arXiv:2603.02086v1 [math.NA] 2 Mar 2026

基于强化学习的对流主导流动滤波器：无参考数据与参考引导训练

ANNA IVAGNES[1], *、MARIA STRAZZULLO[2], * 和 GIANLUIGI ROZZA[1], *

Abstract. 我们提出一种强化学习（RL）框架，用于在不可压缩湍流的演化–滤波（EF）正则化策略中动态选择滤波参数。该框架不再通过启发式方法预先设定滤波半径，而是让强化学习智能体学习随时间自适应地控制滤波强度，从而兼顾数值稳定性与物理精度。

我们在两个动力学特性截然不同的基准问题上评估该方法：圆柱绕流和衰减均匀湍流。研究考察了参考引导和无参考两种奖励函数形式。在参考引导设置中，智能体利用有限时间窗口内的直接数值模拟（DNS）数据进行训练，随后通过外推评估其表现。在无参考数据设置中，奖励仅依赖基于物理的量，不使用参考解，从而完全消除了与直接数值模拟相关的计算成本。

结果表明，所提出的 RL-EF策略能够防止数值爆炸，同时避免标准 EF 方法中基于固定科尔莫戈洛夫长度尺度时常见的过度耗散。学习得到的策略能够准确再现各尺度上的相关流动动力学，并保持大尺度结构与小尺度耗散之间的正确平衡。值得注意的是，无参考奖励取得了与有参考方法相当的性能，这表明即使没有直接数值模拟数据，也能学习到稳定且具有谱一致性的滤波策略，从而大幅降低训练阶段的计算成本。

所提出的框架为人工调节正则化参数提供了一种稳健而灵活的替代方案，使湍流模拟中的滤波控制能够自适应地实现，并符合物理规律。

1. Introduction

在欠分辨或边缘分辨的模拟中，对流主导流动面临重大挑战，也就是说，所使用的计算网格未细化至科尔莫戈洛夫尺度。在这类情况下，纳维–斯托克斯方程的经典离散方法（如有限元或有限体积策略）可能会产生伪振荡、失稳以及非物理的能量累积 [5, 34]。为解决这些问题，通常会采用稳定化机制，包括大涡模拟（LES） [10, 35], 、雷诺平均纳维–斯托克斯（RANS）建模 [5, 34, 47, 22, 2, 36, 70], 以及基于正则化的方法。正则化（这些方法的概述见 [34]）是一类利用空间滤波来提高数值流动模拟准确度的数值格式；与直接数值模拟（DNS）相比，它使采用较粗糙的基于网格的离散成为可能，同时保持准确度并降低计算成本。在各种正则化方法中，演化–滤波（EF）策略已成为一种有效且计算成本低廉的框架。其核心思想是在模拟的每个时间步重复执行：(i) 演化 步骤，例如采用标准的纳维–斯托克斯离散方法推进速度；(ii) 空间滤波 阶段，对演化后的速度进行平滑，以减轻伪振荡。微分滤波器（DF）是广泛应用的一类空间滤波器。微分滤波器的成功归功于其数学稳健性 [5]。

[1] mathLab, Mathematics Area, SISSA, via Bonomea 265, I-34136 Trieste, Italy[2] Politecnico di Torino, Department of Mathematical Sciences “Giuseppe Luigi Lagrange”, Corso Ducadegli Abruzzi, 24, 10129, Torino, Italy* INdAM-GNCS group member

EF 型方法可以轻松集成到现有求解器中，这解释了它们为何广泛应用于各种离散化范式 [8, 9, 12, 14, 15, 17, 18, 19, 20, 65, 39, 27, 40, 42, 49,57, 61, 53, 55] 和众多应用领域 [19, 21, 7, 66, 43, 67, 60, 34]。然而，EF公式中的一个核心问题是滤波半径 [6, 19, 55, 54, 61, 5, 69, 68]的选择。该方法对这一参数高度敏感：滤波不足可能导致不稳定，而滤波过度则会造成过度扩散的解，并使具有物理意义的结构丢失。大多数现有方法根据网格尺寸或科尔莫戈洛夫长度尺度等启发式依据，为空间和时间指定恒定的参数值。近期研究通过滤波器中的非线性指示函数考虑了空间依赖参数 [19], ，并通过针对参考DNS数据进行优化来考虑时间依赖参数。然而，前一种方法面临如何恰当选择指示函数的问题；后一种方法则严重依赖每个时间点的数据（通常意味着极高的计算成本），且无法推广到校准窗口之外。本文将参数优化重新表述为一个序贯决策问题。我们不再独立优化每个时间步的滤波参数，而是通过强化学习（RL）策略学习一个随时间变化的控制策略。

强化学习（RL）是一类机器学习方法，智能体通过与环境交互，并最大化累积奖励信号 [64, 26] 来学习策略。与监督学习不同，强化学习不需要带标签的输入—输出对；相反，它通过连续的试错交互来发现最优动作。这一范式尤其适用于控制和决策问题，因为动作会影响未来状态，因此它自然适合用于随时间变化的参数选择。在计算流体力学中，强化学习近期已应用于流动控制、湍流建模、亚格子尺度闭合建模和自适应数值稳定化 [30, 29, 28, 44, 45, 58, 16, 13, 62, 48, 41]。这些应用表明，即使在复杂的非线性状态下，强化学习也能够直接从高维流场状态中学习非平凡的控制策略。在本研究中，我们利用这些优势，在演化-滤波框架内动态调节滤波强度，将参数选择构建为由流动动力学演变驱动的自适应控制问题。

以往研究中的强化学习智能体训练阶段大多依赖参考数据，而本文同时考虑基于 DNS 参考数据和无 DNS 参考数据的形式。下文将基于强化学习的无参考数据滤波器形式称为无数据（ DF-RL-EF），这意味着奖励构造不使用任何直接数值模拟参考数据；与之相对的是基于参考数据的数据驱动设置（DD-RL-EF）。具体而言，在 DD-RL-EF 设置中，奖励会在有限的训练区间内纳入参考数据，随后对学习得到的策略进行外推评估。在 DF-RL-EF 中，奖励完全根据具有物理意义的量构造，不依赖参考解。该形式能够灵活设计奖励，包括纳入全局不变量或保结构约束。由此得到的框架以自适应的状态依赖控制取代静态参数校准。对具有代表性的对流主导流动开展的数值实验表明，学习得到的策略在数值稳定性与物理精度之间实现了良好平衡，并且在不同动力学状态下均表现出稳健性。本研究的主要创新点包括：

- 提出了一种新颖的、由强化学习选择的滤波半径选取方法，用于演化-滤波策略，从而保证随时间对滤波过程进行非平凡的自适应选择；与参考数据相比，该方法能够保留大尺度结构，并维持流动模拟的主要特征；• 更为重要的是，研究了一种不依赖直接数值模拟数据的强化学习-EF算法，该算法基于方程残差、基于梯度的指标或保持结构的指标等物理量，能够以显著降低的计算成本提供稳定且可靠的滤波操作。

本文结构安排如下。第2节回顾 EF公式并讨论参数敏感性。第3节介绍强化学习理论、强化学习-EF方法以及所提出的奖励函数形式。第4节介绍二维流体动力学基准问题的数值研究

：(i) $Re = 1000$条件下的圆柱绕流，以及 (ii) $Re = 40\,000$条件下的衰减均匀湍流。结论与未来研究方向将在第5节概述。

2. 问题表述与演化-滤波策略

本节介绍我们所研究方法的核心内容。具体而言，我们将在第 2.1 节简要介绍离散纳维–斯托克斯方程（NSE）和演化与滤波（EF）策略。感兴趣的读者可参阅 [63, 55, 20, 6]，了解 EF 的更多详情。

2.1. NSE 的 EF 算法。在所提出的数值研究中，我们聚焦于不可压缩纳维–斯托克斯方程。在连续层面，对于物理区域 $\Omega \subset \mathbb{R}^d$，其中 $d = 2,3$，我们通过以下方程描述速度为 $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t) \in \mathbb{U}$ 、压力为 $p = p(\boldsymbol{x}, t) \in \mathbb{Q}$ 的不可压缩流体的运动：

$$\begin{cases} \dfrac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} - \nu \Delta \boldsymbol{u} + \nabla p = 0 & \text{in } \Omega \times (t_0, T), \\ \nabla \cdot \boldsymbol{u} = 0 & \text{in } \Omega \times (t_0, T), \\ \boldsymbol{u} = \boldsymbol{u}_D & \text{on } \Gamma_D \times (t_0, T), \\ -p\boldsymbol{n} + \nu \dfrac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}} = 0 & \text{on } \Gamma_N \times (t_0, T), \end{cases} \tag{1}$$

给定区域 $\Omega \times \{t_0\}$内的初始条件 $\boldsymbol{u} = \boldsymbol{u}_0$，其中 $\Gamma_D \cup \Gamma_N = \partial\Omega$，$\Gamma_D \cap \Gamma_N = \varnothing$，且$\nu$ 为运动黏度，其中 $\mathbb{U}$ 和 $\mathbb{Q}$ 为适当的希尔伯特函数空间。在 Γ_D 上，给定函数 $\boldsymbol{u}_D$ 被指定为狄利克雷边界条件，而在 Γ_N 上施加“自由流动”边界条件。流动状态由无量纲雷诺数决定，其定义为

$$Re = \frac{UL}{\nu}, \tag{2}$$

其中，U 和 L 分别是当前问题的特征速度和特征长度。对于较大的 Re，惯性力占主导，黏性力相对较弱，此时的流动状态称为对流主导。在所提出的数值测试中，我们分别采用 $\mathbb{P}^2$-$\mathbb{P}^1$ Taylor-Hood 有限元方法（*FEM*）和一阶后向差分公式（BDF1）进行空间和时间离散。对于欠分辨或临界分辨的数值模拟，标准空间离散技术可能会产生伪数值振荡。为解决这一问题，我们采用基于 EF 的算法。给定时间步 Δt，我们考虑时间实例 $t_n = t_0 + n\Delta t$，其中 $n = 0, \ldots, N$，最终时间为 $T = N\Delta t$。在时间 t_n，$\boldsymbol{u}_n \in \mathbb{U}^{N_h^{\boldsymbol{u}}}$和$p_n \in \mathbb{Q}^{N_h^p}$ 分别表示半离散 FEM 速度解和压力解，其中 $N_h^{\boldsymbol{u}}$和 N_h^p是与有限元空间近似 $\mathbb{U}^{N_h^{\boldsymbol{u}}}$和 $\mathbb{Q}^{N_h^p}$相关的自由度。对于预先确定的滤波半径 δ，时间 t_{n+1} 处的 EF 算法如下：

(I) Evolve:

$$\begin{cases} \dfrac{\boldsymbol{w}_{n+1} - \boldsymbol{u}_n}{\Delta t} + (\boldsymbol{w}_{n+1} \cdot \nabla)\boldsymbol{w}_{n+1} - \nu \Delta \boldsymbol{w}_{n+1} + \nabla p_{n+1} = 0 & \text{in } \Omega \times \{t_{n+1}\}, \\ \nabla \cdot \boldsymbol{w}_{n+1} = 0 & \text{in } \Omega \times \{t_{n+1}\}, \\ \boldsymbol{w}_{n+1} = \boldsymbol{u}_D & \text{on } \Gamma_D \times \{t_{n+1}\}, \\ -p_{n+1} \cdot \boldsymbol{n} + \dfrac{\partial \boldsymbol{w}_{n+1}}{\partial \boldsymbol{n}} = 0 & \text{on } \Gamma_N \times \{t_{n+1}\}. \end{cases}$$

(II) Filter:

$$\begin{cases} -2\delta^2 \, \Delta \boldsymbol{u}_{n+1} + \boldsymbol{u}_{n+1} + \nabla \lambda_{n+1} = \boldsymbol{w}_{n+1} & \text{in } \Omega \times \{t_{n+1}\}, \\ \nabla \cdot \boldsymbol{u}_{n+1} = 0 & \text{in } \Omega \times \{t_{n+1}\}, \\ \boldsymbol{u}_{n+1} = \boldsymbol{u}_D & \text{on } \Gamma_D \times \{t_{n+1}\}, \\ \dfrac{\partial \boldsymbol{u}_{n+1}}{\partial \boldsymbol{n}} = 0 & \text{on } \Gamma_N \times \{t_{n+1}\}. \end{cases}$$

该过程分为两个步骤：**(I)**针对时间 t_{n+1}处给定的输入 $\boldsymbol{u}_n$ 求解标准纳维–斯托克斯方程（NSE），得到演化后的速度 $\boldsymbol{w}_{n+1}$。步骤**(II)** 应用滤波半径为 δ的微分滤波器（DF），该半径即滤波操作的空间长度尺度。DF 利用椭圆算子平滑中间流动速度 $\boldsymbol{w}_{n+1}$的小尺度（即高频），从而发挥空间滤波器的作用。最后，滤波后的速度 $\boldsymbol{u}_{n+1}$ 将作为时间积分格式的下一步输入。在该算法中， λ_{n+1} 是一个辅助变量，用于保证 $\boldsymbol{u}_{n+1}$的无散约束。按照 [55], 的记法，我们用缩写**noEF** 表示未经正则化的离散 NSE 解。

注释 2.1（斯托克斯滤波器）。为了在步骤 **(II)**中保持不可压缩约束，我们采用斯托克斯 DF [12, 33]。也可以采用其他滤波操作。例如，可以不强制满足无散性质 $\nabla\cdot\boldsymbol{u}_{n+1}$，而是采用散度-梯度稳定化 [23, 25]: 具体而言，在 DF 方程中加入如下形式的一项

$$\tilde{\gamma}\nabla(\nabla\cdot\boldsymbol{u}_{n+1}), \tag{3}$$

并引入 $\tilde{\gamma}$，根据测试案例进行校准。关于这一主题的概述，读者可参阅 [11, 25, 32]。尽管采用散度-梯度稳定化的 DF 更易于编写代码，但本文采用斯托克斯滤波器，因为它与纳维–斯托克斯方程模型相一致且具有鲁棒性（除 δ 外，无需进行参数校准，见注释 2.3）。

