol{S}^f_{k+1})^T + \boldsymbol{R}_{k+1}\right]^{-1}$$

7 状态矩阵更新：

$$\boldsymbol{x}^a_{i,k+1} = \boldsymbol{x}^f_{i,k+1} + \boldsymbol{K}_{k+1}(\boldsymbol{y}_{i,k+1} - \boldsymbol{s}^f_{i,k+1})$$

2.4.3. 膨胀

经典集合卡尔曼滤波存在一些缺点，例如集合中的可用成员数量有限而导致抽样误差。在流体力学应用中，尤其是计算流体力学领域，这一点尤为明显，因为每次模拟都可能需要大量计算资源和存储空间。因此，对于三维运行而言，实际可接受的集合成员总数约为 $N_e \in [40, 100]$。由于这种误差会在同化步骤中累积，减少这一问题的一种方法是将误差协方差矩阵 $\mathbf{P}_{k+1}$ 乘以因子 λ^2 [6]，以此进行膨胀。

该系数 $\lambda > 1$ 控制所谓的乘性膨胀，可将其应用于分析状态矩阵。它会增加

状态估计的变异性:

$$\mathbf{x}_i^a \longrightarrow \overline{\mathbf{x}^a} + \lambda(\mathbf{x}_i^a - \overline{\mathbf{x}^a}) \tag{19}$$

显然，对于 $\lambda = 1$ ，可得到经典集合卡尔曼滤波的结果。

同样，通过集合卡尔曼滤波对推断参数集 θ进行优化，可能会迅速收敛到局部最优解，从而得到次优结果。膨胀可用于缓解模型参数描述过快收敛的问题，通过人为增加参数的变异性，使其能够趋向全局最优解。

2.4.4. 局部化

经典集合卡尔曼滤波建立了观测与模型自由度之间的相关性，但这种相关性不受二者间距离的影响。在集合规模有限（如计算流体力学中所采用的规模）时，这可能会导致大计算域中的状态矩阵更新出现虚假效应。实际上，有限集合近似所带来的误差可能远大于真实的物理相关性，而在连续系统中，物理相关性会自然地随距离增加而衰减。由于计算资源有限，无法在集合中使用更多成员，因此避免这些虚假效应的一种方法是对协方差矩阵$\mathbf{P}_{k+1}$ 的取值乘以一个修正乘数项，该项考虑观测传感器与状态网格单元之间的物理距离。这种策略称为协方差局部化。与膨胀一样，局部化能够有效提高计算准确度，并降低集合卡尔曼滤波发散的概率。协方差局部化的原理是将协方差矩阵$\mathbf{P}_{k+1}$ 与一个修正矩阵$\mathbf{L}$进行逐元素相乘。这种运算称为舒尔积，因此该方法也称为舒尔局部化。由此可得局部化卡尔曼增益的表达式，即公式 20。

$$[\mathbf{P}_{k+1}]_{i,j}[\mathbf{L}]_{i,j} \longrightarrow \mathbf{K}_{k+1}^{loc} = [\mathbf{L}]_{i,j}[\mathbf{X}_{k+1}^f(\mathbf{S}_{k+1}^f)^T]_{i,j}\left([\mathbf{L}]_{i,j}[\mathbf{S}_{k+1}^f(\mathbf{S}_{k+1}^f)^T]_{i,j} + \mathbf{R}_{k+1}\right)^{-1} \tag{20}$$

由于矩阵 $\mathbf{R}_{k+1}$ 对运行的影响有限，为方便起见，算法中简化了该表达式。局地化的卡尔曼增益变为:

$$[\mathbf{P}_{k+1}]_{i,j}[\mathbf{L}]_{i,j} \longrightarrow \mathbf{K}_{k+1}^{loc} = [\mathbf{L}]_{i,j}[\mathbf{K}_{k+1}]_{i,j} \tag{21}$$

矩阵 $\mathbf{L}$ 的结构必须由用户设定，并且应体现真实的物理相关性。在连续系统中，物理变量之间的相关性会随空间距离增加而迅速减弱。因此，局地化矩阵通常采用指数衰减形式：

$$\mathbf{L}(i,j) = e^{-\Delta_{i,j}^2/\mu} \tag{22}$$

其中，$\Delta_{i,\ j}$ 是给定观测传感器与模型评估点之间的距离（在计算流体力学中，该点为网格单元的中心）。 μ 是衰减系数，可根据测试案例的特征进行调整。

2.5. CONES：将 *OpenFOAM* 与数值环境耦合

CONES（将 OpenFOAM 与数值环境耦合）是一个 C++库，旨在将计算流体力学软件 OpenFOAM 与其他任何类型的开源代码耦合。目前，该库用于实施序贯数据同化技术，尤其是基于集合卡尔曼滤波的先进数据驱动方法。基于集合卡尔曼滤波的代码与 OpenFOAM 之间通过 CWIPI（带插值的并行耦合接口） [36], 进行通信。CWIPI 是一个开源代码耦合器，适用于大规模并行的多物理场/多组件应用和动态算法。

CONES 在使用 OpenFOAM 进行数据同化时具有以下主要优势：

- 无需修改 OpenFOAM 的安装，只需编译用户自定义函数。

- 不同代码之间的侵入式耦合是在保留现有计算流体力学求解器原有结构的前提下实现的。所有与 CONES 相关的函数都包含在经过修改的 Pstream（OpenFOAM 的一部分）库中，因此，数据交换通过调用特定函数在求解器循环结束时进行，而计算循环保持不变。

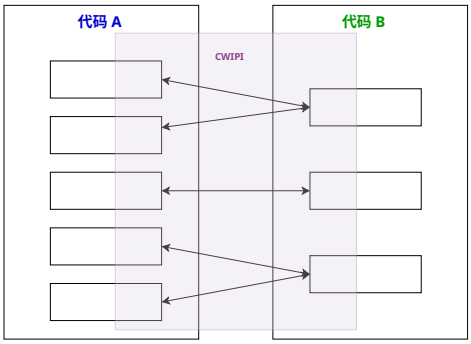
- 物理状态和网格信息的交换效率很高（可同时快速发送和接收包含数百万个元素的数组）。

- 多个处理器之间的直接高性能计算通信，这些处理器负责数值模拟和数据同化过程的分区。

- 模拟和数据同化同时在线 运行，也就是说，无需在每个分析阶段停止模拟。这一点能大幅降低所需的计算资源。

ONERA 和 CERFACS 开发的耦合器 CWIPI 因其能够高效管理基于分布式网格定义的完全并行数据交换，并且能够在不重合的网格之间进行插值（这对某些基于集合卡尔曼滤波的高级工具非常有用，例如 MGEnKF [29]），因此被选用。它的大多数应用都与燃气涡轮设计[37, 38], 有关，但最近也应用于使用 OpenFOAM 的气动声学领域 [39]。

CWIPI 通信协议基于消息传递接口库。因此，必须在代码中初始化 MPI 和 CWIPI 环境。这样便可使用 CWIPI 的基本功能，在两个代码之间交换信息。图 4 展示了两个代码通过 CWIPI 进行直接通信的方式，以及其中一些主要的基本功能。



(a)两个代码之间的 CWIPI

代码 A

```
MPI_Init(&argc, &argv);
cwipi_初始化('CodeA', localcomm);

cwipi_创建_耦合('cplAB',...);

cwipi_define_网格('cplAB',...);

for (i=0; i<nit;++i){
        cwipi_exchange('cplAB', ex i,...);
}

cwipi_delete_耦合('cplAB',...);

cwipi_finalize;
MPI_Finalize;
```

