arXiv:2304.05029v1 [physics.flu-dyn] 11 Apr 2023

使用小型局部神经网络进行湍流闭合：受迫二维流动和 β 平面流动

KAUSHIK SRINIVASAN、MICKAËL D. CHEKROUN 和 JAMES C. MCWILLIAMS

摘要。 我们使用仅含O(1000)个参数的简单2层卷积神经网络（CNN），对高雷诺数（Re∼25000）下 β-平面上的正弦强迫二维湍流中的亚格子尺度（SGS）通量进行参数化；其参数量比近期采用8–10层更深层的CNN 的研究小两个数量级；在 16 倍降尺度因子下，我们获得了稳定、准确且长期的在线或 后验 解。我们的方法显著提高了在线大涡模拟（LES）运行的训练效率和速度，同时为此类湍流中的闭合物理机制提供了深入见解。我们的方法得益于在学习率和权重衰减系数空间中进行广泛的超参数搜索，以及采用周期性学习率退火；与固定学习率相比，这带来了更稳健、更准确的在线解。我们的CNN以粗尺度速度场或涡度和应变场作为输入，并输出偏应力张量的两个分量，即 $\boldsymbol{S}_d$。我们最小化亚格子尺度涡度通量散度（由高分辨率求解器计算）与CNN 建模的 $\boldsymbol{S}_d$ 所得到结果之间的损失，而无需施加能量或涡量平方守恒约束。浅层 CNN能够准确地对这类湍流进行参数化，这一成功表明，亚格子尺度应力对粗尺度场具有较弱的非局部依赖；这也与我们的物理认识一致，即小尺度由涡旋及其受应变丝状结构等大尺度局部控制。此外，由于其固有的低维性，2 层 CNN 参数化更有可能具备可解释性和良好的泛化能力。

Contents

1. 引言1.1.相关工作1.2. 当前工作2. 湍流模型与数据生成2.1. 动力学方程与湍流状态 2.2. 降尺度 2.3.亚格子尺度应力2.4.已解析尺度与亚格子尺度相互作用的局域性2.5.伽利略不变亚格子尺度模型2.6.基准参数化 3. 机器学习框架3.1. CNN架构与非局域性3.2. 数据生成与损失函数3.3.学习率退火与模型选择 3.4.超参数网格搜索 3.5.诊断 4. 结果4.1.离线准确率4.2.在线预测5.讨论与总结5.1. 亚格子尺度近局域性的物理与计算影响 5.2.理论后果与迈向可解释性5.3. 机器学习与物理设计选择 致谢 参考文献

2
3
4
4
4
6
7
8
8
9
10
10
11
12
12
13
14
15
15
22
22
23
24
24
24

1. 引言

物理系统中的湍流涵盖极其广泛的空间和时间尺度；例如，在地球的海洋中，作为海洋中动能的主要储库，中尺度涡旋的尺度为O(100km)，而在海气界面处，由表面加热/冷却、风作用和地球自转共同驱动的小尺度三维运动，其尺度为 O(1m)。如此广泛的空间尺度远远超出了当前数值海洋模型的求解范围，即使拥有现有的最大规模计算设施也是如此。气候模型面临着更大的挑战，因为它们需要运行数年或数十年，以研究长期时间范围内的气候变化。在实践中，这些模型以受可用计算资源限制的某一分辨率运行，同时未解析湍流过程则通过有效表示或参数化来处理。低于数值模型分辨率、因而需要进行参数化或表示的湍流运动，通常被称为亚网格尺度（SGS）运动；它们的影响则表示为有效通量（亚格子尺度通量），可将其加入数值模型已经求解的运动方程中。

历史上，未解析湍流流动的参数化是通过结合经验数据、基于物理的建模，以及利用短时间运行的高分辨率过程模型计算得出的简化代数模型来实现的。例如，海洋表面附近的三维小尺度湍流通常采用 K剖面参数化（Large 等人，1994）框架进行参数化。该框架是一种特设的经验方法与基于物理的方法相结合的方案，用于模拟海气界面处的各种湍流过程，包括表面对流以及边界层中的旋转分层剪切湍流。参数化也可以通过数值过程研究（本质上是针对特定现象开展的理想化数值模拟）获得，通常采用高分辨率大涡模拟（LES），结合量纲分析和物理模型来拟合简化的代数模型（Souza 等人，2020）。

随着机器学习方法的发展，尤其是基于神经网络的深度学习方法日益成熟，加之数值模拟产生的大量数据日益可得，一种替代方案是利用高分辨率过程研究生成亚格子尺度通量，从而无需借助简化的代数模型，也能对其进行准确建模。如今已有框架支持将使用 Python 编程语言训练的神经网络（NN）模型整合到 Fortran 中（Ott 等人，2020）。Fortran 是大多数天气、海洋和气候模型所使用的语言，因此原则上可以在气候模型中准确模拟亚格子尺度通量，从而显著提升其可预测性和保真度，并减少偏差。在实践中，一个突出的主要问题是，整合了基于神经网络模型的数值模型会面临数值稳定性方面的挑战（Brenowitz 等人，2020）。

神经网络是由更简单的模块化单元构成的层次化非线性函数，其中包含大量可调参数，原则上，只要有足够的数据，就能表示任意非线性函数（Hornik 等，1989；Zhou，2020）。通常，这类复杂模型可能会对给定数据集产生过拟合，无法学习数据中实际存在的通用关系。然而，过去十年间已经发现了大量正则化技术，使神经网络能够学习在未见数据上也具有良好泛化能力的非线性关系（Srivastava 等，2014；Ioffe 和 Szegedy，2015；Ba 等，2016；Loshchilov和 Hutter，2016）。将神经网络应用于动力学建模，尤其是求解支配气候系统的复杂动力学方程时，还面临额外要求：方程的解必须在数值上稳定，并且在长时间范围内不能累积非物理偏差。

动力系统中的基于神经网络的参数化可根据学习框架的设计方式进行分类。在线或后验 学习（Ma 等人，2018；Rasp，2020）将神经网络参数化直接纳入运动方程；对运动方程进行时间步进，并在满足以下约束的条件下学习神经网络的参数：由神经网络驱动的流动轨迹与高分辨率数值解的输出相匹配，而该数值解已截断至基于神经网络的解所采用的分辨率。这种方法属于轨迹优化的一种形式，在模仿学习中很常见（Hussein 等人，2017），其优点是神经网络驱动的解能够感知物理机制，并且经过充分优化的轨迹不会受到高分辨率数值解中不存在的偏差影响。然而，这种方法的一个缺点是，数值框架必须完全可微，因为轨迹中的误差需要通过数值求解器本身进行反向传播；这一思路与数据同化中的伴随模型密切相关。例如，天气和气候系统中的可微数值模型，无法通过简单修改现有气候和天气模型来构建，必须从头重写，工作十分复杂。此外，对于复杂的非线性动力系统，需要对轨迹进行多个时间步的展开，并优化多步损失，以确保所学习的基于神经网络的方程具有准确性和稳定性（Kochkov 等人，2021；Keisler，2022；

Frezat 等人，2022；List，2022）；这会大幅增加训练流程的计算负担和复杂性。

离线或 先验学习利用高分辨率数值模型来诊断低分辨率下缺失的所有亚格子尺度通量。然后，使用适当的神经网络架构对诊断得到的亚格子尺度通量进行直接建模，该架构以粗网格流动变量作为输入；神经网络训练采用标准的监督式训练（通过最大似然估计完成）。从方法论上看，这是一种显著更简单的方法，并且不会对数值求解器本身施加任何限制，因为数值求解与神经网络的学习是解耦的。因此，训练成功的神经网络模型可以直接整合到现有的气候和天气模型中。然而，相应地，这种方法的一个严重缺点是，神经网络模型并不了解底层系统的动力学，而且神经网络亚格子尺度模型产生的小误差在被纳入数值求解器后可能会累积，从而导致数值不稳定或有偏解；数值模型运行的时间越长，这一问题就越严重（Frezat 等人，2020； Rasp，2020；Frezat 等人，2022）。与使用固定的神经网络模型不同，对亚格子尺度通量进行概率建模可以缓解部分数值稳定性问题（Perezhogin 等人，2023年）。

无论是离线训练还是在线训练的模型，都必须在合理的长时间范围内进行在线或事后稳定性和精度测试。换言之，训练或学习过程可以离线进行（即直接监督学习），也可以借助可微数值求解器在线进行；但所有训练好的模型都必须部署到粗粒化方程中，并接受在线稳定性和保真度测试。我们再次强调，离线或在线学习过程与模型实际部署到方程中的过程有所不同；按照定义，后者始终是在线的，或称事后。基于神经网络的离线和在线学习方法还面临其他挑战，尤其是神经网络针对图形处理器（GPU）进行了优化，而大多数气候系统模型目前并非为在 GPU 上运行而设计。特别是，现代深度神经网络可能包含数百万个参数，在中央处理器（CPU）上运行会给数值求解器带来大量额外开销，甚至可能抵消构建精准亚格子尺度模型所带来的收益。

1.1. **相关研究**。 衰减二维湍流是用于发现和测试数据驱动闭合与参数化方法的最常见模型湍流问题之一（Maulik 等人，2019）。衰减湍流的物理机制已广为人知（Brachet 等人，1988；McWilliams，1984；Carnevale 等人，1991，1992），其过程包括涡旋缓慢合并并相互作用，形成越来越大的涡旋，直至整个流动域由一个大型涡旋占据。近期研究表明，即使在高雷诺数（Re∼20000）条件下，仅凭数量适中的数据样本，也能准确建模该问题中的亚格子尺度应力（Guan 等人，2022a；Frezat 等人，2022；List 等人，2022），从而在网格降尺度因子达到 16 倍或更高时，获得准确且稳定的在线粗网格解。

然而，持续受迫的二维湍流是一个复杂得多的动力学问题（Xiao 等人，2009），其能量存在持续的双重级联：能量从强迫尺度逆向级联至更大尺度，而涡量平方则正向级联至更小尺度（Kraichnan，1971）。过去两年间，只有少数近期研究成功地为这一问题及相关问题获得了准确、稳定的在线闭合。 Guan 等人(2022b)研究了正弦强迫的二维湍流流动问题，该流动在双周期通道中求解，在经典流体力学中也称为 Kolmogorov 流，并证明离线学习可有效地利用深层 CNN（10 层，约 80 万个参数）对亚格子尺度应力进行建模，从而实现稳定的在线闭合；前提是通过约 200 万个高分辨率模型时间步生成足够的训练数据。实际上，他们发现只需使用 2000 个快照进行训练即可，这些快照每隔 1000 个时间步选取一次，以确保彼此不相关。 Guan 等人(2022b)还发现，利用旋转等变卷积来发挥双周期计算域的几何对称性，可将数据需求大幅降低 40 倍。在一项较新的研究中，Ross 等人(2023)考察了一个与地球物理流动相关的理想化动力学问题的离线闭合，即 β平面上的双层准地转流；有关 β平面近似的简要说明，见第 2.1节。在这项范围广泛的研究中，Ross 等人(2023)考察了多种选择如何影响其离线训练卷积神经网络的在线准确率。这些选择包括输入是粗粒化速度还是速度梯度、亚格子尺度场的输出形式，以及降尺度方法的具体选择；这些网络有 8 层，约含 30 万个参数。

与使用神经网络对亚格子尺度应力进行参数化的标准方法不同，Kochkov 等人(2021通过“神经离散化”方法直接对非线性平流项进行参数化。该方法此前已在较简单的动力学问题中得到展示。Kochkov 等人(2021采用完全在线或事后训练方式，通过轨迹优化进行训练。同一团队随后发表的论文（Dresdner 等人，2022）发现，对于同一问题，在线训练的参数化方法也能取得准确结果；该研究采用谱方法（而非有限体积方法）求解，并使用直接对亚格子尺度进行参数化的卷积神经网络，而不是（Kochkov 等人，2021）所用的神经离散化方法。一个密切相关的

研究由Frezat 等人(2022开展；该研究也采用在线学习的参数化方法，研究受迫的 β平面湍流以及标准二维湍流，但雷诺数远高于（Dresdner 等人，2022）的研究。Dresdner 等人(2022）和Frezat 等人(2022）均使用了含 8 至 16 层的深层 CNN；前一项研究采用了编码器—处理器—解码器模型，这类模型如今在神经湍流预测模型中日益受到重视（Stachenfeld等人，2021；Keisler，2022）。

1.2. **当前工作**。 本研究提出的方法源自前述研究：这些研究通过对亚格子尺度应力进行纯离线训练，在高雷诺数（Re∼25000）下，对 β-平面上受正弦强迫的二维地转湍流进行参数化。然而，我们在此所作的选择与前述研究有诸多不同之处，从而显著提升了湍流闭合的训练效果和效率。这些差异不仅涉及技术层面，也为此类湍流流动闭合问题的物理和数学方面带来了新的认识。下面列出本研究的主要贡献。

(i) **高 Re 下精准且高效的浅层卷积神经网络闭合**。 我们证明，简单的双层卷积神经网络足以在高 Re-湍流中，以 16 倍降尺度获得稳定且长期准确的在线解。由此得到的亚格子尺度通量卷积神经网络参数化仅包含 $\mathcal{O}(10^3)$个参数；其规模比前述近期研究小两个数量级，从而使在线 CNN-LES 运行显著更快、更高效。

(ii) **超参数空间探测**。 我们通过对损失函数（3.1）优化过程中涉及的学习率和权重衰减系数进行广泛的超参数搜索，找出最佳模型。得益于本文采用的小型卷积神经网络，这一探测过程大为简化。

(iii) **优化策略**。 我们证明，与传统采用固定学习率训练的卷积神经网络相比，采用周期性学习率退火训练的卷积神经网络能够得到更稳健、更准确的在线解。

(iv) **物理变量**。 卷积神经网络的输入采用粗网格速度场（$\bar{u}, \bar{v}$）或涡度和应变¯场（$\bar{\omega}, \bar{\sigma}_s, \bar{\sigma}_n$），输出为偏应力张量的两个分量；我们不考虑（$\psi, \bar{\omega}$）¯输入，因为在复杂空间区域上的数值模型中难以获得流函数 ψ。

(v) **损失函数**。我们的卷积神经网络通过最小化均方误差损失函数进行优化，该损失函数衡量其相对于以偏应力张量表示的亚格子尺度的参数化缺陷，并通过神经网络权重的平方和进行惩罚；见下文（3.1）。我们不需要使用近期工作中采用的更复杂的物理机制约束损失函数（List 等人，2022；Guan 等人，2022b）。

(vi) **弱非局地参数化及其物理意义**。 在本文所研究的这类湍流问题中，浅层 CNN 能够准确地对湍流进行参数化，这一成功直接影响了我们对问题本身物理机制的理解。首先，由于 CNN 的空间非局域性会随着网络深度增加，因此我们基于浅层 CNN 得出的闭合结果意味着，亚格子尺度应力对粗尺度场的依赖仅具有弱非局域性。浅层 CNN 参数化具有较弱的空间非局域性，这也意味着它们更容易嵌入气候模型；气候模型通常依赖空间域分解技术，在大型集群上进行计算。相应地，深层 CNN 则需要在计算节点之间交换规模大得多的数据。

最后，鉴于通常需要具有大量参数的深度神经网络来逼近复杂的非线性函数，而我们的卷积神经网络较为浅层，因此我们推断，亚格子尺度应力在此处对粗尺度流场的依赖具有相对简单的非线性关系。这一简单观察为我们浅层卷积神经网络参数化背后的非线性映射提供了有利基础，使其具有可解释性（Π_{CNN} 见下文公式 5.1），对此我们留待未来研究。Ross 等人 （2023年）将混合遗传编程与稀疏线性回归方法相结合所获得的近期解析算子形式相对简单，这与我们的发现一致。

2. 湍流模型和数据生成

2.1. **动力学方程和湍流状态**。 我们采用一个广泛使用的流体流动问题模型，该模型表现出复杂的湍流现象，并且与地球系统和行星大气中的大尺度湍流密切相关——即 β平面上的正弦强迫二维湍流。流动域在（x， y）坐标中定义，由一个大小为（L， L）的正方形域组成，其中 $L = 2\pi$。控制运动方程是纳维-斯托克斯（N-S）方程，用于描述流速矢量场的演化， $\boldsymbol{v}(x, y, t) = (u(x, y, t)\ v(x, y, t))$，在每个

