WaveSim：一种用于天气与气候场的基于小波的多尺度相似性度量

Gabriele Accarino[1,7], **Viviana Acquaviva**[2,3],**Sara Shamekh**[4],**Duncan Watson-Parris**[5],**David Lawrence**[6]

[1]哥伦比亚大学地球与环境工程系，纽约州纽约市，美国 [2]纽约市立大学纽约市技术学院，布鲁克林，纽约州纽约市，美国 [3]哥伦比亚大学拉蒙特-多尔蒂地球观测站，帕利塞兹，纽约州，美国 [4]纽约大学库朗数学科学研究所，纽约州纽约市，美国 [5]斯克里普斯海洋学研究所和哈利乔格鲁数据科学研究所，加利福尼亚大学圣迭戈分校，加利福尼亚州，美国 ˇ美国国家大气研究中心（NCAR），博尔德，科罗拉多州，美国 [6]CMCC基金会——欧洲-地中海气候变化中心，莱切，意大利

arXiv:2512.14656v1 [physics.ao-ph] 16 Dec 2025

通讯作者：Gabriele Accarino， ga2673@columbia.edu

摘要

我们介绍了WaveSim，这是一种用于天气和气候应用中空间场评估的多尺度相似性度量。WaveSim利用小波变换将输入场分解为特定尺度的小波系数。该度量通过将由这些系数导出的三个正交分量相乘构建而成：幅度，用于量化系数能量分布的相似性，即场的强度；位移，通过比较归一化能量分布的质心来捕捉空间偏移；以及结构，用于评估与位置和振幅无关的模式组织。每个分量都会产生一个范围从0（完全不相似）到1（完全相似）的尺度特定相似性得分，随后在不同尺度间进行组合，以得到总体相似度度量。我们首先使用合成测试案例评估WaveSim，通过施加受控的空间和时间扰动，系统地评估其敏感性及预期行为。随后，我们通过地球系统模式中气候变率关键模态的物理相关案例研究，展示其适用性。传统的逐点度量缺乏将误差归因于物理尺度或不相似性模态的机制。WaveSim在小波域中运行，并沿独立轴分解信号，从而避开了这些局限，为复杂场的相似性评估提供了一个可解释且具有丰富诊断能力的框架。此外，WaveSim框架允许用户强调特定尺度或分量，并适用于用户定制的模型比较、模型评估，以及预测系统的校准和训练。我们提供了可直接用于PyTorch的WaveSim实现及所有评估脚本，地址为： `https://github.com/gabrieleaccarino/wavesim`。

通俗摘要

WaveSim 是一种用于天气和气候应用的新方法，可通过衡量两个空间场在多个空间尺度上的相似程度来进行比较。它不再逐点比较地图，而是将每幅地图分解为大尺度特征和更精细的细节。随后，它从三个互补的方面比较这些地图：模式的强度（幅度）、模式是否位于相同位置或发生了偏移（位移），以及它们的形状和整体组织是否相似（结构）。我们首先使用合成示例，在不同的受控测试情形下检验 WaveSim，以更好地了解其表现。随后，我们将 WaveSim 应用于与重要气候模式相关的地球系统模型模拟数据。结果表明，WaveSim 能够检测传统方法往往会遗漏的多尺度相似性，因此可用于模型比较与评估，以及预测系统的校准。

1 引言

由于地球系统过程本身具有多尺度和异质性，准确比较网格化气候场仍是一项重大挑战（Ojima 等，1992）。比较场的主要目的是量化两个空间场之间的“接近”程度，通常比较的是模型预报或模拟输出与参考场或“真实值”，例如分析场或观测数据。这项任务称为空间检验（Briggs 和 Levine，1997），对模型评估和业务预报等诸多应用至关重要。改进空间检验指标也将直接惠及不断发展的数据驱动模拟器领域（Bi 等，2023；Lam 等，2023；Chen 等，2023；Kochkov 等，2024），并有助于推动当前对天气和气候模型预报进行评估和基准测试的工作，例如 WeatherBench2（Rasp 等，2024）、ESMValTool（Eyring 等，2020）和 ClimateBench（Watson-Parris 等，2022）。均方误差（MSE）等传统逐点误差指标因数学形式简单、计算效率高而得到广泛应用。然而，

这些指标往往不适用于空间检验，因为它们将空间上复杂的信息压缩为单一分数，因此无法有效反映空间结构，也无法体现待比较场的多尺度特性（Briggs 和 Levine，1997）。

这些局限性在评估降水场时尤为明显，因为降水场呈现出 t
独特的检验挑战。降水呈现出稀疏且间歇性的空间模式，大面积区域没有降水，而小片区域可能包含强烈且局地化的事件（Buschow 等人，2019）。这些特征导致空间结构呈现斑块状且高度间歇性，使得使用逐点度量评估降水尤其困难。即使场发生很小的位移，也会导致“双重惩罚 ”
效应（Ebert，2008年；Brown 等人，2011；Kapp 等人，2018年）：预报未能预测实际发生降水的位置会受到一次惩罚，而即使整体模式与参考场十分相似，在错误位置预测降水也会再次受到惩罚。比较高分辨率地图时，这一问题更加突出。为解决这些局限，人们提出了多种空间检验方法。Gilleland 等人（2009）；Ahijevych 等人（2009）；Gilleland 等人（2010）；Buschow 等人（2019）进行的全面综述将这些方法分为四大类：邻域法、基于特征的方法、场变形法和尺度分离技术。此外，Dorninger 等人（2018年）又将距离度量列为一类。基于距离的方法通过评估原始空间中所有网格点之间的地理距离，或对地图进行阈值处理后在二值地图上评估这些距离，从而衡量预报场与观测场之间的距离（Skok，2022年）。典型例子包括降水平滑距离（Skok，2022年）、降水归因距离（Skok，2023年）及其全球版本（Skok 和 Lled´o，2025年），以及分数技巧评分的应用（Skok 和 Roberts，2016年、2018年；Skok 和 Kosovelj，2025年）。在这些类别中，尺度分离技术与本研究尤为相关。这种方法通过应用单频带空间滤波器（例如傅里叶变换或小波变换），将场分解为不同尺度的空间分量，并分别对每个尺度分量应用传统检验指标（Briggs 和 Levine，1997；Casati 等人，2004；Casati 和 Wilson，2007；Jung 和 Leutbecher，2008年；Casati，2010年；Weniger 等人，2017；Kapp 等人，2018年；Buschow 和 Friederichs，2020年、2021年；Casati 等人，2023年）。除了空间检验方法外，我们还认为，评估两个空间场接近程度的任务，可以与评估感知相似度的任务联系起来；这一概念在计算机视觉和图像质量评估（IQA）领域已得到广泛研究。在这一背景下，结构相似性指数（SSIM）（Wang 等人，2004年）等指标并非通过比较单个像素值来评估图像相似度，而是量化亮度（即明暗）、对比度和结构等局部特征。然而，即便将空间场视作二维地图，并将其数值解释为像素强度，将基于相似度的指标直接应用于天气和气候数据仍并非易事。与自然图像不同，这类地图由带有物理单位且动态范围不受限的浮点值构成。其统计和结构属性与传统图像质量评估所假设的属性存在显著差异。尽管如此，近期研究已通过开发数据结构相似性指数（DSSIM）指标（Baker 等人，2023年），将这些方法应用于气候数据。该指标与结构相似性指数（SSIM）类似，但可直接用于浮点数据，旨在评估数据质量。基于尺度分离技术以及感知相似度框架带来的启示，我们提出 WaveSim，这是一种用于评估空间场的多尺度相似性度量。WaveSim 利用小波变换将场分解为不同的空间尺度，并通过相应的小波系数明确分离不同分辨率上的信息。该指标由三个互补分量组成，每个分量均由小波系数导出，并捕捉相似度的一个不同方面：（i）幅度量化小波系数能量的振幅差异；（ii）位移 通过比较小波系数的边缘能量分布，捕捉相干结构的空间偏移；（iii）结构评估特征的空间组织形式，不受其位置或振幅影响，从而检测模式层面的

相似性。每个分量在每个单独尺度上都会给出一个介于 0（不相似）和 1（完全相似）之间的相似度分数。随后，逐尺度组合这些分量的评分，得到整体相似度分数。与在原始数据空间中计算的传统逐点评分不同，WaveSim 在小波域中运行，因此能够整合多尺度、多分量信息。由此得到的复杂空间场相似度评估更具可解释性，也能提供更丰富的诊断信息。

本文其余部分安排如下：第 2 节介绍离散小波变换，并阐述 WaveSim 方法。第 3 节报告将 WaveSim 及其他相似性度量应用于合成测试案例和不同气候变率模态所得的结果。最后，第 4 节总结通过这些测试案例展示的 WaveSim 主要能力，并提出结论性意见及未来的发展方向。

2 方法

本节介绍用于多尺度分解和尺度分离的离散小波变换背景知识。我们还将介绍 WaveSim 的数学表述，以及用于评估其在合成测试案例和不同地球系统模式中气候变率模态上的表现的数据集。

2.1 离散小波变换

离散小波变换（DWT）是一种多分辨率分析技术，可将信号表示为同时在空间（或时间）和尺度（频率）上局部化的分量。该变换由一个母小波 ψ构造而成，通过二进尺度变换（以 2 的幂为比例）和离散平移，可得到一整族小波：

$$\psi_{s,l}(x) = \frac{1}{\sqrt{2^s}}\psi\Big(\frac{x - 2^s l}{2^s}\Big), \tag{1}$$

其中，$s \in \mathbb{Z}$ 是尺度索引（较大的s 对应较粗的分辨率），$l \in \mathbb{Z}$ 是平移索引（控制空间或时间上的平移）。因此，每个$\psi_{s,l}(x)$都是母小波经过重新缩放和平移后的版本。如果选择 ψ 使其满足某些可容许性和正交条件，则集合$\{\psi_{s,l}(x)\}$ 构成 $L^2(\mathbb{R})$的标准正交基。这意味着任何 $f \in L^2(\mathbb{R})$都可以表示为这些基函数的和：

$$f(x) = \sum_l \sum_s d_s[l]\psi_{s,l}(x), \tag{2}$$

其中，尺度 2^s 上的小波（或细节）系数 $d_s[l]$ 通过将 f 投影到该基上获得。

在实际应用中，无法计算公式（2）中的无限个尺度。为了使分解可行，引入尺度函数 ϕ 来表示信号平滑的低频部分。

$$f(x) = \sum_l a_S[l]\phi_{S,l}(x) + \sum_l \sum_s^S d_s[l]\psi_{s,l}(x), \tag{3}$$

其中 $a_S[l]$是最粗尺度上的近似系数 S，而ϕ_S.