代码 B

```
MPI_Init(&argc, &argv);
cwipi_init('CodeB', localcomm);

cwipi_create_耦合('cplAB',...);

cwipi_define_网格('cplAB',...);

for (i=0; i<nit;++i){
        cwipi_交换('cplAB', ex i,...);
}

cwipi_delete_耦合('cplAB',...);

cwipi_finalize;
MPI_Finalize;
```

(b)CWIPI 中的主要基本功能

图 4：CWIPI 的运行方式

在本研究中，CONES 将求解器 SimpleFOAM 与团队开发的序贯数据同化库耦合；SimpleFOAM 用于模拟采用雷诺平均纳维–斯托克斯方法的流动。图 5 展示了单次运行的结构示例。两个代码（OpenFOAM 和数据同化库）中均初始化了 MPI 通信和耦合器 CWIPI。

尽管基于集合卡尔曼滤波的工具的应用与解的时间推进相关，但将其应用于定常流动十分直接。分析所用的虚拟时间窗以代码的迭代步数为基准设定。当这些时间步数

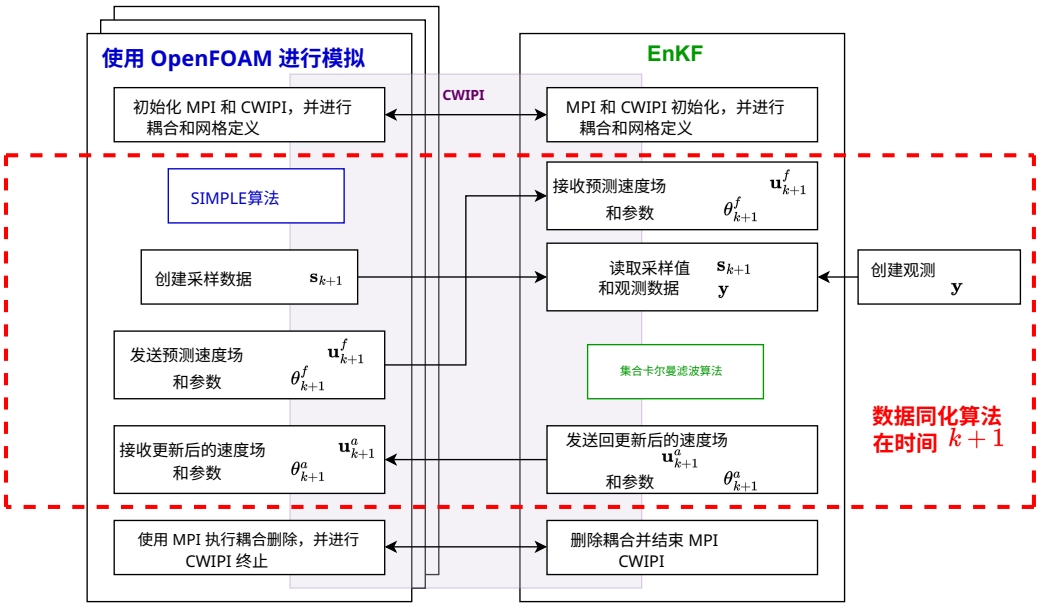


图 5：用于稳态模拟的 CONES 示意图

由 N_e 集合成员同时执行（计算流体力学计算）后，它们会将信息发送至集合卡尔曼滤波代码，并在线等待更新后的流场／参数化描述。目前交换的信息包括速度场 $\mathbf{u}^f$ 以及模型的参数 θ^f。因此，由集合中与成员（计算流体力学模拟）数量相同的状态向量组成的状态矩阵，就是数据同化循环在时间 $k+1$ 时由式 23 表示的状态矩阵。

$$\mathbf{x}_{i,k+1} = \begin{bmatrix} \mathbf{u}_{i,k+1} \\ \theta_{i,k+1} \end{bmatrix} \tag{23}$$

模拟提供的一项额外信息是数值集合$\mathbf{s}_{i,\ k+1}$，即针对每个集合成员，将模型解投影到传感器坐标上所得的结果。考虑到传感器布置未必与网格单元的中心重合，因此必须对流场进行插值。OpenFOAM 提供了多种函数，可将单元中心量转换为特定点处的量。已考虑这些插值方法的准确度 [40] 。

观测值$\mathbf{y}$ 的性质将在第 3 节中作更详细的分析，但 CONES 可以使用同时测量压力 p 和速度场$\mathbf{u}$的传感器。在这个涉及定常模拟的特定案例中，

观测值是恒定的，只需加载一次；但如果分析非定常流，则可以在每个分析阶段都将其纳入。因此，数据同化代码会接收来自模型和传感器的信息，并生成一组更新后的状态和参数 ($\mathbf{u}^a$, θ^a)，然后将其发送回 OpenFOAM 模拟。每个集合成员的压力 p 都会通过泊松方程更新，并使用这组完整数据启动一组新的迭代步骤。当模型参数收敛至用户设定的阈值后，耦合将被解除，并终止 MPI 和 CWIPI 环境。

3. 数据同化实验

本文使用 CONES，基于第 2.3 节所述的数值测试案例研究建筑物周围的流动。具体而言，使用数据同化工具优化驱动 $\mathcal{K}-\varepsilon$ 模型的五个全局常数的取值，以尽可能减小雷诺平均纳维–斯托克斯结果与所提供的高保真观测之间的差异。Zhao等人 [41], 最近进行了类似分析，但他们使用的是高保真模拟的数值结果。本文的观测取自实验的时间平均测量值。

首先需要确定合适的先验状态，包括速度场和压力场，以及参数描述。对于后者，我们优先采用 Margheri等人 [42]通过认知不确定性传播得到的数值，而非 Launder 和 Sharma [14]提出的经典数值。表 1 列出了这些基准值，它们是 N_e 集合模拟的初始均值。每次计算流体力学计算中各参数的取值，初始时均根据有界高斯分布 $\mathcal{N}(\mu_N,\ \sigma_N)$确定，其中 μ_N为参数均值，σ_N 的选取则参考 Margheri等人 [42]的研究，以确保参数空间具有足够大的初始变化范围。正态分布的取值范围介于 σ_N 和 $7\sigma_N$之间，其上下限根据系数的敏感性通过经验设定。例如，$C_{\varepsilon 1}$ 的取值范围为 $1.25\sigma_N$，而 σ_ε 的取值范围为 $7\sigma_N$。每个集合的初始物理状态均由一次计算获得，该计算使用了参考文献[42]中的模型常数。

集合成员的数量 $N_e = 40$ 是参考文献中其他将计算流体力学用于集合卡尔曼滤波器[43, 44, 29]模型部分的研究选定的。

观测来自总计 120 个传感器的时间平均数据。其中，90 个传感器是测压孔，30 个传感器是热线，用于测量速度场的两个分量：流向速度u_x和展向速度 u_z。总计得到 150 个时间平均观测值。数据在分析阶段开始时以以下格式加载：$\mathbf{y} = \begin{bmatrix} u_{x1} & \dots & u_{x30} & u_z & \dots & u_{z30} & p_{31} & \dots & p_{120} \end{bmatrix}^T$，并在整个计算过程中保持不变。同样，协方差矩阵$R_{k+1}$也视为常量，因此得到 $R = \sigma_m I$，其中 σ_m 表示对测量的置信度。

