arXiv:2611.31212v1 [physics.flu-dyn] 23 Nov 2026

使用物理引导条件去噪扩散概率模型进行洪水预报

Omar Sallam[a], *、David Coe[a]、Jeremy Feinstein[a]、Hugo Torres[a]、Sofia Sahagun Covarrubias[a]、Eugene Yan[a]

[a]阿贡国家实验室环境科学部，美国伊利诺伊州 *Lemont*，*9700 S Cass Ave*，邮编 *60439*

摘要

洪水是全球最常见且破坏性最强的自然灾害之一，但即使在现代图形处理器架构上，使用完整的二维浅水方程（SWE）求解器进行高分辨率洪水预报，计算成本仍然很高。本研究开发了一种条件去噪扩散概率模型（DDPM），用于预报美国伊利诺伊州东北部德斯普兰斯河流域未来四天内 30 m 水深场的时间演变。该模型采用 SynxFlow 生成的水动力模拟进行训练。SynxFlow 是一款多GPU有限体积浅水方程求解器，模拟的驱动数据为 2003 年至 2024 年间在美国中西部各州观测到的 113 场历史风暴，每场风暴均在三种前期湿润情景下进行模拟。为提高物理一致性并抑制生成式幻觉，DDPM 还可选择以 100 m 或 150 m 分辨率的粗分辨率水动力模拟作为条件输入。

由于反向扩散过程具有随机性，该模型从构造上就是概率性的，本文报告的每个预报都是从同一训练好的模型中抽取的20成员集合。粗分辨率条件约束使集合平均的四天平均水深均方根误差从无条件约束时的0.170 m降至采用100 m和150 m条件约束时的0.054 m和0.048 m。较粗的150 m条件约束还取得了最佳的临界成功指数（0.839）、虚警比、连续排序概率得分和布里尔评分，同时运行成本更低，这表明更精细的条件化分辨率并不能保证更好的技巧。与单次实现相比，对成员取平均可使水深误差降低5–11%，而集合离散度则提供具有空间明确性的洪水超越概率。不过，这些集合的离散度不足，离散度-误差比为0.35–0.73：对扩散噪声进行重采样能够反映代理模型自身的不确定性，但无法替代对气象强迫的扰动。相较于单GPU 30 m模拟，单成员预报在100 m、150 m和无条件配置下分别快 $7.8\times$，$11.0\times$和 $16.3\times$ 倍；这种优势会随集合规模和预报时域的增大而提升，最高可达

$48\text{–}58\times$ 对于20成员的14天预报。这些结果表明，物理引导扩散模型有望成为区域尺度洪水预报的概率代理模型。

关键词： 洪水预报、生成式人工智能、扩散模型、集合预报、不确定性量化、气候韧性

1. 引言

洪水是全球发生频率最高的自然灾害和气候相关灾害 [1]。根据灾害流行病学研究中心（CRED）的数据，过去二十年间，洪水平均每年造成169起重大灾害事件，超过了所有其他气候相关危害，包括平均每年发生108次的风暴 [1]。同期，洪水平均每年导致5,544人死亡，并在全球造成约456亿美元的年均经济损失 [1]。这些统计数据凸显了洪水灾害给发达地区和发展中地区带来的重大社会与经济负担。

*通讯作者电子邮箱： osallam@anl.gov (Omar Sallam)

洪水的高发与多种水文气候过程密切相关。随着全球气温升高，大气能够容纳更多水汽（根据克劳修斯–克拉佩龙关系，每升温 ℃，水汽容量约增加7%），从而导致极端降水更为强烈，风暴系统更加强劲 [2, 3]。此外，热带气旋、大气河、融雪事件和海平面上升都可能直接或间接导致洪水。因此，洪水往往是多种气候相关危害最终且影响最广泛的表现形式，包括强降雨引发的暴雨洪水、河流溢流造成的河流洪水、在响应迅速的集水区发生的山洪，以及由风暴潮和海平面上升引发的沿海洪水 [4]。