注释 2.2（松弛步骤）。文献中还采用了另一种有价值的算法来处理对流主导情形：演化-滤波-松弛（EFR）策略。事实上，在演化-滤波方法中，若滤波半径 δ 选择不当，可能会导致过度扩散模拟 [54, 24]。EFR 策略沿用 EF 算法，但新的输入 $\boldsymbol{u}_{n+1}$ 是演化后的速度场 $\boldsymbol{w}_{n+1}$ 与滤波速度场 $\boldsymbol{w}_{n+1}$ 的凸组合，即 —

$$\boldsymbol{u}_{n+1}=(1-\chi)\boldsymbol{w}_{n+1}+\chi\overline{\boldsymbol{w}}_{n+1},$$

对于选定的松弛参数 $0\leqslant\chi\leqslant 1$。然而，在 [54] 中，作者质疑其应用，因为在二维绕圆柱流动测试案例的松弛步骤中出现了伪数值振荡；该案例也是本文所研究的测试案例之一。此外，正如在 [24], 中已经讨论过的，必须谨慎选择参数 χ，以避免非物理模拟。此外，在 [24] 中同时优化 δ 和 χ，所得结果与在演化-滤波方法中仅优化 δ 的结果相当（见注释 2.3）。基于这些原因，本文拟仅依靠 EF 算法，采用基于强化学习的策略来寻找最优的 δ。

注释 2.3（ δ 的经典取值）。 在计算流体力学应用中， δ 的最优取值仍是一个悬而未决的问题。文献中提出了若干种尺度分析方法。例如，在[55, 6], 中，作者将 δ 选为最小单元直径 h_{min}；对于非结构网格，则选为柯尔莫哥洛夫尺度 $\eta=L\cdot Re^{-\frac{3}{4}}$。在 [38], 中，取值则基于能量分析。

演化-滤波可能会受到未经调优的 δ 参数影响： δ 取值过大将产生过度扩散结果[54], 而取值过小则可能不足以平滑伪振荡。在 [24],中，作者提出对滤波半径随时间进行优化，利用少量DNS数据点作为参考，以妥善处理数值模拟中层流与湍流状态之间的转捩。然而，他们的数据驱动策略严重依赖数据，且未给出外推结果。接下来，我们提出一种基于强化学习的增强型参考引导的数据驱动策略，能够随时间预测滤波操作且具有稳健性。

3. Reinforcement Learning for Evolve-Filter

本节简要概述强化学习（RL）作为一种决策框架，介绍深度 Q 网络（DQN）算法作为一种强化学习方法（第3.1节），并说明如何将这些概念应用于开发基于强化学习的演化–滤波（RL-EF）算法，以优化 EF 模拟中的滤波（第3.2节）。从本节开始， $\boldsymbol{u}_n$ 将不再表示速度场的半离散有限元近似，而是表示其向量，该向量的各个分量为 $\mathbb{R}^{N_h^{\boldsymbol{u}}}$ 中有限元展开的系数。

3.1. 强化学习简介。强化学习（RL）是一种以通过交互进行学习为原则的范式，智能体通过与动态环境交互来学习如何做出决策。学习过程通常在马尔可夫决策过程（MDP）框架下进行建模。具体而言，MDP由一个元组 $(\mathcal{S},\ \mathcal{A},\ P,\ r,\ \gamma)$定义，其中 $\mathcal{S}$表示状态空间，$\mathcal{A}$ 表示动作空间，P 表示状态转移概率，r 表示奖励函数，而 $\gamma \in (0,\ 1]$ 表示折扣因子。智能体与环境之间的交互以回合为单位组织，每个回合由有限个训练决策步骤组成。在给定的决策步骤 t_n，环境由状态 $s_n = s(t_n) \in \mathcal{S}$刻画。智能体根据对当前状态的观察，按照策略 $\pi(a_n \mid s_n)$选择动作 $a_n = a(t_n) \in \mathcal{A}$；该策略定义了在给定状态 s_n时选择动作 a_n 的概率。由于所选动作的作用，环境遵循转移规律 $P(s_{n+1} \mid s_n,\ a_n)$转移到新状态 s_{n+1}，智能体则获得一个标量奖励 r_{n+1}。智能体的目标是学习最优策略，使其随时间推移最大化期望累积（可能经过折扣的）奖励。具体而言，对于给定回合，累积奖励（或回报）定义为

$$R_{\text{train}} = \sum_{n=0}^{N_{\text{train}}-1} \gamma^n r_{n+1},$$

其中，N_{train} 表示回合长度，即每个回合中的时间步数；$\gamma \in (0,\ 1]$为折扣因子。通过与环境反复交互，智能体逐步改进其策略，从而能够泛化到训练期间可能未曾明确遇到的状态，并因此具备时间外推能力。

3.1.1. 深度Q网络算法。 在本文中，我们采用深度 Q网络（DQN） [37]，这是一种基于价值的强化学习算法，用于建模 Q函数，即动作价值函数:

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k r_{n+k} \,\middle|\, s_n = s, a_n = a\right].$$

在上述表达式中，$\mathbb{E}_{\pi}[\cdot]$ 表示相对于策略 π 和环境动态所引起的随机性的期望，即对从 $(s,\ a)$开始遵循策略 π 所生成的所有可能未来状态–动作轨迹取期望。无穷级数表示沿未来轨迹收集的折扣累积奖励，其中折扣因子 γ 确保收敛，并降低较晚时间步所获奖励的重要性。在DQN中，Q函数通过深度神经网络进行近似。 Q网络表示一种非线性映射

$$Q_{\theta} : \mathcal{S} \times \mathcal{A} \to \mathbb{R},$$

该函数估计与每个状态–动作对相关的期望累积奖励。策略并不是显式学习得到的，而是通过贪心选择规则从 Q-函数中诱导得到的，

$$a_n = \pi(s_n) = \arg\max_{a \in \mathcal{A}} Q_{\theta}(s_n, a),$$

在训练过程中还可能结合探索策略。在 DQN 中，使用了两个神经网络：参数为 θ的在线 Q-网络，该网络在训练过程中更新；以及参数为 θ^{-}的目标Q-网络，该网络会定期与在线网络同步，以使学习过程更加稳定。通过最小化时序差分损失，对 Q-网络参数 θ 进行优化

$$\mathcal{L}(\theta) = \mathbb{E}\Big[\big(r_{n+1} + \gamma \max_{a'} Q_{\theta^{-}}(s_{n+1}, a') - Q_{\theta}(s_n, a_n)\big)^2\Big],$$

其中 θ^{-} 表示目标网络的参数。若要更深入地了解强化学习，读者可参阅 [56]。

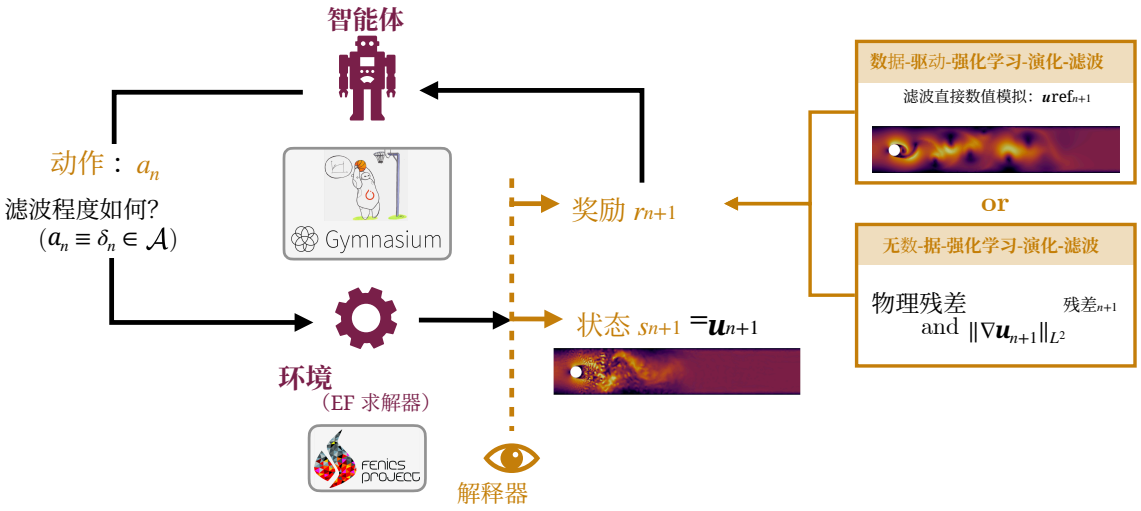


FIGURE 1. RL-EF训练回合中的示意性时间步进机制及相关软件。

3.2. 强化学习-EF算法。 我们将第3.1节所述的强化学习技术应用于演化-滤波模拟，提出基于强化学习的演化-滤波算法（**RL-EF**）。更具体地说，RL-EF是一种优化的EF模拟，其中滤波半径 δ_n 随时间变化，并在每个时间步由预训练强化学习智能体更新。在所提出的设定中，动作与时刻 t_n所采用的滤波半径一致，即 $a_n = \delta_n$，而观测（状态） s_{n+1} 对应于使用强化学习智能体所选滤波半径得到的向量值滤波速度场 $\boldsymbol{u}_{n+1}(\delta_n)$ 。因此，动作空间 $\mathcal{A}$ 取50个离散值，由 $[10^{-10},\ 10^{-7}]$ 中的四个对数间隔值以及$[10^{-6},\ 10^{-3}]$中的四十六个对数间隔值组成，如图2所示。这一选择背后的逻辑在于滤波对速度场的影响：量级从 $\mathcal{O}(10^{-10})$ 到 $\mathcal{O}(10^{-7}$的较小滤波值会产生相似的EF解，例如 $\boldsymbol{u}_n(10^{-10}) \approx \boldsymbol{u}_n(10^{-7})$，而较大的值对解有显著影响，例如 $\boldsymbol{u}_n(10^{-6}) \not\approx \boldsymbol{u}_n(10^{-3})$。因此，我们在范围 $[10^{-6},\ 10^{-3}]$内考虑更多取值。

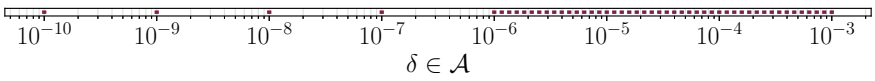


FIGURE 2. RL-EF中的动作空间。

观测空间则取决于EF模拟所采用离散化的自由度数量，即 $\mathcal{S} \equiv \mathbb{R}^{N_h^{\boldsymbol{u}}}$ 。在我们的设定中，一个强化学习回合对应一系列物理时间步，也就是EF模拟在指定训练时间窗口内的时间步，如图1所示。训练过程完成后，强化学习智能体会确定一条优化轨迹

$$\{(\boldsymbol{u}_1, \delta_1), \ldots, (\boldsymbol{u}_N, \delta_N)\}$$

在整个时间段内 通过从缩短的训练窗口 $(0,\ T_{\text{train}}]$ 中学习，并结合 $T_{\text{train}} \ll T$。

注 3.1（近端策略优化）。 也可以采用其他强化学习策略，例如近端策略优化（PPO）算法 [50], ，它使我们能够考虑连续动作

空间。我们在测试案例中也尝试了这种方法；然而，在解的精度和训练稳定性方面，我们得到的结果较差。我们并不排除，对网络架构进行更深入的实证研究和微调可能会得到更准确的结果。然而，这超出了本文的主要范围。

3.2.1. *RL-EF*奖励。 在强化学习的背景下，自然会产生一个问题：应采用什么样的奖励，才能让系统学习到合适的滤波操作？ 在我们的数值研究中，我们考虑两类奖励：一类是**数据驱动**的，其依据是RL-EF解相对于某些预先收集的参考数据的准确度；另一类是**无数据**的，其依据是物理残差和RL-EF的速度梯度。因此，通用的强化学习-EF方法可具体分为**DD-RL-EF** （基于数据的强化学习演化-滤波）和**DF-RL-EF** （无数据的强化学习演化-滤波）。在数据驱动设置中，我们使用参考直接数值模拟的数据，该模拟在细化网格上进行，然后将其投影到RL-EF所使用的基本分辨网格上。从现在起，我们将参考速度解称为**滤波直接数值模拟**，并将其表示为 $\boldsymbol{u}_n^{\mathrm{ref}}$在时间步 t_n。

关于奖励，如 [30, 29, 28], 所示，我们考虑每个时间步的奖励的一般形式如下 t_{n+1}:

$$r_{n+1}^{\mathrm{DD}} = 2\exp\left(-\frac{1}{\alpha}\mathrm{e}_{n+1}\right) - 1, \tag{4}$$

其中，e_{n+1} 是一个度量指标，用于衡量 RL-EF 解 $\boldsymbol{u}_{n+1}(\delta_n)$ 相对于 DNS 参考解 $\boldsymbol{u}_{n+1}^{\mathrm{ref}}$的误差。表达式（4）在强化学习中被广泛用于将误差 e_{n+1} 映射到区间$[-1,\ 1]$。这一变换可提高稳定性，并促进训练阶段的收敛。

备注 3.2。 在文献中，将强化学习应用于流体动力学时，所考虑的误差度量指标通常是前几个（能量较高的）波数处的相对能谱差，即:

$$\mathrm{e}_{n+1}^{\mathrm{spectrum}} = \frac{1}{\kappa_{\max}}\sum_{\kappa=1}^{\kappa_{\max}}\left(\frac{\mathcal{E}_{n+1}(\kappa) - \mathcal{E}_{n+1}^{\mathrm{ref}}(\kappa)}{\mathcal{E}_{n+1}^{\mathrm{ref}}(\kappa)}\right)^2, \tag{5}$$

其中， $\mathcal{E}(\kappa)$和 $\mathcal{E}^{\mathrm{ref}}(\kappa)$分别表示时间步 t_{n+1}、波数 κ 处的 RL-EF 能量和滤波后直接数值模拟的能量。 $\kappa_{\max}$ 是误差计算中考虑的最大波数，通常通过经验测试选取，并取决于测试案例。初步实验表明，在我们的测试案例中，能谱误差度量指标（5）对 RL-EF 的效果不佳。因此，我们研究以下数据驱动误差，它等同于 RL-EF 速度与参考解之间的相对误差:

$$\mathrm{e}_{n+1}^{\mathrm{DD}} = \left(\frac{\|\boldsymbol{u}_{n+1} - \boldsymbol{u}_{n+1}^{\mathrm{ref}}\|_2}{\|\boldsymbol{u}_{n+1}^{\mathrm{ref}}\|_2}\right)^2, \tag{6}$$

其中，符号 $\|(\cdot)\|_2$ 表示区域 Ω 中的 L^2 范数。

在我们的测试中，**DD-RL-EF** 奖励由公式（4）给出，其中 $\alpha = 1$ 和 $\mathrm{e}_{n+1} = \mathrm{e}_{n+1}^{\mathrm{DD}}$。关于 DD-RL-EF，自然会产生两个问题:

(i) 一个强化学习回合应考虑多少个时间步？(ii) 每个回合都应从模拟的起点开始吗，例如 $t = 0$?