进行了三项相互独立的数据同化实验。这些实验的差异与模型或观测的细节无关，而是考察数据同化流程的不同特征。具体而言，分析的情形如下：

- 情形 A：经典集合卡尔曼滤波。
- 情形 B：采用协方差局地化的集合卡尔曼滤波。
- 情形 C：采用协方差局地化和膨胀的集合卡尔曼滤波。

参数	$\mathcal{K}$-ε **模型默认值**	**集合卡尔曼滤波的先验**		
		$\mu_{\mathbf{N}}$	$\sigma_{\mathbf{N}}$ 适用于情形 A ,B	$\sigma_{\mathbf{N}}$ 适用于情形 C
C_μ [-]	0.09	0.1	0.01	0.005
$C_{\varepsilon 1}$ [-]	1.44	1.575	0.1	0.05
$C_{\varepsilon 2}$ [-]	1.92	1.9	0.1	0.05
$\sigma_{\mathcal{K}}$ [-]	1.0	1.0	0.1	0.05
σ_ε [-]	1.3	1.6	0.1	0.05

表1：RANS $\mathcal{K}-\varepsilon$ 模型中的常规常数与集合卡尔曼滤波所采用的初始参数（$N_e = 40$）对比

对于案例研究 A 和 B，虚拟同化窗口的宽度固定为 150 次迭代。对于案例研究 C，考虑到膨胀导致参数进一步演化，窗口宽度固定为 100 次迭代。这些数值是根据初步分析的结果选定的；分析结果表明，至少需要 50 次迭代，才能在物理量上清晰地观察到新参数设置的特征。

对于 B 和 C 两种情况的局部化，计算域也会在观测值周围足够大的范围内进行裁剪，这与 Moldovan 等人

在 BARC 几何中采用的方法类似。与物理裁剪类似，这种做法减少了集合卡尔曼滤波中的自由度，从而降低了计算成本；局部化裁剪的目的也在于确保去除远离传感器、且只会受到数据同化流程轻微更新的区域。图 6 显示了为物理局地化定义的裁剪区域。局地化矩阵 $\mathbf{L}$ 的系数 μ 设置为使矩阵在裁剪边界处的值接近 0。这样可以避免物理场出现不连续问题。模型的参数 θ 不受局部化影响。

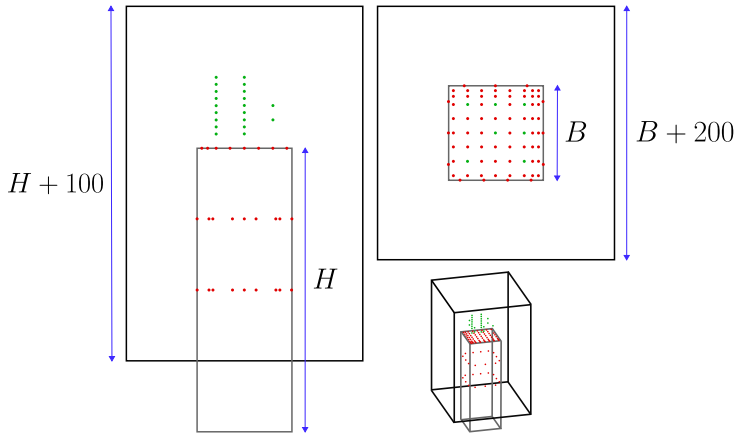


图6：用于局部化的裁剪框：压力传感器以红色表示，速度传感器以绿色显示（数据见 mm）

3.1. 情形 *A*：经典集合卡尔曼滤波

在这个第一个案例中，使用经典的集合卡尔曼滤波。参数收敛后，计算便告结束；本例中，这发生在经过 60 个分析阶段之后（即总计 6000 次计算流体力学迭代）。图 7 展示了 $\mathcal{K} - \varepsilon$ 模型五个参数均值的变化。可以看出，集合卡尔曼滤波得到的最终结果与基准值有显著差异，而且各参数的收敛速度也有显著差别。尤其是，σ_ε 的变化值得讨论。该系数控制方程中湍流扩散项的大小，该方程描述的是与非均匀条件相关的 ε，$D_\varepsilon = \nu_t/\sigma_\varepsilon + \nu$（见第 2 节）。集合卡尔曼滤波执行的优化在计算过程中使 σ_ε 趋向非常低的值，从而提高了方程中 D_ε 的重要性。然而，卡尔曼

增益传播的噪声可能会使某些集合成员的取值变为负值，从而导致计算发散。因此，我们对该参数施加了约束，使其取值不得低于预先设定的一个较小正值。对于其他参数，可以看出， C_μ 和 $C_{\varepsilon 1}$收敛到接近初始估计值中 $1/3$ 的值； $\sigma_{\mathcal{K}}$ 没有出现较大变化，而 $C_{\varepsilon 2}$ 则大了三倍。

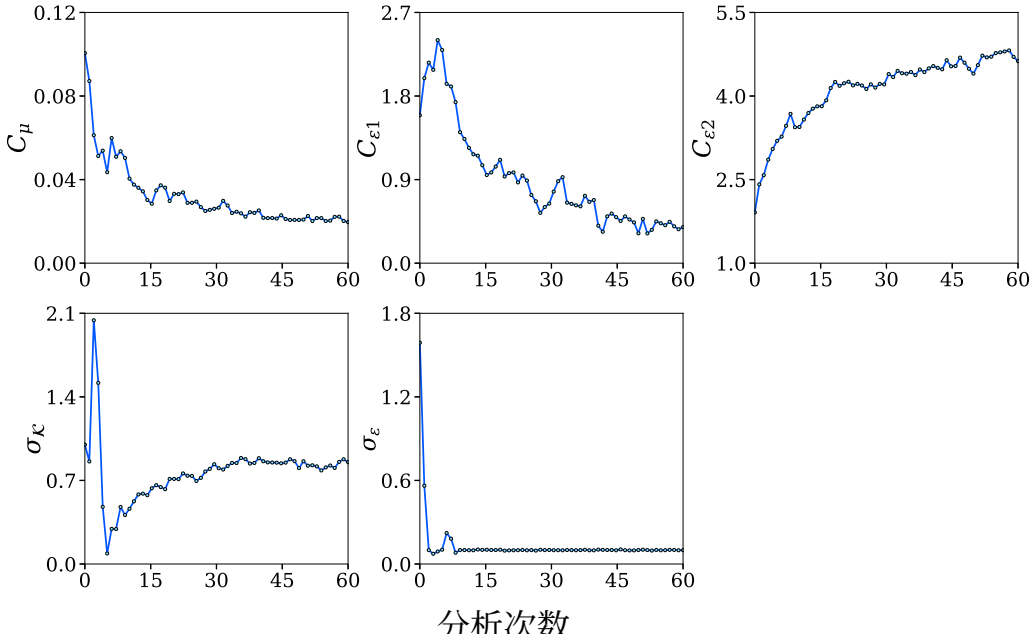


图 7： $\mathcal{K}-\varepsilon$ 在未进行膨胀和局部化时的系数收敛情况

3.2. 情形 *B*：采用协方差局地化的集合卡尔曼滤波

在此情形下，计算采用协方差局地化。五个系数的演变如图 8 所示。其趋势，尤其是 σ_ε 的演变，与情形 A 中观察到的情况相似。情形 A 中针对 σ_ε 的说明在此同样适用。不过，在收敛之前可以观察到更大的波动。此外，需要更多次迭代（100 个分析阶段）才能使参数良好收敛。

