arXiv:2602.02096v1 [physics.ao-ph] 2 Feb 2026

WADEPre：一种结合多尺度学习的极端降水临近预报基于小波的分解模型

刘百田
sonderliu@hdu.edu.cn 杭州电子科技大学 中国浙江杭州

张海平
zhanghp@hdu.edu.cn 杭州电子科技大学 中国浙江杭州

Huiling Yuan
yuanhl@nju.edu.cn 南京大学 中国江苏南京

王冬景
dongjing.wang@hdu.edu.cn 杭州电子科技大学 中国浙江杭州

李莹 陈峰
ying.li@mail.bnu.edu.cn
chenfeng@mail.iap.ac.cn 浙江省气象科学研究所 中国浙江杭州

Hao Wu*
wuhao@hdu.edu.cn 杭州电子科技大学 中国浙江杭州

摘要

降水强度的重尾特性阻碍了精确的降水临近预报。采用逐像素损失进行优化的标准模型容易产生向均值回归偏差，从而模糊极端值。现有的基于傅里叶的方法也缺乏解析瞬态对流单体所需的空间定位能力。为克服这些固有局限，我们提出WADEPre，这是一种针对极端降水的基于小波的分解模型，可将建模过程转移至小波域。通过利用离散小波变换进行显式分解，WADEPre 采用双分支架构：近似网络用于建模稳定的低频平流，将确定性趋势与统计偏差分离；空间定位的细节网络则用于捕获高频随机对流，从而解析瞬态奇异结构并保留锐利边界。随后，细化模块动态地将这些解耦的多尺度分量重构为最终的高保真预报。为解决优化不稳定性，我们引入多尺度课程学习策略，逐步将监督从粗尺度转向精细细节。在 SEVIR 和上海雷达数据集上的大量实验表明，WADEPre 达到了最先进的性能，在捕捉极端阈值和保持结构保真度方面均取得显著提升。代码可在 https://github.com/sonderlau/WADEPre 获取。

CCS 概念

• **应用计算** → **地球与大气科学**； •**信息系统** → 空间-时间系统； • **计算方法** → 神经网络。

关键词

降水临近预报、物理感知机器学习、极端降水、小波变换、多尺度学习

*通讯作者。

1 引言

强对流风暴具有暴雨、冰雹和破坏性大风等特征，对公共安全和经济稳定构成重大威胁 [22, 27]。准确的高分辨率降水临近预报对于灾害减缓和城市水文管理至关重要 [3, 23, 28]。然而，数值天气预报（NWP）存在“启动”延迟和高昂的计算成本 [27]，而光流方法（例如 PySTEPS [21]）往往无法捕捉非线性演变。因此，由于降水呈重尾分布，可靠地预测极端事件仍是一项艰巨挑战。这导致了一个根本性的优化困境：标准的逐像素目标函数（例如均方误差）在统计上受大量低强度样本主导 [42]。因此，模型倾向于回归均值，系统性地模糊掉罕见但影响巨大的极端事件 [31]。重尾分布造成的这一根本性优化陷阱 在图1（面板 a）中有所展示。

为了解决逐像素回归的局限性，该领域已转向频谱域建模。近期方法引入了频谱损失（例如，基于快速傅里叶变换的目标函数 [37]），以捕捉全局模式。然而，一个关键的不匹配问题依然存在：这些模型仍主要在像素空间中运行，迫使网络隐式地学习从空间输入到频谱目标的复杂映射，并且在训练过程中往往会平均掉高频纹理 [2]。此外，尽管傅里叶域模型（例如 AlphaPre [17]）能够显式地对频率建模，但它们从根本上受到海森堡不确定性原理 [19]的限制。傅里叶基在频率上具有全局局部性，却在空间上非局部，因此本质上不适合解析瞬态、空间局部化的对流单体，而这对于极端天气临近预报至关重要。如图1（b栏）所示，现有范式无法清晰捕捉极端事件：逐像素方法会模糊细节，而傅里叶变换则会引入重影伪影。

因此，现有的极端降水临近预报受到三项科学挑战的制约：

（1）**C1：重尾分布中的回归偏差**。低强度背景噪声主导了来自梯度的

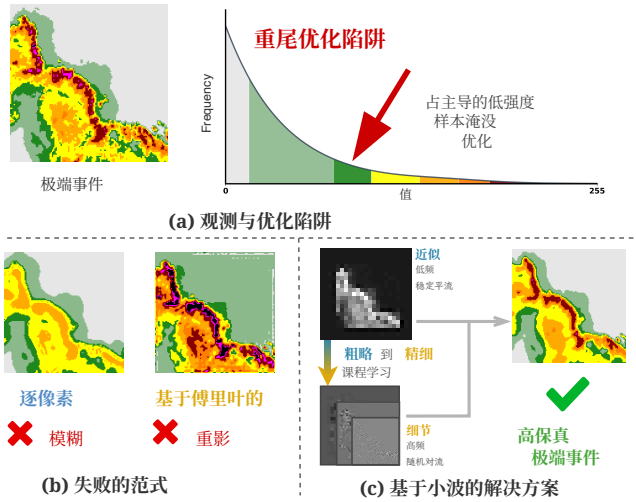


图1：打破预测困境。(a) 重尾分布会造成优化陷阱，使模型忽视罕见极端事件。(b) 现有范式存在不足：逐像素方法会模糊细节，而傅里叶方法会引入重影伪影。(c) WADEPre 通过小波分解解决这一问题，实现对极端事件的清晰、准确预测。

标准优化中的极端事件，导致高强度核心 [35] 持续衰减并变得模糊。(2) **C2：频谱域中的空间定位失效**。要捕捉对流单体的锐利边界，需要一种同时具备时间局域性和频率局域性的表示 [9]。傅里叶变换无法保留这种空间局域性，从而导致重影伪影和结构不连贯。

(3) **C3：频谱—物理不一致性与优化不稳定性**。独立建模各个频带会造成“频谱壁垒”，往往导致重建场缺乏物理一致性 [37]。此外，在没有引导的情况下直接学习高频波动并不稳定，会导致优化发散 [42]。

为解决这些固有局限性，我们提出了**WADEPre**（基于小波的极端降水临近预报分解模型），如图1（面板c）所示。该物理感知框架将建模过程转移至小波域。为应对上述三项挑战（C1-C3），我们采用了三项具体设计：**(1)用于确定性平流的近似网络（A-Net）（解决C1）：**我们不再采用单一回归，而是使用离散小波变换（DWT）分离稳定的低频分量。我们提出近似网络来显式建模确定性运动趋势，避免重尾极端值影响背景优化，并解决向均值回归偏差。**(2)用于随机对流的细节网络（D-Net）（解决C2）：**为解决傅里叶方法的非局部性问题，我们设计了细节网络，使其作用于高频小波系数。借助小波的多尺度空间局部性，D-Net能够显式捕捉瞬态对流单体和锐利梯度，从而保留结构保真度并消除重影伪影。**(3)结合课程学习的物理感知细化器（解决C3）：**为跨越频谱壁垒并稳定训练，我们提出了细化器模块，并将其与多尺度课程学习策略相结合。细化器协调解耦分支，以确保物理一致性；课程策略则逐步将监督从粗尺度转向精细纹理，从而确保稳健收敛。

本文的主要贡献总结如下：

(1) 双域瓶颈的识别：我们将重尾分布造成的回归偏差（导致模糊）以及傅里叶方法中的空间定位失效（导致重影）识别为制约极端临近预报的双重瓶颈。

(2) 基于小波的解耦模型：我们提出了**WADEPre**，将降水明确分解为稳定的确定性平流和瞬态随机对流。细化器对这些解耦的流进行协调，以确保物理一致性，从而避免背景趋势过度平滑高频极端值。**(3) 稳定的多尺度课程优化：**我们引入了一种由粗到细的课程学习策略。借助小波的层次特性，我们逐步引入高频监督，从而解决极端值预测中固有的优化不稳定性。**(4) 极端事件预报的最先进水平表现：**在 SEVIR 和上海雷达数据集上进行的大量实验表明，WADEPre 达到了新的最先进水平；尤其是在高阈值下显著提升了临界成功指数（CSI），同时保持了出色的结构保真度（SSIM）。

2 相关工作

2.1 数据驱动气象预报

深度学习改变了天气预报，使其从数值积分转向数据驱动的模式识别 [1, 6, 15]。在降水临近预报领域，早期研究将这一任务视为时空序列预测。循环神经网络（RNN）开创了这一领域，而 ConvLSTM [25] 和 PredRNN [32, 33]引入了卷积结构，以捕捉时空相关性。近年来，为了降低循环结构带来的计算开销，SimVP [12]、TAU [29], 和 MAU [4] 等简化的编码器-解码器架构相继出现，在提升推理效率的同时实现了具有竞争力的性能。

尽管架构取得了这些进步，这些确定性模型本质上仍在优化均方误差（MSE）等逐像素目标函数。从统计学角度来看，最小化 MSE 假定目标分布为单峰高斯分布，这并不适合降水的混沌、多模态特性 [20, 35]。因此，为了降低误差，这些模型往往会回归到各种可能未来状态的模糊平均值，导致高频极端事件受到严重抑制 [9, 38]。

2.2 频域预报

为解决局部卷积感受野的局限性，研究人员日益采用频域建模。傅里叶神经算子（FNO） [16]率先通过在谱域求解偏微分方程（PDEs），开创了这一方法。在此基础上，AlphaPre [17] 采用基于快速傅里叶变换的解耦来建模全局演变。最近，WaveC2R [24]展示了小波在稳健卫星雷达融合中的有效性，表明人们日益认同多分辨率谱建模对于气象边界的价值。