在区域和时间中的点。对于此处选定的具体问题，边界条件取为周期性的，即

$$u(x,y,t) = u(x+L,y+L,t), \tag{2.1}$$
$$v(x,y,t) = v(x+L,y+L,t). \tag{2.2}$$

请注意，该区域在拓扑上等价于二维环面。在二维情况下，存在背景旋转时的 N-S 方程是速度场两个分量的演化方程，$\boldsymbol{v} \equiv (u,\ v)$，其中 u 和 v 分别是沿 x方向和 y方向的速度。无量纲形式下，方程变为

$$\partial_t \boldsymbol{v} + \boldsymbol{v}\cdot\nabla\boldsymbol{v} + f(\boldsymbol{k}\times\boldsymbol{v}) = -\mu\boldsymbol{v} + \frac{1}{Re}\nabla^2\boldsymbol{v} + \boldsymbol{F_v}(x,y)\,. \tag{2.3}$$

此处，f，科里奥利参数是地球或其他行星大气的局部旋转速率，$\boldsymbol{k}$ 是垂直于$(x,\ y)$平面的单位向量，而 $\boldsymbol{F_v}(x,\ y) = [-\sin(k_f y),\ \sin(k_f x)]$ 是持续驱动流动的正弦时间不变强迫场；这一特定形式与（Guan 等人, 2022b）中的形式相同，其中强迫波数为 $k_f = 4$。系数 μ表示线性阻力系数，在地球物理流动中用于表示底部摩擦的影响，而 Re 是衡量非线性平流项相对于黏性项强度的雷诺数（左侧第二项相对于右侧第二项）；我们取 $Re = 25000$。

速度场还必须满足通过连续性方程体现的质量守恒，$\nabla\cdot\boldsymbol{v} = \partial_x u + \partial_y v = 0$。在二维情况下，可以通过定义一个称为流函数的标量场来隐式满足连续性方程，$\psi(x,\ y,\ t)$

$$u = -\psi_y\ , \qquad v = \psi_x \tag{2.4}$$

在二维情况下，这两个方程连同连续性约束可以表示为单个标量场——涡度的演化方程，涡度定义为速度场的二维旋度。

$$\omega = \partial_x v - \partial_y u = \nabla^2\psi, \tag{2.5}$$

并捕捉流体微团的局部旋转。因此，二维涡度演化方程可写成标准的涡度-流函数（$\omega,\ \psi$）形式，即

$$\partial_t\omega + J(\psi,\omega) - \beta v = -\mu\omega + \frac{1}{Re}\nabla^2\omega + F_\omega(x,y), \tag{2.6}$$

其中雅可比项 $J(\psi,\omega) = \psi_x\omega_y - \omega_x\psi_y$ 表示非线性平流项，而强迫项的涡量形式则变为

$$F_\omega(x,y) = k_f[\cos(k_f x) + \cos(k_f y)] \tag{2.7}$$

βv项是一个与地球物理系统相关的近似，在切平面近似（$\beta = \frac{\partial f}{\partial y}$）下，能够捕捉地球系统中海洋和大气所经历的微分旋转效应。β-项的存在使流动能够表现出行星系统中常见的一类特定波动，称为罗斯贝波；相较于 $\beta = 0$时产生的流动，这些波动可以显著改变湍流流动动力学。

式 (2.6) 采用标准伪谱方法在傅里叶空间中求解（Orszag 和 Israel，1974）；时间推进采用半隐式 Crank-Nicholson 二阶 Adams-Bashforth（CN-AB2）方法，时间步长为$\Delta t = 5\times10^{-5}$。我们的基准高保真模型在一个正方形 $N_0\times N_0$ 网格上求解，其中 $N_0 = 1024$，因此物理空间中的网格间距为 $\Delta_{DNS} = 2\pi/N_0$。按照标准术语，我们将这些解称为直接数值模拟（DNS），因为除二次黏性外，没有使用任何参数化来表示网格尺度以下的湍流。我们的黏性系数取值，即无量纲形式中的 $1/Re$，选择时确保：固定 Re 并减小网格尺寸，或固定网格尺寸并增大 Re （减小黏性），都会不可避免地导致能量迅速堆积在最小波数处，最终造成数值不稳定。作为对比，Dresdner 等（2022）在 (Dresdner 等，2022) 中采用的 $Re\sim1000$ 使作者能够在最高达 128×128的较粗网格上获得稳定解，且不使用任何参数化；他们的问题也涉及从 1024×1024到 64×64的 16 倍降尺度。Frezat 等人（2022）在其 32 倍降尺度在线闭合中使用的 Re 大于本研究所用值；而 Ross 等（2023）在其双层 QG 湍流研究中（降尺度因子为 4 倍，从 256×256到 64×64）使用的 Re 值可能与 Dresdner 等（2022）的取值范围相近。

在没有 β项的情况下，流动在时空上具有统计均匀性和各向同性。图1a）展示了经过 $t = 2\times10^6\Delta t$ 后涡度场的单个瞬时快照；大型相干涡旋通常具有不小于

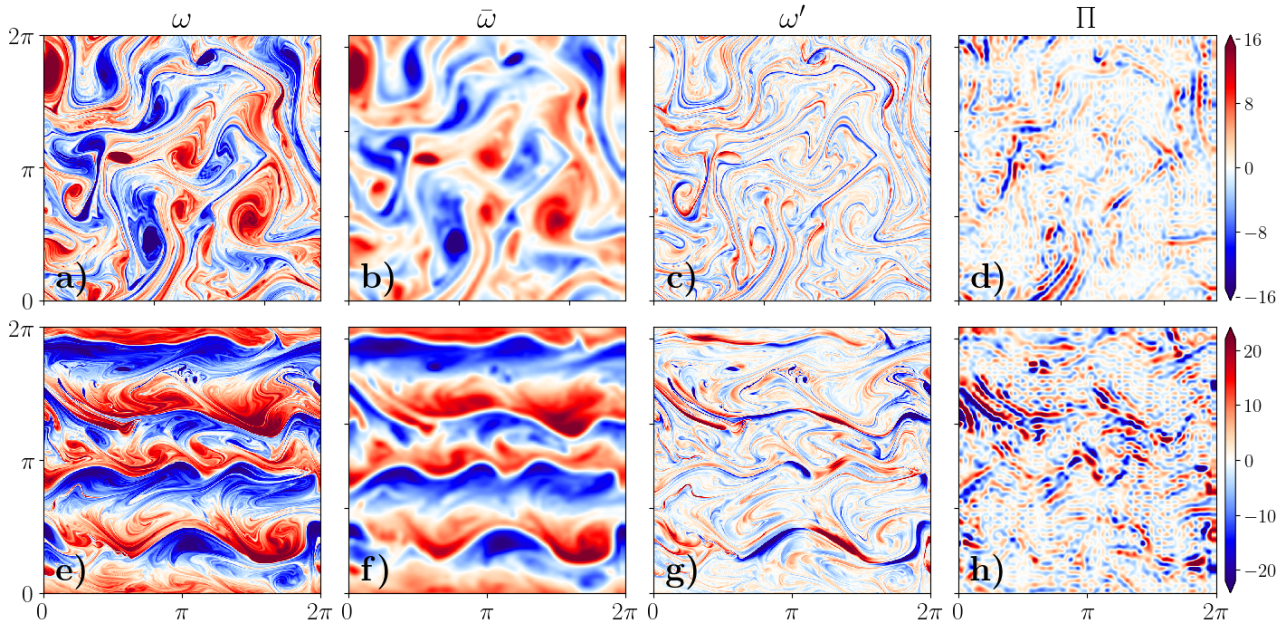

图 1. $t = 2{\times}10^6\Delta t$ 后 $\beta = 0$ 时的快照：a) 直接数值模拟涡度 ω；b) 滤波直接数值模拟涡度，$\bar{\omega}$；c) 亚格子尺度涡度场，$\omega' = \omega - \bar{\omega}$；d) 亚格子尺度应力散度，$\Pi$。e)-f) 为 $\beta = 20$时的对应结果；注意沿 x轴的射流。

强迫尺度的长度尺度，在细密的丝状结构构成的海洋中十分醒目。作为能量向更大尺度级联的结果，流动动能（$\mathcal{E} = u^2 + v^2$）集中在相干涡旋中；而涡旋核心和丝状结构都是流动涡量平方（$\mathcal{Z} = \omega^2$）的重要储库，后者是涡量平方向小尺度直接级联的结果（Kraichnan，1971）。

然而，当 β取有限值时，湍流会在 y方向上发生对称性破缺不稳定性，从而形成交替的带状纬向急流，即沿 x方向排列。急流形成的机制并不简单，最好将其描述为一种湍流不稳定性，称为纬向急流不稳定性（Farrell 和 Ioannou，2007；Marston 等，2008；Srinivasan 和 Young，2012；Bakas 等，2015）。这类类似急流的结构常见于行星大气中，最引人注目的例子是木星上肉眼可见的带状急流结构。图 1（e）展示了在这一特定 $\beta = 20$ 条件下形成的急流快照；此时，在 $\beta = 0$ 情形中观察到的相干涡旋已不再单独可见，而是与急流本身紧密相互作用，不过仍可清楚地看到细尺度丝状结构。

2.2. 降尺度。 本节介绍我们如何将在上一节中描述的高保真直接数值模拟解降尺度或进行粗粒化，使其适用于低分辨率网格。降尺度在谱空间中使用截断滤波器实现，该滤波器也称为锐谱滤波器，其简单地将高于截断波数的模态设为零。对于降尺度因子 n_d，降阶模型的网格间距因此为 $\Delta_c = n_d\Delta_0$，网格点数量则减少为 $N_c = N_0/n_d$。我们选择 $n_d = 16$，从而将直接数值模拟解从 1024×1024 网格投影到 64×64网格。在实际操作中，在应用截断滤波器之前，遵循Guan 等人(2022b的做法，我们首先对场进行一定程度的高斯滤波。选择这一做法有三个原因：首先，Zhou 等人(2019) 证明，这一过程产生的亚格子尺度通量与粗粒化场具有更高的相关性，从而使学习问题更加容易。其次，这使谱学习问题更接近地球系统模型中常见的有限差分方法和有限体积方法，因为局部化的有限差分模板可以表示为一种有效的指数截断滤波器（Lele，1992）。最后，仅使用截断滤波器会在物理空间中产生小尺度的空间非局部特征，这些特征几乎没有物理结构，并且由于神经网络的谱偏差而难以学习；神经网络更容易学习低频信息，而不是高频信息（Rahaman 等人，2019），尤其是在离线学习的参数化方案中，因为其中不存在固有的动力学感知；参见第1节的讨论。这些原因解释了为什么使用高斯滤波器和指数滤波器成为

近期涉及基于卷积神经网络的参数化的二维湍流研究中，截断之前的一种常见做法；例如（Guan 等人，2022a，b；Dresdner 等人，2022；Frezat 等人，2022）。记 $\widehat{\omega}(k_x,\ k_y)$ 为涡度的傅里叶变换，$\omega(x,\ y)$，粗粒化场（用上横线表示）写为

$$\overline{\widehat{\omega}}(k_x,k_y)=\widehat{\omega}*G_{\Delta_c},\ \ |k_x|,|k_y|<k_c, \tag{2.8}$$

其中，$*$ 表示卷积算子，$k_c=\pi/\Delta_c=N_c/2$ 是截断波数（对于 $N_c=64$、$k_c=32$），高斯滤波器 G_{Δ_c} 给定为

$$G_{\Delta_c}(k_x,k_y)=\exp\left(-\left(\frac{k_x^2+k_y^2}{24}\right)\Delta_F^2\right) \tag{2.9}$$

$$=\exp\left(-\frac{\pi^2}{6}\frac{k^2}{k_c^2}\right), \tag{2.10}$$

与 $\Delta_F=2\Delta_c$；参见（Guan 等人，2022a）。在低分辨率网格上从直接数值模拟解中滤波得到的场称为滤波直接数值模拟（FDNS）场。图 1b）和 f）分别展示了与图 1a）和 e）中的直接数值模拟快照相对应的 $\beta=0$ 和 $\beta=20$ 情况下的 16X 下采样场（即 FDNS 场）的快照。有趣的是，Ross 等人（2023年）发现，上述高斯滤波方法使学习问题更具挑战性，但我们发现卷积神经网络的离线学习或其在线部署都很少出现问题。

2.3. **亚格子尺度应力**。我们将式（2.8）中的滤波步骤应用于动量方程（2.3）和相应的涡度方程（2.6）。随后将动量方程改写为

$$\partial_t\bar{\boldsymbol{v}}+\bar{\boldsymbol{v}}\cdot\nabla\bar{\boldsymbol{v}}+f(\boldsymbol{k}\times\bar{\boldsymbol{v}})=-\mu\bar{\boldsymbol{v}}+\frac{1}{Re}\nabla^2\bar{\boldsymbol{v}}+\boldsymbol{F_v}(x,y)+\nabla\cdot\bar{\boldsymbol{S}}, \tag{2.11}$$

其中亚格子尺度动量通量张量（有时简称应力张量）是

$$\boldsymbol{S}=\begin{bmatrix}\tau_{uu} & \tau_{uv}\\ \tau_{uv} & \tau_{vv}\end{bmatrix}, \tag{2.12}$$

其中 $\tau_{uu}=\overline{u^2}-(\bar{u})^2$、$\tau_{uv}=\overline{uv}-\bar{u}\bar{v}$ 和 $\tau_{vv}=\overline{v^2}-(\bar{v})^2$ 是三个张量分量，因为 $\bar{\boldsymbol{S}}$ 显然是一个对称张量。我们将式（2.12）分解为

$$\boldsymbol{S}=\underbrace{\begin{bmatrix}\frac{\tau_{uu}-\tau_{vv}}{2} & \tau_{uv}\\ \tau_{uv} & -\frac{\tau_{uu}-\tau_{vv}}{2}\end{bmatrix}}_{\boldsymbol{S}_d}+\underbrace{\frac{\tau_{uu}+\tau_{vv}}{2}\boldsymbol{\mathcal{I}}}_{\boldsymbol{S}_0}, \tag{2.13}$$

其中，$\boldsymbol{\mathcal{I}}$ 是单位矩阵，$\mathcal{E}=\frac{\tau_{uu}+\tau_{vv}}{2}$ 是亚格子尺度动能。右端（RHS）的第一项 $\boldsymbol{S}_d$ 称为偏应力张量，只有两个独立分量；第三个独立分量则表现为对角张量的大小 $\boldsymbol{S}_0$。请注意，强迫项不受滤波操作的影响，因为我们始终选择了截断尺度 $k_c>k_f$。

将滤波器（式2.8）应用于涡度方程（式2.6），得到对应于式（2.11）的“滤波”方程，其形式为

$$\partial_t\bar{\omega}+J(\bar{\psi},\bar{\omega})-\beta\bar{v}=-\mu\bar{\omega}+\frac{1}{Re}\nabla^2\bar{\omega}+f(x,y)+\Pi, \tag{2.14}$$

其中

$$\Pi=\overline{J(\psi,\omega)}-J(\bar{\psi},\bar{\omega}), \tag{2.15}$$

是亚格子尺度涡度通量的散度，它还可以用另外两种形式表示。第一种形式将 Π 表示为亚格子尺度涡度通量的散度：

$$\Pi=\nabla\cdot(\overline{\boldsymbol{v}\omega}-\bar{\boldsymbol{v}}\bar{\omega}). \tag{2.16}$$

第二种形式与本文采用的方法更为直接相关，它将 Π 与亚格子尺度动量通量张量 $\boldsymbol{S}$ 联系起来

$$\Pi=\nabla\times(\nabla\cdot\boldsymbol{S})=\nabla\times(\nabla\cdot\boldsymbol{S}_d) \tag{2.17}$$

其中，上述旋度运算是二维旋度。上述第二个等式成立，是因为 $\boldsymbol{S}_0$ 是一个对角张量，因此可得 $\nabla\times(\nabla\cdot\boldsymbol{S}_0)=0$。图 1d) 和 h) 展示了 Π 在 $\beta=0$ 和 $\beta=20$ 这两种情况下的快照；同一图中还展示了相应的亚格子尺度涡度场 $\omega'=\omega-\bar{\omega}$ 以供参考。ω' 与 Π 之间存在很强的相关性，但显然，这种依赖关系并不简单。