$2^{-S/2}\phi\ 2^{-S}x-l$
$\phi_{S,\ l}$

$_l\ x\ \ =(\ )(\ )$ 是经过缩放和平移的尺度函数。与. 类似，
$\psi_s\ l$

构成函数的正交基，这些函数在给定尺度 2^S 上更加平滑，可用于分解小波（细节）未捕获的平滑残差（Mallat，1999；Ha 等人，2021）。母小波应根据场的空间特征进行选择，以确保在平滑性、对称性和紧致性之间取得适当平衡，从而有效捕捉其物理结构。离散小波变换（DWT）最重要的性质之一是，在尺度 2^{s+1} 上的近似系数和细节系数都可以由前一尺度 2^s 上的近似系数计算得到（Mallat，1989；Meyer，1992）。DWT 可用于高效分解二维信号，例如图像或浮点值矩阵，如表示气候场的地图。首先，沿每个轴（行/列）分别应用公式（3）。此操作会生成四个矩阵，分别表示近似部分（对于 Haar 母小波，这只是分辨率降至原来一半的地图）以及尺度 2^s 上沿三种不同方向（水平、垂直和对角）的细节系数。随后，对包含尺度 2^s上近似系数的矩阵迭代重复这一过程，直至达到所需分辨率 2^S 。中间尺度和最终尺度上包含细节系数的所有矩阵也都会被存储，从而实现无损表示。图1展示了一个欧洲降水地图的二维三级 DWT 示例。

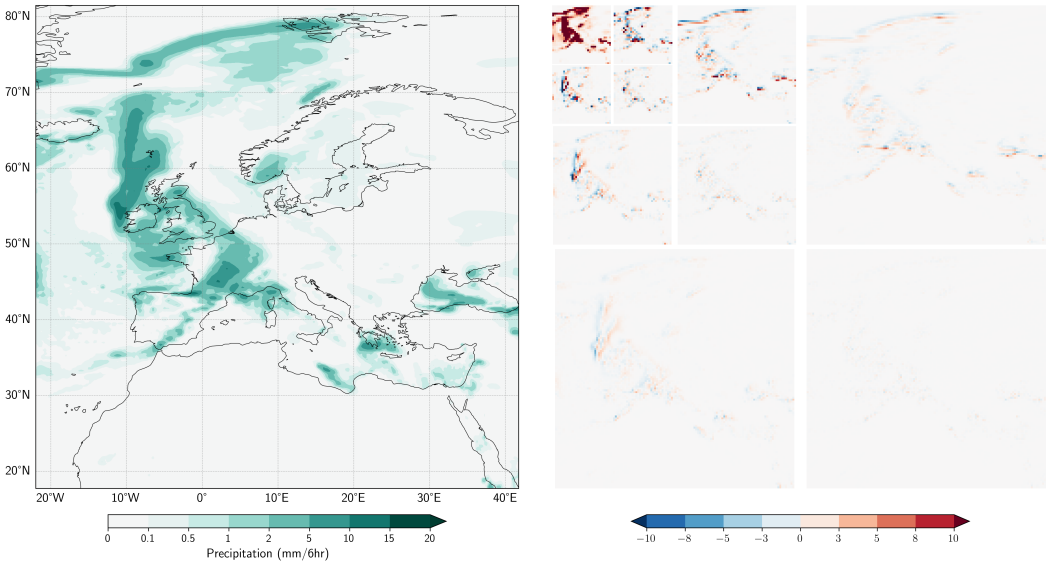

图1. 欧洲的参考6小时降水图（毫米/6小时）（左图），以及使用 Haar 小波进行的相应二维三尺度离散小波变换（DWT）（右图）。该参考地图取自 ERA5 于 2016年1月2日 06:00:00 的数据，裁剪至欧洲范围（17.5–81.5°N,–22–41.75°E），由一个 256 × 256 网格组成。在小波分解中，每一级都会生成三个系数矩阵，分别捕捉水平、垂直和对角细节，且分辨率在连续尺度上减半（128 × 128、64 × 64 和 32 × 32 个网格点）。左上象限包含近似系数，它们保留了场的大尺度低频结构（即地图平滑后的版本，不再进行进一步分解）。

2.2 尺度分离

在每个尺度上，DWT 将一个场分解为近似系数（粗尺度、低频结构）和细节系数（细尺度、高频变化）。这种分解遵循二进层级的尺度结构，在该层级中，系数的空间分辨率在每个后续尺度上减半（例如，{v1}{v2}等）。

在每个尺度上，细节会进一步分为水平、垂直和对角方向，从而对高频内容进行方向性细分。DWT 的一项基本性质是可逆性：使用逆离散小波变换（IDWT），可以根据近似系数和细节系数完美重建原始场。换言之，这种分解不会丢失或舍弃信息，而是将信息重新组织到不同的尺度和方向上。可逆性使得尺度分离过程能够分离出特定尺度的贡献：

- 选择特定尺度下的一种方向（水平、垂直或对角），并将所有其他系数矩阵置零；
- 仅使用保留的系数应用 IDWT，以重建与该特定方向和尺度对应的空间场，其分辨率与原始地图相同；
- 在同一尺度下对三种方向重复此过程，并将得到的地图相加。这样便得到该尺度对应的细节图，它整合了所有方向上的贡献，同时仍局限于该尺度；
- 在各个尺度上重复这一过程，得到一张近似图（包含尚未解析的最粗尺度信息）以及每个尺度对应的一张细节图，所有地图均采用原始空间分辨率。

按照这一方法，如图1所示，在每个尺度（或层级）上，我们会得到四个小波系数矩阵：三个细节矩阵（分别对应尺度1、2和3上的全部细节）以及一个近似矩阵。该近似矩阵会在下一个（更粗的）细节尺度上进一步分解为细节矩阵和近似矩阵。此分解过程会持续进行，直至达到所需的最粗尺度。这些矩阵共同清晰、直观地呈现了场在不同尺度和方向上的结构分离情况。举例来说，3尺度 DWT 会生成三张细节系数图，每张图都提取特定尺度上的信息；此外，还会生成一张近似系数图，表示尺度分离过程结束时的最粗尺度信息。

2.3 WaveSim

WaveSim 是一种用于二维地图（例如气候场）的基于小波的多尺度相似性度量。WaveSim 不会直接在原始空间中比较两个场 X 和 Y，而是利用它们的小波分解来分离不同尺度的信息。随后，在尺度隔离的重建结果上评估相似性，确保不同尺度的贡献仍可解释且相互独立。WaveSim 将总体相似度分数分解为三个正交分量：

- 幅值相似性（$\mathcal{M}$）：衡量不同尺度上的能量幅度或亮度相似性，而不考虑能量集中在何处；
- 位移相似性（$\mathcal{D}$）：描述能量分布在空间维度（例如纬度和经度）上的对齐程度。此分量旨在不受全局幅度差异影响；
- 结构相似度（$\mathcal{S}$）：衡量小波系数结构模式的保留程度，不受幅度或空间位置影响。

总体 WaveSim 分数将这三个分量结合起来，并通过可调权重汇总不同尺度的信息。形式上，设 X，$Y \in \mathbb{R}^{H \times W}$ 为两个输入场，表示为形状为高度（H）和宽度（W）的地图。对于每个尺度 s，我们通过第

2.2 节所述的逆离散小波变换（IDWT）过程，重建小波系数 和 对应的特定尺度地图 。得到的系数地图与图 1（右图）所示的地图相似，但形状与原始地图相同，即 $H \times W$。随后将 WaveSim 分数定义为：

$$WaveSim(X,Y) = \sum_{s=1}^{S} w^s \cdot \left(\mathcal{M}(\tilde{X}^s, \tilde{Y}^s)^\alpha \cdot \mathcal{D}(\tilde{X}^s, \tilde{Y}^s)^\beta \cdot \mathcal{S}(\tilde{X}^s, \tilde{Y}^s)^\gamma \right)$$

其中， α、 β、 $\gamma \in [0,1]$ 控制各分量的相对重要性（其中 $\alpha = \beta = \gamma = 1$ 表示各分量具有相同的重要性）。参数 w^s 为尺度相关权重，其归一化方式满足 $\sum_{s=1}^{S} w^s = 1$；根据具体应用，可调整这些权重以突出特定尺度。WaveSim框架支持细节系数和近似系数地图，但在本分析中，我们仅使用细节系数，以聚焦于各层级的尺度特定空间特征。接下来，我们给出集成于WaveSim中的各分量的形式化定义。

2.3.1 幅度分量（$\mathcal{M}$）

~ ~在每个尺度 s上，给定两个场 X^s 和 Y^s的小波系数能量：

$$E_{\tilde{X}^s} = (\tilde{X}^s)^2, \qquad E_{\tilde{Y}^s} = (\tilde{Y}^s)^2$$

我们计算每个尺度的平均能量：

$$\bar{E}_{\tilde{X}^s} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} E_{\tilde{X}^s}(i,j), \qquad \bar{E}_{\tilde{Y}^s} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} E_{\tilde{Y}^s}(i,j)$$