然而，傅里叶方法受限于其全局基函数；单个系数扰动会影响整个区域 [8], , 因此不适合处理尖锐对流边界等局部奇异性。这会导致吉布斯现象 [13]和空间泄漏，这是由海森堡不确定性原理 [19]所决定的一种权衡，该原理使空间域和频域无法同时达到高分辨率。

2.3 物理信息与生成式学习

连接深度学习与大气物理已成为一个关键的研究前沿。PhyDNet [14]和 EarthFarseer [36], 等混合框架将偏微分方程嵌入循环单元，以模拟物理动力学。近期研究 [7] 进一步探索将数值求解器与神经网络耦合，以增强鲁棒性。与此同时，为解决回归均值模糊问题，生成式建模日益受到重视。DGMR [22] 利用生成对抗网络（GAN）使预测结果更加清晰，而后续的基于扩散的模型，如 NowcastNet [41],、PreDiff [10],和 Dif-FCast [39] , 通过迭代精炼高斯噪声，实现了更出色的纹理合成。

实际部署面临精度—延迟权衡 [10]。采用软正则化的物理信息方法可能产生不一致状态 [41]。生成模型也带来挑战：扩散技术需要耗费大量资源进行采样 [39], , 而生成对抗网络则存在不稳定和模式崩溃问题 [22], , 从而降低了可靠性。

3 方法

3.1 预备知识：小波分解

二维离散小波变换（离散小波变换，DWT）递归地将输入信号 A^{l-1}（初始为 $A^0 = X$）分解为多分辨率频率子带，使用低通（L）和高通（H）滤波器 [19]。该变换沿空间维度应用可分离一维卷积，随后进行二进下采样（$\downarrow 2$），得到粗略近似 A^l 和三个细节分量 $\{D_h^l,\ D_v^l,\ D_d^l\}$，它们捕捉第 l 级的水平、垂直和对角纹理：

$$\begin{cases} A^l = (L \otimes L)A^{l-1} \downarrow 2 \\ D_h^l = (L \otimes H)A^{l-1} \downarrow 2 \\ D_v^l = (H \otimes L)A^{l-1} \downarrow 2 \\ D_d^l = (H \otimes H)A^{l-1} \downarrow 2 \end{cases} \tag{1}$$

其中 $\otimes$ 表示可分离卷积。这种层次化表示明确地将大尺度风暴骨架（由 A^l捕获）与精细的强度波动（由细节系数分离）分解开来。

3.2 问题定义

遵循以往研究 [11, 17, 39],，我们将降水临近预报表述为一个序列到序列预测问题。给定一系列雷达观测 $X = \{X_{t-T_{\text{in}}+1},\ \ldots\ldots,\ X_t\} \in \mathbb{R}^{T_{\text{in}}\times C\times H\times W}$，其中 T_{in} 表示输入时段，C表示通道数（雷达回波通常为 $C = 1$），H，W表示空间分辨率，目标是学习一个映射函数 f_θ，其可学习参数为 θ，用于预测随后 T_{out} 个时间步的未来序列 $Y = \{Y_{t+1},\ \ldots\ldots,\ Y_{t+T_{\text{out}}}\} \in \mathbb{R}^{T_{\text{out}}\times C\times H\times W}$。这一过程可形式化表示为 $Y = f_\theta(X)$。

3.3 所提出的 WADEPreModel

我们将极端降水临近预报形式化为小波域中的时空序列预测任务。令 X_{seq} 表示输入雷达序列。我们采用离散小波变换将其分解为一个低频近似分量（A_{seq}）和一组分层的高频细节分量（D_{seq}）。

如图2所示，WADEPre 通过三个专用模块处理这些分量。**近似网络**旨在捕捉确定性全局趋势，确保大尺度平流的稳定性。同时，**细节网络**专注于建模随机局部波动，以保留标准回归中经常丢失的高频极端事件。最后，**细化器** 整合这些多尺度输出，以确保频谱一致性并校正空间对齐伪影。

3.4 近似网络

近似网络（A-Net）捕捉降水场缓慢且确定性的演变，以表征物理云系，如图3所示。给定输入小波近似 A_{seq}，A-Net 预测未来状态 A_{pred}。

为有效建模长期依赖关系并保持空间一致性，我们提出了一种时空膨胀注入架构。其工作流程包含三个阶段：(1) 编码与时间注入，(2) 时空膨胀演化，以及 (3) 解码与平稳纹理重建。

(1) 编码与时间注入。低频系数具有较高的空间冗余和较强的时间连续性。我们首先使用二维卷积编码器将序列映射到高维潜在空间。我们将时间维度 T 视为通道，以压缩时间维度：

$$Z_{\text{enc}} = \text{Conv2d}(A_{\text{seq}}) \in \mathbb{R}^{\text{dim}\times H\times W}. \tag{2}$$

同时，为了显式保留帧间动态特性（例如气团的平移和旋转），我们引入了*TemporalInjector*。该模块采用三维卷积从序列中提取体积时空特征：

$$Z_{\text{inj}} = \text{Conv3d}(A_{\text{seq}}) \in \mathbb{R}^{\text{dim}\times H\times W}. \tag{3}$$

随后，将注入特征 (Z_{inj}) 与编码特征 (Z_{enc}) 相加，并通过 1×1 卷积进行融合，形成初始隐状态 Z_0。

（2）时空空洞演化。核心演化由一系列时空块（STBlocks）驱动。为使网络能够感知大尺度平流模式，同时不损失分辨率，我们将混合过程分解如下：

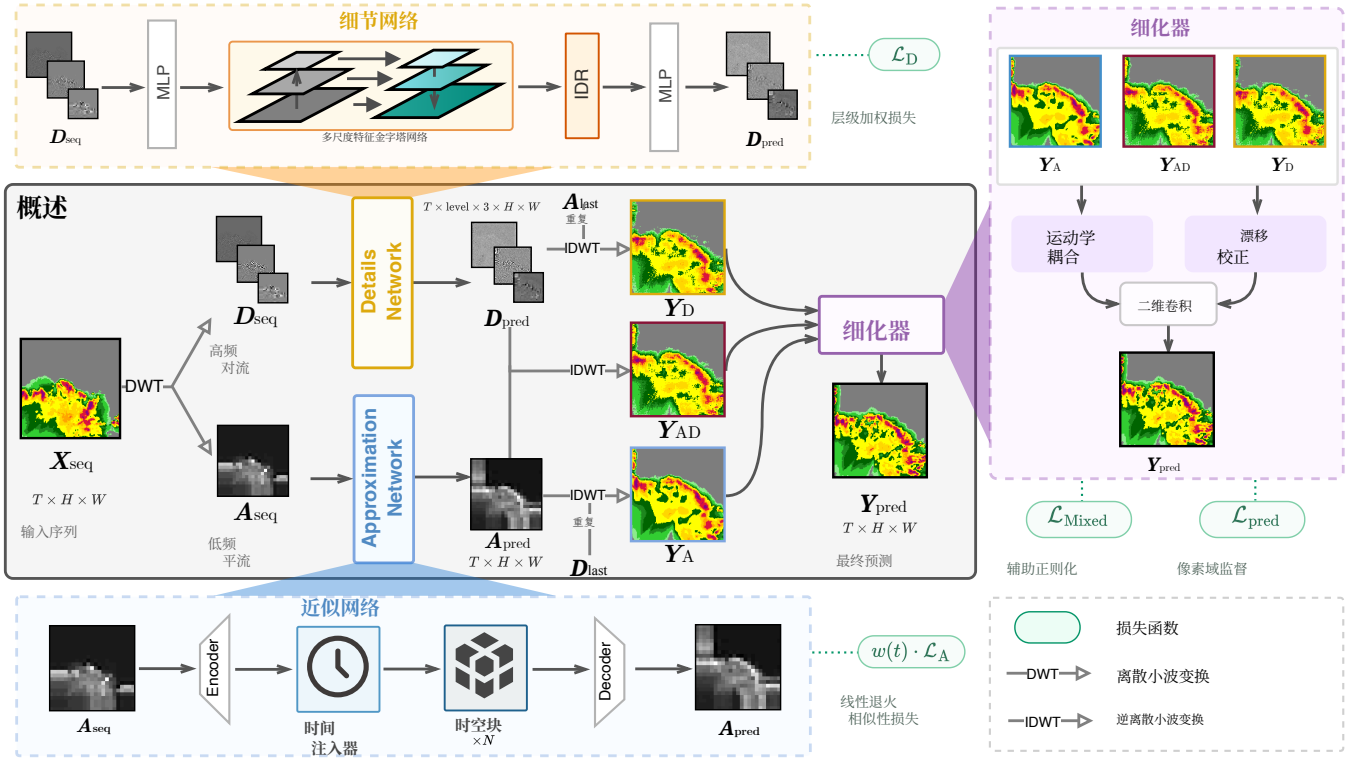


图2：WADEPre 架构示意图。输入序列通过离散小波变换分解为近似（A_{seq}）和细节（D_{seq}）系数。这些分量分别由专用的近似网络（编码器-混合器-解码器）和细节网络（多尺度特征金字塔网络）处理。预测得到的系数通过逆离散小波变换重构，并由细化器融合，以生成最终预测 Y_{pred}。绿色胶囊表示训练期间使用的损失函数。

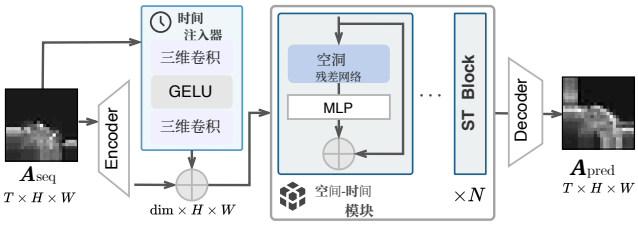


图3：近似网络（A-Net）的架构。旨在对确定性低频平流进行建模。该网络采用时间注入器（通过三维卷积）从输入序列 A_{seq} 中提取帧间动态。核心演化由堆叠的时空块（STBlocks）驱动，这些模块在不损失分辨率的情况下捕捉天气尺度空间依赖关系。