原则上，拥有一个关于 Π 的数据驱动模型，使我们能够求解方程（2.14），但问题在于应采用上述哪种形式的 Π 进行建模。最常用的是方程（2.15）中的形式，即使用神经网络直接对 Π 进行建模（Maulik 等人，2019；Guan 等人，2022b,a；Frezat 等，2022）。另一种方法是对亚格子尺度涡度通量矢量 $\boldsymbol{v}\omega - \bar{\boldsymbol{v}}\bar{\omega}$ 的两个分量进行建模，再通过计算散度求得方程（2.16）中的Π 。本文采用的方法受 Zanna 和 Bolton（2020）的启发，即对亚格子尺度动量通量 $\boldsymbol{S}$ 的分量进行建模，并使用方程（2.17）计算 Π 。Zanna 和 Bolton（2020）使用单个卷积神经网络对 $\boldsymbol{S}$的三个分量进行建模，分别为 τ_{uu}、 τ_{vv} 和 τ_{uv}。本研究提出了这种方法的一种变体：通过学习偏应力张量的两个分量（$\tau_{uu} - \tau_{vv}$）/2 和 τ_{uv}的卷积神经网络近似值来进行建模。正如前文所述，这足以得到 Π；见方程（2.13）。

2.4. **解析尺度与 SGS 尺度相互作用的局部性**。 解析场与 SGS 运动之间的相互作用仅具有弱非局部性（Eyink，2005；Eyink 和 Aluie，2009）。例如，可以通过考察 $\bar{\omega}$ 和 Π之间的二维空间互相关来推断这一性质。出于效率考虑，我们对每个快照使用快速傅里叶变换和互相关定理（维纳-辛钦定理的推广）计算空间互相关（Fisher，2008），然后对其进行时间平均。在图 2中，我们展示了互相关的 x向和 y向截面。当 $\beta = 0$时，互相关基本上是各向同性的，因此考察这些截面中的任意一个即可；从图 2 可以看出，空间中超过 7 个网格点的区域后，相关性就很弱了。当 $\beta = 20$时，沿射流方向的相关性与 $\beta = 0$ 结果相似，但横跨射流方向的相关长度更大；这与构成流动的各射流之间的相互作用和耦合有关。不过，即使在这种情况下，高度相关的区域仍然相对局部。这一观察结果促使我们采用本文中所用的神经网络架构，具体内容将在后文详述。

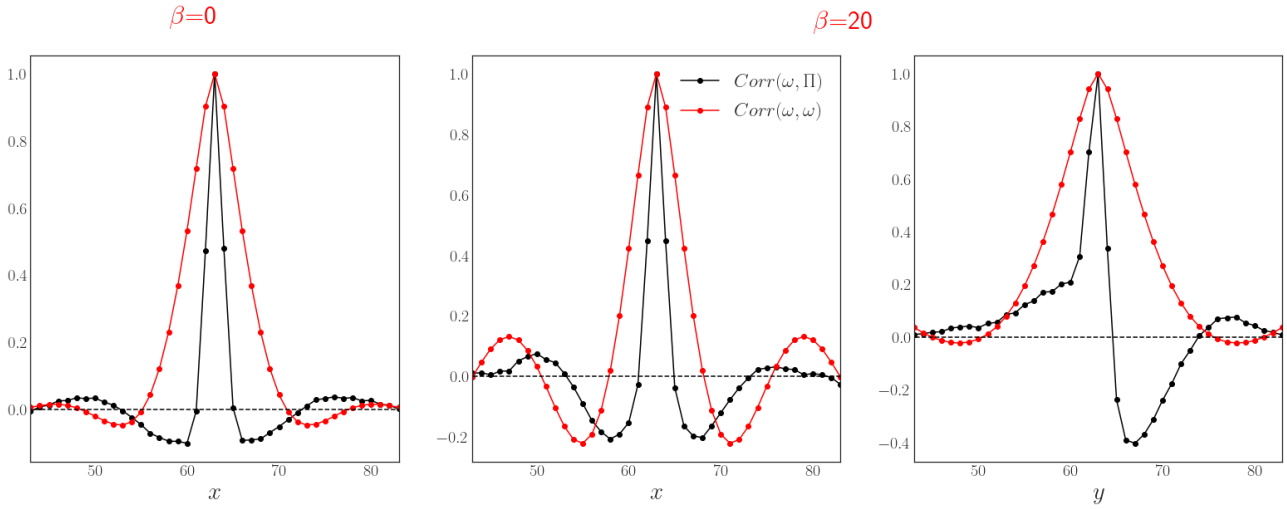

图 2. Π 与解析涡量场 $\bar{\omega}$之间空间互相关的 x和 y截面。对于 $\beta = 0$，两个方向上的相关性几乎对称，因此仅显示其中一个方向。对于有限的 β，横向于喷流的相关长度（沿 y）大于沿喷流方向的相关长度；后者与 $\beta = 0$情况下的相关长度相当。

顺带一提，我们强调，解析涡度的自相关突显了 $\beta = 0$ 情况下涡旋的平均尺度，以及 $\beta = 20$ 时的喷流尺度（图中红色曲线 2）。

2.5. **伽利略不变亚格子尺度模型**。 运动方程，方程（2.3）和（2.6）具有伽利略不变性，即在惯性参考系变换下保持不变。为说明这一点，且不失一般性，我们可以在以恒定速度平移的参考系中重写这些方程（用撇号表示），$\boldsymbol{V} \equiv (U, V)$ 沿着 x 和 y方向，即

$$x' = x - Ut, \quad y' = y - Vt, \quad t' = t. \tag{2.18}$$

and

$$u' = u + U, \quad v' = v + V, \quad \psi' = \psi + Vx - Uy, \quad \omega' = \omega. \tag{2.19}$$

请注意，涡度对于均匀平移是不变的，而其他量则不是；这是因为涡度由速度的梯度构成；一般而言，速度的全部四种可能梯度对于伽利略变换都是不变的

$$v'_x = v_x, \quad v'_y = v_y, \quad u'_x = u_x, \quad u'_y = u_y, \tag{2.20}$$

这些速度梯度的线性组合也是如此，包括 ω 以及两个应变分量，即分别定义为法向应变和剪切应变的分量，

$$\sigma_n = v_x - u_y, \tag{2.21}$$

and

$$\sigma_s = v_x + u_y. \tag{2.22}$$

伽利略不变性的一个重要推论是，滤波后的方程（2.11）和（2.14）也必须满足伽利略不变性（滤波算子与时间无关），因此亚格子尺度通量也必须满足这一性质，$\boldsymbol{S}$ 和 Π。

湍流建模或闭合是指对亚格子尺度通量进行建模：可以直接建模 Π，也可以对 $\boldsymbol{S}$ 建模，再利用式（2.17）求得 Π，并将其表示为粗尺度场变量的函数；不过，如上所述，输入变量也可以有不同的选择。求解滤波涡度方程（2.14）时，一个自然的选择是将粗尺度¯场变量（ψ，$\bar{\omega}$）作为输入。在（Maulik 等人，2019；Guan 等，2022a,b；Frezat 等，2022）中采用的这一选择，是将 Π¯ 直接建模为 $\Pi = \Pi_\theta(\psi, \bar{\omega})$；其中 θ 按照惯常说法，表示构成模型的一组参数，其模型可以是简单的线性回归模型，也可以是神经网络等更复杂的形式。另一种选择是将粗粒化速度作为输入（Dresdner 等人，2022），即 $\Pi = \Pi_\theta(\bar{u}, \bar{v})$。

Zanna 和 Bolton (2020) 使用 $(\bar{u}, \bar{v})$ 作为输入，并对 $\boldsymbol{S} = \boldsymbol{S}_\theta(\bar{u}, \bar{v})$ 的三个分量进行建模，以通过方程（2.17）求得 Π。然而，值得注意的是，这些选择都不满足伽利略不变性，因为改变惯性¯参考系会导致 $(\psi, \bar{u}, \bar{v})$ [按照方程（2.19）发生变化]

$$\bar{u}' = \bar{u} + U, \quad \bar{v}' = \bar{v} + V, \quad \bar{\psi}' = \bar{\psi} + Ux + Vy. \tag{2.23}$$

因此，从先验角度看，以上述变量为函数的模型不满足伽利略不变性；不过，如果数据充足，参数数量足够多的神经网络很可能会学到这种物理对称性。然而，我们可以采用一种简单且充分的建模选择，隐式地假设伽利略不变性。对于偏应力张量的两个分量 $\boldsymbol{S}_d$，我们的目标是找到以下参数化形式

$$\boldsymbol{S}_d = \boldsymbol{S}_d^\theta(\bar{\omega}, \bar{\sigma}_n, \bar{\sigma}_s), \tag{2.24}$$

其中，θ 表示卷积神经网络的所有参数。由于 $(\bar{\omega}, \bar{\sigma}_n, \bar{\sigma}_s)$ 表示速度梯度的所有可能线性组合，因此加入均匀速度后，它们显然保持不变[1]。下文也将按照 Dresdner 等人 (2022)；Zanna 和 Bolton (2020)；Ross 等人 (2023) 学习非伽利略不变模型

$$\boldsymbol{S}_d = \boldsymbol{S}_d^\theta(\bar{u}, \bar{v}). \tag{2.25}$$

¯我们选择不学习将以下内容作为输入的神经网络模型 $(\psi, \bar{\omega})$ (例如 Guan 等人 (2022a); Frezat 等人 (2022))，原因很简单：在现实的地球物理模型中，流函数通常无法获得，因此这类选择在实践中不易推广。确实值得注意的是，在谱空间中，涉及的微分算子往往是对角的，但通常需要求解一个计算代价高昂的泊松方程才能得到 ψ。

2.6. 基准参数化。遵循近期研究（Guan 等人，2022b；Frezat 等，2022；Ross 等，2023），我们选择 Smagorinsky 和 Leith 参数化作为基准。这些模型广泛用于当代大涡模拟研究，以及当前气候系统中使用的现实数值模型（Bachman 等人，2017；Pearson 等人，2017）。Smagorinsky 参数化通常在动量方程（2.11）中实现，并以粗尺度应变张量来参数化 $\boldsymbol{S}_d$。然而，我们选择的 Smagorinsky 版本直接以粗尺度涡度场 $\bar{\omega}$ 为基础实现（Maulik 等人，2019），且采用散度形式（Frezat 等，2022）；请注意，Smagorinsky 的动量形式和涡度形式并不等价。上述两种参数化均为扩散型参数化，并假设 Π 可表示为扩散形式：

$$\Pi = \nabla \cdot (\nu_e \nabla \bar{\omega}), \tag{2.26}$$

[1]请注意，第四个线性组合，即散度，$\bar{\delta} = \bar{u}_x + \bar{v}_y = 0$ 是由连续性关系决定的。

其中 唯一未知量是“涡流”扩散率 ν_e。Smagorinsky 参数化将 ν_e a 建模为 s

$$\nu_e = c_s \Delta_c^2 |\bar{S}|, \tag{2.27}$$

其中，$|\bar{S}|^2 = \sigma_n^2 + \sigma_S^2$ 是应变幅值，Δ_c 是粗网格的网格尺寸。Leith 参数化的形式为

$$\nu_e = c_l \Delta_c^3 |\nabla \bar{\omega}| \,. \tag{2.28}$$

请注意，如上一节所述，上述两种参数化均为伽利略不变。按照（Ross 等人，2023年），我们改变 c_l 和 c_s 在 $[0.01, 1.0]$ 中的取值，并使用 Smagorinsky 和 Leith 参数化运行粗网格解，使其演化时间与 CNN-LES 同样长，以便进行直接比较。先透露一下后续结果：我们注意到，在上述涡度形式下，当 $c_s = c_l$时，Leith 和 Smagorinsky 模型得到的流动极为相似。

3. 机器学习框架

3.1. CNN 架构与非局域性。本文采用完全离线的机器学习流程来学习亚格子尺度通量。如前所述，为求解滤波涡度方程（式（2.14））， 我们使用卷积神经网络对 $\boldsymbol{S}_d$ 进行建模，该网络以滤波场变量作为输入，输入形式为（$\bar{u}$， $\bar{v}$）或（$\bar{\omega}$， $\bar{\sigma}_n$， $\bar{\sigma}_s$）；随后根据式（2.17）由 $\boldsymbol{S}_d$ 得到 Π。卷积层具有平移等变性，是自然的选择，因为纳维–斯托克斯方程也具有相同的对称性。近期研究（Guan 等，2022a,b；Dresdner 等，2022；Frezat 等，2022；Ross 等，2023）采用深层 CNN 进行参数化，其参数数量从几十万到一百万不等。然而，我们发现，浅层 CNN，尤其是仅含 $\mathcal{O}$（1000）个参数的简单两层神经网络，足以对亚格子尺度应力进行建模。我们的具体架构如图 3 所示。输入层包含 2 个或 3 个输入，具体取决于选择速度，还是涡度与应变的组合作为输入；唯一的隐藏层包含 N_f 个卷积滤波器，两个输出层则仅包含构成 $\boldsymbol{S}_d$的两个独立场，即（$\boldsymbol{S}_d^{00}$， $\boldsymbol{S}_d^{01}$） = （$\frac{\tau_{uu}-\tau_{vv}}{2}$， τ_{uv}）；见式（2.13）。每层中每个卷积滤波器的宽度（通常称为卷积核尺寸）设为 $ks = 5$。 N_f 是一个关键超参数，我们将其设为 8、16 或 32。对于（$\bar{u}$， $\bar{v}$）输入，这几种取值对应的参数总数分别为 800、1600 和 3200。无论哪种情况，卷积滤波器的数量都远低于常见水平；我们的目标是找到能产生准确且稳定解的最小神经网络，并且我们证明这样规模的网络确实足够。我们在第一层之后使用一个 Swish 激活函数 $\mathcal{S}$；即 $\mathcal{S}$（ξ） $= \xi H$（ξ），其中 H（ξ）是 Sigmoid 函数。在许多具有挑战性的数据集上，Swish 的表现往往优于 ReLU（Ramachandran 等，2017）。

支持我们选择浅层卷积神经网络的两个重要依据中，第一个是有效非局域性。在计算机视觉文献中，这一量也称为卷积神经网络的感受野（RF）。它被定义为输入域中的一组点，这些点与卷积神经网络当前处理的点区域中心相连接。换言之，这就是卷积神经网络的有效滤波器宽度。例如，一个具有 5×5 滤波器的单个卷积层具有一个简单的 5×5 =25 点感受野，即有效滤波器宽度为 5。由于神经网络是由包含卷积滤波器的层组合而成，感受野是滤波器数量的线性函数。因此，一个具有 5×5 个滤波器的两层卷积神经网络，其感受野为 9×9 个点，即有效滤波器宽度[2]为 9。作为比较，Guan 等人(2022b) 使用的 10 层卷积神经网络具有41×41的感受野，在 64× 64 域上，其非局域性几乎达到全局范围。Maulik 等人(2019) 的研究使用空间局域神经网络，采用全连接神经网络而非卷积神经网络的形式，其感受野仅为 3×3；对于当前问题而言太小了。Ross 等人(2023年) 使用了一个具有 21×21感受野的 8 层卷积神经网络；他们针对其 $\beta = 0$ 情形，考察了单个点处输出亚格子尺度通量相对于输入的梯度，发现非零值仅出现在一个9×9空间区域内，这与本研究所选两层卷积神经网络的感受野一致。然而，我们还发现了一个并非显而易见的结果：同一个两层 CNN 架构也足以准确模拟 $\beta = 20$ 射流情形，后续章节将对此进行说明。

RF 与数值离散中的 模板尺寸密切相关，在求解偏微分方程的语境下可直接将其视为模板尺寸。尽管我们对卷积神经网络的选择隐含地假设了非局域性，这与其他关于亚格子尺度闭合的研究相同；但与之相反，我们在下文展示的使用两层卷积神经网络得到的闭合结果表明，即使在高雷诺数下，亚格子尺度参数化中固有的非局域性程度也很弱。这一观察结果对在海洋和大气模型中部署亚格子尺度参数化具有影响，因为这些模型通常采用空间域分解

[2]注意，对于由 $ks = 5$ 个滤波器组成的一般 n 层卷积神经网络，其有效宽度为 $2n - 1$。

以便在大型计算集群上求解。具有较大空间局部性的卷积神经网络会严重限制区域分解，而选择浅层 CNN 可缓解这一问题。

需要注意的第二点是有效的非线性。由于具有层次结构，层数更深的神经网络能够表示比具有相同参数数量的浅层神经网络更复杂的非线性函数族（Bengio 和Delalleau，2011）。因此，网络中的较低层学习简单的特征，而较高层则在此基础上构建，以表示更复杂的模式和抽象表示。因此，我们选择两层卷积神经网络，实际上隐含了这样的假设：亚格子尺度通量与输入粗粒化流动变量之间关系的复杂性和稀疏性较为有限。通过 1×1“卷积”层可以在不增加其感受野或非局域性的情况下提高卷积神经网络的非线性程度；严格来说，这些层并不是卷积层，而是沿特征维度连接的全连接层。然而，我们并不认为有必要采用这类选择。