以及它们的相对差异：

$$\delta^s = \frac{|\bar{E}_{\tilde{X}^s} - \bar{E}_{\tilde{Y}^s}|}{\bar{E}_{\tilde{X}^s} + \bar{E}_{\tilde{Y}^s} + \varepsilon}.$$

随后，我们将幅度分量定义为：

$$\mathcal{M}(\tilde{X}^s, \tilde{Y}^s) = 1 - \delta^s$$

~ ~其中，$\mathcal{M}(X^s, Y^s) \in [0,1]$，且 ε是一个较小的正数，用于确保数值稳定性。

2.3.2 位移分量 *($\mathcal{D}$)*

为考虑能量谱的位移，我们考察小波系数地图各行（纬度）和各列（经度）上能量的归一化边缘分布：

$$\rho^{lat}_{\tilde{X}^s}(i) = \frac{\sum_j E_{\tilde{X}^s}(i,j)}{\sum_{i,j} E_{\tilde{X}^s}(i,j)}, \qquad \rho^{lon}_{\tilde{X}^s}(j) = \frac{\sum_i E_{\tilde{X}^s}(i,j)}{\sum_{i,j} E_{\tilde{X}^s}(i,j)}$$

~（ Y^s也同理）。归一化确保边缘分布不受整体幅度影响，而整体幅度已由幅度分量捕获。我们使用 Jensen-Shannon 散度（JSD）（Men´endez 等，1997）来衡量 $\rho^{(\cdot)}_{\tilde{X}^s}$ 与 $\rho^{(\cdot)}_{\tilde{Y}^s}$ 之间边缘分布的差异：

$$JSD(\rho^{lat}_{\tilde{X}^s}, \rho^{lat}_{\tilde{Y}^s}) = \frac{1}{2}KL(\rho^{lat}_{\tilde{X}^s} || m^{lat}_{\tilde{X}^s,\tilde{Y}^s}) + \frac{1}{2}KL(\rho^{lat}_{\tilde{Y}^s} || m^{lat}_{\tilde{X}^s,\tilde{Y}^s}),$$

其中，纬度的混合分布 m 定义为 $m^{lat}_{\tilde{X}^s,\ \tilde{Y}^s} = \frac{1}{2}$（$\rho^{lat}_{\tilde{X}^s} + \rho^{lat}_{\tilde{Y}^s}$），而 KL（·）表示 Kullback-Leibler散度（Shlens，2014）。经度对应的公式定义方式相同。与 KL 散度不同，JSD 的取值范围有界于 $[0, 1]$, ，其中 0 表示正在比较的两个分布完全相同。由于在我们的相似性框架中，1 表示完全相似，因此我们将散度转换为位移相似度分数，方法如下：

$$\mathcal{D}(\tilde{X}^s, \tilde{Y}^s) = (1 - JSD(\rho^{lat}_{\tilde{X}^s}, \rho^{lat}_{\tilde{Y}^s})) \cdot (1 - JSD(\rho^{lon}_{\tilde{X}^s}, \rho^{lon}_{\tilde{Y}^s})).$$

$\tilde{}$ $\tilde{}$这确保了$\mathcal{D}(X^s, Y^s) \in [0, 1]$，因为它是两个有界量的乘积$[0, 1]$。

2.3.3 结构分量（$\mathcal{S}$）

为了计算结构分量，我们首先将尺度 s 上的小波系数矩阵展平为向量并进行排序。我们将尺度 s上排序后的系数向量定义为 $\tilde{x}^s$、 $\tilde{y}^s \in \mathbb{R}^{H \cdot W}$，分别对应于 X^s 和 Y^s。这一步保留了结构模式，同时去除了位置信息。随后，我们对它们进行中心化和归一化，以消除对能量幅度的敏感性：

$$\bar{x}^s = \frac{\tilde{x}^s - \mu(\tilde{x}^s)}{\|\tilde{x}^s\|_2 + \varepsilon}, \qquad \bar{y}^s = \frac{\tilde{y}^s - \mu(\tilde{y}^s)}{\|\tilde{y}^s\|_2 + \varepsilon}$$

其中 ε 是一个为保证数值稳定性而添加的小正数。从 $\bar{x}^s$ 和$\bar{y}^s$出发，我们计算它们的余弦相似度，该相似度衡量向量的方向，因此能够捕捉其结构模式的相似性，而不受绝对幅度影响：

$$cos(\bar{x}^s, \bar{y}^s) = \frac{\bar{x}^s \cdot \bar{y}^s}{\|\bar{x}^s\| \|\bar{y}^s\|} \in [-1, 1],$$

并将其平移至 $[0, 1]$ 范围，方法为 $cos^*(\bar{x}^s, \bar{y}^s) = \frac{1}{2}(1 + cos(\bar{x}^s, \bar{y}^s))$。然而，仅依赖余弦相似度会使该指标过于宽松：两个幅度差异很大但排序形状相似的系数组，仍可能被视为结构相同。为避免这种情况，我们引入幅度变化惩罚，以衡量排序分布中变化的相对一致性：

$$P^s = 1 - \frac{\|\tilde{x}^s - \tilde{y}^s\|/n}{\mu(|\tilde{x}^s|) + \mu(|\tilde{y}^s|) + \varepsilon}$$

其中 n 为向量长度。当分布的相对振幅存在显著差异时，该项会降低结构相似度，同时对幅度分量已捕捉到的整体缩放保持不敏感。具体而言，它可防止余弦相似度对形状相符、但振幅被任意拉伸或压缩的模式给出高分。尺度 s 下的最终结构相似度定义如下：

$$\mathcal{S}(\tilde{X}^s, \tilde{Y}^s) = cos^*(\bar{x}^s, \bar{y}^s) \cdot P^s$$

从而确保结构分量与幅度和位移均正交，同时对不成比例的振幅失真保持鲁棒。

3 个应用示例

3.1 合成测试案例

为展示 WaveSim 的行为，我们基于 2016年 1st 月 ERA5 总降水量（mm）的一幅参考地图生成了合成测试案例，该数据累积时长为 6 小时（图2，面板 a）。我们从全球地图中提取了欧洲的一个 256×256 区域（与图1左侧面板中的区域相同），并施加了一组空间扰动，包括：空间位移（面板 b–d），即将区域在纬度和/或经度方向上移动若干网格点；对角翻转（面板 e），即沿对角线翻转区域，在保留数值的同时改变其方向；以及升尺度（面板 f），即先将地图降采样至原分辨率的一半（尺度因子为 0.5），再升采样回原始形状，从而去除小尺度变率并引入块状伪影。尽管某些扰动（例如翻转，或如面板 d 所示较为一致的位移）并不代表空间验证任务中的实际情形，但它们提供了受控的测试案例，可用于评估 WaveSim 的各个分量是否表现符合预期。除合成空间扰动外，我们还评估了相对于参考地图（面板 a）的 1、2 和 3天提前量降水场的相似度（面板 g–i），并将其视为参考地图在时间上的扰动版本。

每对参考数据场和扰动场都使用 Daubechies（db）母小波分解为四个空间尺度，并据此得出 WaveSim 的分量和总体相似度分数。Daubechies 小波族（ Daubechies，1992 ）由具有指定数量消失矩的正交小波组成，这些消失矩决定了它们高效表示平滑多项式趋势的能力。具有n个消失矩的小波会去除或忽略最高至n-1次的所有多项式趋势，同时捕捉底层场中的高频变化或细节。此外，不同阶数的db小波具有不同的支撑，即小波函数非零的区间。高阶小波的支撑较大，更适合描述较大的场，而低阶小波的支撑则更紧凑。对于合成测试案例中使用的高分辨率地图，我们采用具有db 母小波，其消失矩数量为n=4（$db4$），以满足对较大支撑的需求，并适应底层快速变化的场（用 4 阶及以上的高次多项式描述效果更好）。在本次评估中，所有分量的重要性相同（$\alpha = \beta = \gamma$ =1），尺度权重均匀设为 $w_s = 0.25$，确保四个尺度上的权重之和为 1。

我们将 WaveSim 与另外两种指标进行比较：一种是受 (Casati 等人, 2004) 提出的技巧评分启发而改进的均方根误差版本，称为归一化均方根误差 (NRMSE)；另一种是 (Wang 等人, 2004; Baker 等人, 2023) 定义的数据结构相似性指数 (结构失真度)。

为构建 NRMSE，我们对无上界的均方根误差值进行归一化，并将其映射到 0 到 1 之间的相似度尺度，如下所示：

$$NRMSE(X,Y) = \exp\left[-\left(\frac{RMSE(X,Y)}{RMSE(X,X^*)}\right)\right],$$

其中 $RMSE(X,\ X^*)$ 表示参考地图 X 与随机打乱版本 X^* 之间的均方根误差，后者代表由无技巧模型生成的地图。

结构失真度由三个感知分量的加权组合得出：亮度（l）、对比度（c）和结构（s），其权重分别为 α、β、$\gamma > 0$:

$$SSIM(X,\ Y) = [l(X,\ Y)]^{\alpha} \cdot [c(X,\ Y)]^{\beta} \cdot [s(X,\ Y)]^{\gamma}.$$

从概念上讲，WaveSim 的灵感来自 SSIM 对多个分量进行整合和加权的方式，不过它在小波空间的多个尺度上运行，并通过位移分量捕获额外的正交信息。这两个指标返回的相似度分数均在 $[0, 1]$ 范围内，其中 1 表示完全相似，因此可以直接与 WaveSim 进行比较。