洪水事件可能导致人员伤亡、社区流离失所、财产损失、交通中断、电力和供水中断、供应链中断、土壤侵蚀、水体污染、生态系统退化，以及疾病暴发、霉菌生长、微生物污染和有毒化学物质暴露等公共卫生风险 [5]。因此，可靠的业务化洪水预报已成为灾害风险降低和气候适应策略的重要组成部分。准确的中长期洪水预报有助于开展疏散规划、识别高风险区域、保护关键基础设施和能源生产设施，并保障基本公共服务的连续性 [6]。

业务化洪水预报系统通常整合气象预报、水文降雨–径流模型和水动力淹没模型 [7]。对于城市和区域尺度的洪水预报应用，包括暴雨积涝、河流洪水和山洪，基于浅水方程（SWE）的数值模型得到广泛应用。这些模型可采用一维河道演算模型、二维淹没模型，或旨在降低计算成本的简化水动力公式[8]。

在这些方法中，完整的二维浅水方程模型具有很高的物理保真度，因为它们能够显式解析空间变化的水深、流速、洪水波传播、干湿锋、壅水效应和洪泛区连通性 [8]。然而，对于业务预报应用而言，这些模型的计算成本仍然很高，尤其是在流域面积较大、涉及区域或州级计算域、需要以 30 米或更精细的空间分辨率进行高分辨率淹没制图，以及采用旨在量化风暴路径、降水强度、流域前期状态和边界条件不确定性的集合预报框架时 [9]。

为弥合高保真度洪水预报与大规模数值模拟所受计算能力限制之间的差距，并应对现场观测数据匮乏的问题，机器学习已成为一种颇具前景的代理建模方法，可实现高效计算、大尺度和高分辨率的洪水预报。通过从历史数据和数值模拟中学习水文气候驱动因素、流域特征与洪水响应之间的关系，这些模型能够快速进行预测，适用于业务预报和预警应用 [10, 11]。

研究人员已探索多种机器学习方法用于洪水预报，包括随机森林 [12]和 XGBoost [13, 14]等基于树的方法；物理引导模型，包括物理信息神经网络（PINNs） [15], 、傅里叶神经算子（FNOs） [16, 17, 18, 19], 和条件潜在动力学网络（CLDNet） [20]；图神经网络（GNNs） [21, 22, 23, 24, 25, 26]等基于图的方法；卷积视觉架构，包括卷积神经网络（CNNs） [27, 28], 、残差网络 [29], 和 U-Net模型 [30, 31, 32, 33]；长短期记忆（LSTM）网络 [34, 35, 36]等序列学习模型；以及生成式概率方法，包括生成对抗网络（GANs） [37, 38, 39, 40, 41], 、去噪扩散概率模型（DDPMs） [42, 43, 44], 和混合式LSTM视觉生成架构 [45, 46, 47,48, 49]。

基于树的机器学习方法是一类应用广泛的预测模型，因为它们能够捕捉复杂的非线性关系、处理异质输入特征，并且只需相对有限的特征工程即可实现稳健的性能。随机森林 [50] 采用由独立训练的决策树组成的集合，而梯度提升及其广泛采用的实现 XGBoost [51, 52]则按顺序构建决策树，以迭代方式降低预测误差。这些方法已成功应用于洪水预报问题，例如街道级沿海洪涝 [12] 和洪水峰值水深预测 [14]。然而，现有应用主要关注点预测、街道路段预测或峰值预测，而非连续的时空淹没预报。由于基于树的模型作-

用于表格特征，且不会显式学习空间或时间依赖性，因此它们在大规模区域计算域和高分辨率二维洪水预报中的可扩展性仍然有限。

物理引导机器学习模型通过将控制方程、守恒定律或其他形式的科学知识纳入学习过程，力求提高预测准确性和物理一致性。最具代表性的方法包括物理信息神经网络（PINN） [53]和傅里叶神经算子（FNO） [54]。PINN 将偏微分方程残差直接纳入损失函数，从而在减少对大型训练数据集依赖的同时，实现物理一致的预测。相比之下，FNO 通过傅里叶域中的谱卷积学习函数空间之间的映射，从而能够高效地近似复杂的解算子。PINN 具有很强的物理可解释性，但通常需要耗费大量计算资源的优化过程和较长的训练时间，尤其是在大型计算域中；此外，在高度非线性流态和未见过的水文条件下，其表现可能不佳。FNO 通常能显著加快推理速度并提升可扩展性；然而，它们需要大量由数值模拟生成的训练数据，在应用于未见过的事件或边界条件时，性能也可能下降 [16, 17, 18, 19]。