3.3. 情形 *C*：同时采用膨胀和局部化的集合卡尔曼滤波

此处的数据同化计算采用确定性膨胀处理模型参数，并进行协方差局地化。这是最

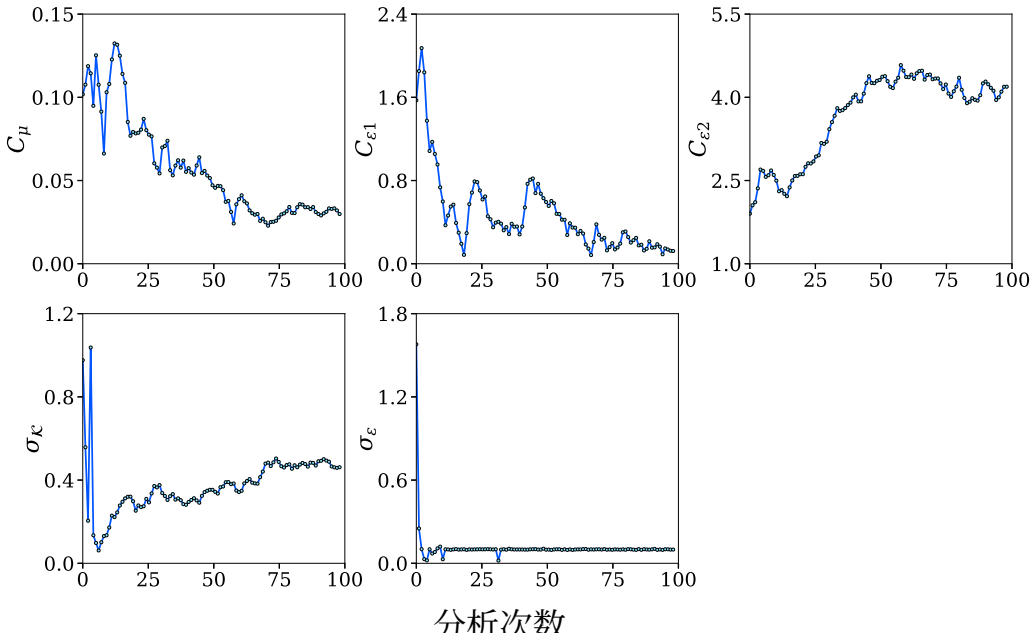


图 8： $\mathcal{K}-\varepsilon$ 系数在无膨胀但采用局部化时的收敛

复杂的数据同化算法运行。五个参数的演变如图 9 所示。为确保模拟在最初几个时间步中的稳健性，膨胀因子 λ 从 1.05 逐渐增加至 1.3，之后再予以移除，以实现收敛（ $k \in [1,40]$对应$\lambda = 1.05$， $k \in [41,120]$对应 $\lambda = 1.1$， $k \in [121,160]$对应 $\lambda = 1.2$， $k \in [161,200]$对应$\lambda = 1.3$ ，以及 $k > 200$对应 $\lambda = 1$ ）。诸如C_μ 和 $\sigma_\mathcal{K}$ 这样的系数对 λ取值变化表现出更高的敏感性，这凸显了膨胀对于确定适合优化的大型参数空间的重要性。因此，此处的收敛明显较晚。此外，此处提高了 σ_ε 的阈值，以避免参数膨胀带来的较高变异性容易引发的稳定性问题。

第4节将研究表2所列三种不同参数化描述的物理预测影响。

4. 结果

现在将三次数据同化运行获得的结果与先验（使用 $\mathcal{K}-\varepsilon$ 模型的经典雷诺平均纳维–斯托克斯）以及时间平均实验值

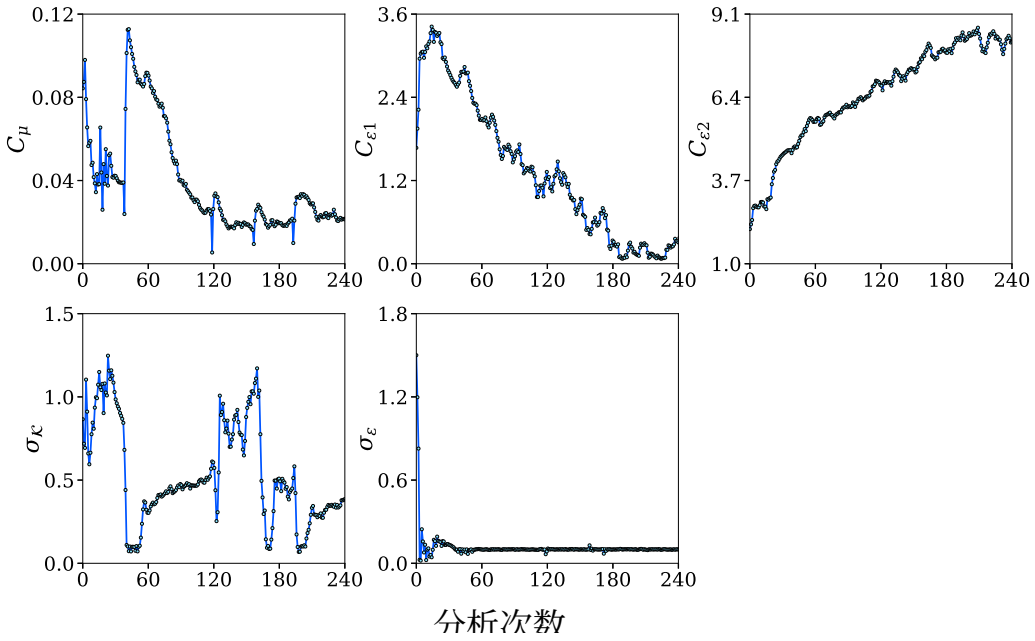


图 9： $\mathcal{K}-\varepsilon$ 系数在膨胀和局部化作用下的收敛

参数	优化值		
	情形 A	**情形 B**	**情形 C**
C_μ [-]	0.021	0.032	0.020
$C_{\varepsilon 1}$ [-]	0.418	0.165	0.162
$C_{\varepsilon 2}$ [-]	4.629	4.080	7.978
$\sigma_\mathcal{K}$ [-]	0.837	0.476	0.355
σ_ε [-]	0.1	0.1	0.1

表2： $\mathcal{K}-\varepsilon$ 模型系数的集合卡尔曼滤波最终结果

结果进行
比较。

4.1. 速度场

首先对速度场进行分析。在不可压缩流的分离式求解器中，速度是一个显式变量。因此，可以通过对该量的预测所获得的定性改进来评估数据同化策略的性能。

图10显示了对应于热线位置的若干位置处的流向速度剖面 u_x 以及法向速度剖面 u_z

。对于 u_x，可以看到，通过数据同化得到的预测场在每个位置的准确度都有明显提升。比较三种不同的数据同化策略时，只能观察到非常细微的差异。另一方面，法向速度 u_z 的预测与先验非常相似。考虑到先验与实验数据吻合良好，这并不令人意外。在点20处可以观察到一个有趣的结果：先验与实验数据之间的最大差异出现在 u_z。在这种情况下，可以看到数据同化预测更接近实验结果，这证实集合卡尔曼滤波能够提供统计上更准确的流动预测，并达到不同信息来源所指示的置信度

图11进一步评估了速度场的特征，其中展示了穿过高层建筑中心的垂直平面上的流线。这里将先验和数据同化运行结果与经过验证的大涡模拟研究进行比较。可以看出，在房间区域整合局部实验信息后得到的雷诺平均纳维–斯托克斯参数变化，使建筑物后方的回流区显著缩小。尽管该区域现在比参考大涡模拟中的区域更小，但在中部高度仍可观察到类似的拓扑结构，而雷诺平均纳维–斯托克斯先验未能捕捉到这一结构。图12展示了屋顶区域的放大图，表明对速度场进行同化有助于改进对回流区的预测，且改进效果显著。