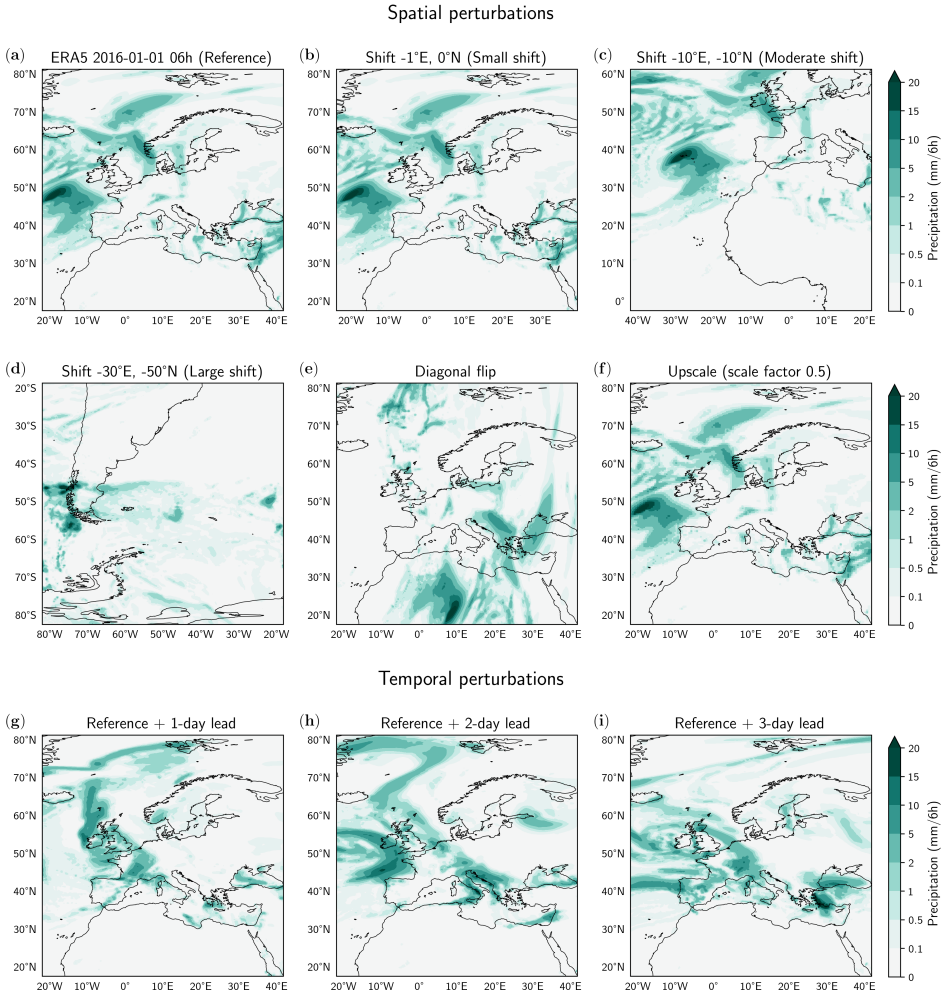

图2. 合成测试案例包括空间扰动和时间扰动（参见第3.1节）。空间扰动通过对单幅参考地图（面板 a）应用仿射变换（面板 b–e）和降尺度—升尺度过程（面板 f）生成。该地图展示了 ERA5 在 2016年1月 1^{st}的降水模式。时间扰动（面板 g–i）表示参考数据之后连续三天内不断变化的降水模式。所有案例均由分辨率为 256×256 个点的地图组成。

合成测试案例的结果见表1。正如预期，纬度和经度方向的空间偏移会改变扰动地图中降水模式的位置。由于扰动地图中的降水特征相对于参考数据发生了偏移，WaveSim的每个分量都会受到一定程度的影响。这在案例

（b）中显而易见：经度方向上的小幅偏移 $-1^\circ E$ （纬度方向无偏移）使各分量均保持较高的相似度，幅度和结构几乎没有变化，但位移分量的得分较低（介于.86至.93之间）。WaveSim的总体相似度为.87，与扰动地图中引入的小幅空间偏移相符。归一化均方根误差为.71，同样反映出中等偏高的相似度；而结构失真度则对这一扰动给出了过低的分数（.49），表明它对哪怕很小的地理位移也过于敏感。案例（c）考虑了 $-10^\circ E$、 $-10^\circ N$的中度空间偏移，相似度分数的下降也变得更加明显。幅值相似性仍然较高（高于.8）；这是预期之中的，因为在这种中度偏移下，系统中的大部分能量仍得以保留。相反，位移值显著下降（在最精细尺度上低至.28），准确反映了降水锋面在纬度和经度方向上的位移。结构相似度保持在.92以上，说明尽管扰动地图发生了空间位移，仍能识别出相似的结构模式。WaveSim的位移分量使总体分数降至.28，与结构失真度（.22）的表现相近，而归一化均方根误差（.37）则表明相似度略高。案例（d）代表一种极端情景：扰动造成了非常大的偏移，使目标地理区域与参考地图没有重叠。在此案例中，WaveSim的位移相似度在所有尺度上都急剧下降（低于.4），表明小波谱在纬度和经度方向上发生了显著位移，且缺乏空间对应关系。幅度分量表现不一：在粗尺度和粗至中尺度上相似度较低（分别为.18和.32），但在中至大尺度上数值较高（分别为.97和.75）。结构相似度仍然中等偏高（高于.7），这表明尽管两个场代表不同区域，其小波系数的内部组织结构仍具有可比性。尽管如此，幅度和位移分量的低相似度值使WaveSim的相似度分数降至.15，这与两幅地图之间的显著不匹配相符。结构失真度也报告两幅地图之间的相似度较低（.19），而归一化均方根误差则得出略高的数值（.43）。

与空间位移不同，像案例 (e) 中那样使降水模式沿对角线翻转的扰动，预计会保留结构和幅度，但会显著改变位移。这是因为翻转会重新分配小波系数边际能量分布在各位置上的分布，同时保持其振幅不变。WaveSim 清晰地捕捉到了这一行为：在所有尺度上，幅度和结构分量均获得完美相似度值（即 1.0），而位移相似度则大幅下降，尤其是在最精细的尺度上（0.14）。整体 WaveSim 低分 0.18 是由三个相似度分数的乘法组合按构造得出的，但 WaveSim 能够正确地将不相似性归因于其根本来源（位移）。对于这种扰动，DSSIM 给出的相似度分数更低（0.09），而 NRMSE 得分则更高（0.35）。这两个指标都未能考虑到结构和幅度得以保留这一点。案例 (f) 对应的扰动涉及分辨率粗化（降尺度），随后再通过插值恢复到原始尺寸（升尺度）。这一过程会产生块状插值伪影。正如预期，WaveSim 分量显示，幅值相似性随尺度增大而提高（从最精细尺度的 0.43 上升至最粗尺度的近 1.0），因为粗尺度变率在此过程中大多得以保留。位移也呈现类似趋势，从 0.81 起步，并迅速在粗尺度上趋近完美相似度（0.99）。结构表现出相近的模式：最精细尺度上的得分较低（0.65），而后续尺度上的相似度则逐渐升高（$0.89–0.99$）。整体 WaveSim 相似度为 0.71，既捕捉到小尺度模式的退化，也识别出大尺度组织结构得到了良好保留。相比之下，NRMSE（0.88）无法充分反映尺度依赖差异，因而给出更高的相似度；DSSIM（0.79）的得分略高于 WaveSim，但无法指出受到影响的是哪些尺度。这凸显了 WaveSim 的优势：能够区分细尺度变率的损失与较粗尺度特征的保留。

案例 (g–i) 表示时间扰动，其中将 ERA5 在提前一天、提前两天和 3天提前量的降水图与同一张参考地图（图 a）进行比较。与前面讨论的应用场景不同，这些扰动体现了时空演变模式，类似于现实世界空间检验任务中通常遇到的挑战。在案例 (g) 中，即提前一天的预报时效，WaveSim 给出的整体分数为 0.38。各尺度上的幅度和结构相似度均保持较高水平（两者都高于 0.8），而位移在所有尺度上都低于 0.5。这表明，尽管总降水量及其内部结构大体一致，各地图中的降水系统仍出现了明显偏移；仅经过一天，位置差异便已显现。在提前两天的预报时效（案例 h）下，整体 WaveSim 相似度降至 0.32。幅值相似性有所减弱，尤其是在较精细尺度上（在 $\sim$100 km 处降至 0.69），而较粗尺度仍保持中等程度的一致性。各尺度上的位移仍然较低（介于 0.43 至 0.53 之间），反映出降水系统的错位日益加剧。结构相似度也略有下降，但仍高于 0.8，说明扰动地图保留了降水特征的整体组织形态，只是空间位置有所偏移。在对应提前三天预报时效的案例 (i) 中，时空演变的影响变得更加明显。整体 WaveSim 分数降至 0.25。幅值相似性进一步下降，尤其是在细尺度上（分别在 $\sim$100 和 50 km 处降至 0.59 和 0.64）；各尺度上的位移得分始终处于中低水平（0.37–0.48），凸显了降水模式存在显著的位置错配。然而，结构相似度仍然较高（$0.81-0.86$），尽管与提前一天和提前两天的预报时效相比有所下降。与 WaveSim 和 DSSIM 相比，NRMSE 值仍相对较高（案例 g–i 分别为 0.43、0.39 和 0.4），而 DSSIM 给出的分数略低（0.26、0.23、0.24）。总体而言，WaveSim 似乎能够更有效地刻画降水系统的时间演变。根据结果及对其分量的解读，WaveSim 不仅反映了降水特征逐渐发生的位移，还捕捉到了它们随时间变化的尺度依赖结构和幅度变化。