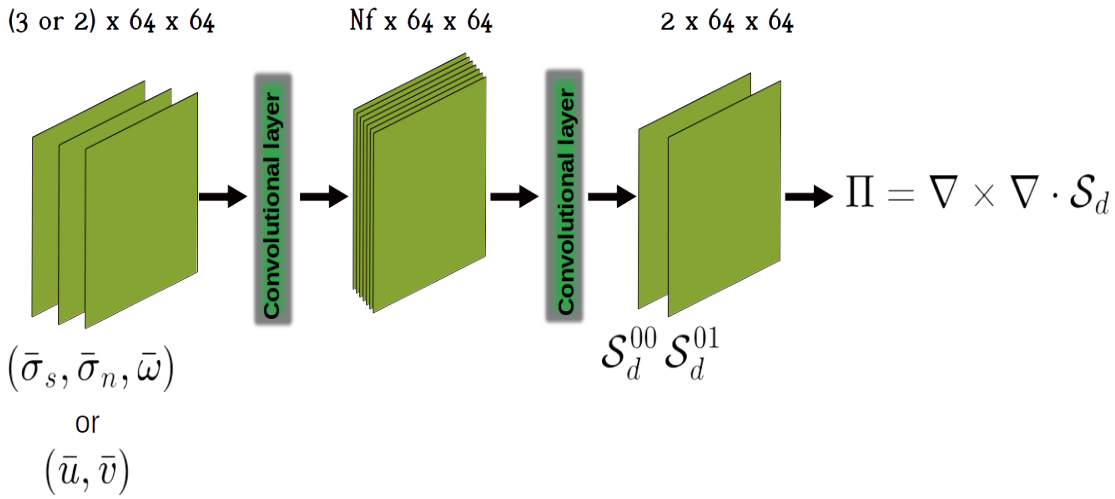

图3. **用于模拟亚格子尺度应力的卷积神经网络流程**。输入场可以是粗粒化场组合 $(\bar{\omega}, \bar{\sigma}_n, \sigma_s)$ 或 $(\bar{u}, \bar{v})$，输出滤波器则是偏应力张量的两个相关分量，$\boldsymbol{S}_d^{00}$和 $\boldsymbol{S}_d^{01}$。亚网格尺度涡量通量首先通过张量散度计算，随后进行二维旋度运算。

3.2. **数据生成与损失函数**。由于我们的卷积神经网络输出为偏应力张量，$\boldsymbol{S}_d$，我们使用式（2.17）计算式（2.14）中使用的亚格子尺度涡度通量散度。我们的损失函数是一个简单的 L_2 损失，用于衡量卷积神经网络对 Π 的参数化结果与 FDNS 得到的结果之间的差异，并按如下方式进行正则化

$$\mathsf{L} = \|\nabla \times (\nabla \cdot \boldsymbol{S}_d^{\theta}) - \Pi_{\mathrm{FDNS}}\|_2^2 + \mathrm{wd} \sum \theta_j^2. \tag{3.1}$$

这里，范数 $\|\cdot\|_2$ 指的是流动域上的 L_2 范数，而对神经网络所有权重平方和施加的 L_2 惩罚系数称为“权重衰减”系数。与近期研究（List 等人，2022；Guan 等人，2022b）相比，使用额外的物理信息损失来进一步约束离线训练，并不是在浅层卷积神经网络框架中推导出精确闭合模型的重要因素。正如后续章节所示，无需在损失函数中加入更多物理约束，我们也能得到高保真度的解。

为了训练我们的 2 层卷积神经网络，我们生成了四条独立轨迹，其高分辨率解具有不同的随机初始条件。按照 Guan 等人 (2022b 的方案，我们采用间隔 $1000\Delta t$ 次迭代的弱相关快照进行训练和测试，以提高训练数据集的多样性。我们从前两条轨迹中各取 200 个快照作为训练集，共计 400 个快照，对应 400,000 次高分辨率模型迭代；从第三条轨迹中取 400 个快照作为测试集。第四条轨迹（验证轨迹）则用于初始化我们的 CNN-LES 运行，并与相应的 FDNS 演化进行比较。由于我们希望表征参数化解在长时间范围内的保真度，因此将第四条轨迹运行 $T = 6 \times 10^6 \Delta t$，在线 CNN-LES 解的验证运行时长也与此相同。

这里一个细小但重要的细节是，在将卷积神经网络应用于输入和输出数据之前，我们不会对其进行归一化，也不会使用任何归一化层，例如批归一化（Ioffe 和 Szegedy，2015）。

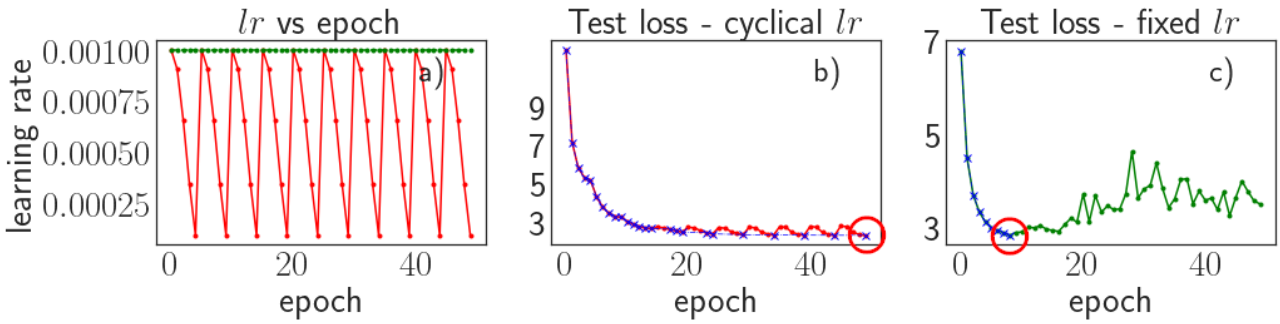

图 4。左：余弦退火情况下（红色）的学习率与固定情况下（绿色）的学习率随训练轮数的变化。中：循环退火学习率情况下的总测试损失（红色）及测试损失的下界曲线（蓝色）。请注意，蓝色曲线是单调递减函数。右：与中图相同，但针对固定 lr 情况。红色圆圈表示整个训练过程中测试误差最低的模型状态；在该点保存卷积神经网络模型，并将其用于驱动在线 CNN-LES 运行。

3.3. **学习率退火与模型选择**。 学习率是神经网络优化中的关键超参数。虽然最简单的训练方法是在整个训练过程中保持学习率不变，但特定形式的 lr-退火可以显著提高测试准确率。我们选择带热重启的余弦退火（Loshchilov和 Hutter，2016），因为它往往能在广泛的超参数范围内提供强大的隐式正则化和稳健解。

图4(a) 显示了循环退火的 lr 随训练周期的变化。在此优化框架中， lr 以最大值（lr=0图 4 中的 .0014）开始，并在5 个训练周期内通过余弦函数递减至零，随后突然恢复到其原始值。图 4b 显示了典型的测试损失随训练周期的变化（退火情况，红色曲线），图 4c 显示了固定 lr 情况（绿色曲线）。对于退火情况，在初始下降之后，测试损失也会升高，但每个周期后都会达到一个新的局部最小值，其损失函数值低于此前达到的值。

对于固定lr 情况而言，然而，测试损失会下降到一个最小值，此时卷积神经网络模型被选为最优模型，但随着训练周期数增加，测试损失随后变得不稳定。我们还展示了每种情况下测试损失的下包络线（蓝色）；请注意，在循环退火情况下，这条曲线在更多训练周期内呈单调下降趋势。在每种情况下，我们选择整个训练过程中测试损失最小的卷积神经网络模型，这对应于蓝色曲线最右侧的点，如图4b-c所示。由于我们采用了特定的模型选择标准，因此无需监控训练过程；其结果是，我们的测试集需要足够多样，以反映动力学的变化性，同时也需要足够大。

3.4. **超参数网格搜索**。 模型优化过程中的另外两个核心超参数是批次大小和权重衰减。通常会将批次大小设为大于某个阈值，但我们在所有模型训练中都将批次大小设为 *1* ，这通常会导致训练极其缓慢且噪声很大。这里的主要原因是，如上文第 3.1 节所述，此处选用的 2 层卷积神经网络的有效非局域性为 9×9 ，也就是卷积神经网络能够独立于域中其他部分进行运算的粗网格区域。由于我们的粗网格域大小为 64×64，这意味着我们的有效 批次大小实际上为 $64^2/9^2 \approx 50$。对于足够深的卷积神经网络，有效批次大小将与实际批次大小相同。

为了找到适用于我们闭合方案的最佳卷积神经网络参数化，我们在（lr，wd）搜索空间中进行了大规模超参数搜索。对于每个隐藏层规模值 N_f，我们在 $\{8,16,32\}$ 中改变学习率 lr（范围为 $\{10^{-4},10^{-3},10^{-2}\}$）以及权重衰减系数 wd（范围为 $\{10^{-5},10^{-4},10^{-3}\}$）。对于每组参数选择，我们分别训练模型，采用固定的 lr 和周期性退火的 lr；对于退火情形，所选的 lr 值表示其最大值

的 lr。对于上述每种选择，我们考虑两类输入：$(\bar{u},\ \bar{v})$ 和 $(\bar{\omega},\ \bar{\sigma}_n,\ \bar{\sigma}_s)$，如第 2.5 节所述。因此，我们总共训练并测试 $3^3 \times 2 \times 2 = 108$ 个卷积神经网络模型（2 种优化方式——退火或不退火；2 种输入类型），分别用于 $\beta = 0$ 和 $\beta = 20$ 情形。请注意，由于我们的卷积神经网络参数数量很少，训练时间和 CNN-LES 运行时间都非常短。

在 V100 GPU 上将每个 CNN 模型训练 50 个 epoch 只需约 1.5 分钟，而运行 CNN-LES 模型 $T = 6 \times 10^6 \Delta t$ 则需约 20 分钟。由于此处采用了 16 倍降尺度，我们降尺度模拟的时间步长增大了 16 倍，因此这相当于约 375,000 Δt_{LES}。值得注意的是，在 $\beta = 0$ 时，108 个案例中有 84 个在整个在线运行过程中数值稳定；在 $\beta = 20$ 时，108 个案例中有 87 个数值稳定。我们还观察到另外两个解在 t=5$\times 10^6 \Delta t$ 后变得数值不稳定，这说明评估在线稳定性的挑战。

3.5. **诊断**。 我们使用多种指标评估在线 CNN-LES 运行的准确性和保真度，这些指标用于量化由 CNN 参数化的降尺度方程的结构和动力学，并将其与真实值 FDNS $\hat{}$ 解进行比较。动能能谱 $E(k)$ 是量化跨越不同空间尺度的湍流和动力学状态的基本指标，其中$\hat{}$ 符号表示该量处于谱空间，且$k^2 = k_x^2 + k_y^2$ 为径向波数。

我们还可以写出与亚格子尺度通量相关、在动力学上更具意义的跨尺度动能和涡量平方通量，其在谱空间中定义为

$$E_{flux}(k) = \Re\left(-\Pi(k)^* \hat{\bar{\psi}}\right), \tag{3.2}$$

$$Z_{flux}(k) = \Re\left(-\Pi(k)^* \hat{\bar{\omega}}\right), \tag{3.3}$$

其中，$\Re$ 表示实部，$\Pi(k)$表示谱空间中的亚格子尺度涡度通量散度，z^* 表示 z的复共轭。[3] 请注意，上述通量并不代表 CNN-LES 解中动能和涡量平方的总通量，而仅代表与 Π相关的贡献。$\Pi(k)$本身也是一项指标，可通过比较 FDNS 与相应的 CNN-LES 运行来检验保真度。

我们还构建了粗粒化涡度场的概率分布函数，$P(\bar{\omega})$以及亚格子尺度通量的概率分布函数，$P(\Pi)$。对于湍流，这些量并不服从高斯分布，而且可能具有很重的尾部。由于二维湍流和地转湍流中长寿命涡旋持续存在，间歇性问题会因此加剧。间歇性意味着，若没有极长时间的模拟，就难以准确捕捉分布的尾部，所得结果也会带有噪声。通常，计算湍流这些量的标准方法是核密度估计（KDE）（Guan 等人，2022b；Ross 等人，2023年），这种方法取决于底层核函数的选择以及用于获得最佳拟合分布的方法（Botev 等人，2010）。然而，我们将在线验证模拟运行在长度为 $T = 6 \times 10^6 \Delta t$ 的长时间范围内，只需估计直方图，无需采用 KDE，即可获得定义良好的概率密度函数（PDF）估计。我们使用 3000 个间隔为 $2000\Delta t$ 的快照，计算在线 CNN-LES 解及相应 FDNS 解的谱、通量和概率密度函数。

对于上述各项谱、通量和概率密度函数量，我们定义了若干指标，用于评估 FDNS 与 CNN-LES 解之间的相似性。这些指标的选择与 Ross 等人 (2023年) 所采用的指标相关。对于能量通量，我们直接使用 FDNS（上标 ^{D}）与在线 CNN-LES（上标$^\theta$）量之间的决定系数来衡量二者的差异，定义如下：

$$\texttt{energy-flux-diff} = 1 - R^2\left(E^D_{flux}(k), E^\theta_{flux}(k)\right), \tag{3.4}$$

$$\texttt{enstrophy-flux-diff} = 1 - R^2\left(Z^D_{flux}(k), E^\theta_{flux}(k)\right), \tag{3.5}$$

在线解的能量谱和亚格子尺度散度谱也以类似方式评估，不过我们将量本身替换为其对数，以确保充分体现较高波数处的误差。因此，

$$\texttt{spectral-diff} = 1 - R^2\left(\log[E^D(k)], \log[E^\theta(k)]\right), \tag{3.6}$$

$$\texttt{spectral-sgs-diff} = 1 - R^2\left(\log[\Pi^D(k)], \log[\Pi^\theta(k)]\right), \tag{3.7}$$

用于比较概率分布的指标通常称为散度，且有多种选择。我们选择一种 L_2-散度，它简单来说就是 FDNS 与在线 CNN-LES 的概率密度函数之间平方差的积分

[3]请注意，为简化记号，我们省略了能量、涡量平方和 Π 符号上方的$\hat{}$符号。

分布，

$$\texttt{distrib-diff} = \int_{-\infty}^{\infty} [P(\omega^D) - P(\omega^\theta)]^2 \mathrm{d}\,\omega, \tag{3.8}$$

$$\texttt{distrib-sgs-diff} = \int_{-\infty}^{\infty} [P(\Pi^D) - P(\Pi^\theta)]^2 \mathrm{d}\,\Pi. \tag{3.9}$$