为回答这些问题，我们考虑了 DD-RL-EF 的两个变体:

- 一种标准的 DD-RL-EF，我们简称为**DD-RL-EF**，其中每个回合都从 $t = 0$ 开始，持续 T_{train} （对应 N_{train} 个时间步）。在我们的实验中，设定 $T_{\mathrm{train}} = T/4$。在这种情况下，每个回合的累积奖励为:

$$R^{\mathrm{DD}} = \sum_{n=1}^{N_{\mathrm{train}}} r_n^{\mathrm{DD}}. \tag{7}$$

• **DD-RL-EF-rand**，其中每个回合都从训练窗口 $n_{\text{rand}} \in \{1,\ 2,\ \ldots\ldots,\ N_{\text{train}}-1,\ N_{\text{train}}\}$ 内随机选取的时刻开始。在这种设置下，与标准 DD-RL-EF 情况相比，回合长度有所缩短，设为 $T_{\text{randtrain}} = T/10,$，对应 $N_{\text{train}}^{\text{rand}}$ 个时间步。与从 $t=0$ 初始化的标准回合相比，这种策略有助于探索范围更广的状态。累积奖励为：

$$R_{\text{rand}}^{\text{DD}} = \sum_{n=n_{\text{rand}}}^{N_{\text{train}}^{\text{rand}}} r_n^{\text{DD}}. \tag{8}$$

另一方面，在无数据设定中，我们考虑如下奖励表达式：

$$r_{n+1}^{\text{DF}} = \frac{1}{2}\Big(2\exp\left(-\alpha_{\text{res}}\,\text{res}_n\right)-1\Big) + \frac{1}{2}\left(2\exp\left(-\left|\frac{1}{\alpha_{\text{grad}}\|\nabla \boldsymbol{u}_{n+1}\|_2 - \|\nabla \boldsymbol{u}_n\|_2)}\right|\right)-1\right), \tag{9}$$

其中包含两项贡献：

$$\text{(i)}\ 2\exp\left(-\alpha_{\text{res}}\,\text{res}_{n+1}\right)-1,\ \text{and}\ \text{(ii)}\ 2\exp\left(-\left|\frac{1}{\alpha_{\text{grad}}(\|\nabla \boldsymbol{u}_{n+1}\|_2 - \|\nabla \boldsymbol{u}_n\|_2)}\right|\right)-1,$$

其中 α_{res} 和 α_{grad} 为正实数。两部分均具有公式（4）所给出的相同形式：

（i）该项贡献与有限元系统的残差res_{n+1}直接相关。当物理残差增大时，奖励随之减小，即当RL-EF预测不是公式（1）中离散纳维-斯托克斯方程的解时。该分量对过小的 δ_n值加以惩罚，因为这些值无法稳定系统。正如在noEF模拟中所观察到的，此类行为会导致欠分辨方程，并随后使残差爆破。

（ii）该项贡献则与速度空间梯度的时间变化相关。梯度的微小变化表现为过度扩散行为，如在 δ 取值较大（例如 $\delta=\eta$）的EF模拟中所见。与残差项不同，该分量对过大的 δ_n 值加以惩罚。事实上，当速度梯度随时间变化更大时，智能体会获得更高的奖励，从而更好地捕捉由其时间变化所预期的流动特征。

常数 α_{res} 和 α_{grad} 在初步经验研究后进行调优。在我们的测试中，**DF-RL-EF** 奖励为 r_{n+1}^{DF}，其中 $\alpha_{\text{res}} = 1\times10^5$ 和 $\alpha_{\text{grad}} = 1\times10^4$。因此，我们将每个回合的累积奖励表示为：

$$R^{\text{DF}} = \sum_{n=1}^{N_{\text{train}}} r_n^{\text{DF}}. \tag{10}$$

注记 3.3（为何不用DF-RL-EF-rand？）. 为保持完全无数据的方法，我们不采用在随机时间步 $n=n_{\text{rand}}$初始化回合的策略，因为这将需要访问滤波DNS状态$\boldsymbol{u}_{n_{\text{rand}}}^{\text{ref}}$，才能在不损害预测准确度的情况下开始时间积分。

最后，奖励公式可加以调整，以纳入弱结构保持约束，例如能量守恒或涡量平方守恒；这些约束不需要访问参考数据，而只需RL-EF解。在此情形下，我们考虑DF-RL-EF的结构保持变体（**SP-DF-RL-EF**），其中奖励包含针对能量（$\mathcal{E}_n$，时间 t_n）和涡量平方（$\mathcal{Z}_n$，时间 t_n）增长的若干惩罚项：

$$R_{\text{SP}}^{\text{DF}} = \sum_{n=1}^{N_{\text{train}}}\left[r_n^{\text{DF}} + \mathbf{1}_{\{\mathcal{E}_n>\mathcal{E}_{n-1}\}}\left(\exp\left(-\frac{\mathcal{E}_n-\mathcal{E}_{n-1}}{\alpha_{\mathcal{E}}\,\mathcal{E}_{n-1}}\right)-1\right) + \mathbf{1}_{\{\mathcal{Z}_n>\mathcal{Z}_{n-1}\}}\left(\exp\left(-\frac{\mathcal{Z}_n-\mathcal{Z}_{n-1}}{\alpha_{\mathcal{Z}}\,\mathcal{Z}_{n-1}}\right)-1\right)\right].$$

在上述表达式中，$\mathbf{1}_{\{\mathcal{E}_n>\mathcal{E}_{n-1}\}}$ 和 $\mathbf{1}_{\{\mathcal{Z}_n>\mathcal{Z}_{n-1}\}}$ 分别表示仅在能量和涡量平方积分增长时才会激活的指示函数。在这种情况下，会在奖励中加入与能量和/或涡量平方积分增长相关的额外惩罚项。我们将与惩罚项相关的参数调至 $\alpha_{\mathcal{E}}=\alpha_{\mathcal{Z}}=0.1$。表1总结了本文所介绍的 RL-EF 方法所使用的缩写。

TABLE 1. 所提出的 RL-EF 方法的缩写及主要特征。

缩写	回合开始	回合长度 累计奖励		所需数据
DD-RL-EF	$t=0$	$T_{\text{train}}=T/4$	R^{DD}	$\boldsymbol{u}_n^{\text{ref}}, t_n \in [0, T_{\text{train}}]$
DD-RL-EF-rand	随机 $t \in [0, T_{\text{train}}]$	$T_{\text{randtrain}}=T/10$	$R^{\text{DD}}_{\text{rand}}$	$\boldsymbol{u}_n^{\text{ref}}, t_n \in [0, T_{\text{train}}+T^{\text{rand}}_{\text{train}}]$
DF-RL-EF	$t=0$	$T_{\text{train}}=T/4$	R^{DF}	无需参考
SP-DF-RL-EF	$t=0$	$T_{\text{train}}=T/4$	$R^{\text{DF}}_{\text{SP}}$	无需参考

软件：如图1所示，用于开发 RL-EF 的库包括**Gymnasium** [59, 31]„ 这是一个基于 Python 的标准化库，用于定义强化学习环境；以及**StableBaselines 3** [46], ，这是一个构建于 Gymnasium 之上的库，提供可靠且经过充分测试的强化学习算法。该强化学习环境与开源的、基于 Python 的有限元方法框架 FEniCSX [4, 3, 52, 51],集成，用于定义仿真设置和 EF 时间步进。

4. NUMERICAL RESULTS

本节旨在展示 RL-EF 方法在以下两个测试案例中的潜力：

(i) 雷诺数为 $Re=1000$ 的二维圆柱绕流（第 4.1 节）；(ii) 雷诺数为 $Re=40\,000$ 的二维衰减均匀湍流测试案例（第 4.2 节）。

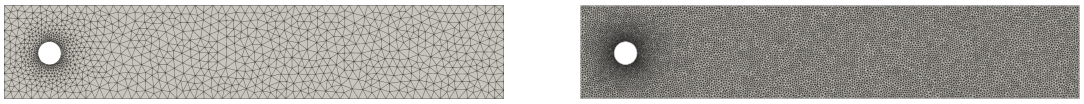
这里回顾一下不同类型模拟所使用的记号：

• **直接数值模拟**：在最精细网格上进行的高成本模拟。• **滤波直接数值模拟**：用于 RL 参考引导训练和对比的参考数据，通过将直接数值模拟插值到粗网格上定义的函数空间中获得。• **noEF**：在粗网格上进行的直接数值模拟；根据网格分辨率的不同，该模拟可能会发散，因此需要进行稳定化处理。• **EF（Kolmogorov）**：应用于粗网格的标准 EF 方法，其中 $\delta_n=\delta=\eta \quad \forall n=1, \ldots\ldots, N$。• **RL-EF**：在粗网格上进行的新型 EF 模拟，其中一个经小时间窗预训练的强化学习智能体会为未见过的时间实例预测时变滤波半径 δ_n。数据驱动和无数据 RL-EF 策略的缩写及相应特征已在表1中介绍。

4.1. 测试案例1：圆柱绕流。 我们考虑在 $Re=1000$ 下绕圆柱的不可压缩流动测试案例，其中入口速度恒定不变。表2列出了速度场（$N_h^{\boldsymbol{u}}$）和压力场（N_h^p）所用的自由度数量，以及网格尺寸的最小值和最大值，分别对应于用于直接数值模拟的细网格，以及用于 noEF、EF 和 RL-EF 方法的边际分辨网格。网格也显示在图3中，其特点是由三角形单元构成。模拟在时间窗口 $[0,\ T]$ 内进行，$T=4$。粗网格模拟所用的时间步为 $\Delta t=4\times10^{-4}$，总计 $N=10\,000$ 个时间实例。

TABLE 2. 测试案例1。自由度数量，以及网格的最小和最大尺寸。

	$N_h^{\boldsymbol{u}}$	N_h^p	h_{min}	h_{max}
粗网格	9792	1274	$4.46e-3$	$4.02e-2$
细网格	43609	11036	$4.44e-3$	$1.13e-2$



(a) 粗网格　　(b) 细网格

FIGURE 3. 测试案例1。模拟中使用的两个网格。noEF、EF 和 RL-EF 模拟使用粗网格，而 DNS 参考模拟使用细网格。

区域为 $\Omega = \{(x,\ y) \in [0,\ 2.2] \times [0,\ 0.41]$，满足 $(x-0.2)^2 + (y-0.2)^2 \geqslant 0.05^2\}$。我们在 $\partial\Omega_{wall}$ 上采用无滑移边界条件（包括顶部、底部和圆柱边界），并在左边界设置入口速度 $\mathbf{u}_{in}$，其表达式如下：

$$\boldsymbol{u}_{in} = \left(\frac{6}{0.41^2}\, y\,(0.41 - y), 0 \right). \tag{11}$$

在区域出口（右边界），我们采用标准自由流边界条件。初始条件为 $\boldsymbol{u}_0 = (0,\ 0)$。黏度设为 $\nu = 1 \times 10^{-4}$。图4展示了滤波直接数值模拟、noEF 和 EF（Kolmogorov）案例中速度大小随时间变化的代表性解，并表明：

- noEF 表现出噪声行为，经过 $t = 1.5$ 后，能量无界增长，模拟发生崩溃；
- EF（Kolmogorov）表现出过度扩散行为，因为它在模拟结束前仅于 $t = 2$之后才在圆柱尾流中呈现涡脱落。

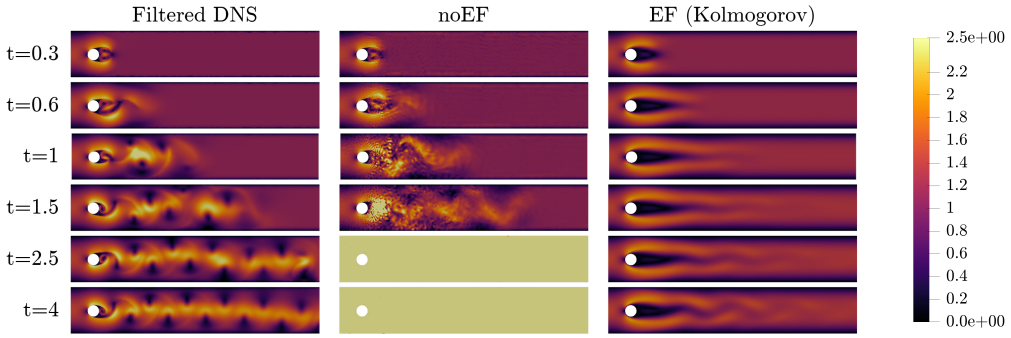


FIGURE 4. 测试案例1。采用 $\delta = \eta$时，滤波直接数值模拟、noEF和演化-滤波的速度大小场。图中展示了不同时间实例的场。

noEF和EF（Kolmogorov）所得的较差结果促使采用RL-EF。在展示RL-EF 模拟的定性结果之前，我们分析图5中的强化学习训练统计数据，尤其是各回合上的累积奖励以及各时间步上的DQN损失函数。首先，需要强调的是，可达到的最大累积奖励与每个回合的时间步数相一致 N_{train}。在DD-RL-EF和DF-RL-EF中，$N_{\text{train}} = N/4 = 2500$, 对应于 $T_{\text{train}} = 1$，而在DD-RL-EF-rand中，如第3.2节所述，$N_{\text{train}}^{\text{rand}} = 1000$对应于 $T_{\text{随机训练}} = 0.4$。可以立即注意到，所有方法都收敛到接近其可达到的最大累积奖励，这表明