• 空间膨胀混合：我们采用膨胀残差网络 [40] 结构。通过在第 k 个模块（$k \in [0, N]$）中指数级增大膨胀率 $d_k = 2^k$，感受野也随之呈指数级扩展。这使网络能够捕捉全局空间依赖关系，同时保持特征图尺寸：

$$Z_k = \text{DilatedResNet}(Z_{k-1}, d_k). \tag{4}$$

• 时间通道混合：通过 1×1 卷积实现，该模块仅沿通道维度dim运行。由于时间动态编码在这些通道中，多层感知机促进潜在时间步之间的密集交互，以传播演化信息：

$$Z_{k+1} = \text{MLP}(Z_k) + Z_k. \tag{5}$$

该操作在保持空间结构的同时对转移动力学进行建模。

(3) 解码与平稳纹理重建。经过N个演化模块后，解码器将隐藏特征（Z_N）投影回预测时域 T，得到预测的近似系数 $A_{\text{pred}} \in \mathbb{R}^{T \times H \times W}$。

为了重建图像空间背景流 Y_A，我们采用平稳纹理假设。由于高频细节（D 系数）代表随流移动、但统计分布变化缓慢的纹理，因此我们使用输入序列（D_{seq}）中最后一个观测到的细节帧，并在整个预测时域内重复该帧：

$$Y_A = \text{IDWT}\left(A_{\text{pred}}, \text{Repeat}(D_{\text{last}}, T)\right). \tag{6}$$

这一策略确保重建的背景保留逼真的纹理锐度，同时严格遵循由 A_{pred} 定义的预测平流路径。

3.5 细节网络

A-Net 捕捉确定性的背景流动，而细节网络（D-Net）则对降水场中易变的高频成分进行建模，具体而言，是对流单体的快速形成与消散。这些成分对应于多个分解层级上的小波细节系数 $\boldsymbol{D}_{\text{seq}} = \{\boldsymbol{D}^l\}_{l=1}^{\text{level}}$。

为应对在预测随机纹理时避免频谱损失的挑战，我们提出一种分层随机精炼架构，如图4所示。该网络分为四个阶段：(1) 时间投影，(2) 跨尺度空间交互，(3) 迭代细节精炼，以及 (4) 解码与重建。

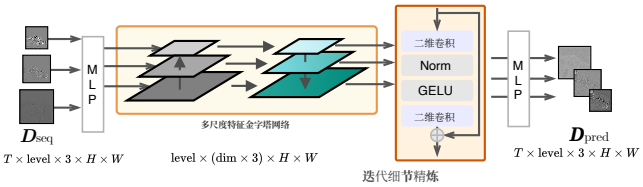


图4：细节网络（D-Net）的架构。该网络旨在解析随机高频对流。网络通过时间多层感知机将细节系数 $\boldsymbol{D}_{\text{seq}}$ 投影到潜在特征空间。以特征金字塔网络（*FPN*）为骨干网络，促进双向跨尺度能量传递；同时，所提出的迭代细节精炼（*IDR*）模块校正频谱不一致性，以在最终预测中保留清晰的局部边界。

(1) 时间投影。与空间上下文占主导的A-Net不同，各层级的细节系数（$\boldsymbol{D}^l_{\text{seq}}$）对短期波动高度敏感。共享的时间多层感知机将输入序列投影到潜在特征空间 $\boldsymbol{V}$：

$$\boldsymbol{V} = \text{MLP}(\boldsymbol{D}^l_{\text{seq}}) \in \mathbb{R}^{\text{level}\times(\dim\times3)\times H_l\times W_l}. \tag{7}$$

该投影将时间演化压缩为通道描述符，为特征进行空间交互做好准备。

（2）跨尺度空间交互。我们采用特征金字塔网络（FPN）[18] 骨干网络，促进不同频率层级之间的信息交换。不同于从单一输入分辨率生成金字塔的标准 FPN，我们采用并行多尺度输入策略，将每个层级的投影细节特征直接注入对应的 FPN 阶段。

注入后，FPN 组织双向信息交换，以协调不同尺度的特征：

- 自底向上聚合：来自更细尺度的高频纹理信息向上传播，以丰富较粗层级的语义表示。
- 自顶向下引导：粗尺度的结构上下文向下传递，以指导细粒度细节的一致演化。

这一设计确保各层的输出特征均可作为 $\boldsymbol{D}^l_{\text{pred}}$ 的前置特征。将其原生层级特有的频率特性与相邻尺度提供的上下文约束结合起来。

（3）迭代细节精炼（IDR）。为进一步提升极端事件的清晰度，我们引入了迭代细节精炼（IDR）模块。如图 4 所示，IDR 作为残差校正模块应用于每个金字塔层级。它由二维卷积-归一化-*GELU*-二维卷积序列组成，旨在校正频谱不一致性：

$$\boldsymbol{V}^l_{\text{IDR}} = \boldsymbol{V}^l_{\text{FPN}} + \text{IDR}(\boldsymbol{V}^l_{\text{FPN}}). \tag{8}$$

其中，$\boldsymbol{V}^l_{\text{FPN}}$ 是 FPN 在层级 l处的输出。通过将细化后的残差加回 FPN 输出，网络能够显式学习恢复下采样过程中丢失的高频边界。

(4) 解码与重建。最后，通过解码多层感知机将细化后的特征（*V*IDR）映射回时间域，得到预测的细节系数 $\boldsymbol{D}_{\text{pred}}$。为重建高频对流场 $\boldsymbol{Y}_D$，我们采用逆离散小波变换。与 A-Net 采用的策略类似，我们使用来自 $\boldsymbol{A}_{\text{seq}}$ 的最后一帧观测近似帧 $\boldsymbol{A}_{\text{last}}$作为静态结构锚点。$\boldsymbol{A}_{\text{last}}$ 会沿预测时域重复，以与 $\boldsymbol{D}_{\text{pred}}$ 的时间维度对齐：

$$\boldsymbol{Y}_D = \text{IDWT}\left(\text{Repeat}(\boldsymbol{A}_{\text{last}},\, T), \boldsymbol{D}_{\text{pred}}\right). \tag{9}$$

这一设计利用 $\boldsymbol{A}_{\text{last}}$ 提供基本空间布局，确保重建后的 $\boldsymbol{Y}_D$ 能够明确捕捉 D-Net 产生的剧烈随机变化。

3.6 细化器

尽管 A-Net 和 D-Net 能有效分解不同频带的演变，但仅通过逆离散小波变换重建最终预测并不足够。正如我们的消融研究（第 4.4 节）所示，由于两个独立子网络之间存在频谱壁垒，直接重建往往会产生缺乏物理一致性的预测。为解决这一问题，我们提出了细化器，这是一种物理感知协调模块，可校正频谱不一致性，并使预测与自然大气演变相一致。

令 $\boldsymbol{Y}_A$ 和 $\boldsymbol{Y}_D$ 分别表示低频背景能量和高频对流能量的部分重建结果。$\boldsymbol{Y}_{AD} = \text{IDWT}(\boldsymbol{A}_{\text{pred}},\ \boldsymbol{D}_{\text{pred}})$ 表示初步的完整重建结果。细化器通过两个并行的物理引导流运行：

- 运动学耦合（f_{KC}）：该模块先通过初始 3×3 卷积，再经由堆叠的残差块融合部分重建结果（$\boldsymbol{Y}_A$和 $\boldsymbol{Y}_D$）。通过学习联合空间表示，网络隐式地促进运动学一致性，将高频对流核在空间上锚定到低频平流骨架，从而解决频谱不一致性。
- 漂移校正（f_{DC}）：为保持时间连续性，我们显式计算与最新观测值的偏差：$\boldsymbol{Y}_{AD} - \boldsymbol{X}\text{last}$。残差块通过全局跳跃连接处理这些差分特征。这种残差校正机制可缓解累积的轨迹误差，有效充当软惯性约束，防止状态发生不合理跳变。

最后，来自两个流的特征图通过带有残差连接的二维卷积层进行融合，并与初步重建结果 Y_{AD}相连。最终预测 Ypred 表示为：

$$Y_{\text{pred}} = Y_{AD} + \text{Conv2D}\left([f_{KC}(Y_A, Y_D), f_{DC}(Y_{AD} - X_{\text{last}})]\right). \quad (10)$$

其中， f_{KC} 和 f_{DC} 分别表示运动学耦合和漂移校正， $[\cdot]$ 表示拼接操作。

该设计确保最终输出（Y_{AD}）能够保持极端事件的尖锐性（通过 D-Net）、风暴系统的结构稳定性（通过 A-Net）以及物理一致性（通过漂移校正和运动学耦合）。

3.7 多尺度课程学习策略

为了在尊重各组件不同物理作用的同时有效训练分解架构，我们引入了一种多尺度课程学习策略。该策略采用动态加权的损失函数，引导优化过程从粗尺度运动逐步过渡到细尺度强度细化。这些损失的具体应用位置在图2中以绿色胶囊标记。总目标函数 L_{total} 定义为：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pred}} + w(t) \cdot \mathcal{L}_A + \lambda_D \cdot \mathcal{L}_D + \lambda_{\text{Mixed}} \cdot \mathcal{L}_{\text{Mixed}}. \quad (11)$$

其中， t表示当前训练步数。L_{pred} =均方误差(Ypred Ytarget) 直接在像素域中监督最终的降水预报，$\lambda_{D,\text{Mixed}}$ 为平衡因子。

近似分支捕捉大尺度平流趋势（风暴骨架）。为实现结构对齐，同时允许局部强度变化，我们采用零归一化互相关（ZNCC）损失：

$$\mathcal{L}_A = 1 - \text{ZNCC}\left(A_{\text{pred}}, A_{\text{target}}\right). \quad (12)$$