3.2 气候变率模态

除了合成测试案例之外，我们还通过地球系统模式（ESM）中具有重要物理意义的气候变率关键模态案例研究，进一步评估 WaveSim，包括太平洋—北美遥相关型（PNA）偏差和厄尔尼诺—南方涛动（ENSO）；后者通过厄尔尼诺减去拉尼娜的合成场估算。我们考察的地球系统模式属于耦合模式比较计划（CMIP）。该计划是旗舰级国际项目，旨在通过公平比较模式输出改进气候模式。具体而言，我们从最新的第6阶段 CMIP6 中选取了若干模式（Eyring 等，2016年；Tebaldi 等，2021年），并从 LEAP-Pangeo（`https://leap-stc.github.io/intro.html`）获取了相应数据。对不同地球系统模式气候变率模态的评估，展示了 WaveSim 在评估相似性以及揭示不同空间尺度差异方面的潜在应用，而这些差异是传统评估指标无法捕捉的。在本研究中，我们无意对地球系统模式进行排名，也不打算确定哪个模式最接近给定的真实值。相反，我们侧重于通过不同的分析分量解读各项得分，从而深入了解相似性或差异性的根本来源。

3.2.1 太平洋-北大西洋偏差

我们使用WaveSim，将三个CMIP6地球系统模式——CESM2-FV2（Danabasoglu，2019年）、CMCC-CM2- HR4（Scoccimarro等，2017年）和GFDL-ESM4（Krasting等，2018年）——的12月至次年2月（DJF）平均500百帕位势高度图（1979–2015年）与NCEP-DOE再分析II（Kanamitsu等，2002年）进行比较，比较范围为图3中的黑色虚线区域。为保持一致，地球系统模式Z500数据被重新映射到与NCEP-DOE再分析资料相同的网格

表1. 参考地图 (a) 与扰动地图 (b-i) 在选定合成测试案例上的相似度分数。每个案例均报告四个尺度上的 WaveSim 分量（$\mathcal{M}$：幅度，$\mathcal{D}$：位移，$\mathcal{S}$：结构），数值按从最精细到最粗尺度的顺序列出。该场的原始空间分辨率为 $0.25^\circ \times 0.25^\circ$，这些尺度大致相当于 50、100、200 和 360 km。相似度分数四舍五入至小数点后两位。

面板	测试案例	归一化均方根误差	结构失真度	**WaveSim（我们的方法）**	WaveSim分量	
(b)	小幅平移	0.71	0.49	0.87	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.96, 0.99, 0.97, 0.99] [0.86, 0.92, 0.89, 0.93] [0.99, 0.99, 0.98, 0.98]
(c)	中度偏移	0.37	0.22	0.28	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.85, 0.97, 0.88, 0.82] [0.28, 0.41, 0.33, 0.30] [0.95, 0.97, 0.93, 0.92]
(d)	Large shift	0.43	0.19	0.15	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.18, 0.32, 0.97, 0.75] [0.07, 0.20, 0.38, 0.35] [0.71, 0.74, 0.86, 0.80]
(e)	Diagonal flip	0.35	0.09	0.18	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[1.00, 1.00, 1.00, 1.00] [0.14, 0.21, 0.17, 0.21] [1.00, 1.00, 1.00, 1.00]
(f)	放大	0.88	0.79	0.71	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.43, 0.80, 0.97, 0.99] [0.81, 0.95, 0.99, 0.99] [0.65, 0.89, 0.99, 0.99]
(g)	+ 1 天提前期	0.43	0.26	0.38	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.88, 0.82, 0.99, 0.86] [0.48, 0.46, 0.42, 0.48] [0.94, 0.91, 0.89, 0.96]
(h)	+ 2 天提前量	0.39	0.23	0.32	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.76, 0.69, 0.83, 0.76] [0.43, 0.52, 0.45, 0.53] [0.92, 0.89, 0.87, 0.82]
(i)	+ 3-day lead	0.40	0.24	0.25	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.64, 0.59, 0.74, 0.80] [0.37, 0.46, 0.45, 0.48] [0.82, 0.81, 0.83, 0.86]

sis II (73×144 点）。所选分析区域是一个 32×32 二进网格，覆盖了部分 PNA（161-240°E，12.5–90°N），使用 Daubechies-2（*db2*）母小波分解为三个空间尺度，约为 520、890 和 1,480 km。我们采用 *db2*，因为它的支撑紧凑；鉴于应用区域相对较小，这一特性很有优势，因为较大的滤波器可能会加剧边界效应。此外，*db2* 在表示 Z500 这类平滑场时也能取得良好平衡。表2报告了数值结果，包括 WaveSim 点得分和各分量的相似度，以及用于比较的 NRMSE 和结构失真度。评估结果表明，WaveSim 对 CMCC-CM2-HR4 中较大的偏差（第二行）尤其敏感，其点得分为 0.66。这表明其相似度中等偏高，但明显低于结构失真度（0.86）和 NRMSE（0.89）的评估结果。重要的是，这一差异与对偏差图（地球系统模式 - 再分析场）的视觉检查结果一致：图中显示，区域中心存在低估，区域北角则有局部高估。将 WaveSim 分数分解为各分量，可以进一步揭示原因。相似度相对较低，主要是由幅度（跨尺度的 $0.71-0.83$ ）驱动，结构（$0.84-0.91$）也有少量影响；这表明多个尺度上的小波系数能量及其空间组织存在一定差异，但程度适中。另一方面，所有尺度上的位移得分始终较高（$0.96-0.97$），这表明沿纬度和经度方向的能量边缘分布没有显著位移，主要问题在于模式的振幅和组织方式，而非其空间位置。

对于 CESM2-FV2（第一行），评估结果表明，在所有指标中，12月至次年2月平均再分析场与模拟结果的整体相似度最高，其中归一化均方根误差为 0.94，结构失真度为 0.95，WaveSim 为 0.84。与 CMCC-CM2-HR4 模式不同，其幅度相似度在所有尺度上都保持较高水平（$0.90-0.96$），结构分量的表现也始终出色（$0.94-0.96$）。这表明，CESM2-FV2 对 NCEP-DOE 再分析场的能量分布和结构组织的再现相似度更高。事实上，图3中 CESM2-FV2 与再分析场的等值线在视觉上非常相似，但在约 $5,200-5{,}000$ 米高度处（左列和中列地图中的绿色至黄色区域）存在一些差异。位移分量也显示出较高的一致性（0.93–0.97）。同样在这种情况下，大尺度特征也得到了较好的捕捉。

最后，GFDL-ESM4 模式（第三行）表现出与 CESM2-FV2 相似的特征，其总体相似度分数（NRMSE $=0.90$、结构失真度 $=0.94$、WaveSim $=0.83$）高于 CMCC-CM2-HR4，但略低于 CESM2-FV2。分解结果表明，与 CESM2-FV2 相比，最粗尺度上的幅度和结构相似性均有所下降（分别为 0.83 和 0.89），这表明该尺度存在更显著的大尺度幅度偏差和结构偏差。相比之下，位移值始终较高（0.94–0.99），与其他地球系统模式相当，这进一步说明，主要差异源于场的幅度不同，而非位置错位。

3.2.2 厄尔尼诺减拉尼娜合成降水异常

作为最后一个应用案例，我们使用 WaveSim 和传统指标来评估厄尔尼诺减去拉尼娜事件的合成降水异常图之间的相似性。我们旨在检验 WaveSim 处理多种背景和情景的灵活性，包括评估正负号相反的异常模式，而非原始场的情况。

具体而言，我们聚焦于不同地球系统模式得出的厄尔尼诺减去拉尼娜事件，并将其与1979–2014年历史时期的参考数据进行比较。参考数据采用全球降水气候学计划（GPCP）月降水分析场2.3版（Adler 等，2016）和美国国家海洋和大气管理局扩展重建海表温度（SST）第5版（Huang 等，2017），以构建1979–2014年的基准气候态。随后，通过减去相应月份的

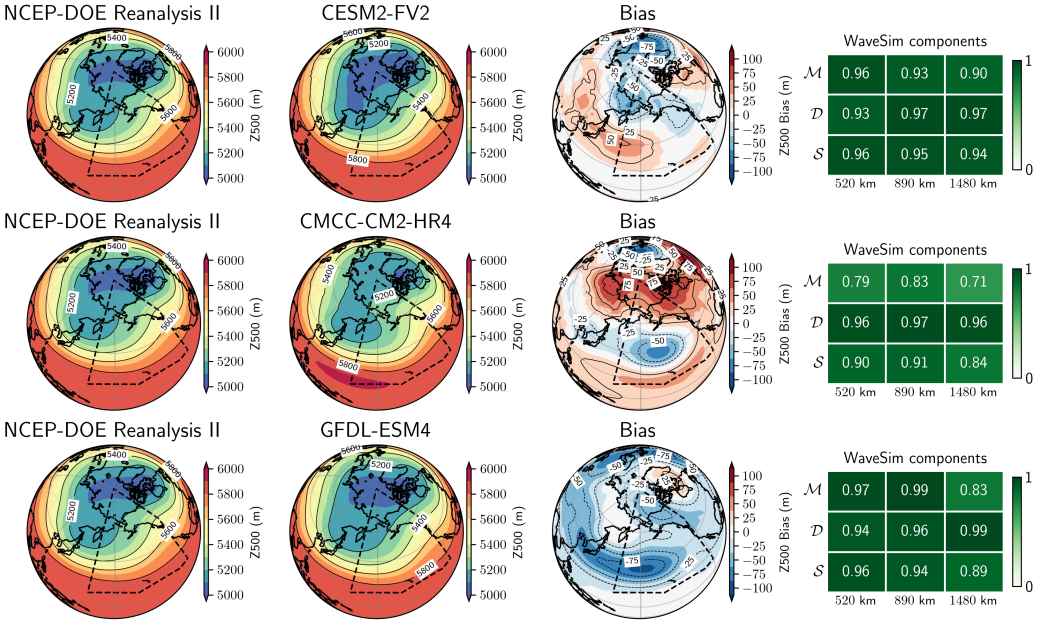

图3. 三个CMIP6地球系统模式在1979–2014年期间的12月至次年2月平均Z500位势高度图与NCEP-DOE再分析II的比较。比较针对图中的虚线区域。偏差图（从右数第二幅）显示了地球系统模式平均地图（从左数第二幅）与相应再分析场（最左侧图）之间的差异，可直观评估历史偏差。Z500值以米为单位，通过等值线表示，并采用正射投影。最右侧图显示了所选空间尺度下WaveSim分量的热图（$\mathcal{M}$：幅度，$\mathcal{D}$：位移，$\mathcal{S}$：结构），便于直观查看不同尺度下的相似度值。相应的相似度分数也列于表2中。