它们成功地学习到近似最优策略。与DD-RL-EF相比，DF-RL-EF更早达到性能平台期，且波动更小，这可以从图5奖励曲线的放大框中看出。这取决于DF-RL-EF所采用的奖励表达式，该表达式包含残差贡献以及速度梯度的时间变化。这种形式连同参数 α_{res} 和 α_{grad}的选取，专门用于获得稳定的强化学习动态。在DD-RL-EF-rand训练中，累积奖励同样呈现较小的振荡，并迅速达到其最大值（1000）。每个回合初始时间的随机选择产生了更大的状态探索，并加速了智能体的学习。各回合和时间步上的损失演化证实了上述讨论：DD-RL-EF-rand和DF-RL-EF的损失比DD-RL-EF衰减得更快，并在第100个回合之前达到 $\sim 1 \times 10^{-4}$。另一方面，DD-RL-EF的损失达到了更大的平台期 1×10^{-3}。

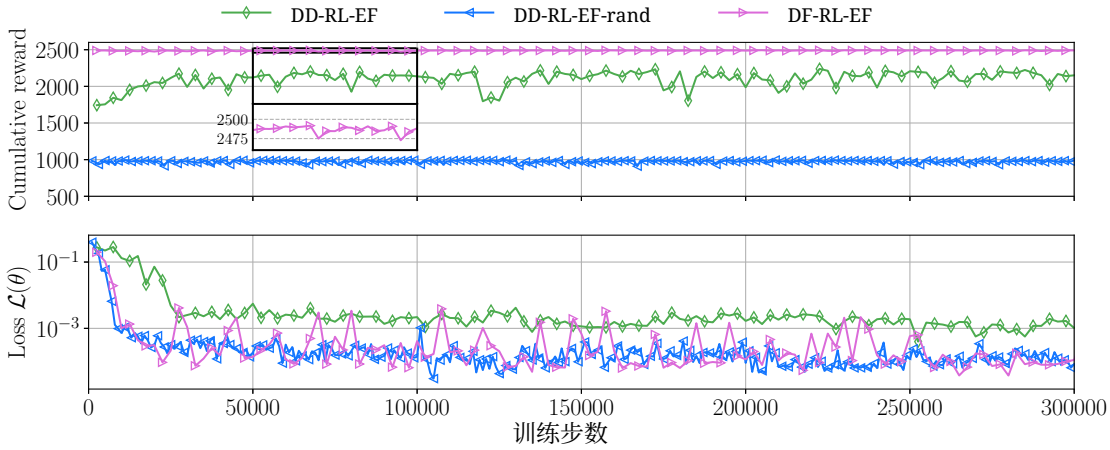


FIGURE 5. 测试案例1。RL-EF方法在训练步数中的累积奖励和损失值。

完成强化学习训练后，我们可以分析RL-EF策略在带有时间外推的在线模拟中的表现。图6展示了：(i) 预训练强化学习智能体在模拟时间内选择的动作 δ_n，$n = 1, \ldots\ldots, N$（左侧）；(ii) 动作频率，即回答以下问题：智能体选择某个动作的时间百分比是多少？

首先，我们可以立即看出，不同的智能体会选择不同的动作，但在所有算法中，动作值随时间呈现振荡且不规则的变化。与 [24], 中进行的初步研究类似，这表明在整个时间窗口内，模拟并不存在唯一且固定的最优 δ；而且根据模拟状态，在某些时间实例中，最优选择是不进行滤波，即 δ_n 的数量级介于 $\mathcal{O}(10^{-10})$ 和 $\mathcal{O}(10^{-7})$ 之间。实际上，强化学习智能体选择的值大多数时候都小于科尔莫戈洛夫长度尺度η，如图6中的灰色虚线所示。此外，与 DD-RL-EF 和 DD-RL-EF-rand 相比，DF-RL-EF 总体收敛到更大的 δ_n 值。这种行为表明，无数据奖励公式隐式地赋予维持模拟稳定性更高的重要性。因此，智能体倾向于更持续地进行滤波，动态调整滤波强度，而不是将其关闭。在奖励中加入基于物理的项会系统性地促进稳定化，防止数值不稳定性增长。

上述讨论得到图7和图8的证实，两图展示了 noEF、RL-EF、EF（柯尔莫哥洛夫）与参考滤波直接数值模拟之间的比较。具体而言，图7从速度范数 $\|\boldsymbol{u}\|_2$ 和 $\|\nabla \boldsymbol{u}\|_2$ 的角度比较了这些方法，而图8分析了阻力和

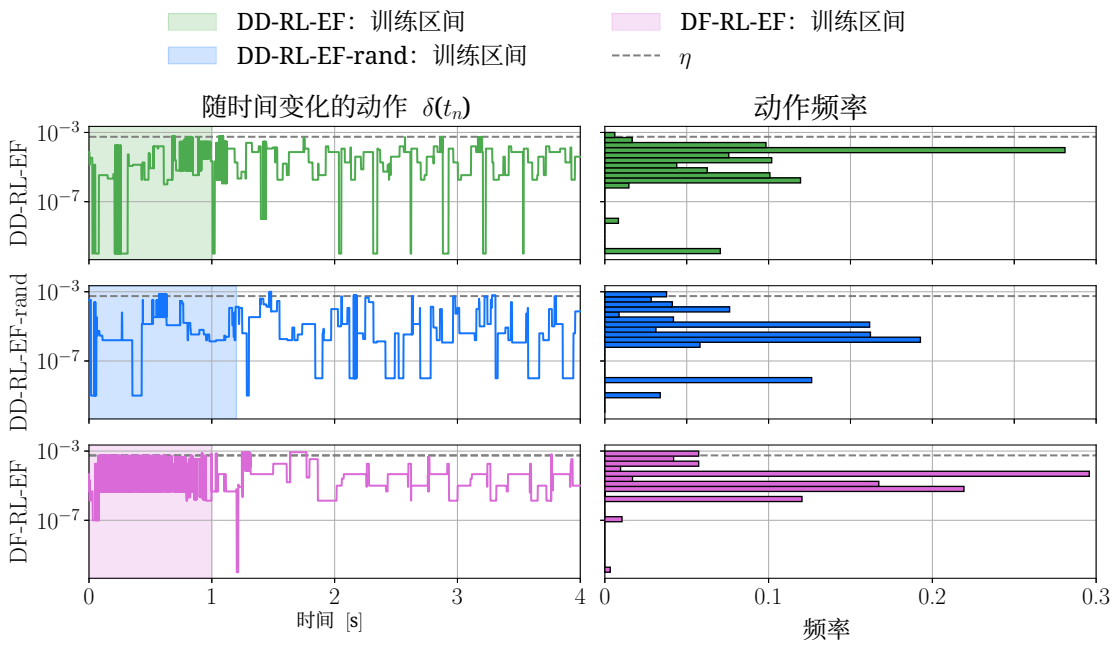


FIGURE 6. 测试案例1。RL-EF方法在测试期间的动作值及其对应频率。

作用于圆柱上的升力。升力和阻力无量纲系数的定义如下：

$$C_D(t) = \frac{2}{U^2 L} \int_{\partial\Omega_C} ((2\nu\nabla \boldsymbol{u}(t) - p(t)\mathbf{I}) \cdot \boldsymbol{n}_C) \cdot \boldsymbol{t}_C \, ds,$$
$$C_L(t) = \frac{2}{U^2 L} \int_{\partial\Omega_C} ((2\nu\nabla \boldsymbol{u}(t) - p(t)\mathbf{I}) \cdot \boldsymbol{n}_C) \cdot \boldsymbol{n}_C \, ds, \tag{12}$$

其中， U 和 L分别是测试案例的特征速度和特征长度，Ω_C 是圆柱边界， $\boldsymbol{n}_C$ 和 $\boldsymbol{t}_C$ 分别是圆柱法向和切向的单位向量。

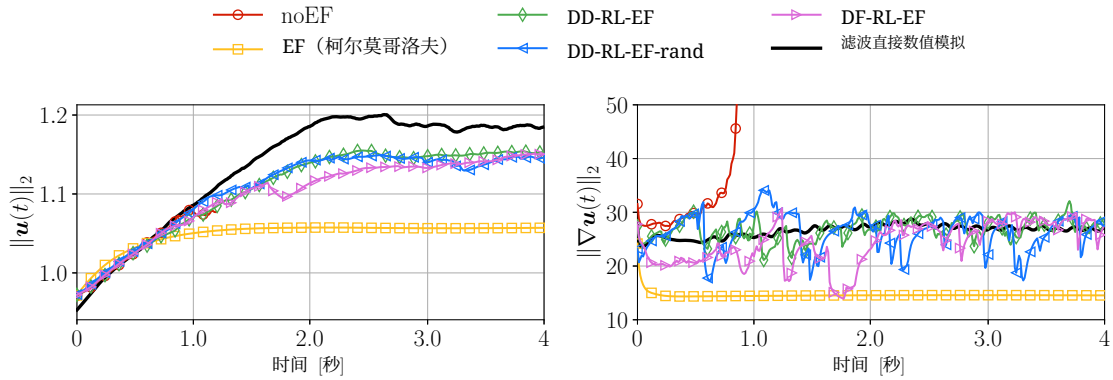


FIGURE 7. 测试案例1。滤波直接数值模拟、noEF、EF（Kolmogorov）以及新型RL-EF方法的速度范数和速度梯度范数随时间的演变。

图7和图8表明，与 noEF 和标准 EF 相比，RL-EF方法在 $\|\boldsymbol{u}\|_2$ 和梯度 $\|\nabla\boldsymbol{u}\|_2$方面都更准确，也更接近滤波DNS参考结果。在所有考察的指标中，标准 EF（柯尔莫哥洛夫）方法的过度扩散特性显而易见：它没有振荡，时间变化也几乎呈平坦状态。虽然这保证了稳定性，却也抑制了流动中重要的动力学特征。另一方面，noEF模拟会变得不稳定，并在 $t = 1$之后发散。因此，图中未展示此时间点之后对应的红色曲线。所有 RL-EF策略都成功恢复了正确的能量演化，并非常接近参考解。然而，速度梯度会围绕参考值随时间振荡。这种表现并不意外，是此类模拟固有的特征。值得注意的是，在 [24],中也观察到了类似的振荡，尽管该研究中的每个动作 δ_n 都是在每个时间步 t_n独立执行一次优化后得到的最优值。因此，振荡的存在不能归因于学习得到的策略不够优。事实上，在 [24] 中，滤波参数是在充分利用参考数据的条件下，通过每个时间步的局部优化确定的，这代表了一种明显更有利的设置。而在本研究中，强化学习智能体仅凭有限的训练数据学习单一策略，随后还必须将其泛化到时间外推区间。在这些更具挑战性的条件下观察到类似的振荡模式，进一步证明了所提出的强化学习框架具有稳健性和有效性。此外，数据驱动方法在升力和阻力系数方面都非常接近参考值，无论是振幅还是频率均如此。无数据策略优于 EF（柯尔莫哥洛夫）方法，但在 [1.5, 2.5] 的某些时间步，略微低估了力的振幅和速度范数。

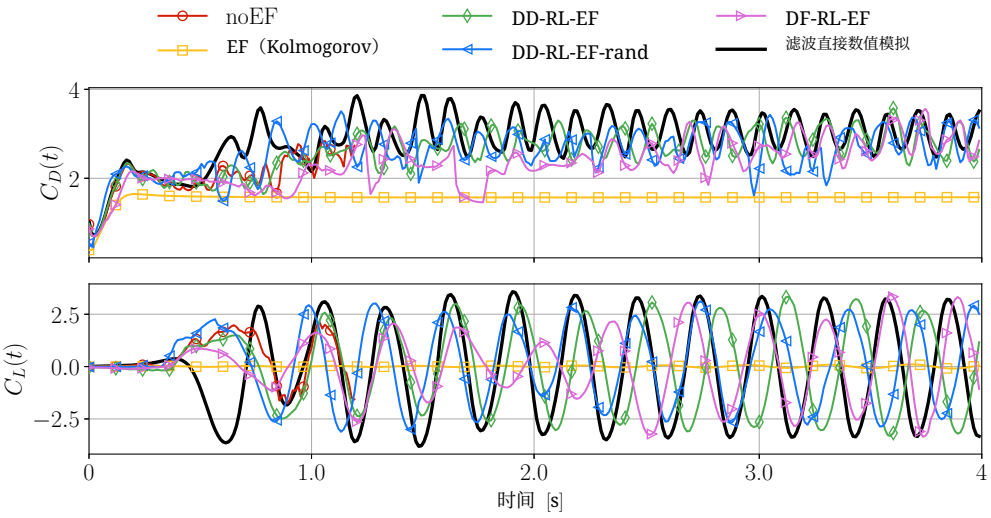


FIGURE 8. 测试案例1。滤波直接数值模拟、noEF、带有 $\delta = \eta$ 的演化-滤波，以及新型 RL-EF方法的阻力和升力系数随时间的演化。阻力和升力是在圆柱边界上计算的。

图9展示了所提出方法在不同波数 $\mathcal{E}(\kappa)$ 处的能谱，并与二维测试案例中的理论湍流趋势 κ 进行了比较。在 $t = 1$时，涡脱落尚未充分发展，所有方法的大尺度能量分布都相近。然而，EF（柯尔莫哥洛夫）方法已明显抑制高波数模态，导致小尺度能量过早衰减。这证实了该方法具有过度扩散的特性。在较晚时刻（$t = 2$ 和 $t = 4$），差异更加明显。noEF模拟在中高波数处积累了过多能量，最终发散。相比之下，EF（柯尔莫哥洛夫）方法强烈抑制小尺度活动，产生的能谱过于陡峭，