此处，ZNCC 指标衡量的是对线性强度变换具有不变性的结构相似性。为促进收敛，我们采用线性退火调度 $w(t)$，在训练早期优先进行粗尺度结构学习：

$$w(t) = \max\left(1 - t/T_{\text{decay}},\ \lambda_{\min}\right). \quad (13)$$

其中， T_{decay} 决定课程阶段的持续时间，$\lambda_{\min}$ 为最小值。

D-Net 对多个尺度上的强度波动进行建模。我们定义一种层级加权损失，以平衡不同频带的贡献：

$$\mathcal{L}_D = \sum_{l=1}^{\text{level}} \frac{1}{2^l} \cdot \text{MSE}(D^l_{\text{pred}}, D^l_{\text{target}}). \quad (14)$$

其中， l表示分解层级。该权重调度可防止模型对噪声过拟合，同时确保恢复尖锐的强度峰值。

为防止中间表示发散，并确保各分支协同学习，我们引入一个辅助正则化项：

$$\mathcal{L}_{\text{Mixed}} = \text{MSE}\left(\left(Y_{\text{A}} + Y_{\text{D}} + Y_{\text{AD}}\right)/3,\ Y_{\text{target}}\right). \quad (15)$$

该项作为一致性约束，鼓励近似分支（Y_A）、细节分支（Y_D）和直接重建分支（Y_{AD}）共同收敛到真实值。

4 实验

4.1 实验设置

4.1.1 基线模型。 为全面评估我们提出的方法，我们选取了五种具有代表性的基线模型进行比较：(1) ConvLSTM [25], (2) MAU [4], (3) SimVP [12], (4) Earth-Farseer [36], 和 (5) AlphaPre [17]。根据先前研究 [17],，这些基线模型分为两类：分解型（D 类）和非分解型（ND 类）。

4.1.2 数据集。 我们在两个降水数据集上进行了实验：SEVIR [30] 覆盖美国，范围为384 km × 384 km，观测间隔为5分钟。实验采用10分钟的分辨率。上海雷达 [5] 记录了面积为501 km × 501 km区域内的降水情况，输入数据间隔为6分钟，输出间隔为12分钟。参数 T_{in} 和 T_{out}均设为6，两个数据集的输入数据均调整为 128×128 。所有模型均从头开始训练。更多细节见附录C。

4.1.3 指标。 我们使用均方根误差（RMSE）衡量数值预测误差，并使用结构相似性指数（SSIM） [34] 衡量结构保持程度。为评估预报技巧，我们计算海德克技巧评分（HSS），该指标衡量与随机预测的差异。临界成功指数（CSI）在特定阈值下进行评估。我们计算六个阈值的平均值（CSI-M），并选取两个阈值（高强度临界成功指数和极端强度临界成功指数）以验证极端值基准。更多细节见附录B。

4.1.4 实现细节。 基线方法采用官方代码库，并按默认设置进行配置。完整的训练流程与超参数详见附录F。

4.2 性能比较

表1汇总了在六个预报时效上取平均后的定量性能。WADEPre在SEVIR和上海雷达两个基准上均达到了新的最优水平，并在关键指标上持续优于各基线方法，包括CSI-M、极端事件指标（如CSI-219、CSI-40）、HSS和SSIM。

为进一步分析模型在更长预报时效下的稳健性，图5展示了SEVIR数据集上强降水（CSI-181）和极端降水（CSI-219）事件的逐帧性能演变。我们将性能比较总结如下：

- 预报技巧：WADEPre在汇总的临界成功指数和海德克技巧评分指标上达到最先进水平。值得注意的是，我们在高阈值下的显著领先（如图5所示）证明了所提出架构的有效性。
- 结构保真度：在两个数据集上均取得最高的结构相似性指数（SSIM），证明了 WADEPre 在拓扑方面的优越性。这进一步证实，基于小波的分解方法将 A-Net 的粗粒度重建与

表1：在SEVIR和上海雷达数据集上对全部六个提前时间取平均的定量比较。类型D和ND分别表示分解模型和非分解模型。↑表示越高越好，↓表示越低越好。最佳结果以粗体突出显示，次佳结果以下划线标出。

模型	Type	SEVIR						上海雷达					
		CSI-M ↑	临界成功指数-181 ↑	临界成功指数-219 ↑	均方根误差 ↓	海德克技巧评分 ↑	结构相似性指数 ↑	CSI-M ↑	临界成功指数-35 ↑	临界成功指数-40 ↑	均方根误差 ↓	海德克技巧评分 ↑	结构相似性指数 ↑
ConvLSTM	ND	0.355974	0.155084	0.041291	1.290777	0.445585	0.717261	0.253558	0.052567	0.001231	3.033739	0.337086	0.770083
MAU	ND	0.378454	0.179911	0.078185	1.290873	0.477095	0.719202	0.346315	0.249759	0.126814	3.234246	0.473638	0.736891
SimVP	ND	0.391180	0.203362	0.073078	1.244711	0.496391	0.668615	0.322941	0.191222	0.074395	3.165812	0.413999	0.738400
EarthFarseer	D	0.394133	0.203624	0.064953	1.238125	0.494665	0.545065	0.362593	0.258890	0.051279	2.607779	0.477972	0.498071
AlphaPre	D	0.408885	0.224541	0.082268	**1.207027**	0.512415	0.749047	0.409432	0.303714	0.191909	2.663889	0.542150	0.726093
WADEPre	D	**0.416419**	**0.238489**	**0.115865**	1.232280	**0.526560**	**0.754846**	**0.421976**	**0.317689**	**0.201965**	**2.595196**	**0.550064**	**0.770512**

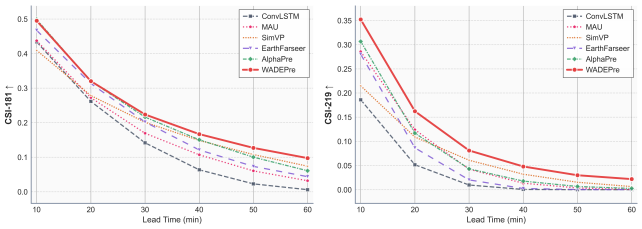


图 5：SEVIR 上极端事件预报技巧的时间演变。曲线展示了 10 至 60 分钟内高阈值（CSI-181 和 CSI-219）下逐帧计算的临界成功指数。与基线方法相比，WADEPre 的长期鲁棒性更强。

D-Net 的多尺度纹理细化相结合，能够准确保留未来降水场的空间分布和结构一致性。

- 数值误差：WADEPre 在 SEVIR 上略高的均方根误差，体现了双重惩罚效应 [26], ：对于空间位置略有偏差的尖锐预测，也会受到惩罚。不同于会使输出模糊的逐像素基线方法，也不同于缺乏类小波空间定位能力的 AlphaPre，WADEPre 保留了奇异性。鉴于其在临界成功指数和海德克技巧评分上的显著提升，这种权衡对气象价值至关重要。

4.3 可视化与案例研究

为评估模型捕捉强天气事件的能力，我们在图6中展示了一条线状飑线的预报结果。该飑线的特点是狭长的高强度回波带，并具有锐利的对流边界。

传统的逐像素回归模型（ConvLSTM、MAU、EarthFarseer 和 SimVP）表现出明显的扩散行为。当预报时效超过 T+30 分钟时，飑线的锐利梯度会变得平滑，导致系统失去线性组织，并使高强度核心退化为弥散云团。AlphaPre 虽能保留部分强度，但存在结构破碎问题，无法维持对流线的连续性。

相比之下，WADEPre 展现出更优的运动学一致性。它成功地将大尺度平流与局地强度变化分解开来，从而保留了飑线的形态完整性，并能够准确预测高强度前缘传播至 T+60 分钟。

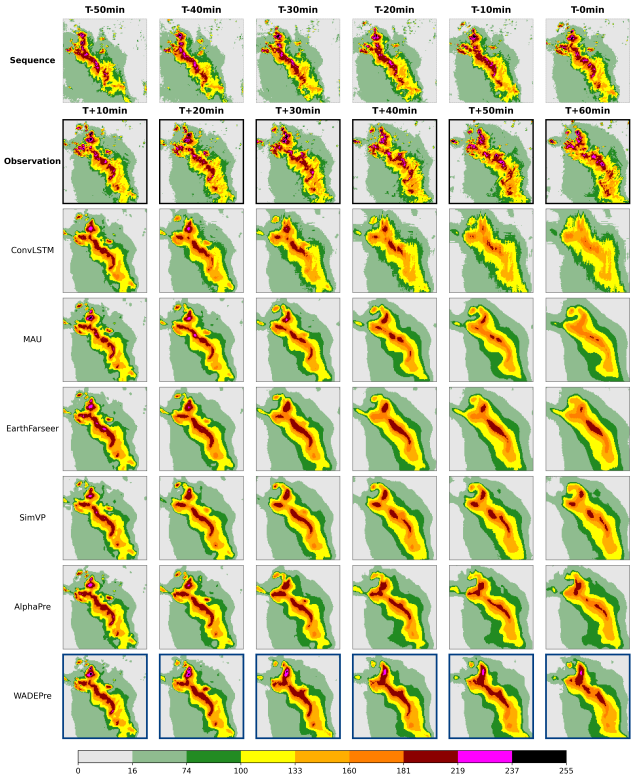


图 6：高强度暴洪事件的定性可视化。T+10 至 T+60 分钟的预报结果对比。基线模型受到严重的频谱平滑和快速强度衰减影响，导致对流核消散为背景噪声；而 WADEPre（蓝色帧）展现出更优的形态一致性，成功保留了锐利边界和高强度峰值（粉色/黑色区域）。