表2. NCEP-DOE再分析II与所选CMIP6地球系统模式的12月至次年2月平均500百帕位势高度图的相似度分数（越高越好）。表中列出了不同尺度下的WaveSim分量值（$\mathcal{M}$：幅度，$\mathcal{D}$：位移，$\mathcal{S}$：结构），尺度从最精细到最粗依次为520、890和1,480千米；分数四舍五入至小数点后两位。

地球系统模型	归一化均方根误差	结构失真度	**WaveSim（本文方法）**	WaveSim分量	
CESM2-FV2	0.94	0.95	0.84	$\mathcal{M}$	[0.96, 0.93, 0.90]
				$\mathcal{D}$	[0.93, 0.97, 0.97]
				$\mathcal{S}$	[0.96, **0.95, 0.94**]
CMCC-CM2-HR4	0.89	0.86	0.66	$\mathcal{M}$	[0.79, 0.83, 0.71]
				$\mathcal{D}$	[0.96, 0.97, 0.96]
				$\mathcal{S}$	[0.90, 0.91, 0.84]
GFDL-ESM4	0.90	0.94	0.83	$\mathcal{M}$	[0.97, 0.99, 0.83]
				$\mathcal{D}$	[0.94, 0.96, 0.99]
				$\mathcal{S}$	[0.96, 0.94, 0.89]

平均值。对于 CMIP6 地球系统模式数据（CESM2-FV2、CMCC-CM2-HR4 和 GFDL-ESM4），也采用相同的方法，相对于各模式在同一时段内各自的气候态计算降水和海表温度异常。此前，我们将具有不同原生空间分辨率的地球系统模式数据重新映射到相同的海表温度（89×180）和降水（72×144）网格，以确保比较的一致性。
观测数据和模式中的 ENSO年份均采用海洋尼诺指数（ONI）一致识别。该指数定义为尼诺3.4区域（$5°S$–$5°N$， $190°$–$240°E$）海表温度异常的三个月滑动平均值。当 ONI 超过 $+0.5°C$ 时，将该年归类为厄尔尼诺年；当 ONI 低于 $-0.5°C$时，则归类为拉尼娜年，阈值均相对于 1979–2014年气候态确定（Trenberth，2025年）。这种阈值设定可确保仅考虑强 ENSO 信号。随后，将厄尔尼诺年和拉尼娜年作为降水异常的替代指标。

对于观测数据和模式，我们分别计算冬季（12月至次年2月）厄尔尼˜诺年和拉尼˜娜年的合成降水距平，随后求得合成降水差值（厄尔尼˜诺减去拉尼˜娜）。图4中的这些合成差值地图为与 WaveSim 及其他指标进行比较提供了依据，并回答了地球系统模式如何表征与 ENSO 相关的降水模式这一问题。计算 WaveSim 时，我们使用 Daubechies-2（$db2$）母小波，将降水参考数据和各地球系统模式的合成场分解为三个尺度，对应的空间分辨率约为550、1,100和2,200千米。其他指标则直接根据合成场地图计算，不进行任何分解。表3报告了三个地球系统模式的厄尔尼˜诺减去拉尼˜娜合成降水距平与 GPCP 分析数据（作为参考数据）之间的相似度分数。需要特别指出的是，此处比较的不是原始降水场，而是厄尔尼˜诺和拉尼˜娜合成场之间的差值。在这种情况下，相似度分数较高表明模式与参考数据的合成降水差值模式相近。三个地球系统模式的归一化均方根误差均处于中等水平（0。48–0。56），说明这些模式的降水差值地图平均振幅相似。然而，这一解读与对图4的视觉检查结果相悖：图4显示降水差值模式在感知上存在明显差异，这表明在此情境下，归一化均方根误差可能高估了相似性。相比之下，结构失真度得分尤其低（0。18–0。22），乍看之下表明这些差值地图的结构组织并不相似。合成测试案例中也观察到了类似现象：当降水模式发生空间偏移时，即使位移相对较小，结构失真度也会降低（例如表1的 b 行和 c 行）。相比之下，WaveSim 通过在不同尺度上分解幅度、位移和结构分量，提供了更细致的分析。

GFDL-ESM4 获得了最高的总体 WaveSim 分数（0.71），反映出其在所有分量和尺度上均取得了中等至较高的分数。CMCC-CM2-HR4（0.60）在较细尺度上的幅值相似性较弱（0.671），但在较粗尺度上具有较强的一致性（0.963），同时表现出较高的结构相似度（0.841–0.923）和中等偏高的位移（$0.780-0.849$）。CESM2-FV2 的总体分数最低（0.428），主要原因是其在最粗尺度上的幅度不匹配（0.555），尽管其位移（0.777–0.854）和结构（0.742–0.880）相对较高。

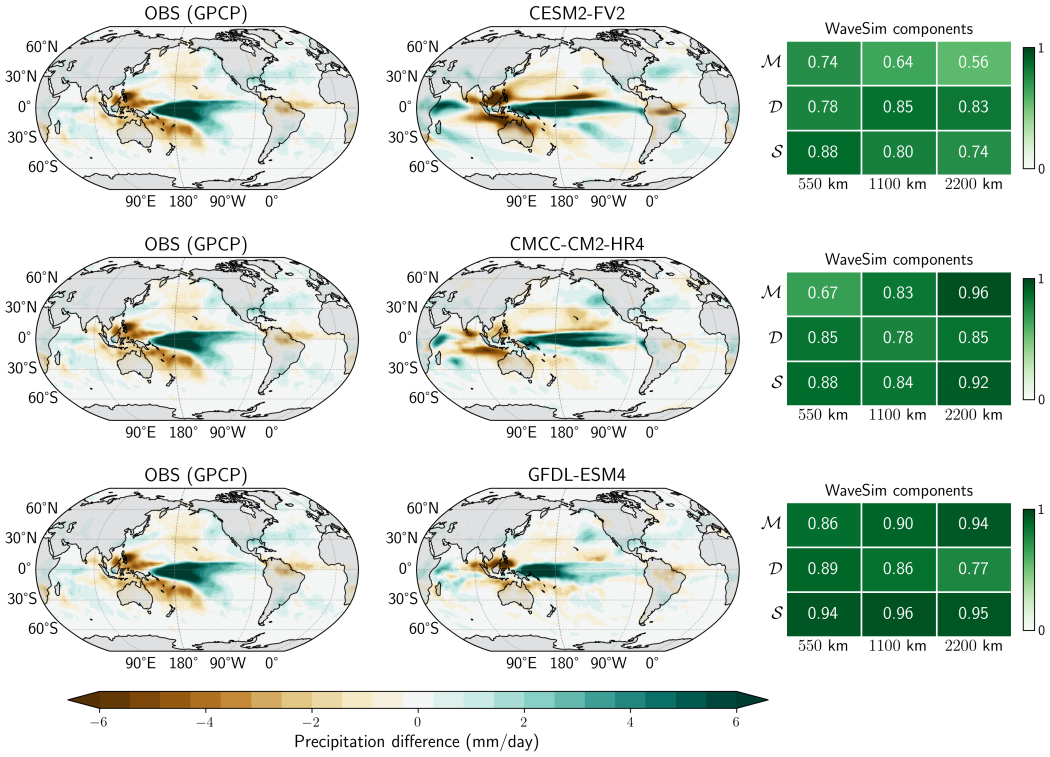

图4. 基于1979年至2014年月气候态，GPCP分析（左图）和选定地球系统模式（中图）在12月至次年2月气候季节的厄尔尼诺减去拉尼娜合成降水异常。色标表示降水差异，单位为毫米/天。右图显示了选定空间尺度上的 WaveSim 分量热图（$\mathcal{M}$：幅度、$\mathcal{D}$：位移、$\mathcal{S}$：结构），可直观查看不同尺度下的相似度值。相应的相似度分数也列于表

3.

4 结论

在本研究中，我们介绍了 WaveSim，这是一种灵活且易于解释的多尺度相似性度量，旨在比较天气与气候地图。

与均方误差等传统逐点误差指标不同，WaveSim 在小波空间中运算，并不基于像素级比较。在 WaveSim 中，地图通过离散小波变换分解到所需数量的空间尺度。在每个尺度上，都会评估相似度的三个正交分量，而整体 WaveSim 分数则是这些分量在各尺度上的加权组合。这种设计能够更丰富、更具诊断性的方式评估相似度，使用户能够辨别差异（因此相似度较低）究竟源于幅度偏差、空间位移还是结构失配。特别需要指出的是，由于 WaveSim 具有乘性特征，即使只有一个分量较低，最终分数也可能相当低，因此务必查看各分量的具体分解情况。

WaveSim is also highly flexible and customizable. Users can adjust component and scale weights to match the requirements of specific applications, and the choice of the mother wavelet can be tailored to the nature of the field and the resolution of data.