4.2. 压力场

现在来考察压力场的表现。预测这一物理量要困难得多，因为数值求解器中求解的泊松方程将压力用作拉格朗日标记。因此，压力场同时被用作物理变量和拉格朗日约束，以保证流动的不可压缩性。因此，分析这一物理量对于评估算法的稳定性和精度至关重要。图13以与实验数据的性能指标对比形式展示了平均压力系数。如果只关注误差为 $> 10\%$ 时表现最佳的区间，可能会错误地认为先验（误差低于10%的情况占12%）优于数据同化运行（误差低于10%的情况占6%至10%）。然而，这一信息具有误导性，因为数据同化运行中有大量情况恰好落在该区间之外。事实上，考虑更大的误差范围时，可以看到

数据同化运行优于先验雷诺平均纳维–斯托克斯。误差阈值为20%时，使用数据同化后，发生比例提升约5%（$22.5-24.5$%对比18.4%）。当误差阈值为30%时，差距扩大至约15%。图14所示的屋顶测压孔处计算得到的平均压力系数也印证了这一趋势。可以看出，数据同化预测几乎总是优于先验，尽管压力场的改善不如速度场显著。

关于三次数据同化运行的性能，有几点值得说明。尽管所采用的技术有所不同，且参数优化得到的结果表面上也不尽相同，但对物理变量的预测却相当相似。模型参数值可能已收敛至一个鲁棒最优解，在该解附近，进一步参数变化对解的敏感性极低。这一方面仍需进一步研究，但或许表明，为观测设置适当的置信区间即可实现鲁棒优化。在此情形下，局部化的应用已被证明是有效的。数据同化过程中自由度的减少显著降低了每个分析阶段所需的计算资源，但并未导致结果退化。另一方面，可能由于所发现的参数最优区域的特性，膨胀技术并未改善结果。

5. 结论

新开发的平台 CONES 已用于对高层建筑周围的流动开展数据驱动研究。具体而言，压力孔和热线数据形式的异质实验样本，与使用开源代码 OpenFOAM 执行的 RANS 计算流体力学计算相结合。该耦合采用基于集合卡尔曼滤波（EnKF）的技术实现，其中包括局部化和膨胀等高级处理。通过 EnKF 得到的增广状态估计还用于优化该湍流模型的五个自由全局模型常数，从而提升模型的预测特征；该湍流模型用于封闭方程组。因此，研究选取了一个相对较小的不确定性空间，与 Ben Ali 等人 [47]使用变分数据同化时采用的空间相同。

结果表明，所研究的物理量整体上均有所改善，而且采用不同数据同化策略所得的结果相当。对于这一点，物理局部化和协方差局地化似乎对复杂流动研究是有效的。减少数据同化问题的自由度并未影响结果质量，同时整体上缩短了数据驱动过程所需的时间。另一方面，采用膨胀并未带来更好的结果，尤其是因为它需要更多的计算资源。

对速度场的分析表明，集合卡尔曼滤波可将流向和法向上的误差显著降低 50% 以上。参数推断的影响也体现在建筑物后方的回流区域。在这种情况下，流动的物理拓扑更接近经实验数据验证的参考大涡模拟结果，尽管回流泡的尺寸被过度缩小。对于压力场，虽然也观察到了改进，但其幅度在量化上不如速度场的改进显著。这可能是由于计算流体力学求解器采用分离式结构，并将压力作为拉格朗日乘子来施加不可压缩约束所致。未来，若在使用压力数据作为观测的数据同化工具中采用更复杂的耦合求解器，或许能够获得更好的结果。

未来的研究将采用更复杂的湍流建模参数化描述，包括数据同化工具与机器学习应用之间的耦合。

本研究获准使用 GENCI 的高性能计算资源，使用的是 A12 项目 A0122A01741 所申请的资源，运行于 IRENE 超级计算机（TGCC）。在 CONES 开发初期，Florent Duchaine 和 Miguel ´Angel Moratilla-Vega 提供了帮助，在此谨致谢意。

参考文献

[1] S. B. Pope，《湍流》，剑桥大学出版社，2000 年。

[2] D. C. Wilcox，《计算流体力学湍流建模》，第 3 版，DCW Indus- tries Inc.，美国加利福尼亚州拉加拿大，2006 年。

[3] S. B. Daley，《大气数据分析》，剑桥大学出版社，1991。

[4] R. E. Kalman，线性滤波与预测问题的新方法，《基础工程学报》82（1960）35–45。[5] G. Evensen，《数据同化：集合卡尔曼滤波》，Springer-Verlag/柏林/海德堡，2009。[6] M. Asch、M. Bocquet、M. Nodet，《数据同化：方法、算法与应用》，工业与应用数学学会，费城，宾夕法尼亚州，2016。 doi:10.1137/1.9781611974546.[7] H. Xiao、J. L. Wu、J. X. Wang、R. Sun、C. Roy，量化并降低雷诺平均纳维–斯托克斯模拟中的模型形式不确定性：一种数据驱动的物理信息贝叶斯方法，《计算物理学杂志》324（2016）115–136。[8] X. Zhang、H. Xiao、T. Gomez、O. Coutier-Delgosha，量化并降低雷诺平均纳维–斯托克斯模拟中的模型形式不确定性：一种数据驱动的物理信息贝叶斯方法，《计算机与流体》203（2020）104530。[9] K. Duraisamy、H. Xiao、G. Iaccarino，数据时代的湍流建模，《流体力学年度评论》51（2019）–。[10] M. Schmelzer、R. Dwight、C. P.，利用稀疏符号回归发现代数雷诺应力模型，《流动、湍流与燃烧》104（2020）579–603。[11] J. Ferziger、M. Peric，《流体动力学计算方法》，纽约：Springer-Verlag，1996。[12] E. Constant、J. Favier、M. Meldi、P. Meliga、E. Serre，OpenFOAM 中的一种浸入边界法：验证与确认，《计算机与流体》157（3）（2017）55–72。[13] M. Meldi、M. V. Salvetti、P. Sagaut，利用多项式混沌量化空间发展混合层大涡模拟中的误差，《流体物理学》24（2012）035101。[14] B. Launder、B. Sharma，将湍流能量耗散模型应用于旋转圆盘附近流动的计算，《传热传质快报》1（1974）131–137。

[15] M. L., M. M., S. M. V., S. P., 雷诺平均纳维-斯托克斯模型自由系数的认知不确定性，《计算机与流体》102 (2014) 315–335。 doi:http://dx.doi.org/10.1016/j.compfluid.2014.06.029。[16] H. Xiao, P. Cinnella，雷诺平均纳维-斯托克斯模拟中的模型不确定性量化：综述，《航空航天科学进展》108 (2019) 1– 31。[17] .-.-. EN，欧洲规范1：结构作用——第1-4部分：一般作用——风作用，欧洲标准委员会，2005。[18] C. Neuhaus，长压力测量软管系统的数值频率相关校准，内部报告，风工程与流体力学，德国波鸿鲁尔大学（2010）。[19] H. Hemida, A. ˇSarki´c Glumac, G. Vita, K. K. Vraneˇsevi´c, R. H¨offer，关于高层建筑周围的流动与风能采集：风速和表面压力的实验研究，《应用科学》10 (2020) 5283。[20]Y. Tominaga, A. Mochida, R. Yoshie, H. Kataoka, T. Nozu, M. Yoshikawa, T. Shirasawa，建筑物周围行人风环境中计算流体力学实际应用的 AIJ 指南，《风工程与工业空气动力学》96 (2008) 1749–1761。[21] T. Norton, J. Grant, R. Fallon, D.-W. Sun，优化自然通风畜禽建筑的通风配置，以改善室内环境均匀性，《建筑与环境》45 (2010) 983e95。[22] D. P. G. Foures, N. Dovetta, D. Sipp, P. J. Schmid，一种用于重建雷诺平均纳维-斯托克斯驱动平均流场的数据同化方法，《流体力学杂志》759 (2014) 404–431。[23] M. C. Rochoux, S. Ricci, D. Lucor, B. Cuenot, A. Trouve，迈向野火蔓延的数据驱动预测模拟——第一部分：基于多项式混沌代理模型的低成本集合卡尔曼滤波器参数估计，《自然灾害与地球系统科学》14 (2015) 2951–2973。