4.4 消融实验

本节将探究模型核心架构组件及所提出的多尺度课程学习策略各自的贡献。

4.4.1 架构组件的贡献。我们首先评估三个核心模块：近似网络（A-Net）、

细节网络（D-Net）和细化器。消融变体定义如下：（1）w/o A-Net：移除 A-Net，细化器直接将 $\boldsymbol{A}_{\text{seq}}$ 作为近似先验；（2）w/o D-Net：移除 D-Net，并以 $\boldsymbol{D}_{\text{seq}}$ 代替细节先验；（3）w/o Refiner：直接通过对预测系数进行 IDWT 生成最终预测。

表2：WADE- Pre 各组件在 SEVIR 数据集上的消融实验结果。↑表示越高越好，↓表示越低越好。最佳结果以粗体标出，次佳结果加下划线。

变体	CSI-M ↑	CSI-181 ↑	CSI-219 ↑	均方根误差 ↓	海德克技巧评分 ↑	结构相似性指数 ↑
不含 A-Net	0.190881	0.019060	0.003977	1.794504	0.244787	0.594154
不含 D-Net	0.365966	0.176532	0.080723	1.402748	0.471082	0.672986
不含细化器	0.362865	0.154449	0.063399	1.383330	0.464147	0.677992
WADEPre	**0.416419**	**0.238489**	**0.115865**	**1.232280**	**0.526560**	**0.754846**

表2证实了 A-Net 作为骨干网络的作用；移除它会导致 CSI-M 骤降 54%，这也证实了全局平流是必要前提。省略 D-Net 会显著降低 CSI-181 和 CSI-219 指标上的性能，凸显了它在消除奇异性方面的作用。细化器对频谱分量至关重要；若没有它，直接叠加会导致频谱泄漏和振铃，从而降低重建质量。

4.4.2 损失函数与课程学习的影响。我们进一步分析了多尺度课程学习策略的影响。变体包括：(1) 不使用 $w(t)$：移除动态加权计划；(2) 不使用 $\mathrm{L_A}$，以及 (3) 不使用 $\mathrm{L_D}$：分别移除 A-Net 和 D-Net 分支的中间监督；(4) $\mathrm{L_{Mixed}}$：排除一致性正则化项。

表3证实了课程学习策略的有效性。移除动态加权 $w(t)$ 会导致最大的性能下降，这表明优先构建低频基础可避免瓶颈。省略中间监督（$\mathrm{L_A}$ 和 $\mathrm{L_D}$）会因缺少针对性引导而降低性能。相比之下，$\mathrm{L_{Mixed}}$ 对于规范频谱-像素一致性至关重要。

表3：WADE- Pre在SEVIR 数据集上损失函数的消融实验结果。↑表示越高越好，↓表示越低越好。最佳结果以粗体突出显示，次佳结果加下划线。

变体	CSI-M ↑	临界成功指数-181 ↑	临界成功指数-219 ↑	均方根误差 ↓	海德克技巧评分 ↑	结构相似性指数 ↑
不含 $w(t)$	0.383020	0.188383	0.092796	1.287580	0.483875	0.704201
不含 $\mathrm{L_A}$	0.395002	0.203690	0.087669	1.270574	0.499125	0.752970
不含 $\mathrm{L_D}$	0.387864	0.191810	0.086628	1.277584	0.488795	0.750401
不含 $\mathrm{L_{Mixed}}$	0.400809	0.218417	0.100016	1.265551	0.508016	0.735994
WADEPre	**0.416419**	**0.238489**	**0.115865**	**1.232280**	**0.526560**	**0.754846**

5 结论

我们采用 WADEPre（一种物理感知小波分解模型），应对极端降水临近预报中长期存在的回归均值困境。该模型将时空演变分解为确定性平流（*A-Net*）和随机波动（*D-Net*），并通过细化器将二者融合，从而缓解谱偏差并确保结构一致性。消融研究证明了这种层级设计的必要性，并强调了各个模块的贡献。在 SEVIR 和上海雷达基准上的评估表明，WADEPre 达到了新的最先进水平，尤其是在预测危险极端事件方面。目前，该模型正在浙江省进行试运行和评估，展现出其在实际天气预报中的应用价值。通过结合数据驱动学习与频谱信号处理，WADEPre 为提升基于人工智能的天气预报模型的物理可靠性提供了一种前景广阔的方法。

局限性与伦理考量

局限性与未来工作。WADEPre目前缺乏严格的热力学一致性，且多尺度变换带来了较高的计算成本。未来版本将通过纳入多变量气象先验来强制满足物理约束，并整合扩散模型以量化概率不确定性，从而解决这些局限。

伦理考量。我们使用公开可用的数据集 [5, 30]，其中不含个人身份信息（PII），也不涉及人类受试者，因此本研究免于机构审查委员会（IRB）的审查。就公平性而言，地理偏差仍是一个值得关注的问题；在代表性不足的地区部署这项技术，需要进行本地微调，以确保结果公平。最后，为防止因过度依赖而导致的误用，本系统仅作为决策支持工具，旨在补充专业人员的专业知识，而非取代既定的操作规程。

致谢

本研究得到国家自然科学基金（U2342218）、浙江省“尖兵”“领雁”研发攻关计划（项目编号：2024C03256）、浙江省自然科学基金联合基金（项目编号：LZJMY24D050007）以及中国气象局（项目编号：FPZJ2025-053）的资助。

参考文献

[1] Kaifeng Bi、Lingxi Xie、Hengheng Zhang、Xin Chen、Xiaotao Gu 和 Qi Tian。2023年。利用3D神经网络实现精准的中期全球天气预报。《自然》 619, 7970 2023 533 538 10 1038 41586 023 06185 3（7月），–。doi：。/

[2] s---Massimo Bonavita。2024。论当前机器学习天气预报模型的一些局限性。《地球物理研究快报》51，12（2024年），e2023GL107377。doi:10.1029/2023GL107377[3] Tim Busker、Bart van den Hurk、Hans de Moel 和 Jeroen C. J. H. Aerts。2025年。降水预报对洪水预警的价值。《美国气象学会公报》106，3（2025年），E473–E491。doi:10.1175/BAMS-D-24-0073.1[4] Zheng Chang、Xinfeng Zhang、Shanshe Wang、Siwei Ma、Yan Ye、Xiang Xinguang 和 Wen Gao。2021年。MAU：用于视频预测及其他任务的运动感知单元。在神经信息处理系统国际会议（*NeurIPS*）论文集中。美国纽约州雷德胡克，文章2064，共13页。https://dl.acm. org/doi/ 10 5555 3540261 3542325 。/。

[5] Lei Chen、Yuan Cao、Leiming Ma 和 Junping Zhang。2020年。一种基于深度学习的方法：利用雷达进行降水临近预报。《地球与空间科学》 7, 2 2020 2019 000812 10 1029 2019 000812 （），e EA。doi：。/ EA

[6] Lei Chen、Xiaohui Zhong、Feng Zhang 等。2023。FuXi：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（2023年），190。doi:10.1038/s41612-023-00512-1[7] Puja Das、August Posch、Nathan Barber、Michael Hicks、Kate Duffy、Thomas Vandal、Debjani Singh、Katie van Werkhoven 和 Auroop R. Ganguly。2024年。物理与人工智能混合方法在极端降水临近预报方面优于数值天气预报。《*npj*气候与大气科学》7，1（2024年）。doi:10.1038/s41612-024-00834-8