图神经网络（GNN）旨在通过将实体表示为节点、将交互表示为边，从图结构数据中进行学习 [55, 56, 57]。借助消息传递操作，GNN 能有效捕捉连通模式和空间依赖性，因此特别适合表示河网和流域系统。因此，GNN 越来越多地应用于洪水预报和水文预测问题 [21, 22, 23, 24, 25, 26]。然而，随着图规模和空间分辨率的增大，训练成本和内存需求也会大幅增加。此外，要表示高分辨率的二维淹没场，通常需要极其庞大的图结构，这限制了其在区域尺度洪水预报中的可扩展性。

长短期记忆（LSTM）网络是一类专门的循环神经网络，旨在捕捉序列数据中的长期时间依赖性，同时缓解梯度消失问题 [58, 59]。通过记忆单元和门控机制，LSTM 能有效学习过去状态与未来状态之间的关系，因此非常适合用于时间序列预测。由此，基于 LSTM 的模型已广泛应用于径流预报、水位预测和水文时间序列分析 [34, 35, 36]。然而，LSTM 主要学习时间关系，并未明确表示洪水的空间动态。因此，相较于在大范围地理计算域内生成高分辨率二维淹没图，LSTM 通常更适用于水文测站的点尺度预报。

卷积神经网络（CNNs）旨在通过卷积操作处理具有空间结构的数据，并自动学习层次化特征表示。广泛采用的一种 CNN 架构是 U-Net [60], ，它结合了编码器-解码器路径与跳跃连接，在重建过程中保留精细尺度的空间信息。CNN 和 U-Net 架构在洪水淹没制图、代理水力建模和空间预测任务中展现出良好的性能 [30, 31, 32, 33]。然而，传统卷积操作本质上具有局部性，因此在捕捉大型流域和洪泛区中的长程依赖方面可能存在困难。

生成对抗网络（GANs）是一类强大的生成模型，能够通过生成器与判别器之间的对抗训练过程生成逼真的合成数据 [61]。凭借生成高分辨率空间场的能力，GANs 已应用于洪水淹没制图、代理水力建模和超分辨率任务 [37, 38, 39, 40, 41]。然而，GAN 训练往往不稳定，且对超参数选择高度敏感。此外，GANs 容易出现模式崩溃 [62, 63, 64], ，即模型只生成部分合理结果，而未能充分表征潜在的洪水分布。这一局限对概率洪水预报和不确定性量化尤其不利。

去噪扩散概率模型（DDPMs）近期已成为一种强大的替代性生成框架，能够通过迭代去噪过程生成高保真度的空间场[65]。与生成对抗网络不同，扩散模型采用基于数据似然变分界推导出的稳定去噪目标进行训练，因此不易发生模式崩溃。近期研究已展示了扩散模型在概率洪水预报和时空预测方面的潜力 [42, 43, 44, 45]。然而，将高分辨率洪水淹没预报、大范围

空间覆盖范围和时间演变整合到单一的基于扩散的框架中，仍基本未得到探索。

尽管由于迭代去噪过程，DDPM 通常需要较高的推理成本，但潜在扩散模型 [66], 、加速采样器和基于一致性的扩散方法 [67]的最新进展已大幅降低了这一计算负担。这些特点使基于扩散模型的方法对于新一代洪水预报系统尤具吸引力，因为这类系统需要高分辨率空间预测、概率预报、不确定性量化以及区域尺度时空可扩展性。