比较天气与气候场时，并不存在绝对的相似性“真实值”；我们依靠具体情境和领域知识来判断何种程度的一致性才有意

表3. 历史时期各地球系统模式和 GPCP 分析中，复合厄尔尼诺减去拉尼娜降水异常的相似度分数（越高越好）。异常值相对于 1979 – 2014，并取冬季（12月至次年2月）的平均值。WaveSim分量（$\mathcal{M}$：幅度，$\mathcal{D}$：位移，$\mathcal{S}$：结构）按尺度列出，顺序从最精细到最粗（∼ 550、1,100 和 2,200 km），分数四舍五入至小数点后两位。

地球系统模型归一化均方根误差		结构失真度	**WaveSim（本文方法）**	WaveSim分量	
CESM2-FV2	0.48	0.22	0.43	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.74, 0.64, 0.56] [0.78, 0.85, 0.83] [0.88, **0.80, 0.74**]
CMCC-CM2-HR4	0.55	0.18	0.60	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.67, 0.83, 0.96] [0.85, 0.78, 0.85] [0.88, **0.84, 0.92**]
GFDL-ESM4	0.56	0.21	0.71	$\mathcal{M}$ $\mathcal{D}$ $\mathcal{S}$	[0.86, 0.90, 0.94] [0.89, 0.86, 0.77] [0.94, 0.96, **0.95**]

义。因此，我们使用合成测试案例对 WaveSim 进行了校准。我们以一张表示降水模式的参考地图为基础，生成了一组受控的空间和时间扰动。通过这一过程，我们验证了各分量彼此区分良好，且总体表现符合预期。例如，在表1所示的“对角翻转”案例中，只有位移分量受到影响，这符合预期，因为场的幅度和结构均未改变。同样，在考察预报中不断增加的预报时效时，也就是该表中的扰动 g-i，所有相似性度量都会随时间降低，这与预期一致。

将 WaveSim 应用于现实世界的地球系统模型输出时，它提供了其他所考察指标无法给出的可解释洞见。例如，在 PNA偏差案例研究中，我们展示了 CMCC-CM2-HR4 的 Z500等值线相对于再分析场的视觉差异，是由目标区域中心的冷偏差（∼ -75 米）以及部分由顶部的暖偏差（∼ +75 米）造成的（见图 3 第二行）。WaveSim 为这些地图给出的总体相似度分数为 0.66，并揭示主要差异源于幅度分量，尤其是在最粗尺度（0.71）上，其次则是结构分量。此解读与视觉感知到的相似性相当吻合。相比之下，NRMSE 和结构失真度为这些地图给出了更高的分数（分别为 0.89 和 0.86），却未能进一步说明差异归因。此外，厄尔尼诺减去拉尼娜的合成分析展示了 WaveSim 直接比较异常图的能力。在这种情况下，异常图之间的高相似度表明各模式具有“相似”的异常模式。

总之，WaveSim 代表了天气和气候应用中空间场评估的一项重大进展。展望未来，我们认为，WaveSim 可以为研究天气和气候模型的数据驱动模拟器所用的新型损失函数提供一个支撑框架。众所周知，基于神经网络的模型往往会抹去小尺度变率，在通常超过一周的时间尺度上进行滚动预报时，生成的估计会越来越模糊。这种退化主要是由于用于向前生成预报的自回归设置，以及使用均值损失函数（如均方误差）所致；这类损失函数以误差分布的均值为优化目标，因而会丢失精细细节和高频分量。这类损失函数

可以明确考虑多个尺度上的信息，以及位移等其他分量；这些内容通常会被基于逐点比较的传统损失函数忽略。WaveSim 还可用于评估数据驱动模型和数值气候和天气模型的性能，也可用于比较模型与数据（例如观测数据或再分析），并将相似性归因于特定尺度和分量。此外，它还能通过将模型与数据不匹配归因于特定空间尺度和分量，助力模型校准，并为代理模型和仿真器框架中的参数调优提供指导（Watson-Parris 等，2021年；Elsaesser 等，2025年）。这将有助于更有针对性地改进模型，并深入了解模型成功或失败的原因及其所在环节。我们公开提供 WaveSim 的开源实现及全部评估代码，地址为：`https://github.com/gabrieleaccarino/wavesim`。

开放研究部分

通过 LEAP-Pangeo平台获取了耦合模式比较计划第六阶段（CMIP6）地球系统模式模拟：`https://leap-stc.github.io/intro.html`。

选取了 NCEP-DOE第二版再分析数据集中的月平均500百帕位势高度（Z500）地图：`https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html`。

月降水数据来自全球降水气候学项目（GPCP）2.3版（Adler 等，2016），此外还从 NOAA PSL档案库获取了美国国家海洋和大气管理局扩展重建海表温度（ERSST）第5版数据集（Huang 等，2017）：分别为 `https://psl.noaa.gov/data/gridded/data.gpcp.html` 和 `https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html`。

我们还在 GitHub 仓库中提供了用于数据获取和预处理的 Jupyter笔记本，从而能够完整复现本文展示的整个流程和所有图表。

致谢

我们感谢美国国家科学基金会通过学习地球与人工智能和物理学（LEAP）科学与技术中心（STC）（奖项编号：2019625）提供资助。VA 感谢 PSC-CUNY 第 55 轮资助（奖项编号：67628）、西蒙斯基金会提供的 PIVOT奖学金（奖项编号：981849）和 PIVOT 研究奖（奖项编号：12871）所提供的支撑。

参考数据

Adler, R., Wang, J.-J., Sapiano, M., Huffman, G., Chiu, L., Xie, P. P., . . . Program, N. C. (2016). 全球降水气候学计划（*GPCP*）气候数据记录（*CDR*），*2.3*版（月度）。National Centers for Environmental Information. 获取自 `https://doi.org/10.7289/V56971M6` （访问日期：[access date]）doi: 10.7289/V56971M6 Ahijevych, D., Gilleland, E., Brown, B. G., & Ebert, E. E. (2009). 将空间检验方法应用于理想化降水预报和数值天气预报（NWP）网格降水预报。《天气与预报》，*24*(6), 1485 - 1497. 获取自 `https://journals.ametsoc.org/view/journals/wefo/24/6/2009waf2222298_1.xml`doi: 10.1175/2009WAF2222298.1 Baker, A. H., Pinard, A., & Hammerling, D. M. (2023). 关于浮点数据的结构相似性指数方法。《*IEEE*可视化与计算机图形学汇刊》，*30*(9), 6261–6274.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). 使用三维神经网络实现精准的中期全球天气预报。《自然》， 619 (7970), 533–538. Briggs, W. M., & Levine, R. A. (1997). 小波与场预报检验。《月度天气评论》， 125 (6), 1329–1341. Brown, B. G., Gilleland, E., & Ebert, E. E. (2011). 空间场预报。载于《预报检验》（第95-117页）。John Wiley & Sons, Ltd. 获取自 https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119960003.ch6doi: https://doi.org/10.1002/9781119960003.ch6 Buschow, S., & Friederichs, P. (2020). 使用小波检验降水预报的尺度结构。《统计气候学、气象学与海洋学进展》, 6 (1), 13–30. 获取自 https://ascmo.copernicus.org/articles/6/13/2020/ doi: 10.5194/ascmo-6-13-2020 Buschow, S., & Friederichs, P. (2021). Sad：检验降水预报的尺度、各向异性和方向。《英国皇家气象学会季刊》， 147 (735), 1150-1169. 获取自 https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3964 doi: https://doi.org/10.1002/qj.3964 Buschow, S., Pidstrigach, J., & Friederichs, P. (2019). 通过随机降水模型评估基于小波的空间检验（wv verif v0.1.0）。《地球科学模型开发》， 12 (8), 3401–3418. 获取自 https://gmd.copernicus.org/articles/12/3401/2019/ doi: 10.5194/gmd-12-3401-2019 Casati, B. (2010). 空间检验方法比较项目中的强度-尺度技术新进展。《天气与预报》， 25 (1), 113 - 143. 获取自 https://journals.ametsoc.org/view/journals/wefo/25/1/2009waf2222257_1.xml doi: 10.1175/2009WAF2222257.1 Casati, B., Lussana, C., & Crespi, A. (2023). 用于比较降水再分析资料的尺度分离诊断与对称有界效率。《国际气候学杂志》， 43 (5), 2287-2304. 获取自https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.7975 doi: https://doi.org/10.1002/joc.7975 Casati, B., Ross, G., & Stephenson, D. B. (2004). 一种用于空间降水预报检验的新型强度-尺度方法。Meteorological Applications， 11 (2), 141-154. 获取自 https://rmets.onlinelibrary.wiley.com/doi/abs/10.1017/S1350482704001239 doi: https://doi.org/10.1017/S1350482704001239 Casati, B., & Wilson, L. J. (2007). 布里尔评分的一种新型空间尺度分解：在雷电概率预报检验中的应用。《月度天气评论》， 135 (9), 3052 - 3069. 获取自 https://journals.ametsoc.org/view/journals/mwre/135/9/mwr3442.1.xml doi: 10.1175/MWR3442.1 Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., & Li, H. (2023). 伏羲：用于15天全球天气预报的级联机器学习预报系统。《npj气候与大气科学》， 6 (1), 190. Danabasoglu, G. (2019). 为CMIP6 CMIP历史试验准备的NCAR CESM2-FV2模式输出。地球系统网格联合会。获取自 https://doi.org/10.22033/ESGF/CMIP6.11297 doi: 10.22033/ESGF/CMIP6.11297 Daubechies, I. (1992).《小波十讲》。SIAM。Dorninger, M., Gilleland, E., Casati, B., Mittermaier, M. P., Ebert, E. E., Brown, B. G., & Wilson, L. J. (2018). mesoVICT项目的设置。《美国气象学会公报》， 99 (9), 1887 - 1906. 获取自 https://journals.ametsoc.org/view/journals/bams/99/9/bams-d-17-0164.1.xml doi: 10.1175/BAMS-D-17-0164.1