[8] I. Daubechies。1990年。小波变换、时频局部化与信号分析。IEEE信息论汇刊 36，5（1990年），961–1005。doi:10.1109/18.57199[9] Ricard Durall、Margret Keuper 和 Janis Keuper。2020年。注意你的上采样卷积：基于卷积神经网络的生成式深度神经网络无法复现谱分布。在IEEE/CVF计算机视觉与模式识别会议（CVPR）中。美国华盛顿州西雅图，7887–7896。doi:10.1109/CVPR42600.2020.00791[10] Zhihan Gao、Xingjian Shi、Boran Han、Hao Wang、Xiaoyong Jin、Danielle C Maddix、Yi Zhu、Mu Li 和 Bernie Wang。2023年。PreDiff：基于潜在扩散模型的降水临近预报。在第37届神经信息处理系统国际会议论文集中。美国纽约州雷德胡克，78621–7865。https://dl.acm.org/doi/ 10 5555 3666122 3669561 。/。[11] Zhihan Gao、Xingjian Shi、Hao Wang、Yi Zhu、Yuyang Wang、Mu Li 和 Dit-Yan Yeung。2022年。Earthformer：探索用于地球系统预报的时空 Transformer。在第36届神经信息处理系统国际会议论文集中。美国纽约州雷德胡克，25390–25403。https://dl. acm.org/doi/ 10 5555 3600270 3602111 。/。[12] Zhangyang Gao、Cheng Tan、Lirong Wu 和 Stan Z. Li。2022年。SimVP：更简单却更优的视频预测。在IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集中。3170–3180。https://openaccess.thecvf.com/content/CVPR2022/papers/Gao_SimVP_ Simpler_Yet_Better_Video_Prediction_CVPR_2022_paper.pdf [13] David Gottlieb 和 Chi-Wang Shu。1997年。论吉布斯现象及其消除。SIAM Rev。 39, 4 1997 644 668 10 1137 0036144596301390 （），–。[14] doi：。/S Vincent Le Guen 和 Nicolas Thome。。从未知因素中解耦物理动力学，实现无监督视频预测。在2020年IEEE/CVF计算机视觉与模式识别会议（CVPR）中。11471–11481。doi:10.1109/CVPR42600.2020.01149[15] Remi Lam、Alvaro Sanchez-Gonzalez、Matthew Willson、Peter Wirnsberger、Meire Fortunato、Ferran Alet、Suman Ravuri、Timo Ewalds、Zach Eaton-Rosen、Weihua Hu、Alexander Merose、Stephan Hoyer、George Holland、Oriol Vinyals、Jacklynn Stott、Alexander Pritzel、Shakir Mohamed 和 Peter Battaglia。2023年。学习高技能的中期全球天气预报。《科学》382，6677（2023年），1416–1421。doi:10.1126/science.adi2336[16] Zongyi Li、Nikola Kovachki、Kamyar Azizzadenesheli、Burigede Liu、Kaushik Bhattacharya、Andrew Stuart 和 Anima Anandkumar。2021年。用于参数化偏微分方程的傅里叶神经算子。在国际学习表征会议（ICLR）中。https://openreview.net/forum?id=c8P9NQVtmnO[17] Kenghong Lin、Baoquan Zhang、Demin Yu、Wenzhi Feng、Shidong Chen、Feifan Gao、Xutao Li 和 Yunming Ye。2025年。AlphaPre：面向降水临近预报的振幅-相位解耦模型。在2025年IEEE/CVF计算机视觉与模式识别会议（CVPR）中。17841–17850。doi:10.1109/CVPR52734.2025.01662[18] Tsung-Yi Lin、Piotr Dollár、Ross Girshick、Kaiming He、Bharath Hariharan 和 Serge Belongie。2017年。用于目标检测的特征金字塔网络。在2017年IEEE计算机视觉与模式识别会议（CVPR）中。936–944。doi:10.1109/CVPR.2017.106[19] Stephane Mallat。1999年。《信号处理中的小波之旅》。爱思唯尔科学出版社。https://books.google.com.hk/books?id=yW2kut44AsMC [20] Mathieu Michael、Couprie Camille 和 LeCun Yann。2016年。超越均方误差的深度多尺度视频预测。在国际学习表征会议（ICLR）中。波多黎各圣胡安。doi: 10 48550 1511 05440 。/arXiv。。[21] S. Pulkkinen、D. Nerini、A. A. Pérez Hortal、C. Velasco-Forero、A. Seed、U. Germann 和 L. Foresti。2019年。PySTEPS：用于概率降水临近预报的开源Python库（v1.0）。Geosci. Model Dev. 12，10（2019年），4185–4219。doi: 10 5194 12 4185 2019 。/gmd---Suman Ravuri、Karel Lenc、Matthew [22] Willson、Dmitry Kangin、Remi Lam、Piotr Mirowski、Megan Fitzsimons、Maria Athanassiadou、Sheleem Kashem、Sam Madge、Rachel Prudden、Amol Mandhane、Aidan Clark、Andrew Brock、Karen Simonyan、Raia Hadsell、Niall Robinson、Ellen Clancy、Alberto Arribas 和 Shakir Mohamed。2021年。利用雷达深度生成模型进行高技能降水临近预报。《自然》597，7878（2021年），672–677。doi:10.1038/s41586-021-03854-z [23] M. G. Schultz、C. Betancourt、B. Gong、F. Kleinert、M. Langguth、L. H. Leufen、A. Mozaffari 和 S. Stadtler。2021年。深度学习能否击败数值天气预报？《皇家学会哲学汇刊A辑：数学、物理与工程科学》379，2194（2021年2月），20200097。doi:10.1098/rsta.2020.0097[24] Chunlei Shi、Han Xu、Yinghao Li、Yi-Lin Wei、Yongchao Feng、Yecheng Zhang 和 Dan Niu。2025年。WaveC2R：小波驱动的由粗到精分层学习，用于雷达反演。在人工智能促进协会会议（AAAI）论文集中。新加坡。doi:10.48550/arXiv.2511.17558[25] Xingjian Shi、Zhourong Chen、Hao Wang、Dit-Yan Yeung、Wai-kin Wong 和 Wang-chun Woo。2015年。卷积长短期记忆网络：一种用于降水临近预报的机器学习方法。在第29届国际神经信息处理系统会议。美国马萨诸塞州剑桥，802– – 810 10 5555 2969239 2969329 。https://dl.acm.org/doi/ 。/ 。[26] Christopher Subich、Syed Zahid Husain、Leo Separovic 和 Jing Yang。2025年。通过改进的球谐损失函数修正数据驱动天气预报中的双重惩罚。载于国际机器学习会议（*ICML*）。加拿大不列颠哥伦比亚省温哥华。https://icml.cc/virtual/ 2025 44912 /poster/[27] Juanzhen Sun、Ming Xue、James W. Wilson、Isztar Zawadzki、Sue P. Ballard、Jeanette Onvlee-Hooimeyer、Paul Joe、Dale M. Barker、Ping-Wah Li、Brian Golding、Mei Xu 和 James Pinto。2014年。数值天气预报在对流性降水临近预报中的应用：近期进展与挑战。美国气象学会公报 95, 3 2014 409 426 10 1175 11 00263 1 ( ), – 。doi: . / [28] BAMS-D- - 。Arnold Tafferner 和 Caroline Forster。. 天气临近预报与短期预报。施普林格·柏林·海德堡，德国柏林，363–380。doi:10.1007/ 978-3-642-30183-4 22_[29] Cheng Tan、Zhangyang Gao、Lirong Wu、Yongjie Xu、Jun Xia、Siyuan Li 和 Stan Z. Li。2023年。时间注意力单元：迈向高效的时空预测学习。载于*2023*年*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）。 18770 18782 10 1109 52729 2023 01800 – 。doi: . /CVPR . 。[30] Mark Veillette、Siddharth Samsi 和 Chris Mattioli。. SEVIR：用于雷达与卫星气象深度学习应用的风暴事件图像数据集。载于神经信息处理系统进展，H. Larochelle、M. Ranzato、R. Hadsell、M.F. Balcan 和 H. Lin（编），第33卷。Curran Associates, Inc.，22009–22019。https://proceedings.neurips.cc/paper_files/paper/2020/file/fa78a16157fed00d7a80515818432169-Paper.pdf [31] Hsing-Jui Wang、Ralf Merz 和 Stefano Basso。2025年。构建重尾洪水分布的地理格局：来自常见径流动态的启示。水文学与地球系统科学29，6（2025年），1525–1548。doi:10.5194/hess- 29-1525-2025[32] Yunbo Wang、Zhifeng Gao、Mingsheng Long、Jianmin Wang 和 Philip S Yu。2018年。PredRNN++：致力于解决时空预测学习中的深度时间困境。载于第*35*届国际机器学习会议论文集，第80卷。PMLR，5123–5132。https://proceedings.mlr.press/v80/wang18b.html [33] Yunbo Wang、Mingsheng Long、Jianmin Wang、Zhifeng Gao 和 Philip S. Yu。2017年。PredRNN：利用时空LSTM进行预测学习的循环神经网络。美国纽约州雷德胡克，879–888。https://dl.acm.org/doi/10.5555/3294771.3294855[34] Zhou Wang、A.C. Bovik、H.R. Sheikh 和 E.P. Simoncelli。2004年。图像质量评估：从误差可见性到结构相似性。*IEEE*图像处理汇刊 13, 4 2004 600 612 10 1109 2003 819861 ( ), – 。doi: . /TIP. 。[35] Peter A G Watson。.天气与气候的机器学习应用需要更加关注极端事件。环境研究快报17，11（2022年），111004。doi:10.1088/1748-9326/ac9d4e[36] Hao Wu、Yuxuan Liang、Wei Xiong、Zhengyang Zhou、Wei Huang、Shilong Wang 和 Kun Wang。2024年。Earthfarseer：在单一模型中实现多功能时空动力系统建模。载于*AAAI*人工智能会议论文集，第 38 15906 15914 10 1609 38 14 29521 卷。– 。doi: . /aaai.v i 。[37] Chiu-Wai Yan、Shi Quan Foo、Van Hoan Trinh、Dit-Yan Yeung、Ka-Hing Wong 和 Wai-Kin Wong。2024年。傅里叶振幅与相关性损失：超越使用L2损失实现高技巧降水临近预报。载于第*38*届神经信息处理系统国际会议。美国纽约州雷德胡克，35页。https: // dl.acm.org/doi/ 10 5555 3737916 3741089 . / 。Shangshang Yang 和 Huiling Yuan。[38] 2023年。用于风暴临近预报的定制化多尺度深度学习框架。《地球物理研究快报》50，13（2023年），e2023GL103979。doi:10.1029/2023GL103979[39] Demin Yu、Xutao Li、Yunming Ye、Baoquan Zhang、Chuyao Luo、Kuai Dai、Rui Wang 和 Xunlai Chen。2024年。DiffCast：基于残差扩散的降水临近预报统一框架。载于*2024*年*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）。27758–27767。doi:10.1109/CVPR52733. 2024.02622[40] Fisher Yu、Vladlen Koltun 和 Thomas Funkhouser。2017年。空洞残差网络。载于*IEEE*计算机视觉与模式识别会议论文集（*CVPR*）。636–644。doi:10.1109/CVPR.2017.75[41] Yuchen Zhang、Mingsheng Long、Kaiyuan Chen、Lanxiang Xing、Ronghua Jin、Michael I. Jordan 和 Jianmin Wang。2023年。利用NowcastNet对极端降水进行高技巧临近预报。《自然》619，7970（2023年），526–532。doi:10.1038/ s41586-023-06184-4[42] Hang Zhao、Orazio Gallo、Iuri Frosio 和 Jan Kautz。2017年。用于神经网络图像复原的损失函数。*IEEE*计算成像汇刊3，1（2017年），47–57。doi:10.1109/TCI.2016.2644865

A 基线模型

A.1 ConvLSTM

ConvLSTM [25] 是一项开创性工作，通过将卷积结构融入循环过渡，将全连接LSTM扩展至时空域。我们选择它作为基础基准，用以量化经典逐像素循环架构向现代频谱设计演进所带来的性能飞跃。