4. 结果

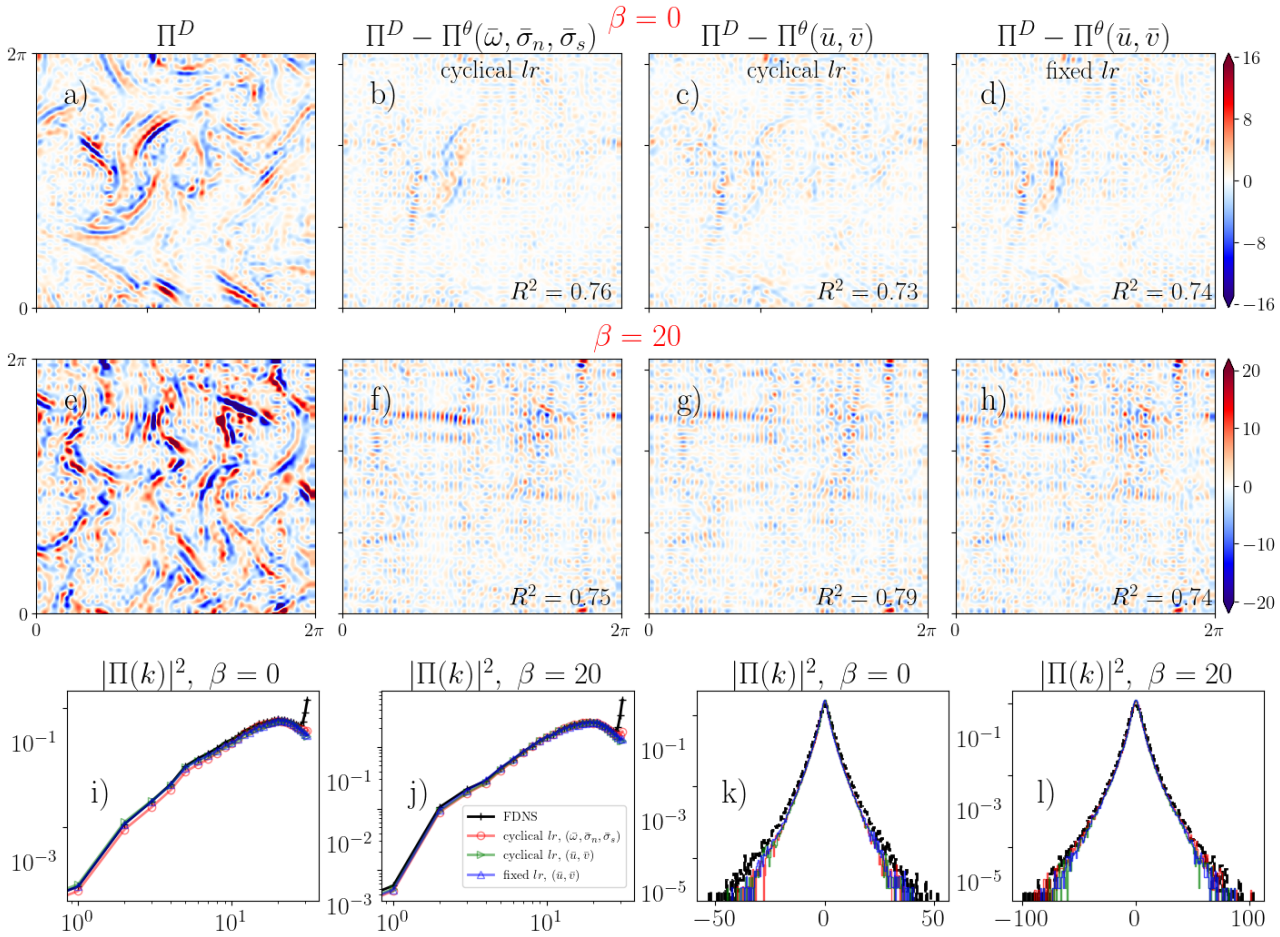

图 5。测试数据亚格子尺度通量散度 Π^D 与经过训练的卷积神经网络模型预测的结果 Π^θ 的比较，这些模型采用了不同的输入和训练方法。**顶行**：单个测试集快照的比较，Π^D 以及卷积神经网络预测误差 $\Pi^D - \Pi^\theta$，具体取决于输入是 $(\bar{u}, \bar{v})$ 还是$(\bar{\omega}, \bar{\sigma}_s, \bar{\sigma}_n)$，或者训练时是否使用了周期性学习率退火，对于 $\beta = 0$。**中行**：与顶行相同，但针对 $\beta = 20$。每个快照还显示了在整个测试集数据集（由 400 个快照组成）上计算得到的决定系数 $R^2(\Pi^D, \Pi^\theta)$。**底行**：功率谱 $\Pi^D(k)$（黑色虚线）与 $\Pi^\theta(k)$ 的比较，以及三个案例中不同 β值对应的概率分布 $P(\Pi^D)$ 和 $P(\Pi^\theta)$，如两行所示。请注意，由于彩色曲线高度接近并持续重叠，因此很难区分它们，这与b)-d) 和 f)-h) 面板中所示的空间误差图一致。为减少杂乱，已移除 k) 和 l) 中的标记。

4.1. **离线准确率**。 我们评估卷积神经网络模型在给定输入（$\bar{u}, \bar{v}$）或（$\bar{\omega}, \bar{\sigma}_s, \bar{\sigma}_n$）的情况下，对测试集中的亚格子尺度通量散度 Π 进行预测的表现。最终，我们的目标是构建计算效率高的神经网络参数化方案，从而获得准确且稳定的在线解。不过，检查离线测试集上的表现，有助于在评估在线性能之前进行预先分析，并可能揭示二者之间的关系。此外，在线解中卷积神经网络-大涡模拟（CNN-LES）与轨迹会因动力学混沌而发散，无法获得真实值；而真实值的存在则使我们能够考察建模误差的性质。

鉴于我们在 (lr, wd)空间进行了大规模超参数搜索，我们选择展示在第 3.5 节定义的在线指标上表现最佳的超参数设置所对应的离线模型比较，更多细节见第4 节。结果见图5，其中我们比较了测试集中的单个 Π^D 快照和卷积神经网络的预测误差 $(\Pi^D - \Pi^\theta)$，并考察了不同输入及是否采用学习率退火的影响。图5 中相应快照标示的平均测试集 R^2 范围为 0.73 至 0.8；然而，无论是 $\beta = 0$ 还是 $\beta = 20$，不同模型及学习率调度方式下的预测误差都呈现出令人意外的相似小尺度结构。相较于 FDNS 测试数据（图5 的底行），从功率谱和概率分布函数中都能观察到不同模型之间的这种高度相似性。从功率谱比较中可以看出，模型预测与数据之间的差异主要局限于频谱末端最大的三个波数。我们发现，即使使用深达 10 层的神经网络，也无法对这些高波数信号进行建模；因此，选择小型参数化卷积神经网络并非问题所在。这些高波数分量的存在，是谱截断算子发生泄漏所致，而高斯滤波器无法消除这种泄漏。

4.2. **在线预测**。

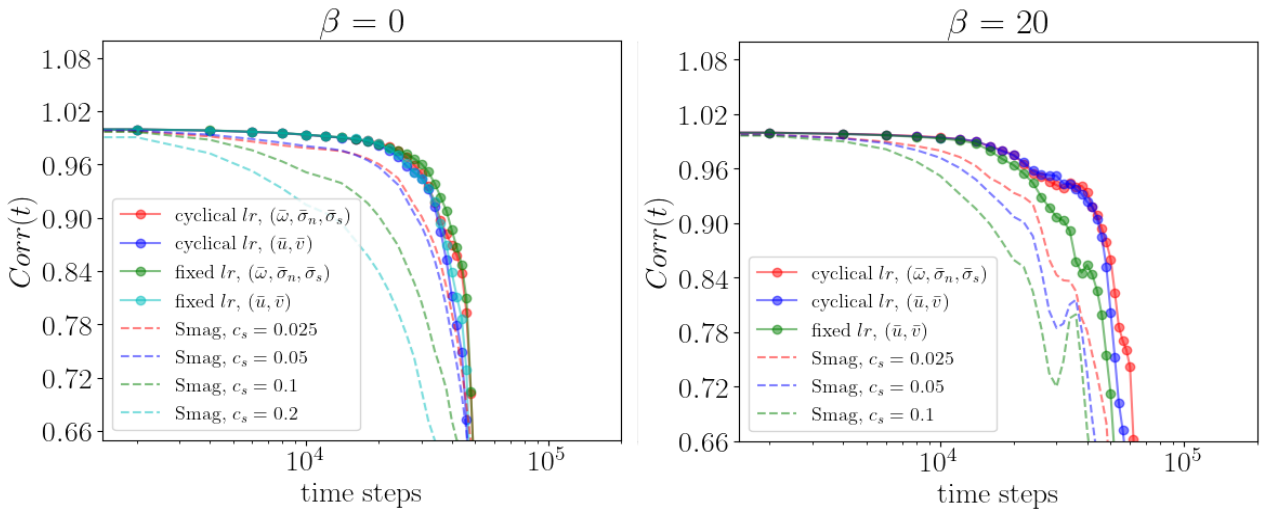

图 6. FDNS 与多种 CNN-LES 模型配置（均以同一 FDNS 快照初始化）之间的相关系数随 DNS 模型时间步数变化的时间演化（带标记的实线）；要得到对应的 CNN-LES 时间步数，请除以 16。图中还显示了采用 Smagorinsky 参数化的模拟与 FDNS 之间的相关性，对应不同的 Smagorinsky 常数，如图所示。此处展示的 CNN-LES 案例，是根据大规模超参数搜索所选取的输入或学习率配置中，长期在线误差最低的案例；见下文第4.2.2 和 4.2.3 节。

4.2.1. 预测准确性。 我们首先考察CNN-LES 模型的短期预测准确性。为此，我们运行涡度闭合方程 Eq. (2.14)，使用 FDNS验证数据集中的同一快照进行初始化，并通过从数据中学习得到的参数化卷积神经网络模型，或 Smagorinsky/Leith 参数化，在时间上向前求解（我们将所有这些模拟统称为大涡模拟）。然后，我们依次比较预先计算的 FDNS 解的短期演化与相应的在线CNN-LES 解以及 Smagorinsky/Leith 模拟。

采用这一协议，大涡模拟运行与真实值 FDNS在时间上保持相关的时间越长，我们就认为大涡模拟的预测准确性越高。我们使用标准的皮尔逊相关系数，在每个时刻计算 FDNS与大涡模拟 $\bar{\omega}$ 快照之间的相关性，并将其可视化为时间的函数，时间以 Δt 的倍数表示，同时考虑到$\Delta t_{LES} = 16\Delta t$。

随后，我们将预测准确性定义为一个单一指标，即去相关时间，也就是FDNS 与大涡模拟之间的相关性降至 0.96 以下之后所经历的时间（Dresdner 等人，2022年）；因此，去相关时间越长，意味着预测准确性越高。鉴于我们研究的是混沌湍流，还可以定义FDNS 本身的内在去相关时间：通过向初始直接数值模拟快照中加入噪声，然后测量流动与未受扰动的 DNS 之间去相关的速度来评估。 Ross 等人（2023年）发现，尽管长期在线性能较差，但在其所有大涡模拟模型中，Smagorinsky 模型的预测准确性接近最佳。因此，我们认为，只需将 CNN-LES 运行的相对预测性能与作为强基准的 Smagorinsky/Leith 大涡模拟运行进行比较，而无需考虑 DNS 本身的内在去相关性。

图 6 展示了由此得到的预测结果。首先，我们注意到，与离线部分（第 4.1节）所示案例相对应的四种 CNN-LES解（实线及实心标记）具有极为相似的相关性曲线。此外，CNN-LES解的表现显著优于 Smagorinsky 案例（虚线），其与 FDNS 模拟结果的去相关时间几乎长一倍（请注意时间轴为对数刻度）。降低 Smagorinsky 常数c_s 可以提升预测性能。然而，正如后续章节将讨论的，将 c_s 降至 0.025 以下（图 6中的紫色曲线）会导致小波数能量堆积，最终使长期统计相较于 FDNS 出现恶化。

对于含有 $\beta = 20$的情况，也有类似的观察结果，但与 $\beta = 0$ 情况不同的是，通过循环 lr 训练的 CNN-LES 模型，其预测准确性明显优于采用固定 lr训练的 CNN-LES 模型。在每种情况下，CNN-LES 运行的预测准确性都优于 Smagorinsky 模拟和 Leith 模拟；而 Leith 的结果与 Smagorinsky 的结果非常相似（未显示）。

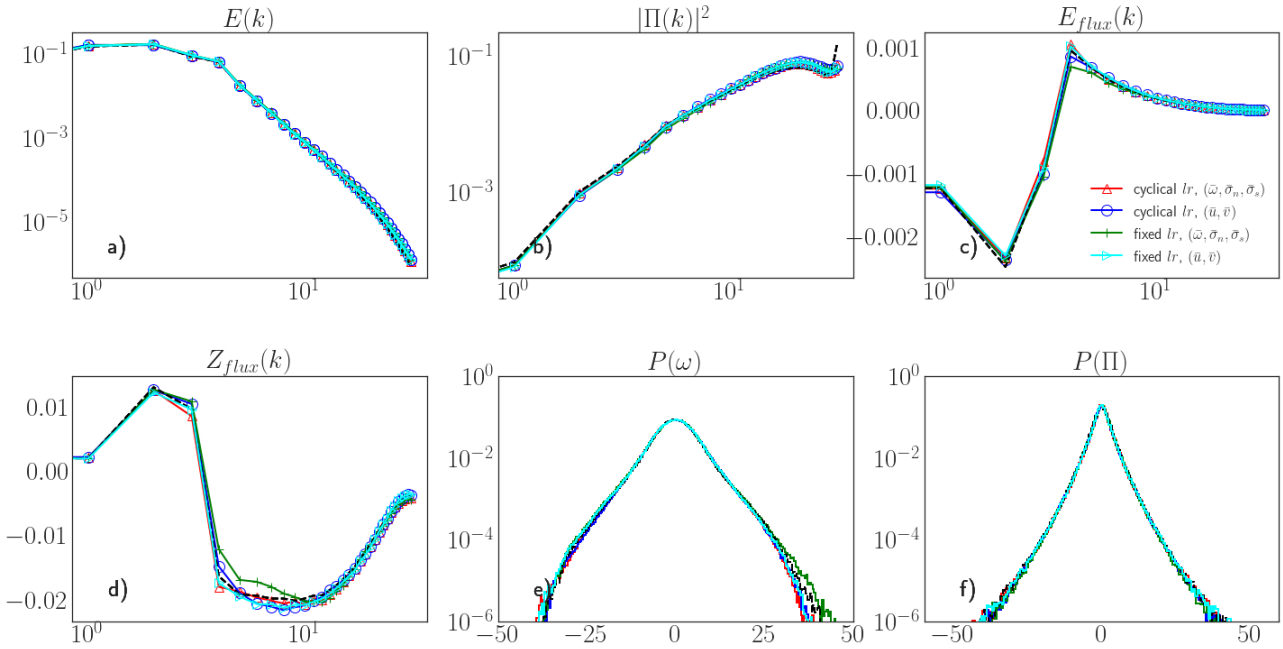

图 7. 图中图 6 和图 5 所示的四种不同情形下，滤波直接数值模拟数据（黑色虚线）与在线模拟的比较，图例中标明了各情形，适用于 β=0 所示的诊断量。请注意，在 a)、b)、f) 中以及一定程度上的 e) 中，不同曲线几乎难以分辨。明显的差异体现在两个跨尺度能量通量 c) 和 d) 上；在这两项中，$(\bar{u}, \bar{v})$对应的两种情形最准确。为减少图面杂乱，e) 和 f) 中已移除标记。

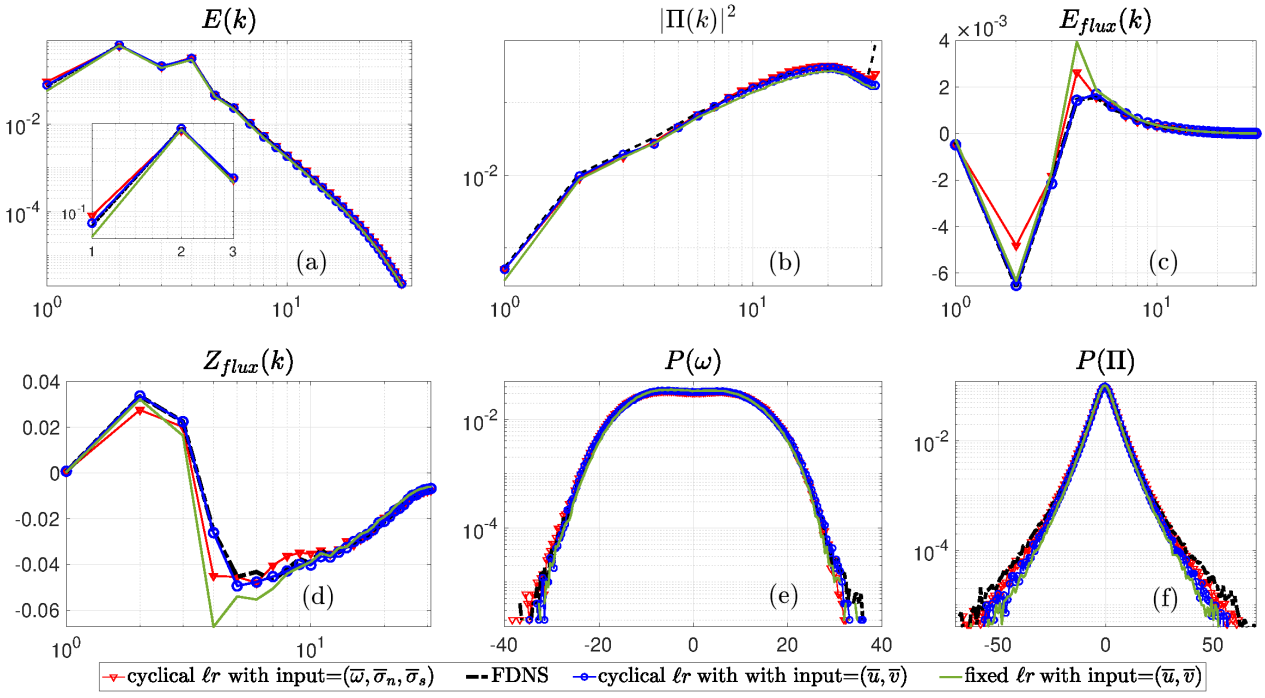

图 8。 与图 7 相同，但针对 β=20。对应于固定 lr 和 $(\bar{\omega}, \bar{\sigma}_s, \bar{\sigma}_n)$ 输入的情况因精度较差而未显示。a) 中的插图展示了每个 CNN-LES 模型的低波数能谱放大图。请注意，与图7不同，通过循环 lr 学习得到的 CNN-LES 模型，其 $\bar{u}, \bar{v})$ 输入（带圆圈的蓝色曲线）的精度显著优于其他 CNN-LES 模型。