[24] M. Meldi, A. Poux，一种基于卡尔曼滤波的降阶模型，用于湍流的序贯数据同化，《计算物理学杂志》347 (2017) 207–234。[25] M. Meldi，通过序贯估计器增强湍流预测：状态估计对可用观测时间采样密度的敏感性，《流动、湍流与燃烧》101 (2018) 389–412。[26] J. W. Labahn, H. Wu, B. Coriton, J. H. Frank, M. Ihme，利用高速测量和大涡模拟进行数据同化，以研究湍流火焰中的局部熄灭事件，燃烧学会会刊37 (2019) 2259–2266。[27] P. Chandramouli, E. Memin, D. Heitz，利用动力学误差模型对湍流进行四维大尺度变分数据同化，《计算物理学杂志》412 (2020) 109446。[28] V. Mons, O. Marquet，通过数据同化进行平均流场重建的线性与非线性传感器布置策略，《流体力学杂志》923 (2021) A1。[29] G. Moldovan, G. Lehnasch, L. Cordier, M. Meldi，一种用于非定常流同化的多重网格/集合卡尔曼滤波策略，《计算物理学杂志》443 (2021) 110481。[30] A. Carrassi, M. Bocquet, L. Bertino, G. Evensen，地球科学中的数据同化：方法、问题与展望概述，WIREs Climate Change 9 (2018)。 doi:10.1002/wcc.535.[31] G. Artana, A. Ca milleri, J. Charlier, E. Memin，用于降阶流体流动建模的强约束与弱约束变分同化，《计算物理学杂志》231 (2014) 3264–3288。[32] V. Mons, J. C. Chassaing, P. Sagaut，用于旋转振荡圆柱绕流非定常流变分数据同化的最优传感器布置，《流体力学杂志》823 (2021) 230–277。[33] X. Zhang, C. Michelin-Str¨ofer, H. Xiao，用于逆问题的正则化集合卡尔曼方法，《计算物理学杂志》416 (2020) 109517。

[34] G. Evensen, 《集合卡尔曼滤波：用于状态与参数联合估计——大型系统数据同化的蒙特卡洛方法》, IEEE Control Systems 29 (2009) 83–104。[35]I. Hoteit、D.-T. Pham、M. E. Gharamti、X. Luo, 《减轻随机集合卡尔曼滤波中的观测扰动抽样误差》, Monthly Weather Review 143 (7) (2015) 2918–2936。 doi:10.1175/MWR-D-14-00088.1。[36] A. Reflox、B. Courbet、A. Murrone、P. Villedieu、C. Laurent 等, 《CEDRE软件》, Aerospace Lab 2 (2011) 1–10。网址: https://hal.archives-ouvertes.fr/hal-01182463[37] F. Duchaine 等, 《使用OpenPALM耦合器分析高性能共轭传热》, Computational Science & Discovery 8 (2015) 015003。 doi:https://doi:10.1088/1749-4699/8/1/015003。[38] P. Legrenzi、K. V. Karpaga、P. J. Gary、T. H. Indi, 《采用低马赫数和可压缩工业计算流体力学求解器的简洁稳健耦合气动热框架》, 美国航空航天学会 1640 (2016)。 doi:https://doi.org/10.2514/6.2016-1640。[39] M. A. Moratilla-Vega、M. Angelino、H. Xia、G. J. Page, 《一种用于气动声学建模的开源耦合方法》, 《计算机物理通讯》278 (2022) 108420。 doi:https://doi.org/10.1016/j.cpc.2022.108420。[40] E. Leonard、H. Qiao、N. Saleh, 《快速流体动力学中插值方法的比较》, 国际高性能建筑会议 341 (2021)。网址: https://docs.lib.purdue.edu/ihpbc/ [41] R. Zhao、S. Liu、J. Lie、N. Jiang、Q. Chen, 《基于集合卡尔曼滤波校准的k-ε模型在城市气流和空气污染物扩散中的泛化能力评估》, Building and Environment 212 (2022) 108823。[42] L. Margheri、M. Meldi、M. V. Salvetti、P. Sagaut, 《雷诺平均纳维–斯托克斯模型自由系数的认知不确定性》, 《计算机与流体》102 (2014) 315–335。

[43]M. Katzfuss、J. R. Stroud、C. K. Wikle, 《理解集合卡尔曼滤波》, 《美国统计学家》70 (2016)。 doi:10.1080/00031305.2016.1141709。
[44] V. Mons、Y. Du、T. Zaki, 《大涡模拟中统计数据的集合变分同化》, 《物理评论：流体》6 (2021) 104607。[45]G. Moldovan、A. Mariotti、G. Lehnasch、L. Cordier、M. Salvetti、M. Meldi, 《数据驱动增强型大涡模拟在BARC测试案例分析中的应用》, 《流体力学杂志》- (2022) 1–26。[46] K. K. Vraneˇsevi´c、G. Vita、S. P. Bordas、A. ˇS. Glumac, 《进一步认识高层建筑周围的流动模式：风能潜力的大涡模拟研究》, 《风工程与工业空气动力学》226 (2022) 105029。
doi:10.1016/j.jweia.2022.105029。[47] M. Ben Ali、G. Tissot、S. Aguinaga、D. Heitz、E. M´emin, 《基于稀疏压力测量和变分数据同化的高层建筑平均风场重建》, 《风工程与工业空气动力学》231 (2022) 105204。