A.2 MAU

MAU（运动感知单元） [4] 引入了一种专用注意力机制，用于捕捉连续帧之间的预测运动动态。我们将 MAU 作为运动中心型循环模型的代表纳入比较。通过这一比较，我们可以严谨地评估基于注意力机制的隐式运动建模能否与我们的显式运动学耦合策略相媲美。

A.3 SimVP

SimVP [12] 将时空建模简化为纯编码器-解码器卷积神经网络框架，通过有效捕捉长程依赖，在多种基准测试中取得了具有竞争力的结果。我们将 SimVP 作为现代基于卷积神经网络的方法的代表性基线。对比结果表明，与可能难以处理高频纹理的 SimVP 不同，WADEPre 能够成功保持清晰对流边界的结构完整性。

A.4 EarthFarseer

EarthFarseer [36] 是一种近期提出的基于Transformer的模型，将地球特定物理约束融入注意力机制，以对时空湍流进行建模。我们选用它作为物理信息Transformer的代表，以评估软物理约束的有效性。

A.5 AlphaPre

AlphaPre [17] 是频域临近预报领域的最先进方法，使用快速傅里叶变换分离振幅与相位，从而将强度与运动解耦。我们将其作为基线，用于比较全局傅里叶变换与局部小波变换的效用。

B 指标

临界成功指数（CSI）。CSI（也称为 ThreadScore）用于评估特定降水阈值下的预报技巧。其定义如下：

$$\mathrm{CSI} = \frac{TP}{TP + FP + FN} \tag{16}$$

其中，TP、FP 和 FN 分别表示对输入张量进行二值化后，真正例、假正例和假负例的数量。

我们报告三个不同强度等级下的 CSI，以评估不同降雨严重程度下的表现：平均强度（CSI-M）、高强度（CSI-H）和极端强度（CSI-E）。所有六个阈值，以及高强度（H）和极端强度（E）的阈值，均列于表 4。CSI 值越高，表示性能越好。

均方根误差 (RMSE)。RMSE 用于衡量预测帧与真实值之间的像素级差异。其定义为

表 4：CSI 和海德克技巧评分（HSS）的阈值。

数据集	阈值	高强度（H）	极端强度（E）
SEVIR	{ 16, 74, 133, 160, 181, 219 }	181	219
上海	{ 20, 30, 35, 40 }	35	40

as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{17}$$

其中， N 表示像素总数， y_i 表示真实值， $\hat{y}_i$ 表示预测值。均方根误差越低，预测准确度越高。

为评估物理域中的均方根误差（RMSE），我们在计算前对模型输出进行反归一化。对于 SEVIR 数据集，预测结果从 $[0, 1]$ 映射回像素整数 $[0, 255]$，随后通过公式 20 转换为物理量（$\mathrm{kg/m^2}$）。同样，对于上海雷达数据集，输出经过反归一化，恢复至原始反射率范围 $0 \sim 70\mathrm{dBZ}$。

海德克技巧评分（HSS）。HSS 衡量相对于随机机会的预报准确性，并将正确负例纳入考量；在稀疏降水场中，正确负例占主导：

$$\mathrm{HSS} = \frac{2 \times (TP \times TN - FP \times FN)}{(TP + FN)(FN + TN) + (TP + FP)(FP + TN)} \tag{18}$$

其中，TP、FN、FP 和 TN 分别表示对输入张量进行二值化后，真正例、假负例、假正例和真负例的数量。

HSS 的取值范围为 0 至 1，并对定义的六个强度阈值（见表 4）取平均；值越高，表示预报技巧越好。

结构相似性指数（SSIM）。SSIM 通过评估亮度、对比度和结构信息来衡量预测图像的感知质量。其计算公式如下：

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{19}$$

其中， x 和 y 是两幅图像， μ 和 σ 分别表示图像的均值和方差，σ_{xy} 表示协方差。SSIM 的取值范围为 0 到 1，值越高表示结构相似性越强。

C 数据集

在两个数据集上， T_{in} 和 T_{out} 均设为 6， C 设为 1（为简洁起见，符号中省略）。

C.1 SEVIR

我们使用 SEVIR 数据集 [30]中的垂直积分液态水含量（VIL）产品。数据最初以 384×384（像元间距为 1 km）提供，随后经双线性下采样至 128×128，并按 10 分钟间隔采样。原始像素值 $x \in [0, 254]$（其中 255 表示缺失数据）会归一化到 $[0, 1]$ 范围内用于训练。

为确保评估结果具有物理意义，均方根误差（RMSE）指标是在反归一化后的物理量 y（单位为 $\mathrm{kg/m^2}$）上计算的，具体遵循官方协议：

$$y = \begin{cases} 0 & x \le 5 \\ (x-2)/90.66 & 5 < x \le 18 \\ \exp\left((x-83.9)/38.9\right) & 18 < x \le 254 \end{cases} \tag{20}$$

C.2 上海雷达

我们使用由上海中心气象台（SCMO）[5], 采集的上海雷达数据集，其中包含长江三角洲的组合反射率（CR）序列。原始数据覆盖一个 460×460 网格，对应的物理区域为 460km $\times$ 398km，反射率值范围为 0 至 70 dBZ。

在我们的实验设置中，输入序列由每隔 6 分钟采样的观测组成，而预报目标由每隔 12 分钟采样的观测组成。按照预处理流程，雷达图被调整为 128×128，并归一化至 $[0, 1]$ 范围。与 SEVIR 一致，均方根误差指标基于反归一化后的值（还原至 0–70 dBZ 范围）计算，以准确反映物理预测误差。

D 定量分析

在图 5 中对极端事件（CSI-181 和 CSI-219）进行重点分析的基础上，本节全面评估模型在所有指标上的表现。展示了六项主要指标——CSI-均值、CSI-High、CSI-极端阈值、均方根误差、海德克技巧评分和结构相似性指数——在六个提前时间上的变化。

图 7 展示了 WADEPre 在 SEVIR 数据集上逐帧性能的对比。WADEPre 始终表现出更优的结果，尤其是在预测的后期阶段。相反，ConvLSTM 和 SimVP 等基线模型的性能会随着提前时间延长而迅速下降，而 WADEPre 则保持着更稳定的趋势。值得注意的是，在临界成功指数-219 和结构相似性指数方面，我们的模型有效减轻了平滑效应，保留了逐像素回归基线通常会丢失的复杂风暴细节。

图 8 展示了上海雷达数据集上的结果。尽管气候特征各异，且强度阈值（CSI-35 和 CSI-40）有所不同，我们的模型仍始终保持显著优势。在 +36 分钟提前时间处，性能出现明显下降，这表明训练参数可能需要针对数据集进行微调；尽管如此，WADEPre 总体上仍是效果最好的模型。从 CSI-40 和结构相似性指数子图可以看出，WADEPre 显著优于基线模型，展现出强大的泛化能力，以及在长期临近预报中的实际应用可行性。

E 案例研究

为进一步展示模型在不同气象天气形势中的泛化能力，我们给出三个具有代表性的案例研究：与龙卷风 生成相关的离散单体事件、一次大尺度暴雨 事件，以及一条线状飑线。

E.1 龙卷风案例：离散单体保持

图 9 展示了一次龙卷风形成事件（ID：S843733），其特点是分散、彼此独立的对流单体，核心区域虽不大却十分强劲。

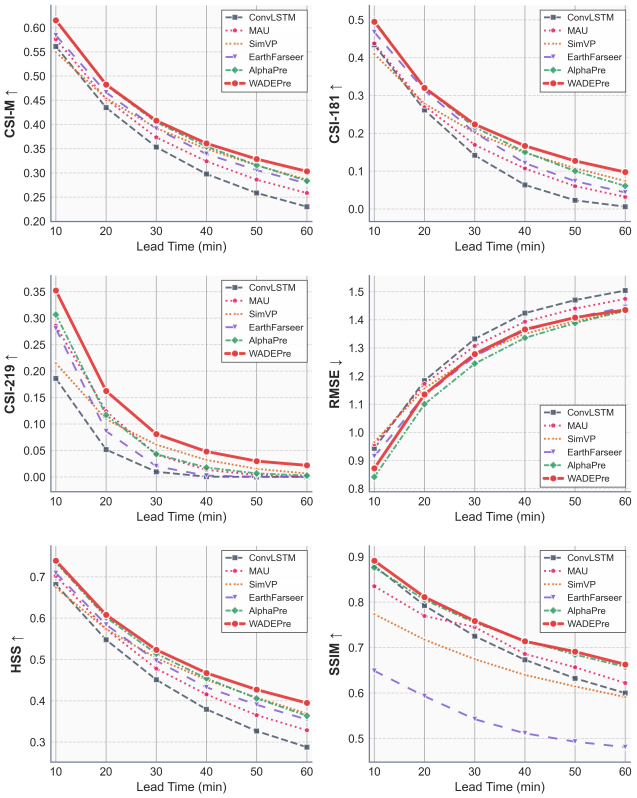


图7：SEVIR基准上的综合时间评估。曲线描绘了六项指标在60分钟预报时效内逐帧的演变。与次优方法AlphaPre相比，WADEPre（红色）在所有技巧指标上保持最高得分，尤其是在高强度区间（CSI-181和CSI-219）以及结构相似性（结构相似性指数）方面。

- **挑战：**主要挑战在于保留这些孤立单体中的高频奇异性。随着提前时间延长，传统基线模型（例如 ConvLSTM 和 EarthFarseer）往往会通过逐像素平均，迅速将这些小目标弱化为微弱的背景噪声。

- **我们的结果：**WADEPre在T+60 分钟时保持拓扑可区分性，利用D-Net对高频细节进行建模，并准确刻画尖锐的对流核边界，从而避免其他方法中常见的“冲刷”效应。