4.2.2. 长期准确性。 短期预报与天气预报等问题直接相关，而长期预报则与气候研究更为相关。没有先验理由可以认为，前者的高准确性意味着后者也有同样的表现，尤其是对于训练过程可能高度针对特定任务的数据驱动模型。事实上，Dresdner 等人(2022) 发现，通过（Stachenfeld等人，2021）所采用的编码器-处理器-解码器框架，一个纯数据驱动的自回归神经网络模型在预测准确性方面与他们的湍流参数化解相当，但该模型最终会在长时间尺度上变得不稳定。我们的目标是验证 CNN-LES 设置在长时间尺度上的数值稳定性和保真度，而预测准确性只是附带结果。鉴于我们对 CNN 尺寸采取了较为激进的选择，我们运行模型的时间远长于近期其他关于神经湍流闭合的研究。Guan 等人(2022b) 计算在线 CNN-LES 运行至[4] $2\times10^6\Delta t$ 而我们则计算至 $T = 6\times10^6\ \Delta t$ ，对应于 375，000 Δt_{LES}。如第3.5节所述，这也确保了概率分布函数等指标收敛，因此现在可以直接计算这些指标，无需借助核密度估计。

本节重点介绍第4.1节和第4.2.1节所述四种特定卷积神经网络模型的长期精度。这些模型针对第3.5节选定的六项指标进行评估，以衡量解的保真度：一维时间平均能谱， $E(k)$；亚网格尺度散度谱， $|\Pi(k)|^2$；跨尺度能量通量和涡量平方通量，$E_{flux}(k)$ 和 $Z_{flux}(k)$；以及在整个在线模拟过程中计算的一维概率分布函数， $P(\bar{\omega})$ 和 $P(\Pi)$。图 7 展示了 FDNS 运行与对应的四次 CNN-LES 运行之间的比较。我们发现，在所示的六项指标上，四种情况均吻合良好；总体精度最高的是采用循环学习率但输入不同的两种情况。在线模拟中，亚网格尺度谱 $|\Pi(k)|^2$ 和概率分布函数 $P(\Pi)$ 的精度与离线测试（图5）同样高，这也与所得结果的高精度相吻合

[4]请注意，我们的计算设置与他们的完全相同，只是我们的 Re 值略大一些。

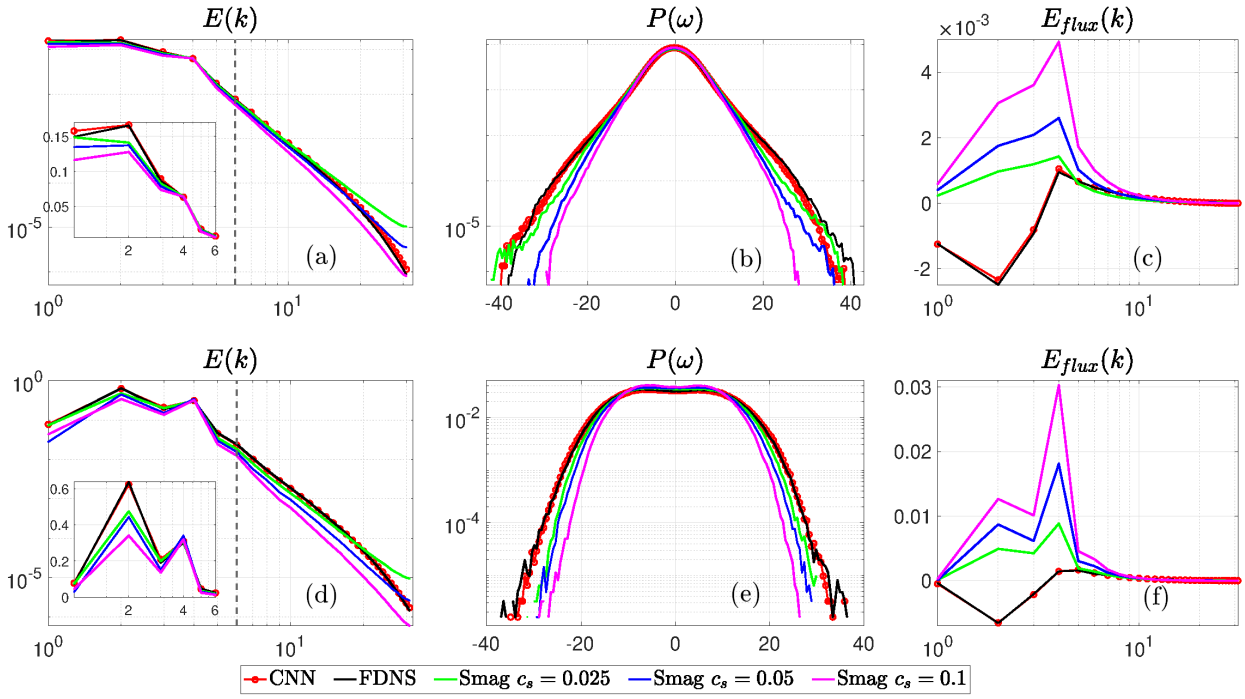

图 9. 图 7 和图 8 中表现最佳的卷积神经网络模型（两种周期性 lr 情形）的长期在线性能对比，以及相应的 Smagorinsky 模拟结果；后者的运行时长与 CNN-LES 情形完全相同。为简洁起见，这里仅展示前文图中几个关键指标。插图突出显示了 CNN-LES 与 Smagorinsky 谱在大波数范围内的对比；请注意，为了进一步凸显差异，插图采用线性 y坐标轴，而非对数坐标轴。(a) 和 (d) 中的竖直虚线表示插图的最右端范围。请注意，与在线 CNN-LES 解相比，Smagorinsky 解未能捕捉到逆级串。

例如 $E(k)$ 和 $P(\omega)$ 这样的结构性流动度量，以及跨尺度通量的动力学度量。这一结果意味着，在这些情况下，伽利略不变性显然并非一种特别难以学习的物理对称性。

然而，这些结果并不适用于 $\beta = 20$ 情形（图 8）：对应于周期性退火lr 且采用（$\bar{u}$，$\bar{v}$）作为输入的模拟（图 8 中的蓝色曲线），除 P（Π）外，在所有指标上的表现都显著更好。周期性退火对模型稳健性和保真度的作用大体上符合预期，但令人意外的是，在建模时先验地选择伽利略不变性（将（$\bar{\omega}$，$\bar{\sigma}_n$，$\bar{\sigma}_s$）作为输入）实际上降低了在线准确率。具体原因尚不清楚，但显然与 $\beta = 20$情形中存在强烈的涡旋驱动纬向急流有关，而$\beta = 0$时则没有这种急流。这一观察似乎表明，对于地球物理湍流情形，最好选择（$\bar{u}$，$\bar{v}$）作为卷积神经网络的输入。同样值得注意的是，尽管 Π（k）预测准确，涡度分布P（$\bar{\omega}$）也预测准确，但分布的尾部 P（Π）却未能捕捉到。

有趣的是，我们观察到，我们的卷积神经网络解在短时间尺度和长时间尺度上（即在“天气”和“气候”状态下）都具有较高的在线准确率。此前的研究并未报告此类解。虽然 Dresdner 等人 (2022) 没有考察长时间保真度，Frezat 等人 (2022) 发现，他们最优的长期准确解（使用其完全在线的方法训练）实际上预测准确性较差，并且在长时间尺度上会输给一个离线训练的卷积神经网络模型，后者在长时间尺度上数值不稳定。与这一结果一致，Ross 等人(2023年) 发现，他们的预测准确性通常与长期保真度不相关。为进一步强调这一结果，我们在图 $\beta = 0$ 和 $\beta = 20$ 所示案例中，选取其中最优的 CNN-LES 解，并在图7 和8）中，于图9中评估其在线性能，同时参考此前图6中展示的各种Smagorinsky 案例。请注意，在捕捉分布方面准确性最佳的 Smagorinsky 解，即 $P(\bar{\omega})$（对应于 $c_s = 0.025$）在小尺度上出现了能量堆积（图9ab）。

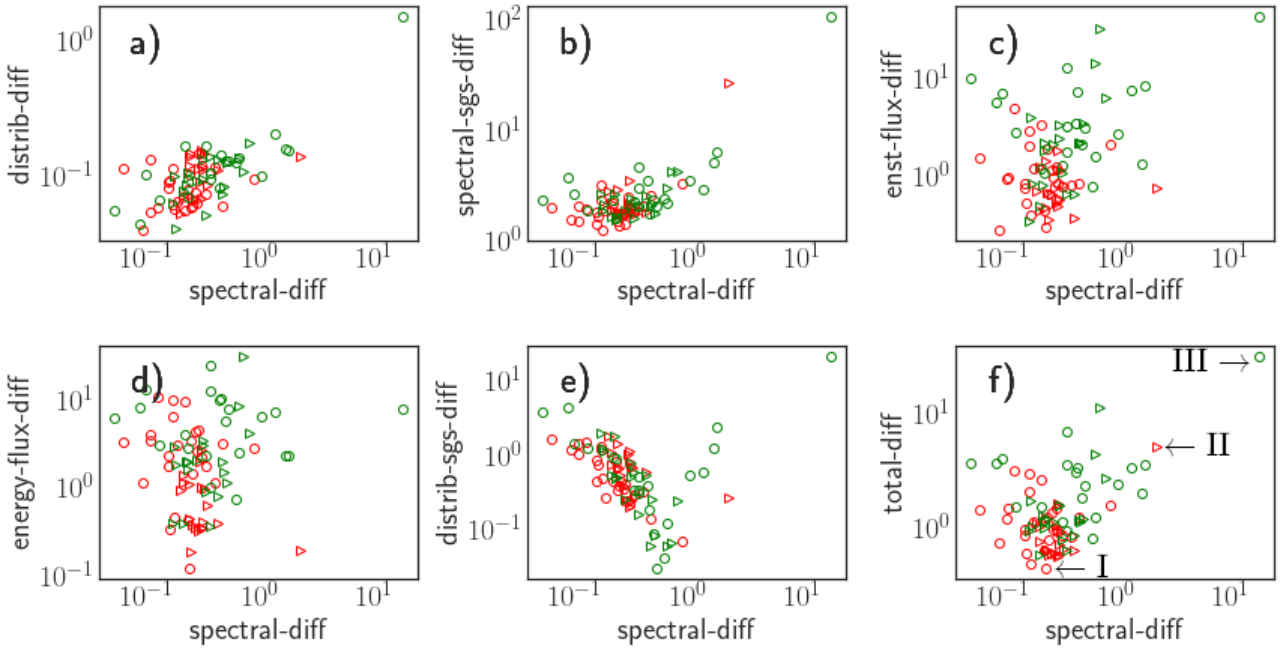

图 10. 第 3.5 节定义的五种差异度量与第六种在线度量 谱差异 的比较。该度量用于衡量基于卷积神经网络训练、且超参数空间中取值范围为 lr、wd) 的在线模拟所得能量谱之间的差异。这里 β=0。红色标记表示采用循环 lr 训练的在线 CNN-LES 模型，绿色标记表示采用固定 lr训练的模型。圆形标记使用 ($\bar{\omega}$、$\bar{\sigma}_n$、$\bar{\sigma}_s$) 作为输入，三角形标记则使用 ($\bar{u}$、$\bar{v}$)。 总差异 度量定义于式（4.1）。标记为 (f) 的情况 I、II 和 III 在图 11中展示。

此外，Smagorinsky 案例无法捕捉涡度分布的尾部，而 CNN-LES解始终能够做到这一点；对于射流情形而言，这一点尤其明显（图9f）。

我们在射流情形中的在线准确率与 Frezat 等人 (2022年) 使用其在线学习框架所获得的结果相当。 Ross 等人 (2023年) 采用与我们类似的离线学习方法对其相应的射流情形进行参数化，但类似的尝试并未成功；他们的卷积神经网络在成功求解其 $\beta = 0$ 情形后，却完全无法对射流进行建模。 Ross 等人 (2023年) 使用这一结果指出，数据驱动的卷积神经网络方法缺乏稳健性。我们的结果与他们的结果相反，因此我们并不认同相同的结论。不过，目前尚不清楚为什么（Ross 等人，2023年）关于 β-情形的结果与我们的结果不同。尽管他们的准地转湍流模型可能是更具挑战性的测试情形，但他们的雷诺数实际上比我们的低得多。我们的在线 CNN-LES解与（Ross 等人，2023年）的结果之间存在差异的原因，也可能在于一些技术因素，例如超参数搜索不理想、所使用的卷积神经网络的深度以及训练过程的选择。我们将在讨论部分进一步讨论这一问题。

最后，本文此处及前两节所展示的最佳模型，都是从在各种超参数设置下训练的大量模型中精选出来的。下一节将根据训练方法或输入选择，考察我们的 CNN-LES 模型在这一超参数空间中的表现。

4.2.3. 模型保真度对超参数的依赖性。 在前面的章节中，我们已经证明，通过在（lr，wd）超参数空间中进行搜索，我们可以利用浅层 CNN 获得高保真度的解。然而，了解这类高保真度解是容易找到还是难以找到非常重要；反过来，鉴于运行次数如此之多，也需要能够轻松找到最准确的在线解。尽管这种评估看似平凡且有些繁琐，但当面对气候系统中观测到的更复杂流动构型的参数化问题时，这一点尤其值得关注，因为届时如此大规模的超参数搜索可能并不可行。

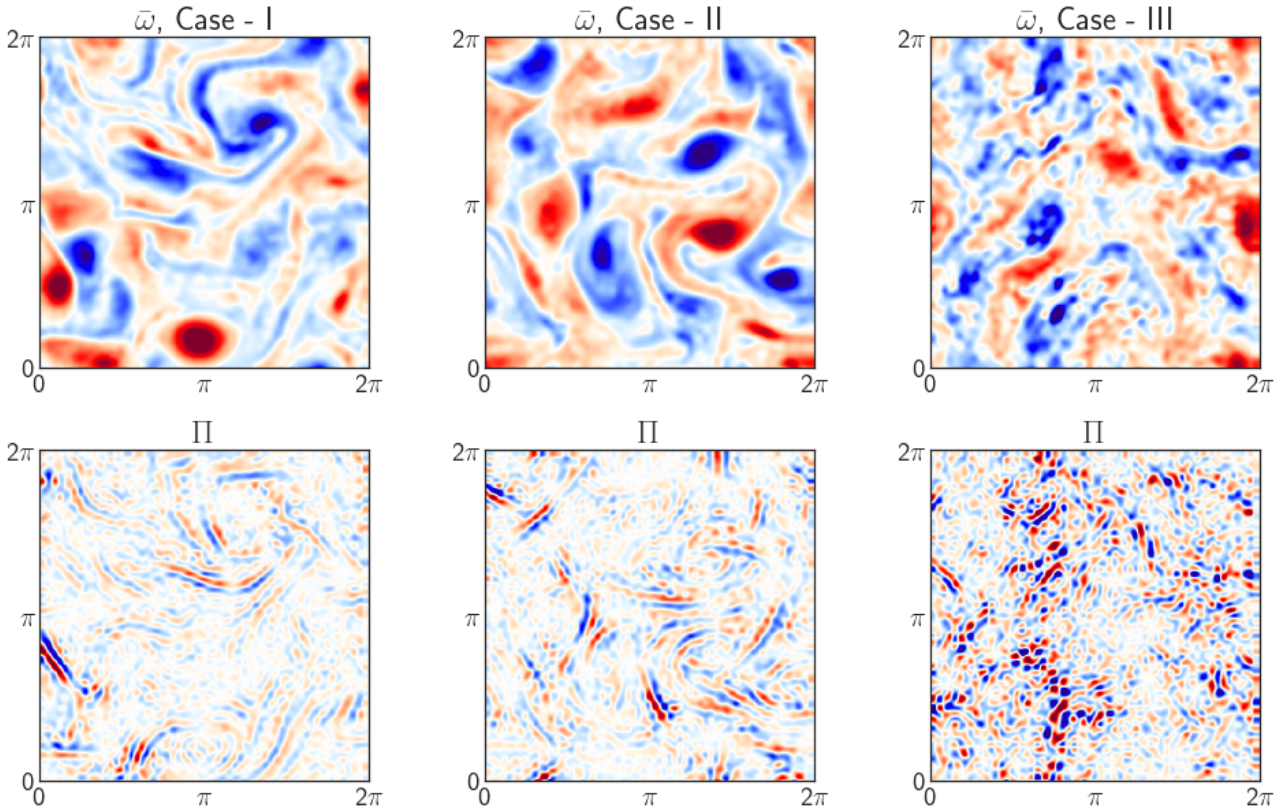

图11. 快照 ω 及其对应的 Π，对应于图 10f 中标记为 I、II 和 III 的三个在线案例。

在第 3.5节中，我们基于图 7 和图 8 所示的六个量定义了六种不同的差异度量，用于比较在线 CNN-LES解相对于FDNS解的误差。在此，我们构造了一个简单地将这六种度量相加的度量，其定义如下：

$$\begin{aligned}\texttt{total-diff} =&\texttt{(spectral-diff + spectral-sgs-diff + enst-flux-diff}\\ &\texttt{+ energy-flux-dim + distrib-sgs-diff)/5}\,.\end{aligned} \tag{4.1}$$