在耗散区显著低于参考解。所有 RL-EF策略都实现了显著改善的平衡：它们在正确保留大尺度能量含量的同时，控制高波数范围，且不会引入过度扩散。所得能谱仍接近参考斜率，包括预期呈现 κ^{-3} 行为的惯性区。此外，在 $t = 1$ 和 $t = 2$时，与 DD-RL-EF策略相比，DF-RL-EF能谱在大波数（即小尺度）处的能量更低，且更接近滤波直接数值模拟。该行为与图6中观测到的较大 δ_n 值直接一致，表明无数据智能体在这些阶段施加了更强的滤波。这证实，基于物理的奖励通过惩罚小尺度能量的不受控增长，有效促进了稳定化。具体而言，它使智能体能够以符合物理规律的方式调节耗散区，防止高波数处能量堆积，同时保留正确的大尺度动力学。由此，小尺度行为得到更好的控制，同时避免了标准 EF（柯尔莫哥洛夫）方法所具有的过度扩散。总之，这表明，纯粹的物理信息奖励足以引导智能体采取谱一致的滤波策略，从而在不同尺度上兼顾稳定性与准确度。

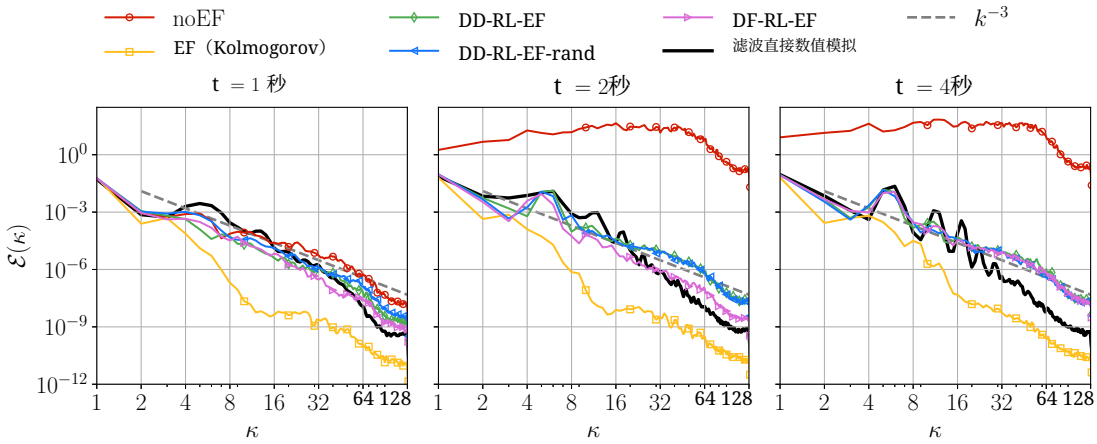


FIGURE 9. 测试案例1。在三个时间实例下，滤波直接数值模拟、noEF、采用 $\delta = \eta$ 的 EF 以及新型 RL-EF 方法的能谱。该能谱是在圆柱尾流处计算的。

此外，图10展示了不同时间实例下的速度大小场。与图4中的noEF和EF（柯尔莫哥洛夫）相比，所有RL-EF在定性上都能准确捕捉涡脱落，既不过度扩散，也不会产生噪声。

最后，我们通过对RL-EF动能的准确度随时间取平均，以及对不同尺度下RL-EF能谱进行平均，比较各方法的全局指标。我们考察全局 量，例如能量和涡量平方，因为传统的 L^2范数相对误差并不适合衡量所提方法的性能。正如 [24, 54], 中所指出的，EF类解与noEF解即使结果不同，其相对 L^2 误差也可能相近。具体而言，我们考察：

$$\mathrm{err}_{\mathcal{E}} = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{\mathcal{E}_n - \mathcal{E}_n^{\mathrm{ref}}}{\mathcal{E}_n^{\mathrm{ref}}}\right|, \quad \mathrm{err}_{\mathrm{spectrum}} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{K}\sum_{\kappa=1}^{K}\log_{10}\left(\frac{\mathcal{E}_n(\kappa)}{\mathcal{E}_n^{\mathrm{ref}}(\kappa)}\right)\right), \tag{13}$$

其中， K 是计算误差时考虑的最大波数。在 $\mathrm{err}_{\mathrm{spectrum}}$ 中，我们采用 $\log_{10}$ 缩放误差，因为它能更好地描述跨越多个数量级的差异相对幅度，并突出频谱中低能量、大波数范围内的差异；这些差异往往对评估湍流和

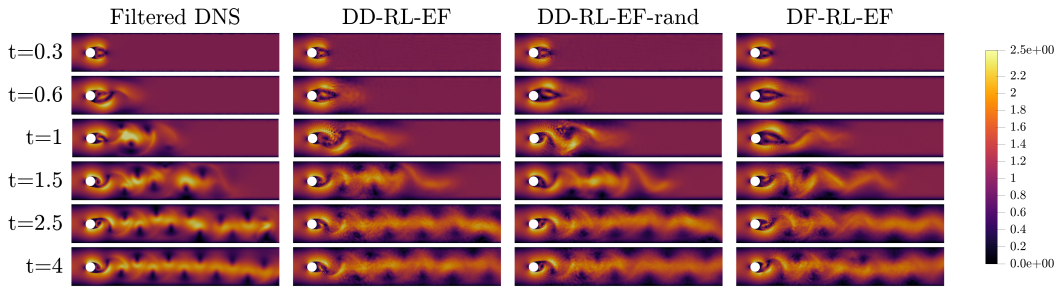


FIGURE 10. 测试案例1。RL-EF方法以及滤波DNS参考结果的速度大小场。图中展示了不同时间实例下的场。

多尺度流动中数值格式的保真度至关重要。图11展示了上述各项指标在所有 RL-EF解中的表现，并与 noEF 和 EF（柯尔莫哥洛夫）进行比较。此外，我们针对两种不同的 K 值考察 $\mathrm{err}_{\mathrm{spectrum}}$：较大的值保留所有高能量（小波数）和低能量（大波数）尺度，较小的值则只保留大能量尺度。这样，我们便能评估 RL-EF 在主导含能尺度（小 K）以及整个频谱范围（大 K）上的准确度。

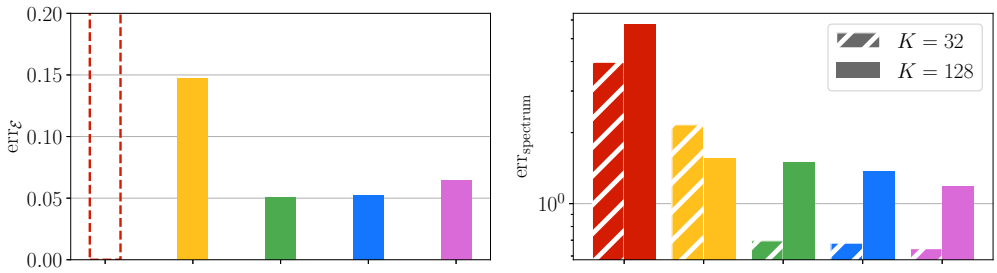


FIGURE 11. 测试案例1。时间平均动能的误差（$\mathrm{err}_{\mathcal{E}}$），以及对波数和时间取平均的能谱对数空间误差（$\mathrm{err}_{\mathrm{spectrum}}$）。能谱误差分别在两个不同的波数范围内计算，即 $K = 32$ 和 $K = 128$。

对于圆柱测试案例（图11），就能量误差 $\mathrm{err}_{\mathcal{E}}$ 而言，所有 RL-EF方法的准确度都显著高于 EF（Kolmogorov）（误差为 15%），其误差仅为 5%。noEF模拟的红线采用虚线表示，因为其解会发散。就能谱误差而言，RL-EF方法，尤其是无参考方法 DF-RL-EF，在两种K 值下的准确度都是最高的。然而，仅考虑大尺度时，误差差异更大（$K = 32$）。这表明，滤波策略的选择对流动主导结构中能量分布准确度的影响更大。具体而言，EF（Kolmogorov）在大尺度能量含量方面表现出更大的误差，因为它无法恢复物理涡脱落。相比之下，基于RL的策略能够恢复相关的流动动力学，并更好地保持跨尺度的能量分布。考虑完整能谱时（$K = 128$），RL方法之间的差异仍然有限，其中 DF-RL-EF 达到

最低的总体误差。EF（Kolmogorov）的性能与 RL-EF方法相近，因为它在小尺度上不会产生伪振荡或噪声行为。总之，需要强调以下几点：

• 在**参考引导**方法中，仅在有限的时间窗口内提供参考数据，即 $[0,\ 1]$（对于 DD-RL-EF-rand，则为 $[0,\ 1.4]$。在此时间区间内，涡脱落尚未完全形成。尽管训练仅限于这一瞬态前训练阶段，强化学习智能体仍成功地仅根据速度状态学习如何调节滤波强度。值得注意的是，学习得到的策略能够推广到时间外推区间，即使流动已经充分发展，仍能保持 RL-EF 模拟的稳定性和准确度。• 在**无参考**设置下，强化学习智能体在完全无法获取参考数据的情况下进行训练，仅依赖基于物理的信息。尽管如此，它仍能自主学习出有效的滤波策略，以保持稳定性和准确度。最终性能与数据驱动方法非常接近，这表明仅凭融入物理信息的奖励，就足以引导学习过程获得稳健且高质量的解，并大幅减少训练所需的计算时间，因为该过程无需使用参考数据，也无需进行直接数值模拟，正如第 4.3 节中更深入分析的那样。

总体而言，这些结果凸显了所提出的强化学习框架的稳健性和泛化能力，表明即使参考数据有限或完全缺失，也能学得有效的稳定化策略。

4.2. **测试案例2：衰减均匀湍流**。第二个测试案例是在边长为 $Re = 40\,000$ 的正方形区域内的二维衰减均匀湍流流动。表 3 列出了速度（$N_h^{\boldsymbol{u}}$）场和压力（N_h^p）场所采用的自由度数量、粗网格和细化网格的网格分辨率，以及相应的网格尺寸。如图 12a 和 12b 所示，两种网格均匀间隔分布，并采用周期性边界条件和四边形单元。

TABLE 3. 测试案例2。所考虑网格的自由度数量、最小尺寸和最大尺寸。

	$N_h^{\boldsymbol{u}}$	N_h^p	网格分辨率	$h_{min} = h_{max}$
粗网格	16641	4225	64×64	$1.56e-2$
细网格	263169	66049	256×256	$3.91e-3$

如 [1], 中所述，速度场按以下步骤初始化：

（1）在谱空间中对速度进行采样，使其在每个波数 κ处与给定能量$\hat{}$ $\mathcal{E}_\kappa$ 相匹配。谱速度的每个分量实际上按 $\|\hat{}\boldsymbol{u}_\kappa\| = \mathrm{a}_\kappa$进行初始化，其中$\mathrm{a}_\kappa = \sqrt{2\hat{}\mathcal{E}_\kappa}\mathrm{e}^{2\pi i\tau_\kappa}$，且 τ_κ 为随机相移。

（2）随后对速度进行投影，使其无散：$\hat{}\boldsymbol{u}_\kappa = \hat{\boldsymbol{P}}_\kappa^{\mathrm{a}_\kappa}\boldsymbol{e}_\kappa\hat{}\boldsymbol{P}_\kappa\boldsymbol{e}_\kappa$，其中$\hat{}\boldsymbol{P}_\kappa$为投影算子：$\hat{}\boldsymbol{P}_\kappa^{\mathrm{a}_\kappa} = I - \frac{\kappa\kappa^T}{\kappa^T\kappa}$，且 $\boldsymbol{e}_\kappa$为随机单位向量。由于项 $\frac{\kappa\kappa^T}{\kappa^T\kappa}\boldsymbol{e}_\kappa$给出了 $\boldsymbol{e}_\kappa$到 κ上的投影，施加算子$\hat{}$ $\boldsymbol{P}_\kappa$可得 $\kappa^T(\hat{}\boldsymbol{P}_\kappa\boldsymbol{e}_\kappa) = 0$，从而$\overline{\kappa^T\hat{}}\boldsymbol{u}_\kappa = 0$。该条件等价于频域中的不可压缩约束。在二维中，$\boldsymbol{e}_\kappa = (\cos(\theta_\kappa), \sin(\theta_\kappa))$，其中 $\theta_\kappa \sim \mathcal{U}[0, 2\pi]$。

（3）最后，通过变换到物理空间（$\boldsymbol{u}_1 = \mathrm{FFT}^{-1}(\hat{}\boldsymbol{u})$）并进一步执行投影步骤，使得 $\boldsymbol{M}_h\boldsymbol{u}_1 = \boldsymbol{0}$, 从而得到速度场；其中，$\boldsymbol{M}_h$为离散散度算子。

每次模拟持续 $T = 2$秒，时间步长为 $\Delta t = 1 \times 10^{-3}$，全局时间步数为 $N = 2000$。作为初步评估，图13展示了滤波直接数值模拟、noEF 和 EF（Kolmogorov）速度场随时间演化的一些代表性解。与圆柱绕流的情况类似，noEF 解仅经过几个时间步后便变得不稳定并发散。相比之下，

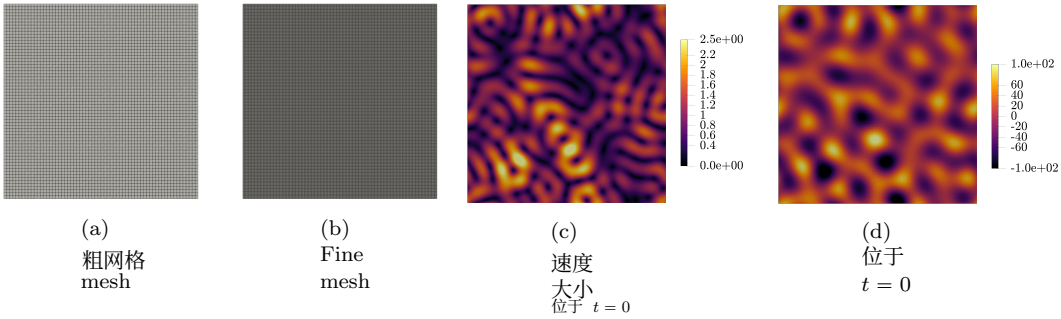


FIGURE 12. 测试案例2。用于衰减均匀湍流测试案例的两个网格（a）和（b）、初始速度场（c）和涡量场（d）。粗网格用于 noEF、EF 和 RL-EF 模拟，而细网格用于 DNS 参考模拟。