E.2 强降雨案例：结构保真

图10展示了一次显著的强降雨事件（ID：S851158），其特点是复杂的中尺度对流系统及其精细的内部结构。

- **挑战：**对于这些范围广泛的系统而言，主要挑战在于防止内部结构细节（纹理）退化为均质、平滑的一团。

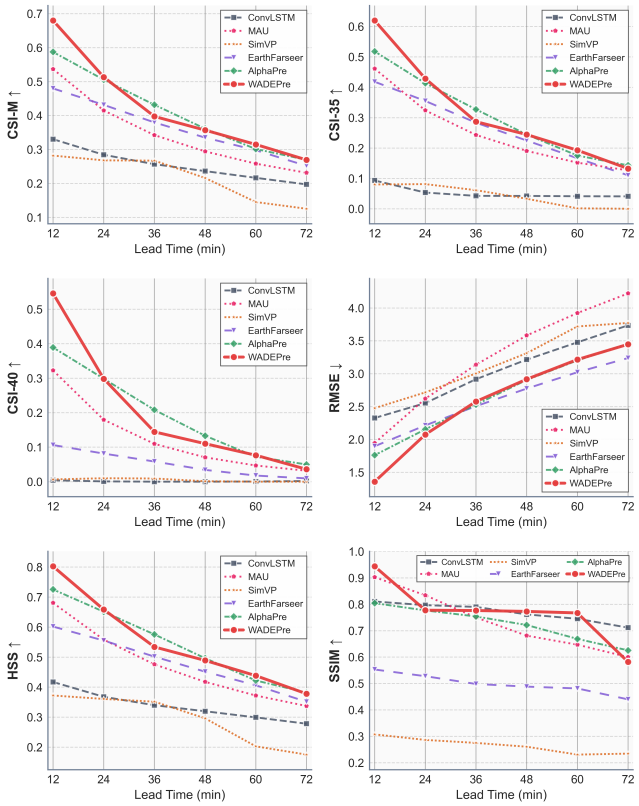


图 8：Shang- hai Radar 基准数据集上的全面时间评估。对 72 分钟提前时间内的全部六项指标进行性能比较。WADEPre（红色）在所有指标上均表现突出，包括最低的均方根误差和最高的极端事件捕获（CSI-40）。

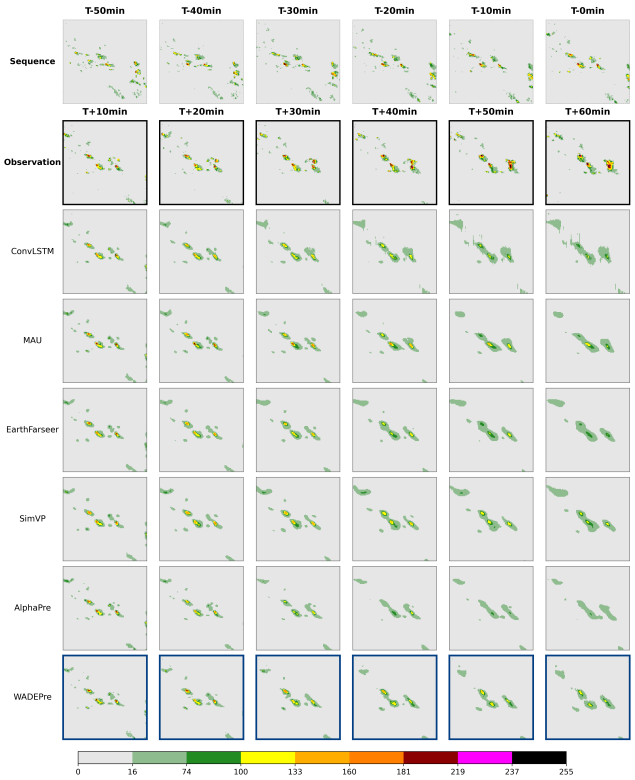


图 9：龙卷风生成事件（ID：S843733）的定性可视化。该案例包含分散的高强度离散单体。基线模型会将这些孤立结构模糊并融入背景，而 WADEPre（蓝色帧）则能在整个 60 分钟提前时间内保留对流核心清晰、锐利的边界。

- **我们的成果：** WADEPre 展现出卓越的纹理保真度。不同于 SimVP 和 MAU 等会生成过度平滑的预测结果、削弱内部强度梯度的模型，WADEPre 能够精准重建风暴复杂的空间层次结构，同时保留主回波带内细粒度的强度波动。

E.3 飑线案例：结构保真度

图 11 展示了一次线状飑线事件的预报结果，其特点是具有一条狭长、高强度的回波带，并带有清晰的对流边界。

- **挑战：**主要挑战在于，在较长的时间范围内保留飑线的线性组织结构和明显梯度。传统的逐像素模型通常会在超过 $T + 30$ 分钟后削弱这些锐利边界，导致有组织系统逐渐退化为弥散、无结构的云团。

- **我们的结果：**WADEPre 展现出卓越的运动学一致性。通过将大尺度平流与局地强度变化区分开来，它成功保持了飑线的形态完整性，并准确预测了高强度前缘的传播，最长可达 $T + 60$ 分钟。

F 实现细节

我们的实验采用 PyTorch Lightning 框架，并在四块 NVIDIA H100 GPU 上进行分布式训练。所有实验中的随机种子均设置为 42，以确保一致性。WADEPre 的具体超参数和训练过程总结如下。

- **小波变换**：小波分解层级 l 设置为 3，并使用 bior2.4 小波。

- **近似网络：** 通道维度设置为 256，并堆叠 3 个时空块（STBlocks）。• **细节网络：** 基础特征维度为 128。FPN 主干网络的通道深度为 $[64, 128, 256]$,，而 IDR 模块的通道深度为 64。• **细化器：** 隐藏维度大小设置为 576。

- **课程训练策略：** 损失平衡系数设为 $\lambda_D = 0.05$ 和 λ混合 $= 0.005$。对于课程学习计划，动态权重会

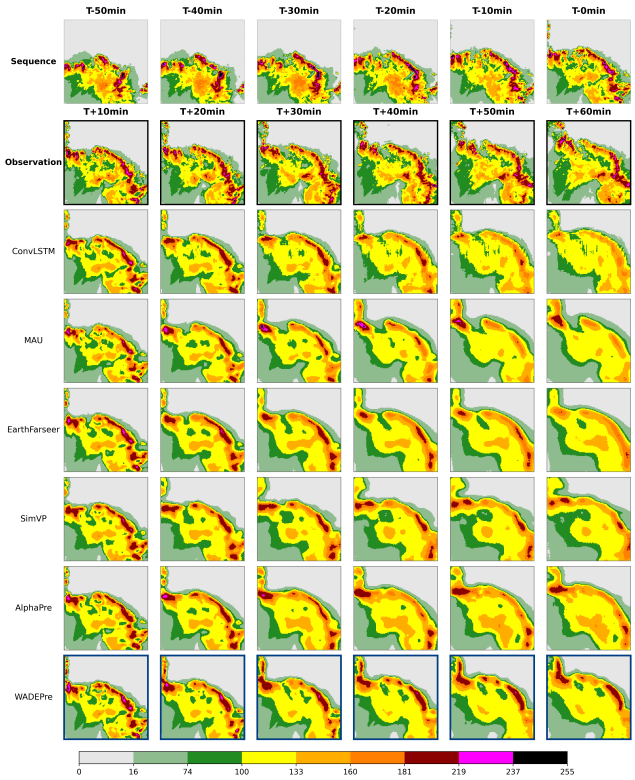


图 11：飑线事件（ID：S825535）的定性可视化。该案例涉及一个具有锐利边界的线性对流系统。WADEPre（蓝色帧）有效缓解了频谱平滑，保留了形态一致性和高强度峰值（粉色/黑色区域）；而在基线预测中，这些特征会消散为背景噪声。

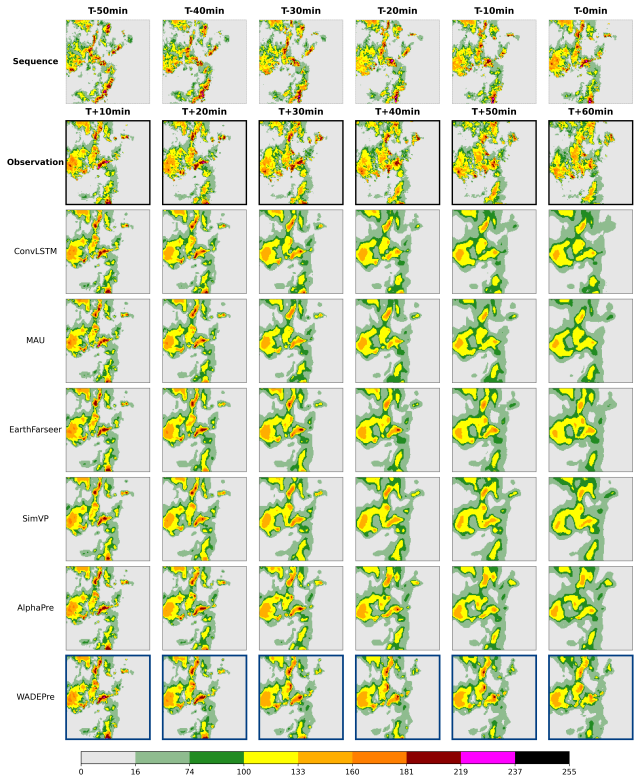


图10：强降水事件（ID：S827031）的定性可视化。该案例涉及一个大尺度对流系统。WADEPre（蓝色帧）有效缓解了平滑效应，保留了丰富的内部纹理和高强度梯度（红色/黑色区域）；而这些特征在基线预测中会丢失。

在 $T_{\text{decay}} = 3000$ 步内线性衰减，且下限为 $\lambda_{\min} = 0.01$。

- **优化：** 我们使用AdamW优化器，学习率为1.5e-4, _betas 0 01 32 (0.9, 0.995)，权重衰减为。。浮点精度为。余弦退火学习率调度器的T max设置为200。