为了解这些度量在超参数空间中的变化情况，我们将除谱差异（用于衡量能谱估计准确度）之外的 5 项度量相对于谱差异进行展示。通过观察其他度量相对于谱差异的变化，我们可以评估这些度量之间隐含的相互关系。图10和图12分别展示了 $\beta = 0$ 和 $\beta = 20$ 的结果。我们逐一考察这两种情形，因为结果显示，这些度量在（lr, wd）空间中的变化各不相同。我们先考察总差异度量（图 10f），它揭示了两个重要特点。首先，我们观察到循环退火情形（红色标记）聚集得更紧密，也更靠近原点（即误差更低）；而固定 lr 情形（绿色标记）的性能分布范围则大得多。这表明，循环 lr 过程不仅能在改变超参数时产生范围更窄的行为，从而总体上带来更稳健的解，而且总体准确度也更高。对于 $\beta = 20$，这些特征更加明显（图 12f），退火 lr 情形在准确度和稳健性上的不足也更为突出。其次，在循环 lr 情形（红色标记）中，使用（$\bar{\omega}, \bar{\sigma}_n, \bar{\sigma}_s$）作为输入的解（红色圆圈）在 $\beta = 0$ 时表现最佳，不过优势不大。然而，当 $\beta = 20$ 时，几乎所有表现最佳的解都来自使用（$\bar{u}, \bar{v}$）作为输入的情形（红色三角形）。与固定lr 情形相比，循环 lr 情形的通量误差也显著更低（图 12cd），不过其他度量则较难区分。值得注意的是，对于 $\beta = 0$ 和 $\beta = 20$，P（Π）的保真度（图 10e 和图 12e）似乎与流动度量呈相反关系，固定 lr 情形的保真度更高

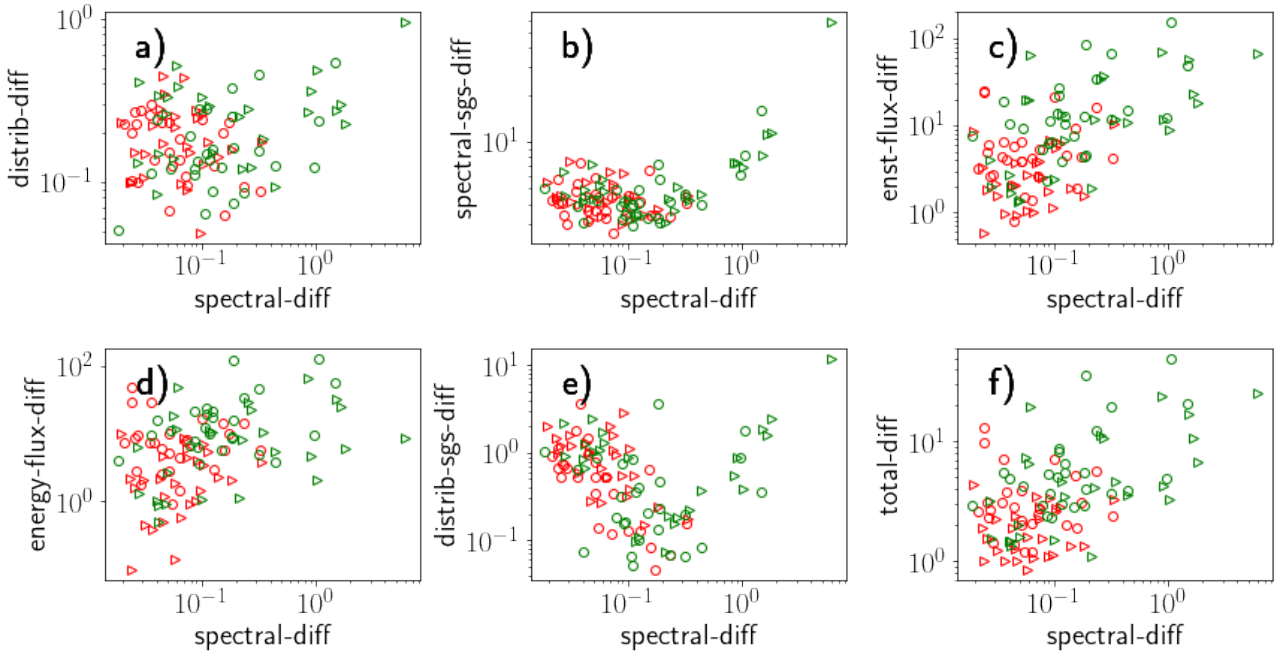

图 12。 与图 10相同，但适用于 β=20。这里再次可以观察到，采用周期性 lr 训练的 CNN-LES 模型总体上比采用固定值训练的在线性能更好（红色与绿色标记）。

该项度量的准确度。我们认为这并非重大问题，因为即便在离线情形下， P（II）的尾部误差也不可忽视（图5k 和5l）。不过，循环 lr 情形在谱-亚格子尺度-差异度量上表现更佳（图10e 和图12e）。

我们详尽的参数搜索使我们能够突出一个在线误差极大的特定案例，其误差比最佳案例大一个数量级，同时在长时间内保持数值稳定（在图 10f 中标记为案例 III）。为了更好地理解这一异常解，我们将三个案例对应的 $\bar{\omega}$ 和 II 的快照可视化。这三个案例分别对应于在线模拟表现最佳、中等以及最差的 CNN-LES 运行。一个令人惊讶的事实是，I（最佳案例）和II（“中等”案例）具有非常相似的流动结构，而 III看起来杂乱且不符合物理规律。然而奇怪的是，这三个案例具有相似的离线准确率，这再次凸显了从离线准确率预测在线准确率所面临的挑战。

上述发现还引发了关于我们在超参数搜索过程中发现的大量在线稳定卷积神经网络参数化之间关系的问题。在图5 中观察到的亚格子尺度误差，在不同输入选择和优化方法下表现出高度的结构相似性，这表明我们可能正在学习一个共同的参数化族。尽管人们可能希望找到一个在某些精度指标下、针对复杂湍流全局最优的单一独特参数化，但要从有限数据集中推断出这一点可能颇具挑战，甚至根本无法实现。人们可能会想知道，在神经网络架构固定但数据量趋于无穷的极限情况下，这些结果会如何变化。这类问题与遍历性质的研究密切相关。如果遍历性成立，那么在无限数据极限下，能够最优地平均掉小尺度的参数化的唯一性（以及存在性）便得到了保证（Chekroun etal., 2020，定理 4）；然而，正如图10 和图12所报告的那样，使用有限采样数据可能会产生多个具有不同评分的可接受解。在潜在受迫流动可能具有遍历性的视角下，对这一现象进行详细研究（Foias et al., 2001；Chekroun and Glatt-Holtz, 2012；Hairer and Mattingly, 2006, 2011）超出了本文的研究范围。

4.2.4. 离线预测与在线预测。 任何数据驱动的参数化方法的最终目标，都是实现高质量的在线保真度和准确率。要达到这一目标，通常需要让模型运行较长时间，并根据前文所述的各种指标对在线解进行评估。当在超参数空间中进行大范围测试时，即使使用本研究中的浅层神经网络，这种方法也会变得尤其

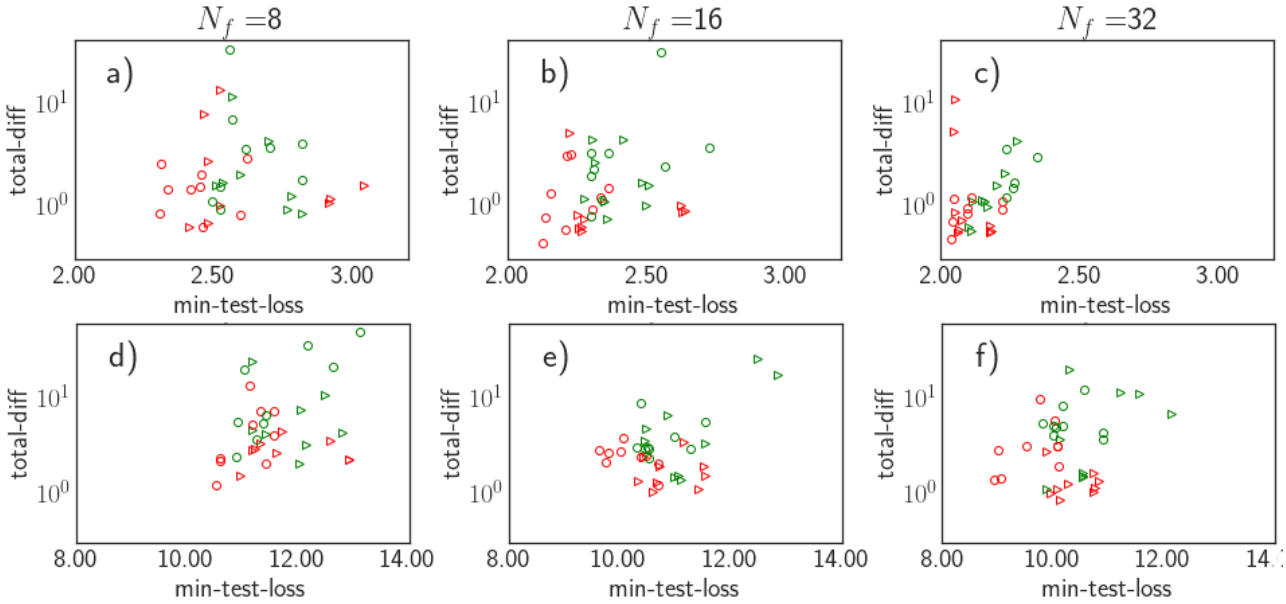

图 13. 在线 `total-diff` 度量与 离线测试误差 在 (lr, wd)- 超参数空间中的比较，针对 a)-c) $\beta = 0$ 和 d)-f) $\beta = 20$。结果按所用卷积滤波器的不同取值分组，以 N_f 表示；CNN 架构的详细信息见第 3.1 节。标记的颜色编码方式与图 12 相同。

昂贵，因为需要在超参数空间中进行大范围测试。理想情况下，人们希望只需计算留出测试集上的离线准确率，并据此估计在线准确率。但这仍然是不切实际的设想。在这方面，图 13 比较了离线准确率与`total-diff`中定义的在线准确率指标（4.1），其范围涵盖 (lr, wd)-超参数空间。结果针对$\beta = 0$ 展示于图 13a-c 中，针对 $\beta = 20$ 展示于图 13d-f 中。与上一节不同，我们将不同隐藏层滤波器尺寸的结果分开呈现。可以观察到，随着 N_f 卷积滤波器数量增加，CNN 参数数量也随之增加，离线误差（ x-轴）会随着 N_f的增加而系统性地减小，这表现为点云沿 x-轴向原点移动。然而，在线误差（ y-轴）的相应下降实际上相当小，因此难以辨别。这一观察结果凸显了一个关键点：离线误差的大幅下降，可能只会带来在线 CNN-LES 性能的小幅提升。这也预示了这样一个结果：更深层的 CNN 虽然必然会提高离线准确率，却不一定能带来更准确的在线解。

与此同时，我们注意到在线性能与离线性能之间的相关性仍然较弱，这意味着，诊断后者不足以了解前者，而前者才是我们的实际目标。原因之一是，我们用于评估离线性能的均方误差是在 400 个快照上计算并通过其中涉及的范数进行空间平均的，但它未必能反映不同 CNN-LES 模型所生成数据的动力学差异。可以设想，采用更基于物理的离线性能度量，或许确实能够提升在线表现，而无需耗费更多资源、长时间运行 CNN-LES；这仍是一个正在积极研究的领域。

5. 讨论与总结

5.1. 亚格子尺度近局部性的物理与计算意义。 本研究的主要成果是，利用浅层双层卷积神经网络，有效地学习到了二维湍流和地转湍流的精确参数化。该成果实质上意味着，这些问题的亚格子尺度应力对粗尺度场的空间依赖几乎是局部的（或弱非局部的）。换言之，我们通过直接构建空间近局部参数化，推断出了这一类问题的物理机制。但这通常并不适用于所有

各类湍流；例如，海洋表层混合层中的对流运动，以及更广泛存在于地球对流层中的对流运动（这些过程还会导致云形成及随后发生的降水），在其发生所涉及的垂直范围内的相当大比例上可能存在很强的相关性。近期成功对大气中的垂直通量进行参数化的机器学习模型，使用了覆盖大部分空气柱的神经网络（Yuval 和 O'Gorman，2020；Yuval 等，2021）。类似地，海洋表层混合层中的传统经验性基于物理的参数化采用简单的垂直非局域通量来表示对流湍流。尽管当前的大气机器学习参数化并未表示水平涡通量，这显然是出于简化的原因，但我们在此的工作表明，更完整的表示很可能在水平方向上具有空间上的（近似）局域性，而在垂直方向上具有更高程度的非局域性；在海洋中，非局域性应仅限于海洋表层混合层的范围，因为海气通量会在该范围内产生主动影响。

这项工作的另一项意义在于海洋、大气和气候模型的数值计算。这些模型通常采用水平区域分解，在大型计算节点集群上进行数值求解；进程间通信通常仅限于子区域边界上的少数几个点。因此，采用数据驱动参数化时，空间非局域性不能过强，否则进程间数据交换的成本会高得令人难以承受；需要注意的是，垂直方向（即垂直于地球运动的方向）不存在这一问题。上述结果表明，水平亚格子尺度通量可以通过浅层卷积神经网络模型进行建模，因此相较于使用更复杂的深层 CNN，更容易将其纳入现有处理流程。

5.2. **理论后果及迈向可解释性**。本研究提出的基于浅层卷积神经网络的神经网络闭合结果，对于惯性区间内的截断尺度以及较高雷诺数均成立。众所周知，这一问题颇具挑战性，因为亚格子尺度上的微小误差通常会因逆级串而放大为大尺度误差（Piomelli 等人，1991；Jansen 和 Held，2014）。因此，如何在控制误差的前提下，舍弃这类湍流在较低截断尺度下的亚格子尺度参数化，是亟待解决的挑战之一。我们的闭合结果在精度和稳定性方面表现良好，强有力地支持了如下非线性函数的存在：Φ_{CNN} 使得亚格子尺度 Π在经过自旋启动后满足如下形式的关系：

$$\Pi = \Pi_{CNN}(\overline{u}, \overline{v}) + \epsilon, \tag{5.1}$$

其中，残差 ϵ 是一个时空函数，其波动在均方意义上受到控制且幅度较小。在公式（5.1）中，Π_{CNN} 表示通过最小化公式（v）中的损失函数训练浅层 CNN 得到的函数。实际上，（5.1）是通过最小化损失函数（Π_{CNN}）及其正则化项构造 3.1的直接结果。

我们已经证明，只要在学习率和权重衰减系数空间中进行适当的超参数搜索，并采用周期性学习率退火，即使使用浅层且局部性较弱的卷积神经网络，也能在线下获得较小的残差，同时得到稳定、准确的在线解。由于 Π_{CNN} 仅是粗粒化变量的非线性泛函，我们出色的闭合效果表明，掌握 Π_{CNN} 已足以取得良好性能，尤其是在 β情形下，与其他近期的神经闭合方法相比更是如此（Ross 等人，2023年）。