为降低生成式幻觉的风险并提升洪水预报应用中的物理一致性，先前研究探索了以计算成本较低的基于物理的模拟输出作为生成模型的条件约束。这些条件输入可能源自粗分辨率或低保真度洪水建模框架；这类框架能够提供物理上合理的大尺度流动结构，同时避免求解完全解析的高分辨率二维浅水方程解所需的计算成本。随后，生成模型学习根据这些融入物理信息的输入，重建更高分辨率、更高频率或更高保真度的淹没动态，从而兼顾计算效率与更高的预测保真度。相关示例包括使用傅里叶神经算子 [19]进行时间超分辨率，以及使用 U-Net [33, 32],残差网络 [29],和生成对抗网络 [40, 41] 架构进行空间超分辨率。

这一范式与湍流和气象建模领域中已得到广泛研究的超分辨率框架非常相似：在这些领域，机器学习模型从粗分辨率或中尺度模拟中重建高分辨率流场，同时保留潜在的低频、大尺度物理结构和动力学特征 [68, 69, 70, 71, 72, 73]。

在本研究中，我们开发了一种用于二维时空洪水淹没预报的条件去噪扩散概率模型。该模型学习洪水动力学的条件概率分布，并预测高分辨率水深的时间演变，而不是预测二元淹没范围或单个洪峰时刻的洪水快照。

为提高物理一致性并降低生成式幻觉的风险，可以选择使用GPU加速的SynxFlow浅水方程求解器生成的低分辨率水动力模拟作为条件约束。这样便将粗分辨率数值模拟的计算效率与扩散模型重建高分辨率淹没模式的能力结合起来。

扩散公式带来的进一步影响，也是本研究的核心之一，是模型从构造上就具有概率性，而不是通过扩展才获得概率性。在气象学 [74] 和洪水预报 [9], 中，集合预报是量化预报不确定性的标准业务途径；近年来，扩散模型已被证明能够为全球天气预报生成具有良好技巧且经过良好校准的集合[75]。由于反向去噪过程是一个随机马尔可夫链，在相同条件约束下重复运行训练好的网络，便可从所学习的洪水状态条件分布中独立抽样。因此，无需重新训练、无需扰动输入，也无需进行任何架构修改，即可生成预报集合；所得集合离散度则提供了空间上明确的概率性洪水信息。我们明确利用了这一特性，为留出的测试事件生成多成员集合，并使用标准概率评分对其进行验证。这使本框架区别于确定性代理模型；对于后者，不确定性量化通常需要输入扰动、多模型构建或辅助方差估计。

所提出的模型使用美国中西部地区观测到的二十年历史降水事件（2003–2024）进行训练。不同于依赖合成降雨强迫的方法，该训练数据集由观测到的二维降雨场构成，保留了自然界降水格局的时空变异性。相应的目标淹没场由经过验证的 SynxFlow 二维浅水方程求解器生成。

该框架在德斯普兰斯河流域进行了评估，空间分辨率为 30 m，对应 4，188，832 个活跃计算网格单元。模型开发与评估使用了共 113 场历史降水事件，每场事件均在三种不同的前期水文条件下进行模拟，因此每种网格分辨率下共有 339 次水动力模拟。另有 6 场事件完全不参与训练，仅用于评估。模型使用大小为 15.36×15.36 km^2 的分区进行训练，而非整个流域范围。虽然本文仅考察了一个流域，但这种基于分区的 formulation 有助于在大型数据集上高效训练，并旨在便于日后纳入其他

流域，并拓展至区域尺度预报。

本研究的主要贡献有四点。首先，我们开发了一种条件去噪扩散概率模型，只需进行一次网络评估，即可预报 30 m 水深的完整四天演变；该模型基于观测到的风暴而非合成风暴的浅水模拟进行训练。其次，我们在受控设置下比较了 100 m、150 m 和无水动力条件约束的情况，并表明较粗的条件约束既可能提高准确性，也可能降低成本。第三，我们使用适当评分规则、离散度—技巧分析、秩直方图和可靠性图对原生扩散集成进行概率评估，并发现其离散度不足。最后，我们在同一台单GPU上比较代理模型与水动力求解器的计算成本，包括加速比与集合规模之间的关系。

本文其余部分安排如下：第 2 节介绍研究区域、SynxFlow 求解器及其验证，以及训练数据集的生成。第 3 节介绍条件去噪扩散概率模型、其条件变量、集合构建和训练流程。第 4节报告预报的确定性技巧和概率技巧及其计算成本；第 5节总结结论并概述未来工作。