Ebert, E. E.（2008年）。高分辨率网格预报的模糊检验：综述与拟议框架。*Meteorological Applications*，*15*(1)，51-64。取自 https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.25doi: https://doi.org/10.1002/met.25 Elsaesser, G. S.、van Lier-Walqui, M.、Yang, Q.、Kelley, M.、Ackerman, A. S.、Fridlind, A. M.、……Strong, J. D. O.（2025年）。利用机器学习生成 GISS ModelE 校准物理集合（CPE）。《地球系统建模进展》，*17*(4)，e2024MS004713。取自 https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024MS004713 (e2024MS004713 2024MS004713) doi: https://doi.org/10.1029/ 2024MS004713 Eyring, V.、Bock, L.、Lauer, A.、Righi, M.、Schlund, M.、Andela, B.、…… Zimmermann, K.（2020年）。地球系统模式评估工具（ESMValTool）第2.0版——用于准业务化和全面评估 CMIP 中地球系统模式的一组扩展大尺度诊断。《地球科学模型开发》，*13*(7)，3383–3438。取自 https://gmd.copernicus.org/articles/13/3383/2020/ doi: 10.5194/gmd-13-3383-2020 Eyring, V.、Bony, S.、Meehl, G. A.、Senior, C. A.、Stevens, B.、Stouffer, R. J. 和 Taylor, K. E.（2016年）。第六阶段耦合模式比较计划（CMIP6）实验设计与组织概述。《地球科学模型开发》，*9*(5)，1937–1958。Gilleland, E.、Ahijevych, D.、Brown, B. G.、Casati, B. 和 Ebert, E. E.（2009年）。空间预报检验方法的比较。《天气与预报》，*24*(5)，1416 - 1430。取自 https://journals.ametsoc.org/view/journals/wefo/24/5/2009waf2222269 1.xml doi: 10.1175/2009WAF2222269 .1 Gilleland, E.、Ahijevych, D. A.、Brown, B. G. 和 Ebert, E. E.（2010年）。对预报进行空间检验。*Bulletin of the American Meteorological Society*，*91*(10)，1365 - 1376。取自 https://journals.ametsoc.org/view/journals/bams/91/10/2010bams2819 1.xml doi: 10.1175/2010BAMS2819.1 Ha, W.、Singh, C.、Lanusse, F.、Upadhyayula, S. 和 Yu, B.（2021年）。通过解释实现神经网络的自适应小波蒸馏。载于 M. Ranzato、A. Beygelzimer、Y. Dauphin、P. Liang 和 J. W. Vaughan（编），神经信息处理系统进展（第34卷，第20669–20682页）。Curran Associates, Inc. 取自 https://proceedings.neurips.cc/paper files/paper/2021/file/acaa23f71f963e96c8847585e71352d6-Paper.pdfHuang, B.、Thorne, P. W.、Banzon, V. F.、Boyer, T.、Chepurin, G.、Lawrimore, J. H.、…… Zhang, H.-M.（2017年）。扩展重建海表温度，第5版（ERSSTv5）：升级、验证与比较。*Journal of Climate*，*30*(20)，8179 - 8205。取自 https://journals.ametsoc.org/view/journals/clim/30/20/jcli-d-16-0836.1.xmldoi: 10.1175/JCLI-D-16-0836.1 Jung, T. 和 Leutbecher, M.（2008年）。集合预报的尺度依赖检验。《英国皇家气象学会季刊》，*134*(633)，973-984。取自 https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.255 doi: https://doi.org/10.1002/qj.255 Kanamitsu, M.、Ebisuzaki, W.、Woollen, J.、Yang, S.-K.、Hnilo, J. J.、Fiorino, M. 和 Potter, G. L.（2002年）。NCEP–DOE AMIP-II再分析（R-2）。*Bulletin of the American Meteorological Society*，*83*(11)，1631 - 1644。取自 https://journals.ametsoc.org/view/journals/bams/83/11/bams-83-11-1631.xmldoi: 10.1175/BAMS-83-11-1631 Kapp, F.、Friederichs, P.、Brune, S. 和 Weniger, M.（2018年，12月）。利用局地小波谱对高分辨率集合降水预报进行空间检验。*Meteorologische Zeitschrift*，*27*(6)，467-480。取自 http://dx.doi.org/10.1127/metz/2018/0903 doi: 10.1127/metz/2018/0903

Kochkov, D.、Yuval, J.、Langmore, I.、Norgaard, P.、Smith, J.、Mooers, G.、……其他作者（2024年）。面向天气和气候的神经环流模型。《自然》，632 (8027)，1060–1066。Krasting, J. P.、John, J. G.、Blanton, C.、McHugh, C.、Nikonov, S.、Radhakrishnan, A.、……Zhao, M.（2018年）。为 CMIP6 CMIP 准备的 NOAA-GFDL GFDL-ESM4 模式输出。地球系统网格联合会。取自 https://doi.org/10.22033/ESGF/CMIP6.1407 doi: 10.22033/ESGF/CMIP6.1407 Lam, R.、Sanchez-Gonzalez, A.、Willson, M.、Wirnsberger, P.、Fortunato, M.、Alet, F.、……Battaglia, P.（2023年）。学习高技巧的中期全球天气预报。《科学》，382 (6677)，1416-1421。取自 https://www.science.org/doi/abs/10.1126/science.adi2336 doi: 10.1126/science.adi2336 Mallat, S. G.（1989年）。l^2（r）的多分辨率逼近与小波正交归一基。《美国数学会汇刊》，315 (1)，69–87。Mallat, S. G.（1999年）。信号处理小波导论。爱思唯尔。Men´endez, M.、Pardo, J.、Pardo, L. 和 Pardo, M.（1997年）。詹森-香农散度。《富兰克林研究所杂志》，334 (2)，307-318。取自 https://www.sciencedirect.com/science/article/pii/S0016003296000634 doi: https://doi.org/10.1016/S0016-0032(96)00063-4 Meyer, Y.（1992年）。小波与算子（第37号）。剑桥大学出版社。Ojima, D. 等（1992年）。地球系统建模。美国科罗拉多州博尔德：UCAR/跨学科地球研究办公室。Rasp, S.、Hoyer, S.、Merose, A.、Langmore, I.、Battaglia, P.、Russell, T.、……其他作者（2024年）。WeatherBench 2：面向下一代数据驱动全球天气模型的基准测试。《地球系统建模进展》，16 (6)，e2023MS004019。Scoccimarro, E.、Bellucci, A. 和 Peano, D.（2017年）。为 CMIP6 HighResMIP 准备的 CMCC CMCC-CM2-HR4 模式输出。地球系统网格联合会。取自https://doi.org/10.22033/ESGF/CMIP6.1359 doi: 10.22033/ESGF/CMIP6 .1359 Shlens, J.（2014年）。关于 Kullback-Leibler散度与似然的笔记。取自https://arxiv.org/abs/1404.2000Skok, G.（2022年）。用于降水检验的新型空间距离度量。《Applied Sciences》，12 (8)。取自 https://www.mdpi.com/2076-3417/12/8/4048 doi: 10.3390/app12084048 Skok, G.（2023年）。降水归因距离。《大气研究》，295，106998。取自 https://www.sciencedirect.com/science/article/pii/S0169809523003952 doi: https://doi.org/10.1016/j.atmosres .2023.106998 Skok, G. 和 Kosovelj, K.（2025年）。全球场的平滑与空间检验。取自 https://arxiv.org/abs/2412.00936Skok, G. 和 Lled´o, L.（2025年）。全球降水预报的空间检验。《英国皇家气象学会季刊》，n/a(n/a)，e5006。取自 https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.5006 doi: https://doi.org/10.1002/qj.5006 Skok, G. 和 Roberts, N.（2016年）。随机降水场和 ECMWF 预报的分数技巧评分属性分析。《英国皇家气象学会季刊》，142 (700)，2599-2610。取自 https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2849 doi: https://doi.org/10.1002/qj.2849 Skok, G. 和 Roberts, N.（2018年）。利用分数技巧评分估计降水预报中的位移。《英国皇家气象学会季刊》，144 (711)，414-425。取自https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3212doi: https://doi.org/10.1002/qj.3212

Tebaldi, C.、Debeire, K.、Eyring, V.、Fischer, E.、Fyfe, J.、Friedlingstein, P.、…… 等（2021年）。CMIP6 情景模式比较计划（ScenarioMIP）的气候模式预测。地球系统动力学，*12*(1)，253–293。Trenberth, K.（2025年）。气候数据指南：尼诺海表温度指数（尼诺 *1+2*、*3*、*3.4*、*4*；*ONI* 和 *TNI*）。 https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni。（由美国国家大气研究中心工作人员编辑。最后修改于 2025-04-29。访问于 2025-09-29）Wang, Z.、Bovik, A. C.、Sheikh, H. R. 和 Simoncelli, E. P.（2004年）。图像质量评估：从误差可见性到结构相似度。*IEEE*图像处理汇刊，*13*(4)，600–612。Watson-Parris, D.、Rao, Y.、Olivi´e, D.、Seland, Ø.、Nowack, P.、Camps-Valls, G.、……等（2022年）。ClimateBench v1.0：数据驱动气候模式预测的基准测试。《地球系统建模进展》，*14*(10)，e2021MS002954。 Watson-Parris, D.、Williams, A.、Deaconu, L. 和 Stier, P.（2021年）。使用 ESEm v1.1.0 进行模型校准——一种开放、可扩展的地球系统模拟器。《地球科学模型开发》，*14*(12)，7659–7672。检索自 https://gmd.copernicus.org/articles/14/7659/2021/ doi: 10.5194/gmd-14-7659-2021 Weniger, M.、Kapp, F. 和 Friederichs, P.（2017年）。使用小波变换进行空间检验：综述。《英国皇家气象学会季刊》，*143*(702)，120-136。检索自 https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2881 doi: https://doi.org/10.1002/qj.2881