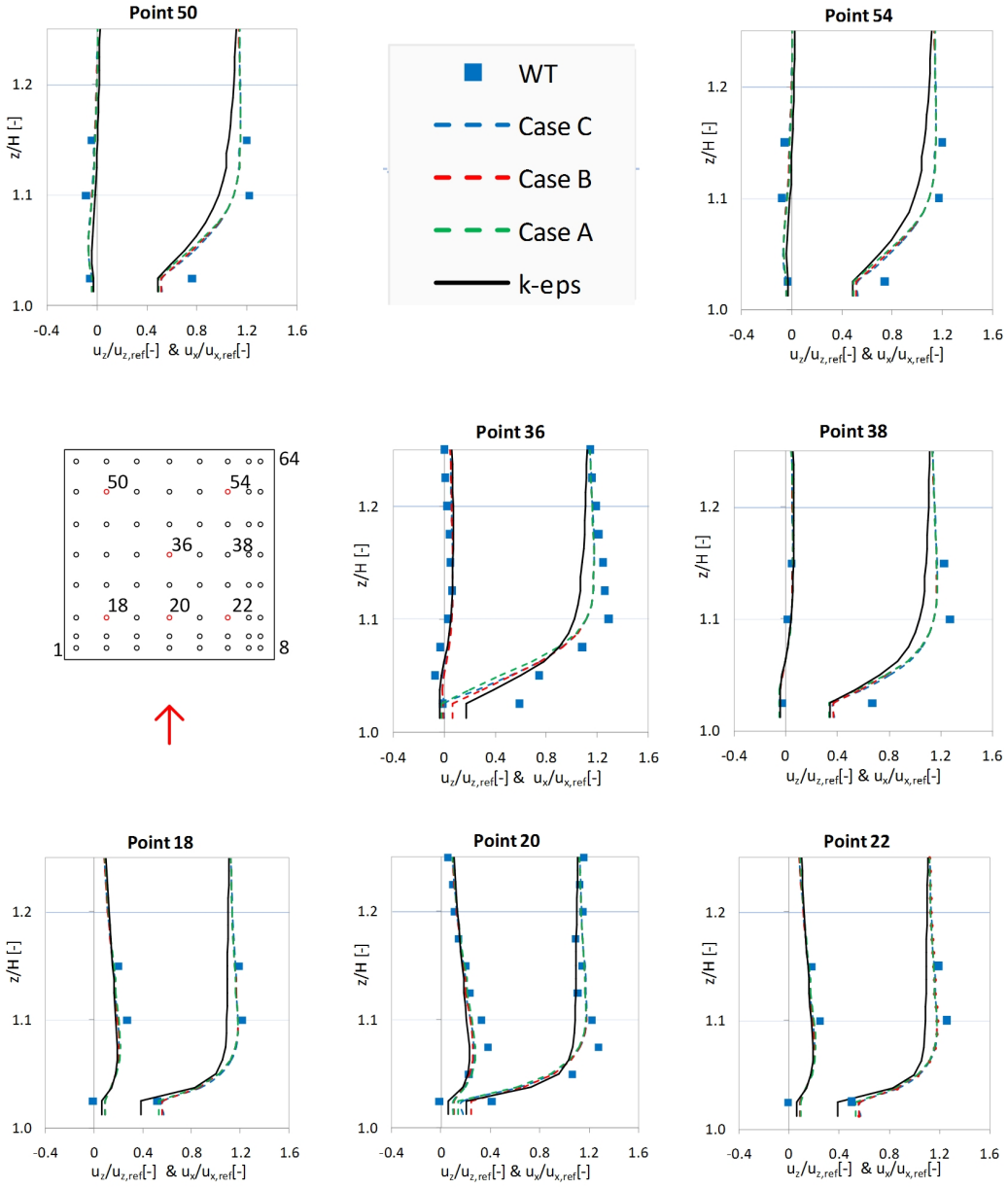


图10：屋顶上标记红色位置处的垂直速度和流向速度剖面（测点：18、20、22、36、38、50、54）；风洞数据（WT）、k-ε模型（k-ε）与三集合卡尔曼滤波案例的比较。