2. 数据集生成

2.1. 研究区域

本研究选取德斯普兰斯河流域作为研究区域，该流域对应水文单元代码（HUC-8）07120004。如图1a 所示，该流域位于伊利诺伊州东北部，靠近芝加哥大都市区，且其上游一小部分延伸至威斯康星州南部。该流域是芝加哥城市化区域的主要河系之一，历史上曾多次发生与极端降水和城市径流条件相关的洪水事件。

该流域的总流域面积约为3,770 km^2, 年平均河流流量约为108 m^3/s（3,800 ft^3/s），河流自北向南流经流域范围，如图1b所示。

本研究使用的地形信息来源于美国地质调查局（USGS）三维高程计划（3DEP）数字高程模型（DEM），空间分辨率为30 m。河网取自美国地质调查局国家水文数据集（NHD），如图1b所示。土地利用和土地覆盖信息取自国家土地覆盖数据库（NLCD）2021数据集，空间分辨率为30 m，如图1b所示。

所有数据集均投影至美国本土阿尔伯斯等积圆锥投影（EPSG:5070）。在所采用的30 m计算网格分辨率下，完整栅格计算域包含5,075 × 1,661个网格单元，其中4,188,832个位于流域边界内，构成有效计算网格单元。用于条件约束的相应粗分辨率计算域在100 m分辨率下包含1,523 × 499 个网格单元，在150 m分辨率下包含1,015 ×333个网格单元。

如图 1b所示，该流域的土地利用分布具有异质性。流域中部以中高密度城市建成区为主，属于芝加哥都会区。相比之下，北部和南部地区的城市化水平较低，土地覆盖主要包括耕地、植被和开放绿地。

2.2. 数值模型

本研究所用的洪水淹没数据集由 SynxFlow（原 HiPIMS）生成。SynxFlow 是一种高性能多GPU水动力模拟框架，旨在基于二维浅水方程（SWEs）开展大规模洪水模拟 [76, 77]。SynxFlow采用Godunov型有限体积格式，在结构化笛卡尔网格上求解深度平均浅水方程的守恒形式，从而能够高效、可扩展地应用于高分辨率流域尺度模拟。

控制方程可写为：

$$\partial_t q + \nabla \cdot \mathbf{F}(q) = R + S_b + S_f, \qquad (t,(x,y)) \in [0,T] \times \Omega, \tag{1}$$

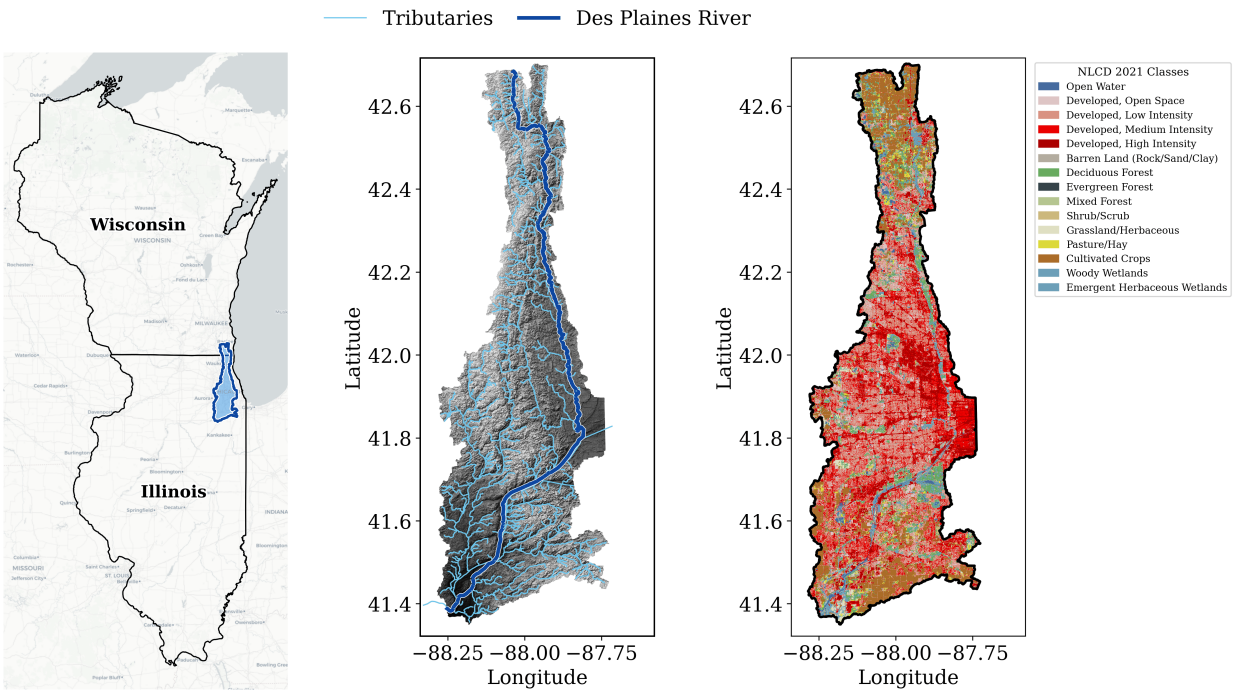
其中， q 表示守恒变量向量，$\mathbf{F}(q)$表示通量张量， R、 S_b和 S_f分别对应降雨/入渗、底坡和摩擦源项。这些量定义如下：