EF（Kolmogorov）解在约 $t = 0.5$ 之前都与滤波直接数值模拟高度吻合。然而，超过这一时间后，由于过度扩散，该解逐渐偏离参考解。这种过度扩散现象从涡结构的大小上可明显看出：与滤波直接数值模拟中的涡结构相比，这些涡结构显得更大。滤波DNS参考结果中的流动最初由小尺度涡结构构成，这些结构逐渐合并并随时间增大，导致湍流呈现出典型的衰减特征，且整体以耗散行为为主。这种演化与圆柱绕流有根本区别。在圆柱绕流中，涡脱落仅在瞬态阶段之后才开始发展，动能先增加，随后进入周期性状态。相比之下，在衰减湍流情形中，动能随时间推移而单调递减，流动动力学完全由不同尺度结构的逐步耗散所主导。这一显著差异引出了一个关键问题：

基于强化学习的滤波策略能否适应不发生持续能量产生的流态，并在不引入人工耗散的情况下，保留正确的衰减速率？以下结果正是针对这一问题。

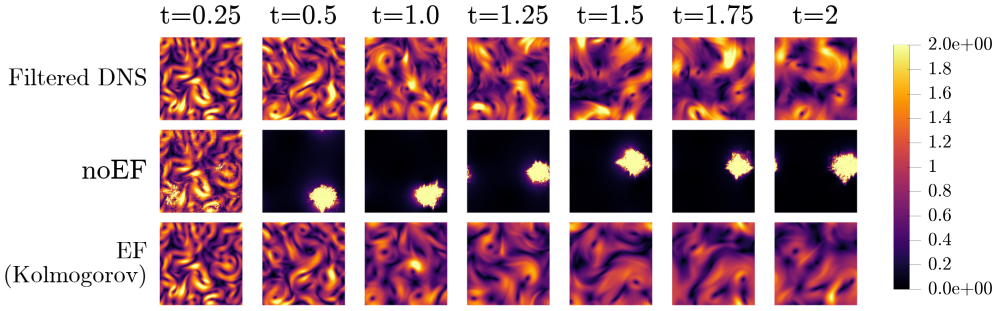


FIGURE 13. 测试案例2。衰减湍流测试案例中，滤波直接数值模拟、noEF 和 EF 模拟的速度大小场。图中展示了不同时间实例的场。

图14展示了不同RL-EF策略在训练回合中的累积奖励和DQN损失函数。由于该测试案例是一个能量和涡量平方保持测试案例，我们

还研究了第3.2节中介绍的**SP-DD-RL-EF**策略。在这种情况下，除DD-RL-EF-rand外，所有RL-EF方法的最大累积奖励均为 $N_{\text{train}} = N/4 = 500$ ($T_{\text{train}} = 0.5$)；DD-RL-EF-rand的回合较短，为 $N_{\text{train}}^{\text{rand}} = N/10 = 200$个时间步（对应$T_{\text{train}}^{\text{rand}} = 0.2$）。所有RL-EF策略的奖励都达到接近其理论最大值的平台期，但损失衰减情况各不相同。首先，与第一个测试案例不同，损失整体衰减得慢得多，只有DD-RL-EF和DF-RL-EF达到 $\sim 1 \times 10^{-4}$。DD-RL-EF-rand和SP-DF-RL-EF表现出更明显的损失振荡。在DD-RL-EF-rand的情况下，这种现象主要与每个回合的随机初始化有关，这一点在第一个测试案例中也已观察到。每个训练回合开始时引入的变化增大了学习动力学的离散程度。而在结构保持设置中，这种现象则可归因于奖励公式中嵌入的额外物理约束。这些约束虽然确保了与底层物理规律的一致性，但也使优化景观更加复杂，导致训练稳定性降低，并使最优滤波参数更难确定。

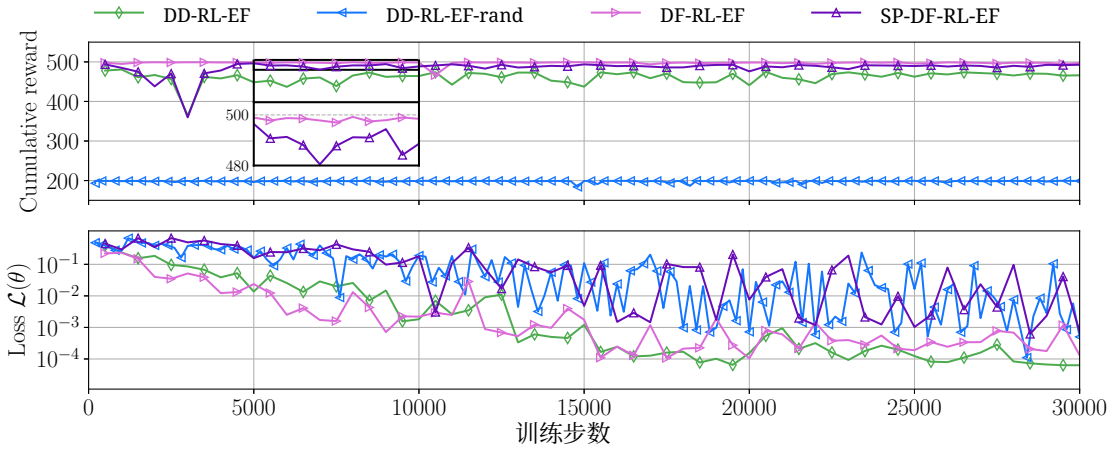


FIGURE 14. 测试案例2。RL-EF方法在训练过程中的累积奖励和损失值。

图15展示了预训练RL智能体在整个模拟过程中如何选择滤波动作，以及最常选滤波值的对应分布。在此测试案例中，与前一情形相比，智能体的行为振荡较少。特别是，所选动作值会在更长的连续时间区间内保持不变。此外，动作频率分布显示，最常选的滤波半径δ_n 在 t_n 处与科尔莫戈洛夫长度尺度一致或非常接近。这一趋势在 DD-RL-EF-rand 和 SP-DD-RL-EF 中尤为明显：智能体分别在 t_n的 40% 和 50% 时间步中选择 $\delta_n = \eta$。这表明，在纯耗散区间内，学习得到的策略会自然趋向于物理上有意义的小尺度截断，从而在稳定性与保持正确衰减动力学之间取得平衡。

关于图16所示的能量和涡量平方随时间的演化，noEF模拟在 $t = 0.25$后不久变得不稳定并发散。相比之下，EF（Kolmogorov）方法表现出能量和涡量平方均平滑且单调的衰减，但代价是过度耗散；与滤波DNS参考结果相比，这导致动力学处于过度阻尼状态。除DD-RL-EF外，所有RL-EF策略都成功捕捉到了正确的能量衰减，同时避免了EF（Kolmogorov）中观察到的过度扩散。其中，DD-RL-EF-rand在能量演化方面与滤波DNS最为一致。相比之下，DD-RL-EF在 $t = 0.5$之前与参考结果吻合良好，这一时段对应于训练时间窗口，但在更晚的时间逐渐变得过度扩散。这种行为与更大的

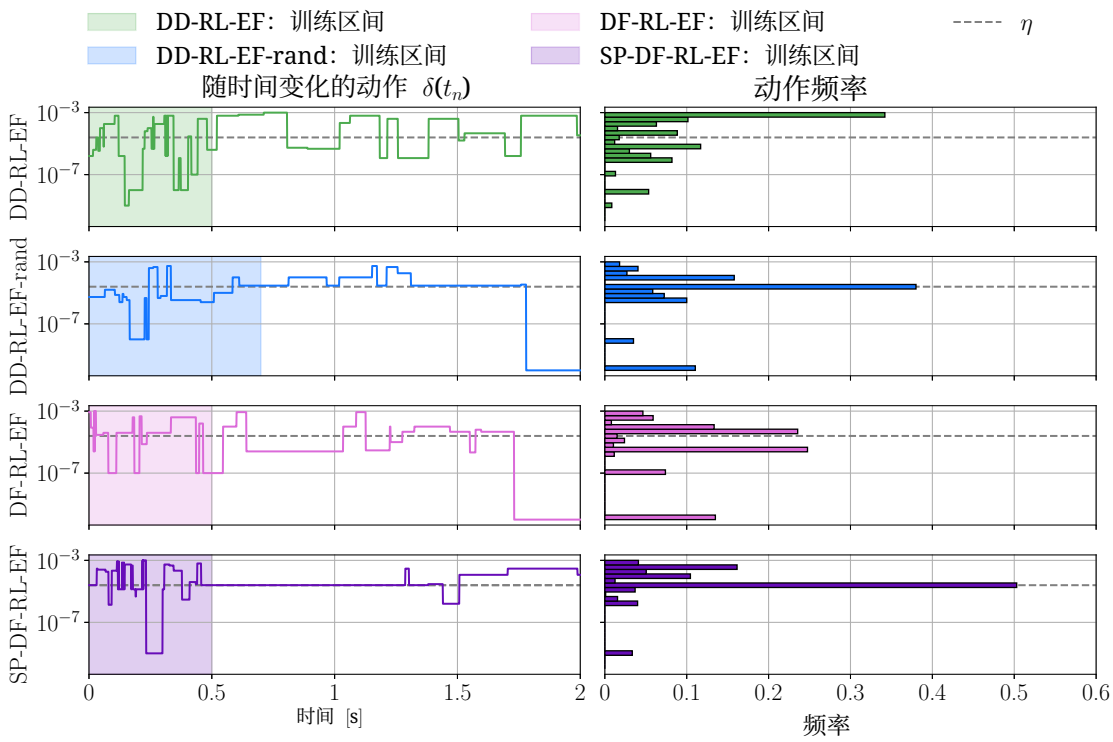


FIGURE 15. 测试案例2。RL-EF方法在测试时间内的动作值及其对应频率。

δ_n 训练时域之外所选取的值一致，如图15所示，表明其在外推区间中的泛化能力有限。关于涡量平方演化，DD-RL-EF、DD-RL-EF-rand和DF-RL-EF均呈现出增加趋势，尤其是在早期。这种行为反映了在耗散区间占主导地位之前，小尺度活动的暂时增强。在SP-DD-RL-EF情况下，涡量平方守恒的弱约束使行为更加受控且平缓。然而，仍然可以观察到涡量平方的小幅增加，这可归因于该约束以弱形式（惩罚项）施加，而不是作为硬约束施加。

图17比较了 $t = 0.5$、1和1。5时的能谱。由于noEF模拟在早期时间变得不稳定，其能谱迅速变得不符合物理规律，在高波数处出现过量能量积聚。相比之下，EF（Kolmogorov）在所有报告的时间点都系统性地低估了小尺度中的能量含量，证实了其过度耗散的特性。在 $t = 0.5$时，所有RL-EF策略产生的能谱彼此非常接近，并且与滤波DNS参考结果吻合良好。然而，在中等和大波数范围内，DD-RL-EF-rand表现出略高的能量水平，表明存在残余的小尺度振荡。这一行为与图16中观察到的较大涡量平方值一致，反映出增强的小尺度活动。在较晚时间（$t \in \{1、\ 1.5\}$）时，DD-RL-EF越来越低估高波数范围内的能量，这与能量和涡量平方变化趋势中已经观察到的更强耗散相符。相比之下，DD-RL-EF-rand、DF-RL-EF和SP-DD-RL-EF与滤波DNS谱保持高度一致，准确再现了大尺度能量含量和小尺度衰减。

从定性角度来看，图18对比了不同时间下 RL-EF 的速度场与滤波直接数值模拟。目视检查证实了上文讨论的趋势。在 $t > 0.5$ 时，DD-RL-EF 明显呈现出过度平滑的涡结构，这与其在能量衰减和谱分析中观察到的过度耗散

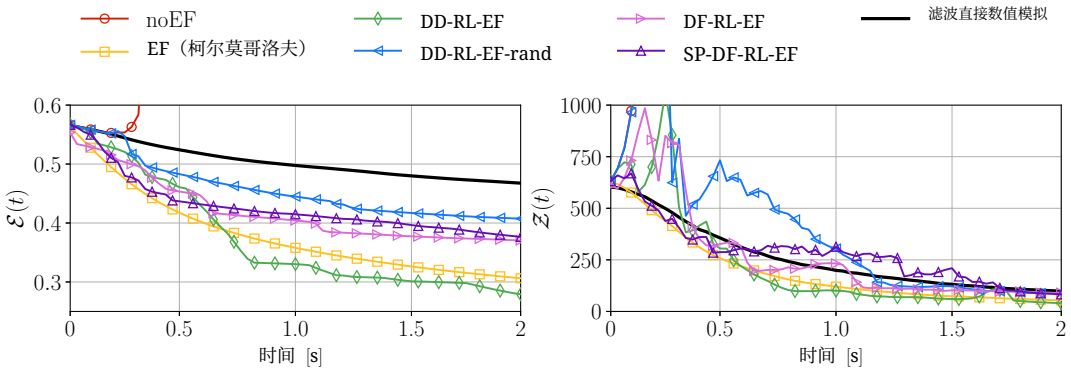


FIGURE 16. 测试案例2。所有RL-EF模拟、标准 EF、noEF以及滤波DNS参考结果模拟中能量与涡量平方的时间演化。

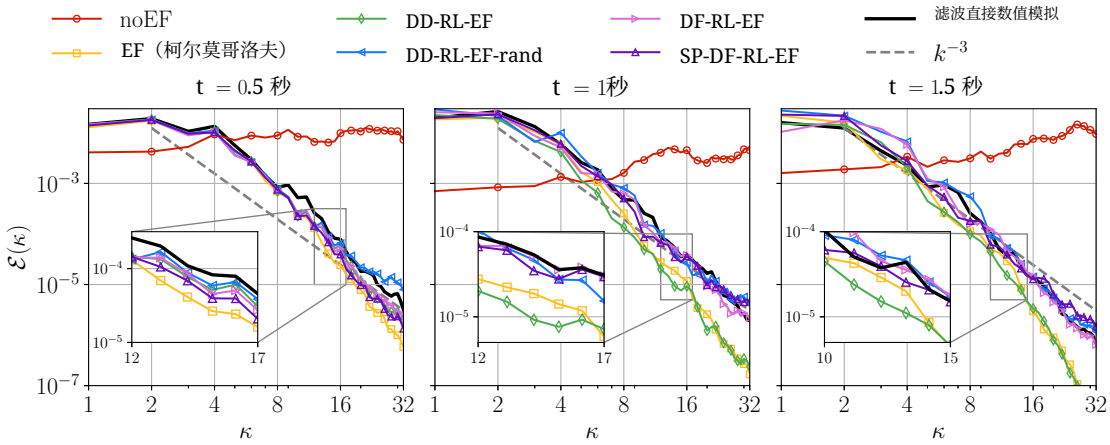


FIGURE 17. 测试案例2。衰减湍流测试案例的能谱，包括所有 RL-EF 模拟、标准 EF、noEF 和滤波 DNS 参考模拟。图中展示了不同时间实例下的能谱。