因此，我们的神经网络闭合结果与（Maulik 等人，2019；Kochkov 等人，2021；Zanna和 Bolton，2020；Subel 等人，2022）一样表明，对于湍流问题，即使在较低的截断尺度下，也应排除在 Mori-Zwanzig（MZ）解释中使用记忆项（Givon 等人，2004；Chorin 和 Hald，2006）；在其他依赖MZ形式体系的闭合研究中，这些记忆项却被不必要地大肆推崇；例如参见（Miyanawala和 Jaiman，2017；Parish 和 Duraisamy，2017；Ma 等人，2019）。

相反，条件期望，即在以粗变量为条件的情况下对未解析变量进行平均的最佳非线性泛函，对于高 Re的受迫二维湍流问题的闭合而言已足够。Π_{CNN} 借助一个小型 ϵ 便可提供这样的近似，因此很可能与底层最优参数化流形（OPM）的存在性相关，该流形以最小二乘意义连接小尺度和大尺度（Chekroun 等，2017，2020，2021），正如最优参数化流形理论所预测的那样（Chekroun 等，2020，定理 5）。

我们的神经网络湍流闭合结果以及近期相关研究，因而重新肯定了Foias 等人(1988、1991)在 20 世纪 80 年代末提出的观点：二维湍流本质上是有限维的，湍流解在均方意义上位于某个有限维流形的一个狭窄邻域内；另见（Chekroun 等人，2020，式（1.5））。事实证明，若依赖Pascal 和 Basdevant，1992）所示，这些观点的适用性被削弱，仅对处于耗散区内或其附近的截断波数成立；而这类方法依赖于传统解析参数化，例如最初在（Foias 等人，1988）中提出的方法。因此，神经网络的应用促使我们重新审视这些基于以下内容得出的结论：

有限一类解析公式，并且实际上为这个老问题带来了新的启示，因为它将（5.1）这类关系的适用范围拓展到了惯性区间内的截断波数。近局部浅层卷积神经网络参数化能够实现这一目标，其内在的低维性很可能使其具有可解释性和可泛化性。因此，我们希望通过寻求能够利用近局部浅层卷积神经网络这一发现的新型解析闭合公式，调和以往解析方法的失败与近期的实证成功。

5.3. **机器学习与物理设计选择**。在他们最近的工作中，Ross 等人 (2023年)表明，基于机器学习的参数化不应过多关注神经网络架构或优化技术等机器学习细节，而应转而关注物理设计选择，在本例中即输入/输出或粗粒化滤波器的选择。根据我们的结果，我们认为实际上应同时关注这两个方面，而且在某些情况下二者密切相关，例如，我们基于亚格子尺度粗场分量与小尺度相互作用所表现出的近局部特性，选择了 2 层 CNN 架构。研究表明，选择诸如循环退火 lr 之类的优化技术可以使学习任务显著更容易，通常也更加稳健；当所考虑的问题不是像本文及本研究所引用的那些研究中处理的简化湍流模型时，这一点将更加重要。此外，超参数搜索虽然繁琐，却能显著提高任务的准确性和效率，因此不应被忽视。

致谢

本研究得到海军研究办公室（ONR）多学科大学研究计划（MURI）资助，项目编号为 N00014-20-1-2023。本研究还部分获得欧洲研究理事会（ERC）在欧盟“地平线2020”研究与创新计划下的资助（MDC） [资助协议编号： 810370]。

参考文献

Ba, J.L.,J.Kiros, and G. E. Hinton,2016：层归一化。*arXiv* 预印本 *arXiv:1607.06450*。Bachman, S. D., B.Fox-Kemper,and B. Pearson, 2017：准地转湍流的尺度感知子网格模型。地球物理研究杂志：海洋，**122 (2)**，1529–1554。Bakas, N. A., N. C.Constantinou,and P.J.Ioannou, 2015：正压β平面湍流均匀态的 S3t稳定性。大气科学杂志，**72 (5)**，1689–1712。Bengio,Y., and O. Delalleau, 2011：深层架构的表达能力。算法学习理论：第*22*届国际会议，*ALT 2011*，芬兰埃斯波，*2011*年*10*月*5*日至*7*日。论文集*22*，Springer，18–36。Botev, Z. I., J.F. Grotowski,andD. P. Kroese, 2010：通过扩散进行核密度估计。统计年鉴，**38(5)**，2916–2957。Brachet, M., M. Meneguzzi,H.Politano,and P. Sulem, 1988：自由衰减二维湍流的动力学。流体力学杂志，**194**，333–349。Brenowitz, N. D., B. Henn, J. McGibbon, S.K.Clark, A. Kwa, W. A. Perkins, O. Watt-Meyer,and C.S. Bretherton, 2020：机器学习气候模型动力学：离线性能与在线性能。*arXiv* 预印本 *arXiv:2011.03081*。Carnevale,G., J.McWilliams, Y.Pomeau,J.Weiss,and W. Young, 1991：二维湍流中涡旋统计量的演化。物理评论快报，**66 (21)**，2735。Carnevale,G., J.McWilliams, Y.Pomeau, J.Weiss,and W.Young,1992：衰减二维湍流中非线性演化的速率、路径与终态：标度理论与选择性衰减。*A*辑流体物理：流体动力学，**4(6)**，1314–1316。Chekroun, M.,H.Liu, and J.McWilliams, 2021：慢流形闭合中快速振荡的随机整流。美国国家科学院院刊，**118 (48)**，e2113650118，doi:10.1073/pnas.2113650118。Chekroun, M.D., and N.E.Glatt-Holtz, 2012：耗散动力系统的不变测度：抽象结果与应用。数学物理通信，**316**，723–761，doi:10.1007/s00220-012-1515-y。Chekroun, M. D.,H.Liu,andJ.McWilliams, 2020：非线性动力系统闭合的变分方法：自治情形。统计物理学杂志，**179**，1073–1160，doi:10.1007/s10955-019-02458-2。Chekroun, M. D., H.Liu,andJ.C. McWilliams,2017：约化原始方程模型中快速振荡的出现及其对闭合理论的意义。计算机与流体，**151**，3–22。Chorin, A., andO. Hald, 2006：数学与科学中的随机工具。第147号，应用数学科学综述与教程，Springer New York。

Dresdner, G., D. Kochkov,P.Norgaard,L.Zepeda-Núñez,J. A. Smith, M.P. Brenner, 和S.Hoyer，2022：学习校正用于模拟湍流的谱方法。arXiv 预印本 arXiv:2207.00556。Eyink, G. L.，2005：湍流级联的局部性。《物理学D：非线性现象》，207 (1-2)，91–116。Eyink,G.L.，和H. Aluie，2009：流体动力学湍流中能量级联的局部性。I. 光滑粗粒化。《流体物理学》，21(11)，115 107。Farrell,B.F.，和P.J. Ioannou，2007：正压湍流中喷流的结构与间距。《大气科学杂志》，64(10)，3652–3665。Fisher, B.，2008：互相关定理与维纳–辛钦定理。《神经科学杂志》，8107–8115。Foias, C., O.Manley,R.Rosa, 和 R. Temam，2001：纳维–斯托克斯方程与湍流，第83卷。《数学及其应用百科全书》。剑桥大学出版社。Foias, C.,O. Manley,和 R.Temam，1988：二维湍流中小尺度涡旋与大尺度涡旋相互作用的建模。《RAIRO 数学建模与数值分析》，22 (1)，93–118。Foias, C.,O.P.Manley, 和R.Temam，1991：湍流中的近似惯性流形与有效黏性。《流体物理学A：流体动力学》，3 (5)，898–911。Frezat, H., J. L. Sommer, R. Fablet,G.Balarac, 和 R. Lguensat, 2022：准地转湍流参数化的后验学习。arXiv 预印本 arXiv:2204.03911。Givon, D.,R.Kupferman, 和 A.Stuart，2004：提取宏观动力学：模型问题与算法。《非线性》，17 (6)，R55–R127。Guan, Y.,A. Chattopadhyay,A. Subel, 和 P. Hassanzadeh，2022a：使用卷积神经网络进行稳定的二维湍流后验大涡模拟：反向散射分析及通过迁移学习推广至更高雷诺数。《计算物理学杂志》，458，111 090。Guan, Y.,A. Subel, A.Chattopadhyay, 和P. Hassanzadeh，2022b：在小数据体制下学习受物理机制约束的亚网格尺度闭合，以实现稳定且准确的大涡模拟。arXiv预印本arXiv:2201.07347。Hairer,M.，和J.Mattingly，2006：具有退化随机强迫的二维纳维–斯托克斯方程的遍历性。《数学年刊》，993–1032。Hairer,M.，和J. Mattingly，2011：半线性随机偏微分方程的次椭圆性与唯一遍历性理论。《电子概率学报》，658–738。Hornik, K.,M.Stinchcombe,和 H.White，1989：多层前馈网络是通用逼近器。《神经网络》，2 (5)，359–366。Hussein, A., M. M.Gaber, E. Elyan,和 C. Jayne，2017：模仿学习：学习方法综述。《ACM 计算调查》（CSUR），50(2)，1–35。Ioffe, S.，和C.Szegedy，2015：批归一化：通过减少内部协变量偏移加速深度网络训练。《机器学习国际会议》，PMLR，448–456。Jansen,M.F.，和 I.M. Held，2014：使用能量一致的反向散射对亚网格尺度涡旋效应进行参数化。《海洋建模》，80，36–48。Keisler, R.，2022：使用图神经网络预测全球天气。arXiv 预印本arXiv:2202.07575。Kochkov, D.,J.A.Smith, A. Alieva, Q.Wang, M. P.Brenner, 和 S. Hoyer，2021：机器学习加速的计算流体动力学。《美国国家科学院院刊》，118 (21)，e2101784118。Kraichnan,R. H.，1971：二维和三维湍流中的惯性区间传递。《流体力学杂志》，47 (3)，525–535。Large, W.,J. C.McWilliams,和S.C. Doney，1994：海洋垂向混合：综述及一种带有非局部边界层参数化的模型。《地球物理学评论》，32，363–403，doi:10.1029/94RG01872。Lele,S.K.，1992：具有类谱分辨率的紧致有限差分格式。《计算物理学杂志》，103，16–42，doi:10.1016/0021-9991(92)90324-R。List, B., L.-W.Chen,和N.Thuerey，2022：使用可微流体求解器进行学习式湍流建模：基于物理机制的损失函数与优化时域。《流体力学杂志》，949，A25。Loshchilov, I.，和F.Hutter，2016：Sgdr：带热重启的随机梯度下降。arXiv 预印本 arXiv:1608.03983。Ma,C., J.Wang, 和 E. Weinan，2019：带记忆的模型约化与动力系统的机器学习。《计算物理学通讯》，25 (4)，947–962。Ma,C., J. Wang,和合作者，2018：带记忆的模型约化与动力系统的机器学习。arXiv 预印本arXiv:1808.04258。Marston, J.,E.Conover,和T. Schneider，2008：基于累积量展开的不稳定正压喷流统计特性。《大气科学杂志》，65 (6)，1955–1966。

Maulik, R.,O. San, A. Rasheed,和P. Vedula, 2019：使用神经网络进行二维湍流的子网格建模。《Journalof Fluid Mechanics》，858，122–144。McWilliams, J. C., 1984：湍流流动中孤立相干涡旋的出现。《Journal of FluidMechanics》，146，21–43。Miyanawala,T.P.,和R. K.Jaiman, 2017：纳维–斯托克斯方程的高效深度学习技术：非定常尾流流动动力学的应用。arXiv 预印本 arXiv:1710.09099。Orszag,S.A., 和 M. Israeli, 1974：黏性不可压缩流动的数值模拟。《Ann.Rev. fluidmechanics》，6，281–318, doi:10.1146/annurev.fl.06.010174.001433。Ott, J., M.Pritchard,N. Best, E.Linstead, M.Curcic, 和 P. Baldi,2020：用于科学计算的 Fortran-Keras 深度学习桥接工具。《Scientific Programming》，2020年。Parish, E. J., 和 K.Duraisamy,2017：使用 Mori-Zwanzig 形式的大涡模拟非马尔可夫闭合模型。《Physical Review Fluids》，2（1），014604。Pascal,F., 和 C. Basdevant, 1992：二维湍流的非线性 Galerkin 方法和子网格尺度模型。《Theoretical andComputational Fluid Dynamics》，3（5），267–284。Pearson, B., B. Fox-Kemper,S. Bachman,和F.Bryan, 2017：全球涡旋丰富模型中尺度感知子网格中尺度涡旋模型的评估。《OceanModelling》，115，42–58。Perezhogin, P.,L. Zanna,和 C.Fernandez-Granda,2023年：理想化海洋模型中用于随机子网格参数化的生成式数据驱动方法。arXiv 预印本 arXiv:2302.07984。Piomelli, U.,W. H.Cabot, P.Moin, 和 S. Lee, 1991：湍流和过渡流动中的子网格尺度反散。《Physicsof Fluids A: Fluid Dynamics》，3（7），1766–1771。Rahaman, N., A. Baratin,D. Arpit, F. Draxler,M. Lin, F. Hamprecht,Y.Bengio, 和 A. Courville, 2019：关于神经网络的谱偏差。《International Conference onMachine Learning》，PMLR，5301–5310。Ramachandran, P., B. Zoph,和 Q.V.Le,2017：搜索激活函数。arXiv 预印本arXiv:1710.05941。Rasp,S., 2020：将耦合在线学习作为应对神经网络参数化中不稳定性和偏差的方法：通用算法与 Lorenz 96 案例研究（v1. 0）。《Geoscientific Model Development》，13（5），2185–2196。Ross, A., Z. Li, P.Perezhogin,C. Fernandez-Granda,和 L.Zanna, 2023年：理想化模型中机器学习海洋子网格参数化的基准测试。《Journal of Advances in ModelingEarth Systems》，15（1），e2022MS003 258。Souza, A. N., G. Wagner, A. Ramadhan, B. Allen, V. Churavy, J. Schloss, J. Campin, C. Hill, A. Edelman,J. Marshall, 和Coauthors,2020：海洋参数化的不确定性量化：应用于穿透性对流的 k-剖面参数化。《Journal of Advances inModeling EarthSystems》，12（12），e2020MS002 108。Srinivasan,K.,和 W. R. Young, 2012：纬向急流不稳定性。《J. Atmos. Sci.》，69，1633–1656，doi:10.1175/JAS-D-13-0246.1。Srivastava,N., G. Hinton, A. Krizhevsky, I. Sutskever, 和 R.Salakhutdinov, 2014：Dropout：防止神经网络过拟合的简单方法。《The journal of machine l earning research》，15（1），1929–1958。Stachenfeld, K., D. B. Fielding, D. Kochkov,M.Cranmer, T.Pfaff, J. Godwin, C. Cui, S. Ho, P. Battaglia,和A. Sanchez- Gonzalez,2021：用于高效湍流模拟的学习型粗粒化模型。arXiv 预印本 arXiv:2112.15275。Subel, A., Y. Guan, A.Chattopadhyay,和P.Hassanzadeh, 2022：解释将数据驱动的子网格尺度闭合迁移学习到不同湍流流动中的物理机制。arXiv 预印本 arXiv:2206.03198。Xiao,Z., M. Wan,S.Chen,和G. Eyink, 2009：二维湍流逆能量级联的物理机制：数值研究。《Journal ofFluid Mechanics》，619，1–44。Yuval,J., 和P.A. O'Gorman, 2020：在一系列分辨率下用于气候建模的稳定机器学习子网格过程参数化。《Nature communications》，11（1），1–10。Yuval, J., P.A.O'Gorman, 和 C. N. Hill,2021：使用神经网络对大气子网格过程进行稳定、准确且物理一致的参数化，并在降低精度下保持良好性能。《Geophysical Research Letters》，48（6），e2020GL091363。Zanna, L., 和 T. Bolton, 2020：海洋中尺度闭合的数据驱动方程发现。《Geophysical ResearchLetters》，47（17），e2020GL088376。Zhou, D.-X., 2020：深度卷积神经网络的普适性。《Applied and computational harmonic analysis》，48（2），787–794。Zhou, Z., G. He, S.Wang,和 G.Jin, 2019：使用人工神经网络的各向同性湍流大涡模拟子网格尺度模型。《Computers & Fluids》，195，104 319。

（KS）气象与海洋科学系，加利福尼亚大学洛杉矶分校，加利福尼亚州 90095 -

（JCM）加利福尼亚大学洛杉矶分校大气与海洋科学系及地球物理与行星物理研究所，加利福尼亚州洛杉矶，邮编 90095-1565，美国