$$q = \begin{bmatrix} h \\ hu \\ hv \end{bmatrix}, \qquad \mathbf{F} = \begin{bmatrix} uh & vh \\ u^2h + \frac{1}{2}gh^2 & uvh \\ uvh & v^2h + \frac{1}{2}gh^2 \end{bmatrix},$$

$$R = \begin{bmatrix} r - I \\ 0 \\ 0 \end{bmatrix}, \qquad S_b = \begin{bmatrix} 0 \\ -gh\partial_x b \\ -gh\partial_y b \end{bmatrix}, \qquad S_f = \begin{bmatrix} 0 \\ -C_f u\sqrt{u^2+v^2} \\ -C_f v\sqrt{u^2+v^2} \end{bmatrix}.$$

其中， h 表示水深， u 和 v 分别表示沿 x 和 y方向的深度平均速度分量， b 表示床面高程。基于曼宁公式的摩擦系数表示为：

$$C_f = gn^2 h^{-1/3}, \tag{2}$$



(a) 德斯普兰斯河流域位置

(b) 叠加在流域地形高程上的主河道和支流网络（左），以及土地覆盖分布，数据来自 2021 年国家土地覆盖数据库（右）。

图 1：德斯普兰斯河流域的地理位置及其对应的地形、水文网络和土地覆盖。

其中， g 为重力加速度， n 为曼宁粗糙系数。根据国家土地覆盖数据库土地覆盖类别，设定空间变化的粗糙度场：开阔水域类别采用 $n = 0.02$ s/m$^{1/3}$ ，陆地类别采用 $n = 0.05$ s/m$^{1/3}$ 。

为确保在地形快速变化及干湿锋存在的情况下保持数值稳定性和稳健的激波捕捉能力，SynxFlow 采用了文献 [78] 提出的表面重构方法（SRM），对通量和河床坡度源项进行离散。此外，采用完全隐式格式处理刚性的摩擦源项 [79]，，从而提高高分辨率城市洪水模拟的数值稳定性。控制方程中使用的符号汇总于表 1。

2.3. 验证

使用2013年4月17日至18日期间发生在伊利诺伊州北部和中部的历史洪水事件，对德斯普兰斯河流域的水动力模型进行了验证。过去几十年来，这次事件是芝加哥都会区遭遇的最严重洪水之一，并已在以往研究中得到广泛记录 [80]。

2013年4月的风暴系统在伊利诺伊州及邻近州的大范围地区造成了极端降水累积。2013年4月18日，伊利诺伊州宣布进入紧急状态；至4月21日，该州102个县中已有44个被指定为灾区。根据美国国家气象局（NWS）的观测，伊利诺伊州、艾奥瓦州和密苏里州的24小时累计降雨量最高分别约为201.4毫米、171.1毫米和136.4毫米 [80]。