行为一致。相反，DD-RL-EF-rand 在早期（例如 $t = 0.25$）的小尺度结构噪声更大，这与其较高的涡量平方水平和略有增加的高波数能量相符。相比之下，无数据方法提供了更为均衡的解。涡结构仍能得到良好解析，且没有出现虚假的振荡或过度平滑，体现了数值稳定化与保留物理衰减动力学之间更有效的平衡。

与测试案例1一样，我们通过比较各方法在动能和能谱方面的全局误差（方程（13）中的表达式）来结束分析。就误差 $\mathrm{err}_{\mathcal{E}}$ 而言，图19证实 DD-RL-EF-rand 是精度最高的方法，其次是无参考方法（SP-DF-RL-EF 和 DF-RL-EF）。由于 DD-RL-EF 过度扩散，其能量误差接近 EF（柯尔莫哥洛夫）方法。对于能谱误差，我们在此也分别考察大尺度的影响（最大波数 $K = 8$）以及完整尺度范围的影响（$K = 32$）。与测试案例1不同，在这里，考虑完整尺度范围时，各方法之间的差异更加明显；而仅考虑最大尺度时，差异则不那么显著。这一现象反映了该测试

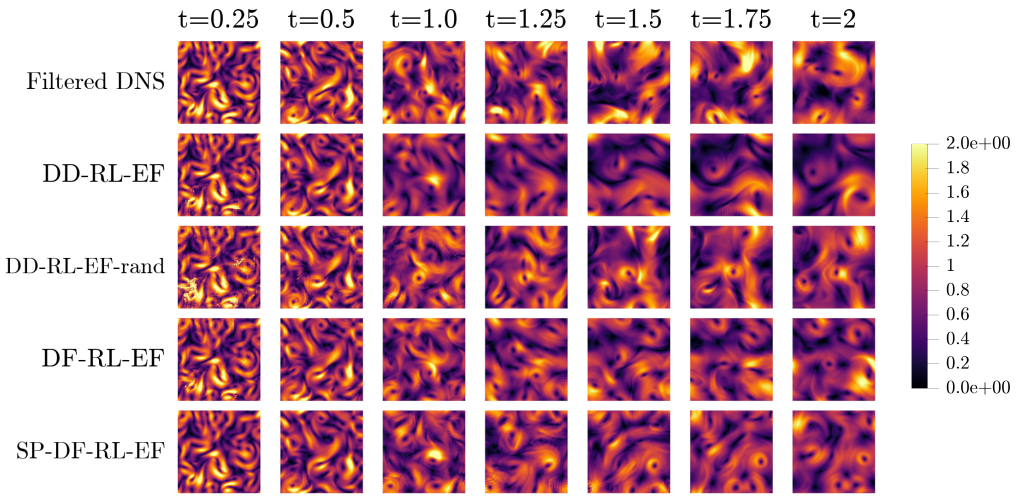


FIGURE 18. 测试案例2。衰减湍流测试案例中，滤波DNS参考结果和所有RL-EF模拟的速度大小场。这些场展示了不同时间实例下的结果。

案例固有的难点：主要挑战在于准确控制速度梯度，同时避免过度耗散。考虑完整尺度范围时，结果显示，过度扩散方法——即 EF（柯尔莫哥洛夫）和 DD-RL-EF——与其他所有方法（DD-RL-EF-rand、DF-RL-EF 和 SP-DF-RL-EF）之间存在显著差异。过度扩散确实会降低大波数处的能量，从而使 $K = 32$ 处的误差增大。考虑较小的 K 时，除保结构策略（SP-DF-RL-EF）外，所有 RL-EF 方法的误差都与 EF（柯尔莫哥洛夫）相近。奖励中的弱约束使得模型能够恢复主导尺度处的能谱。因此，即使训练算法时没有使用参考DNS数据，结果也能更好地符合滤波DNS谱。

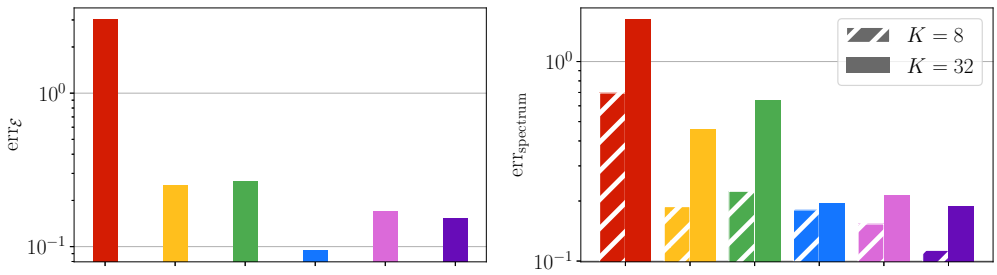


FIGURE 19. 测试案例2。时间平均动能的误差（$\mathrm{err}_{\mathcal{E}}$），以及在波数和时间上取平均的能谱的对数空间误差（$\mathrm{err}^{\mathrm{spectrum}}$）。能谱误差分别在两个不同的波数范围内计算，即 $K = 8$ 和 $K = 32$。

总体而言，该测试案例表明，所提出的基于强化学习的滤波框架能够适应纯耗散区间，这与涡脱落场景有本质区别。具体而言，结果表明：

- 纯**参考引导**训练在动力学演化超出训练窗口时，外推能力可能有限；
- 对训练时间窗口进行随机化可以增强鲁棒性，但可能引入轻微的小尺度噪声；
- **无参考**奖励函数形式显著提升了稳定性，同时保持了频谱精度和能量精度。

4.3. 计算细节。 本节总结了强化学习实验相关的主要计算方面。具体而言，表4报告了各 RL-EF 方法在两个测试案例中的训练时间，表5则列出了便于复现的主要强化学习超参数。关于计算时间，所有实验均在意大利 CINECA 的 Galileo G100集群上，使用 Intel Xeon Platinum 8260节点完成。实验未使用 GPU加速。表4显示，测试案例2中直接数值模拟所需的 CPU时间远长于测试案例1中的直接数值模拟所需时间。这是因为前者的有限元节点更多，对比表2和表3即可看出这一点。一方面，在两个测试案例中，DD-RL-EF-rand 平均每回合所需的 CPU时间都少于其他方法。如表1所述，DD-RL-EF-rand 的回合持续 $N_{\mathrm{randtrain}} = N/10$ 个时间步，而其他所有方法的每回合均包含 $N_{\mathrm{train}} = N/4$ 个时间步。另一方面，DD-RL-EF-rand 收集滤波后的DNS数据所需的 CPU时间更多，因为每个回合都在区间 $[0, T_{\mathrm{train}}]$内随机初始化，也就是说，我们需要 $[0, T_{\mathrm{train}} + T_{\mathrm{train}}^{\mathrm{rand}}]$中的直接数值模拟数据。此外需要指出，无参考方法，即 DF-RL-EF 和 SP-DF-RL-EF，不需要任何直接数值模拟数据。这样既能节省数小时的计算时间，又能获得与参考数据引导方法相当甚至更准确的解（例如，参见图11和图19中的全局谱误差）。表4所列的回合数对应于训练奖励趋于稳定时的节点；随后保存模型，用于进行时间外推测试。

TABLE 4. RL-EF 方法的计算训练时间。

测试案例	算法	每回合平均 CPU 时间	回合数	直接数值模拟 CPU 时间
1	DD-RL-EF	15 分 38 秒	150	63 分 41 秒
	DD-RL-EF-rand	6 分 18 秒	300	89 分 2 秒
	DF-RL-EF	15 分 18 秒	200	不需要
2	DD-RL-EF	32 分 14 秒	90	5 小时 10 分 30 秒
	DD-RL-EF-rand	13 分 11 秒	210	7 小时 13 分 45 秒
	DF-RL-EF	32 分 9 秒	90	不需要
	SP-DF-RL-EF	32 分 14 秒	60	不需要

TABLE 5. 强化学习训练所使用的超参数。

超参数	**值**
折扣因子 γ	0.99
学习率	1×10^{-5}
网络架构	2 个隐藏层，每层 64 个神经元
网络激活函数	ReLU
网络优化器	Adam
最大梯度范数	5
批次大小	128
目标更新间隔	$5N_{\text{train}}$

5. CONCLUSIONS

本文提出了 **RL-EF**：一种强化学习框架，用于在对流占优不可压缩流的 EF 稳定化中动态选择滤波半径。其核心思想是用学习得到的、依赖状态的控制策略取代启发式的恒定滤波宽度设定，使其能够适应不断演变的流动动力学。所提出的强化学习-EF方法将参数调优重新表述为一个序贯决策问题。强化学习智能体在预设训练窗口内学习全局控制策略，并将其应用于时间外推。本文研究了基于参考数据（DD-RL-EF）和完全无参考数据（DF-RL-EF）的奖励函数形式，其中还包括一种结构保持变体，可按具体测试案例的需要纳入较弱的能量和涡量平方约束。在两个动力学特性截然不同的基准问题上进行的数值评估—— $Re = 1000$处的圆柱绕流以及 $Re = 40\,000$处的衰减均匀湍流——揭示了若干重要发现：

- **稳定性与避免过度扩散**。学习得到的策略能够避免欠分辨率 noEF 模拟中出现的数值爆炸，同时避免基于固定科尔莫戈洛夫长度尺度的 EF 方法通常产生的过度耗散。RL 选取的滤波半径随时间呈现非平凡变化，这证实了在整个模拟过程中不存在单一的最优常数值。
- **跨尺度的谱一致性**。RL-EF策略在正确控制耗散区的同时保留了大尺度结构。所得能谱始终接近参考滤波直接数值模拟的能谱，包括惯性区，表明自适应滤波能够恢复能量主导模态与小尺度模态之间的正确平衡。
- **超越训练窗口的泛化能力**。在数据驱动设置中，智能体仅在流动尚未充分发展时的早期时间窗口内接受训练。尽管可用信息有限，学习得到的策略仍能成功泛化至后续时段，包括具有涡脱落特征的状态或长期衰减阶段。这凸显了强化学习推断可迁移控制律的能力，而不只是对参考数据进行插值。
- **物理信息奖励的有效性**。最重要的是，无数据方法取得了与数据驱动方法相当的性能。DF-RL-EF 智能体仅依靠残差控制和基于梯度的指标，而不依赖计算成本高昂的直接数值模拟参考数据，便能自主发现既稳定又具有谱准确性的滤波策略。这表明，即使缺少参考解，具有物理意义的奖励设计也足以引导学习过程形成稳健的稳定化策略。

在两个测试案例中，学习到的滤波策略都会适应流动状态的定性特征：在圆柱尾流中呈振荡且高度动态，在衰减湍流中则更为保守且持续。这一行为证实，强化学习框架捕捉的是依赖于问题的

稳定化需求，而非强制施加统一的耗散机制。所提出的方法为 EF 型方法中的人工参数校准提供了一种稳健而灵活的替代方案。它消除了基于网格尺寸或科尔莫戈洛夫标度的固定调参规则的需要，并以直接从流动演化中学习得到的自适应状态感知控制策略取而代之。

本研究自然引出了若干研究方向。由于我们的首要目标是对所提出方法的稳健性进行初步研究，因此本文尚未探讨这些方面。未来研究将对这些方向展开更深入的考察。首先，拓展至三维湍流是迈向现实高雷诺数应用的关键一步；在这类应用中，尺度分离程度的增加以及更丰富的非线性相互作用会带来大得多的稳定化挑战。其次，一个特别有前景的方向是开发空间依赖滤波策略。在当前的方案中，每个时间步内滤波半径在整个计算域上保持一致。然而，最佳稳定化程度通常具有很强的局部性，尤其是在存在剪切层、涡脱落或边界层的流动中。允许滤波半径随空间变化可显著提高准确度。例如，可以通过多智能体强化学习实现这一点，由不同智能体控制区域中的局部部分；也可以利用能够直接处理网格连接关系并实现局部控制的基于图的神经网络架构。第三，当前框架仍依赖网格，因为观测空间与离散速度自由度相对应。例如，利用图来设计状态的网格无关表示，将提升其在不同网格和分辨率之间的迁移能力，并为与离散化无关的稳定化策略铺平道路。最后，采用连续动作空间可以免去对滤波半径进行离散化的需要，并实现对稳定化机制更精细的控制。尽管使用连续动作算法的初步实验尚未取得有竞争力的性能，但进一步研究策略梯度方法和定制网络架构，或可在适应性和稳健性方面带来更多提升。

此处呈现的结果表明，强化学习为高雷诺数流动模拟提供了一条颇具前景的途径，可用于开发自适应且符合物理规律的稳定化技术。

Acknowledgments

AI、MS 和 GR 感谢 INdAM-GNCS：意大利国家高等数学研究所——国家科学计算组提供的支持。MS 和 GR 感谢“20227K44ME - 耦合系统的全阶与降阶建模：聚焦非匹配方法与自动学习（FaReX）”项目。该项目由欧盟通过下一代欧盟计划资助，属于 PRIN 2022 计划（D.D. 104 - 02/02/2022 意大利大学与研究部）。GR 感谢 iN-EST 联盟（互联的东北部创新生态系统），该联盟由意大利国家复苏与韧性计划（PNRR）——任务4，组成部分2，投资1.5——D.D. 1058 23/06/2022，ECS00000043 支持，并获得欧盟下一代欧盟计划资助。MS 感谢 ECCOMAS EYIC 基金“CRAFT：流动湍流应用中的控制与降阶”。