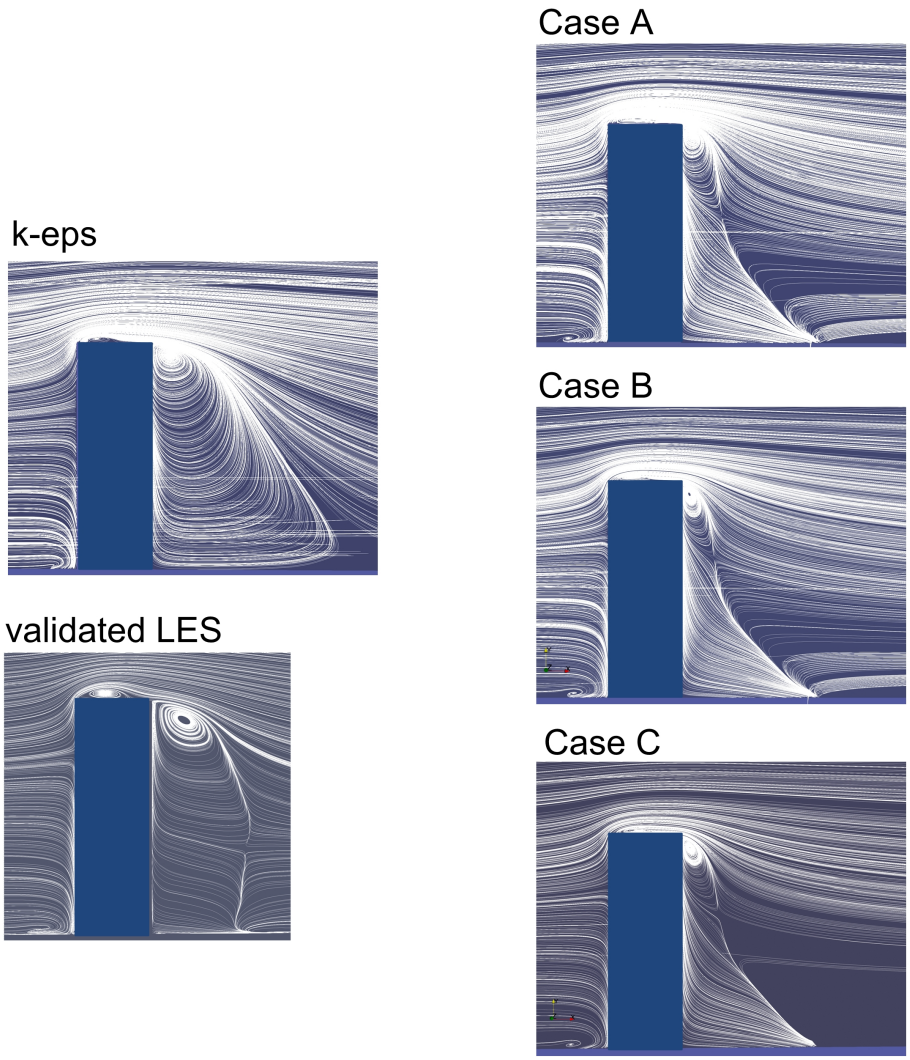


图11：中间平面的流动结构；图左侧为已验证的大涡模拟（已验证的 LES） [46] 与 k-ε 模型（k-ε）的对比，图右侧为三集合卡尔曼滤波案例。

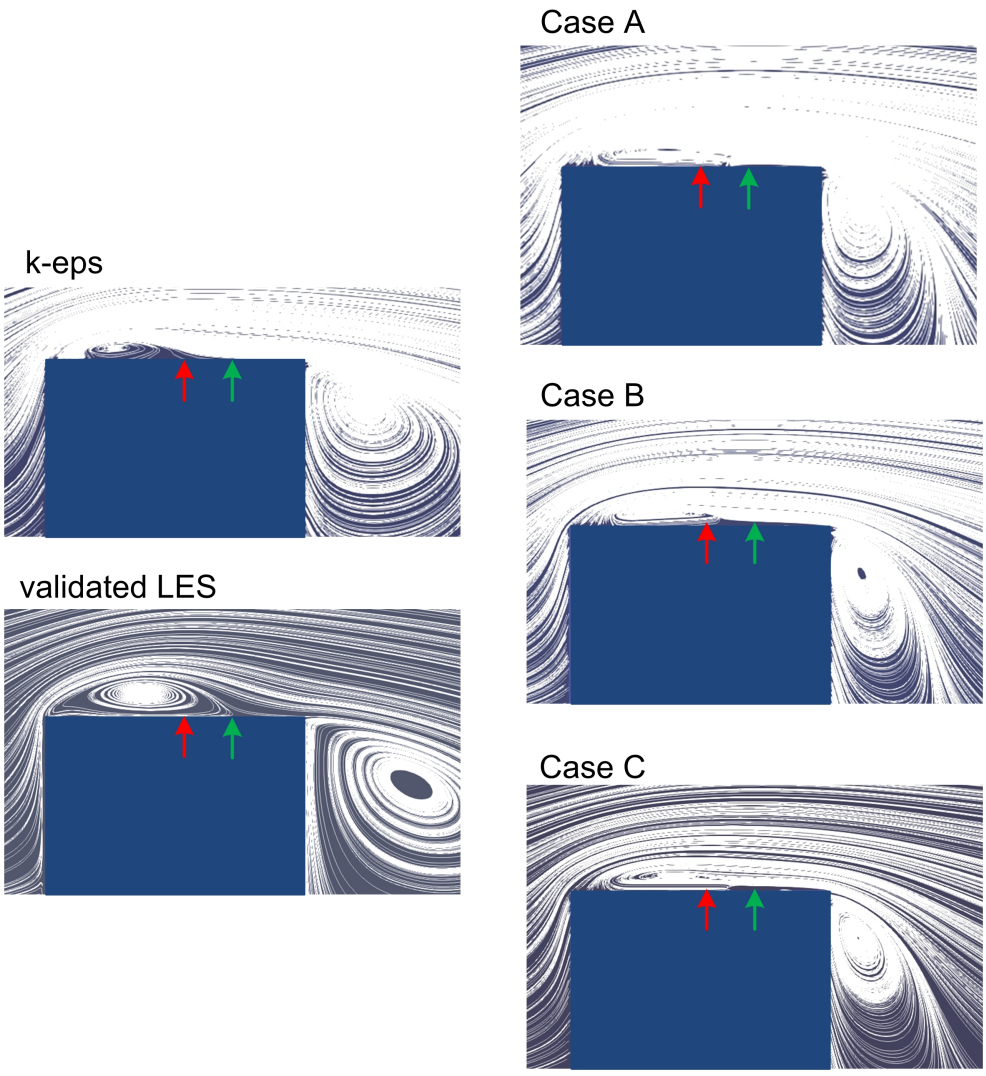


图 12：屋顶上方流动结构的近景，红色箭头表示采用 k-ε模型的模拟中分离泡的再附着位置，绿色箭头表示经验证的大涡模拟中分离泡的再附着位置

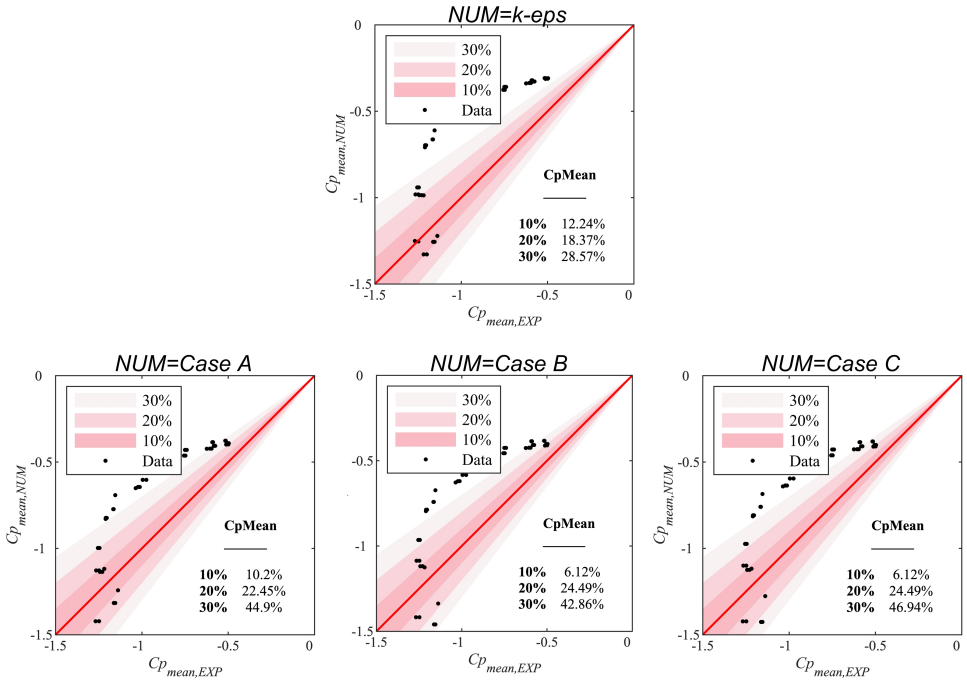


图 13：所有数值算例的平均压力系数散点图及性能指标——与实验数据的比较。

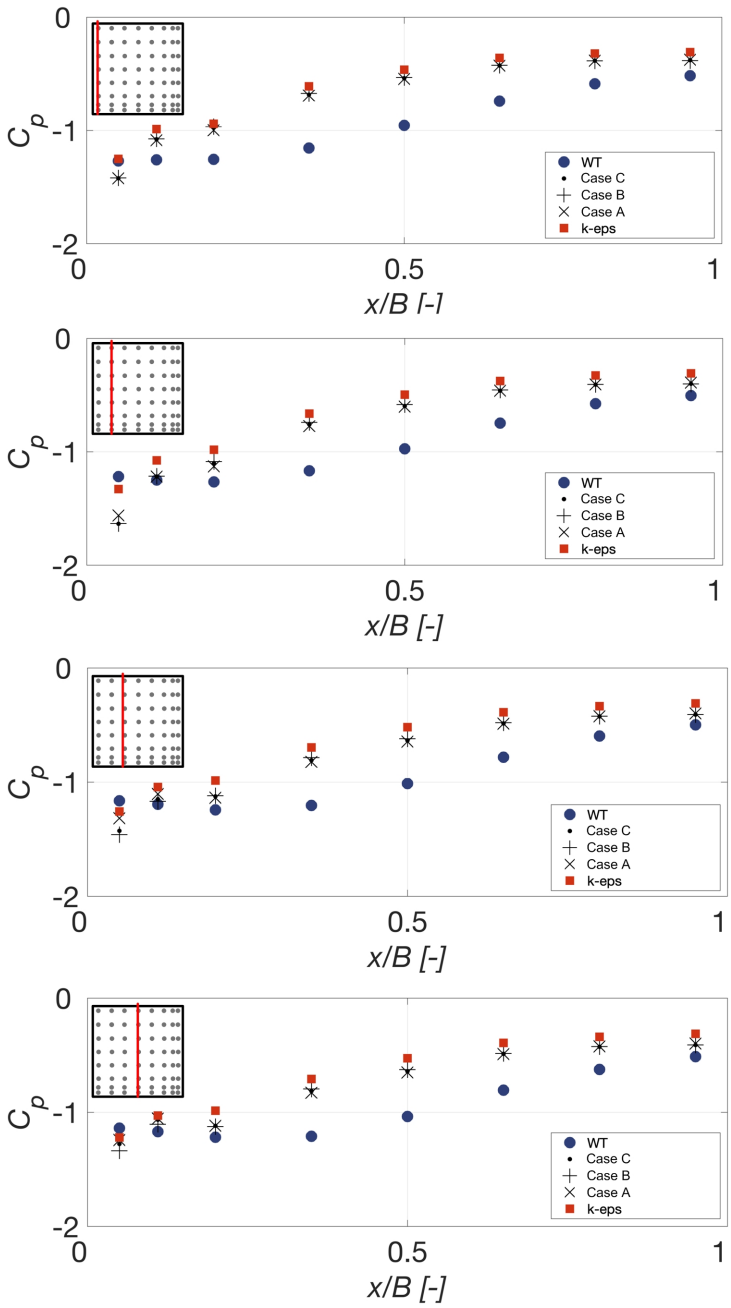


图14：沿屋顶标记的红线分布的测压孔位置处的平均压力系数；风洞数据（WT）、k-ε模型（k-ε）与三集合卡尔曼滤波案例的对比