该洪水事件造成的死亡人数为4人，直接经济损失超过4.65亿美元，同时数千处住宅和商业地产报告了洪水损失 [81]。由于其严重的水文气象影响以及观测数据的可得性，2013年4月洪水事件提供了一个

表 1：式（1）中浅水方程的符号说明。

符号	描述
h	水深
u, v	x 和 y 方向的深度平均速度分量
hu, hv	x 和 y 方向的单位宽流量
g	重力加速度（9.81 m/s^2）
b	床面高程
n	曼宁糙率系数
C_f	摩擦系数（$gn^2h^{-1/3}$）
r, I	降雨和入渗率
q	守恒状态向量
$\mathbf{F}$	通量张量
S_b	河床坡度源项
S_f	摩擦源项

在德斯普兰斯河流域内验证大规模洪水淹没模拟的合适基准。

验证模拟覆盖了2013年4月15日至2013年4月23日期间，使模型能够捕捉洪水过程线的涨水段和退水段。降水强迫数据取自NCEP/NWS Stage IV多传感器（雷达和雨量计）降水分析，空间分辨率约为 4×4 km^2。

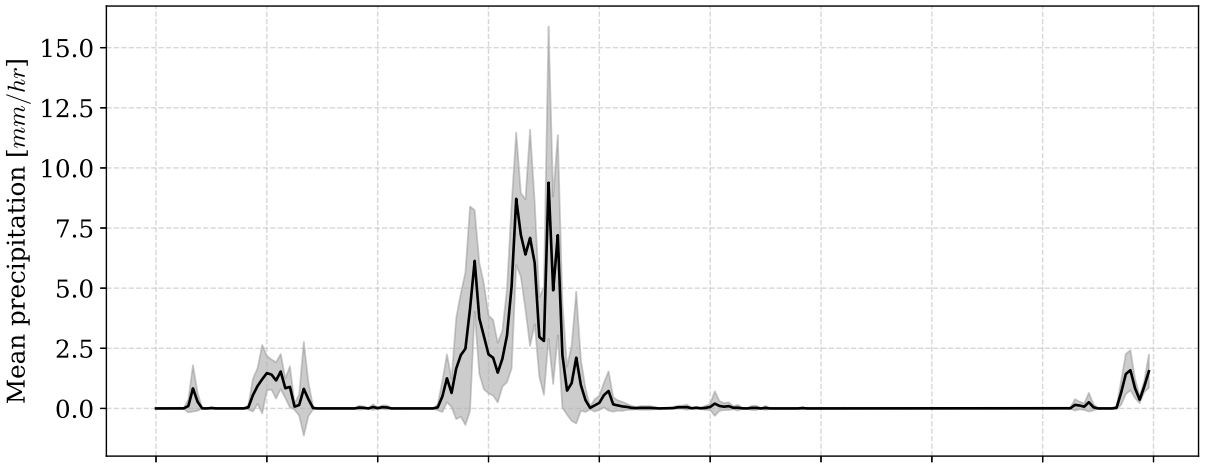
图 2a 展示了2013年4月洪水事件期间德斯普兰斯河流域的空间平均降水时间序列，而图 2b 展示了根据Stage IV雷达产品获得的日累计降水量的相应空间分布。降水等值线表明，大部分降雨累积发生在2013年4月18日，这对应于与历史洪水事件相关的降水峰值时段。

为进行定量验证，我们选取了位于德斯普兰斯河流域内的六个美国地质调查局（USGS）河流水位站，如图 3 所示。其中，四个位于德斯普兰斯河干流沿岸：USGS 05527800（位于拉塞尔）、USGS 05528000（格尼附近）、USGS 05529000（德斯普兰斯附近）和 USGS 05532500（位于里弗赛德）。此外，还纳入了两个支流水文站：USGS 05531500（韦斯特斯普林斯的索尔特溪）和 USGS 05540130（内珀维尔附近的杜佩奇河西支流）。之所以选择这些水文站，是因为在采用 30 m 分辨率的数字高程模型时，相关河道得到了足够精细的分辨。