References

[1] S. D. Agdestein与B. Sanderse。先离散、后滤波：学习用于大涡模拟的散度一致闭合模型。《计算物理学杂志》，522:113577，2025年2月。[2] S. M. Ahmadi、S. Balahang与S. Abolfathi。预测极端洪水事件期间关键交通基础设施的水力响应。人工智能工程应用，133:108573，2024年。[3] M. S. Alnæs、A. Logg、K. B. Ølgaard、M. E. Rognes与G. N. Wells。统一形式语言：一种用于偏微分方程弱形式的领域专用语言。ACM 数学软件汇刊（TOMS），40(2):1–37，2014年。[4] I. A. Baratta、J. P. Dean、J. S. Dokken、M. Habera、J. HALE、C. N. Richardson、M. E. Rognes、M. W. Scroggs、N. Sime与G. N. Wells。DOLFINx：下一代FEniCS问题求解环境。2023年。[5] L. C. Berselli、T. Iliescu与W. J. Layton。湍流大涡模拟的数学，第2卷。施普林格，2006年。[6] L. Bertagna、A. Quaini与A. Veneziani。中等偏高雷诺数不可压缩流的基于反卷积的非线性滤波。《International Journal for Numerical Methods in Fluids》，81(8):463–488，2016年。

[7] L. Besabe、M. Girfoglio、A. Quaini与G. Rozza。双层准地转海洋模型的线性和非线性滤波。Applied Mathematics and Computation，488:129121，2025年。[8] Boyd, John P。切比雪夫与傅里叶谱方法。Courier Corporation，2001年。[9] C. Canuto、M. Y. Hussaini、A. Quarteroni与T. A. Zang。非定常纳维–斯托克斯方程的若干算法。《流体动力学中的谱方法》，第201–239页，1988年。[10] S. Choi、Y. Kim与J. H. Hwang。生成对抗网络生成湍流圆射流的性能。Environmental Fluid Mechanics，25(1):1–19，2025年。[11] V. DeCaria、W. Layton、A. Pakzad、Y. Rong、N. Sahin与H. Zhao。关于grad-div准则的确定。Journal of Mathematical Analysis and Applications，467(2):1032–1037，2018年。[12] V. J. Ervin、W. J. Layton与M. Neda。演化方程基于滤波器的稳定化的数值分析。SIAM Journal on Numerical Analysis，50(5):2307–2335，2012年。[13] D. Fan、L. Yang、Z. Wang、M. S. Triantafyllou与G. E. Karniadakis。实验与模拟中钝体主动流动控制的强化学习。Proceedings of the National Academy of Sciences，117(42):26091–26098，2020年。[14] P. Fischer、J. Kruse、J. Mullen、H. Tufo、J. Lottes与S. Kerkemeier。Nek5000——开源谱元CFD求解器。阿贡国家实验室数学与计算机科学部，伊利诺伊州阿贡，2008年。[15] P. Fischer与J. Mullen。谱元方法的基于滤波器的稳定化。Comptes Rendus de l'Acad´emie des Sciences-Series I-Mathematics，332(3):265–270，2001年。[16] B. Font、F. Alc´antara-´Avila、J. Rabault、R. Vinuesa与O. Lehmkuhl。湍流分离泡中主动流动控制的深度强化学习。Nature communications，16(1):1422，2025年。[17] M. Germano。用于湍流大涡数值模拟的微分滤波器。The Physics of fluids，29(6):1755–1757，1986年。[18] M. Germano。椭圆型微分滤波器。The Physics of fluids，29(6):1757–1758，1986年。[19] M. Girfoglio、A. Quaini与G. Rozza。带非线性滤波稳定化的纳维–斯托克斯方程的有限体积近似。Computers & Fluids，187:27–45，2019年。[20] M. Girfoglio、A. Quaini与G. Rozza。采用大涡模拟滤波方法的流固耦合模拟。Communications in Applied and Industrial Mathematics，12(1):13–28，2021年。[21] M. Girfoglio、A. Quaini与G. Rozza。有限体积框架下准地转方程的一种新型大涡模拟模型。Journal of Computational and Applied Mathematics，418:114656，2023年。[22] M. Guo、D. Goodarzi、J. Pearson、A. Mohammadian与S. Abolfathi。透水挡板对滞留池水力与处理性能的影响。Journal of Water Process Engineering，69:106593，2025年。[23] N. Heavner。不可压缩流问题有限元离散中局部选取的grad-div稳定化参数。SIAM Undergraduate Research Online（SIURO），7:SO1278，2017年。[24] A. Ivagnes、M. Strazzullo、M. Girfoglio、T. Iliescu与G. Rozza。对流主导流动演化-滤波-松弛正则化的数据驱动优化。《工程数值方法国际期刊》，126(9)，2025年。[25] V. John、A. Linke、C. Merdon、M. Neilan与L. G. Rebholz。不可压缩流混合有限元方法中的散度约束。SIAM review，59(3):492–544，2017年。[26] L. P. Kaelbling、M. L. Littman与A. W. Moore。强化学习：综述。Journal of artificial intelligence research，4:237–285，1996年。[27] K. Kaneko与P. H. Tsai。NekROM。`https://github.com/Nek5000/NekROM`。[28] M. Kurz、A. Beck与B. Sanderse。利用等变性：用图神经网络建模湍流。arXiv预印本arXiv:2504.07741，2025年。[29] M. Kurz、P. Offenh¨auser与A. Beck。大涡模拟中湍流建模的深度强化学习。International journal of heat and fluid flow，99:109094，2023年。[30] M. Kurz、P. Offenh¨auser、D. Viola、O. Shcherbakov、M. Resch与A. Beck。高性能计算系统上计算流体力学的深度强化学习。Journal of Computational Science，65:101884，2022年。[31] A. Kwiatkowski、M. Towers、J. Terry、J. U. Balis、G. De Cola、T. Deleu、M. Goul˜ao、K. Andreas、M. Krimmel、A. KG等。Gymnasium：强化学习环境的标准接口。2024年。[32] W. Layton、C. C. Manica、M. Neda、M. Olshanskii与L. G. Rebholz。纳维–斯托克斯方程模拟中旋转形式的准确度。《计算物理学杂志》，228(9):3433–3447，2009年。[33] W. Layton、C. C. Manica、M. Neda与L. G. Rebholz。一种高精度湍流Leray反卷积模型的数值分析与计算测试。《偏微分方程数值方法：国际期刊》，24(2):555–582，2008年。[34] W. J. Layton与L. G. Rebholz。湍流的近似反卷积模型：分析、现象学与数值分析，第2042卷。施普林格柏林海德堡，2012年。[35] Y. Li与T. Okaze。利用前驱体方法优化非等温湍流入流生成：敏感性研究。Building and Environment，269:112465，2025年。[36] B. Mart´ınez、M. Guerra、N. Riviere、E. Mignot与O. Link。障碍物周围超临界明渠流的RANS模拟。Journal of Hydraulic Research，63(1):1–14，2025年。[37] V. Mnih、K. Kavukcuoglu、D. Silver、A. A. Rusu、J. Veness、M. G. Bellemare、A. Graves、M. Riedmiller、A. K. Fidjeland、G. Ostrovski、S. Petersen、C. Beattie、A. Sadik、I. Antonoglou、H. King、D. Kumaran、D. Wierstra，

S. Legg 和 D. Hassabis。《通过深度强化学习实现人类水平控制》。《自然》，518 533，2015 7540 529( ): – .[38] C. Mou、E. Merzari、O. San 和 T. Iliescu。《湍流降阶模型的基于能量的长度尺度》。《核工程与设计》，412:112454，2023年。[39] J. S. Mullen 和 P. F. Fischer。《复杂几何流动的滤波技术》。《工程数值方法通讯》，15(1):9–18，1999年。[40] M. A. Olshanskii 和 X. Xiong。《滤波稳定化与涡黏性模型之间的联系》。《偏微分方程数值方法》，29(6):2061–2080，2013年。[41] R. Paris、S. Beneddine 和 J. Dandois。《利用深度强化学习实现鲁棒流动控制与最优传感器布置》。《流体力学杂志》，913:A25，2021年。[42] R. Pasquetti 和 C. Xu。《关于“谱元方法的基于滤波器的稳定化”的评述》。《计算物理学杂志》，182(2):646–650，2002年。[43] A. Quaini、O. San、A. Veneziani 和 T. Iliescu。《通过空间滤波衔接对流主导流动的大涡模拟与降阶建模：综述与展望》。《arXiv》预印本 arXiv:2407.00231，2024年。[44] J. Rabault、M. Kuchta、A. Jensen、U. R´eglade 和 N. Cerardi。《通过深度强化学习训练的人工神经网络发现主动流动控制策略》。《流体力学杂志》，865 302，2019年 281: – .[45] J. Rabault、F. Ren、W. Zhang、H. Tang 和 H. Xu。《流体力学中的深度强化学习：一种适用于主动流动控制与形状优化的有前景方法》。《水动力学研究与进展》，32(2):234–246，2020年。[46] A. Raffin、A. Hill、A. Gleave、A. Kanervisto、M. Ernestus 和 N. Dormann。《Stable-baselines3：可靠的强化学习实现》。《机器学习研究杂志》，22(268):1–8，2021年。[47] T. C. Rebollo 和 R. Lewandowski。《湍流模型的数学与数值基础及其应用》。施普林格，2014年。[48] F. Ren、J. Rabault 和 H. Tang。《将深度强化学习应用于弱湍流条件下的主动流动控制》。《流体物理学》，33(3)，2021年。[49] E. Rezaian 和 K. Duraisamy。《利用正则化条件参数化图神经网络预测复杂流动》。发表于美国航空航天学会（AIAA）SCITECH 2023 论坛，第1284页。Aerospace Research Central，2023年。[50] J. Schulman、F. Wolski、P. Dhariwal、A. Radford 和 O. Klimov。《近端策略优化算法》。《arXiv》预印本 arXiv:1707.06347，2017年。[51] M. W. Scroggs、I. A. Baratta、C. N. Richardson 和 G. N. Wells。《Basix：运行时有限元基函数求值库》。《开放源代码软件杂志》，7(73):3982，2022年。[52] M. W. Scroggs、J. S. Dokken、C. N. Richardson 和 G. N. Wells。《在多边形与多面体单元网格上构造任意阶有限元自由度映射》。《ACM 数学软件汇刊》(TOMS)，48(2):1–23，2022年。[53] M. Strazzullo、F. Ballarin、T. Iliescu 和 C. Canuto。《纳维–斯托克斯方程对流主导区间下的新型反馈控制与自适应演化-滤波-松弛正则化》。《arXiv》预印本 arXiv:2307.00675，2023年。[54] M. Strazzullo、F. Ballarin、T. Iliescu 和 T. C. Rebollo。《对流主导流动的变分多尺度演化与滤波策略》。《计算方法在应用力学与工程中的应用》，438:117811，2025年。[55] M. Strazzullo、M. Girfoglio、F. Ballarin、T. Iliescu 和 G. Rozza。《对流主导近临界分辨流动的演化-滤波-松弛正则化中全阶与降阶模型的一致性》。《工程数值方法国际期刊》，123(14):3148–3178，2022年。[56] R. S. Sutton 和 A. G. Barto。《强化学习导论》。适应、学习与优化。麻省理工学院出版社，剑桥，马萨诸塞州和伦敦，第二版，2020年。[57] A. Takhirov 和 A. Lozovskiy。《适用于高雷诺数流动的高效模块化非线性滤波器稳定化》。《计算数学进展》，44:295–325，2018年。[58] H. Tang、J. Rabault、A. Kuhnle、Y. Wang 和 T. Wang。《通过深度强化学习训练的大王神经网络实现跨雷诺数范围的鲁棒主动流动控制》。《流体物理学》，32(5)，2020年。[59] M. Towers、A. Kwiatkowski、J. Terry、J. U. Balis、G. De Cola、T. Deleu、M. Goul˜ao、A. Kallinteris、M. Krim-mel、A. KG 等。《Gymnasium：强化学习环境的标准接口》。《arXiv》预印本 arXiv:2407.17032，2024年。[60] P. H. Tsai。《从二维混沌到三维湍流问题的纳维–斯托克斯方程参数化模型降阶方法开发》。博士论文，伊利诺伊大学厄巴纳-香槟分校，尚佩恩，伊利诺伊州，美国，2023年。[61] P. H. Tsai、P. Fischer 和 T. Iliescu。《湍流槽道流的时间松弛降阶模型》。《计算物理学杂志》，521:113563，2025年。[62] R. Vinuesa、O. Lehmkuhl、A. Lozano-Dur´an 和 J. Rabault。《机翼流动控制及通过深度强化学习发现新方法》。《Fluids》，7(2):62，2022年。

[63] D. Wells、Z. Wang、X. Xie 和 T. Iliescu。《对流主导流动的演化后滤波正则化降阶模型》。《International Journal for Numerical Methods in Fluids》，84(10):598–615，2017年。[64] M. A. Wiering 和 M. Van Otterlo。《强化学习》。《适应、学习与优化》，12 729，2012 3( ): .[65] X. Xie、F. Bao 和 C. G. Webster。《随机Burgers方程的演化滤波稳定化降阶模型》。《Fluids》，3(4):84，2018年。[66] H. Xu。《利用基于反卷积的Leray模型对中等偏大雷诺数下的不可压缩流进行高效建模：分析、不确定性量化及其在主动脉夹层中的应用》。在线，2020年。[67] H. Xu、D. Baroli、F. Di Massimo、A. Quaini 和 A. Veneziani。《基于反卷积的大涡模拟建模实现回流稳定化》。《计算物理学杂志》，404:109103，2020年。[68] F. Zoccolan、M. Strazzullo 和 G. Rozza。《最优控制下对流主导偏微分方程的流线迎风 Petrov-Galerkin 降阶方法》。《Computational Methods in Applied Mathematics》，2024年。[69] F. Zoccolan、M. Strazzullo 和 G. Rozza。《具有随机输入的偏微分方程所支配的参数化对流主导最优控制问题的稳定化加权降阶方法》。《Journal of Numerical Mathematics》，2024年。[70] P. Zou、S. Cao 和 D. Cao。《不同湍流模型在矩形圆柱涡激振动模拟中的适用性研究》。《Journal of Wind Engineering and Industrial Aerodynamics》，258:106031，2025年。