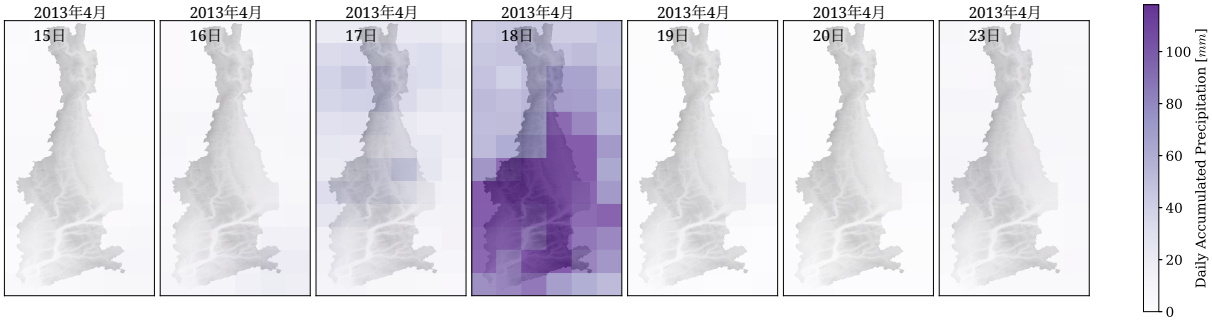
在每个水文站的位置，生成一条垂直于相应 NHDPlus流线的横断面线，并将其映射到计算栅格域。大多数水文站采用 60 m 的横断面宽度，而在四处较宽的河道位置，宽度则扩展至 90 m。模拟流量值通过对每条横断面上相交栅格单元内的流速和水深进行积分计算得到。此外，利用模拟水深和局部床面高程计算水面高程（WSE），并将其与美国地质调查局观测的水位测量值进行比较。

模拟与观测的水动力量在各测站位置总体吻合良好。总体而言，在整个洪水事件期间，模拟的水面高程与观测的水位测量值吻合良好。流量对比中的差异较大，尤其是在 USGS 05528000 和 USGS 05529000 测站，模型往往低估峰值流量。这些差异主要归因于横断面流量积分的不确定性、局部水力控制因素，以及 30 m 空间分辨率下的河道表示局限性。尽管如此，水面高程总体吻合良好，表明该模型能够充分捕捉流域内的大尺度洪水演进和淹没动态。

除定量的基于水文站的验证外，还利用洪水事件期间从社区社交媒体帖子收集的照片开展了定性洪水淹没评估。该



(a) T 2013年4月芝加哥洪水历史事件空间平均降水强度的emporal演变 t.



(b) 2013年4月芝加哥洪水历史事件的每日累积降水场叠加在数字高程模型上。

图2：数值模拟中使用的2013年4月芝加哥洪水降水历史事件的时空特征。

模拟的淹没范围使用Google Earth进行了三维可视化，并与观测到的被淹地点进行了比较，如3图左侧分图所示。数值模拟采用15厘米的水深阈值来界定淹没区域。模拟的淹没多边形与社区照片中可见的被淹区域在总体淹没边界和洪水空间范围方面均吻合良好，且在流域内的多个地点均如此。

2.4. 历史事件和训练数据集

我们生成了一个大规模洪水模拟数据集，用于训练和评估所提出的基于扩散的洪水预报模型。降水强迫取自一个包含 120 个候选历史降水事件的目录，这些事件分布于美国中西部的 12 个州（伊利诺伊州、印第安纳州、艾奥瓦州、堪萨斯州、密歇根州、明尼苏达州、密苏里州、内布拉斯加州、北达科他州、俄亥俄州、南达科他州和威斯康星州），时间跨度为 2003年7月至2024年9月。由于降水档案中有 7 个事件无法完整获取，因此将其剔除，最终保留 113 个历史事件用于整个模拟过程。

降水强迫数据取自 NCEP/NWS Stage IV 多传感器分析，其时间分辨率为每小时一次，空间分辨率约为 4×4 km^2 。每个事件均截取了一个 97 小时的

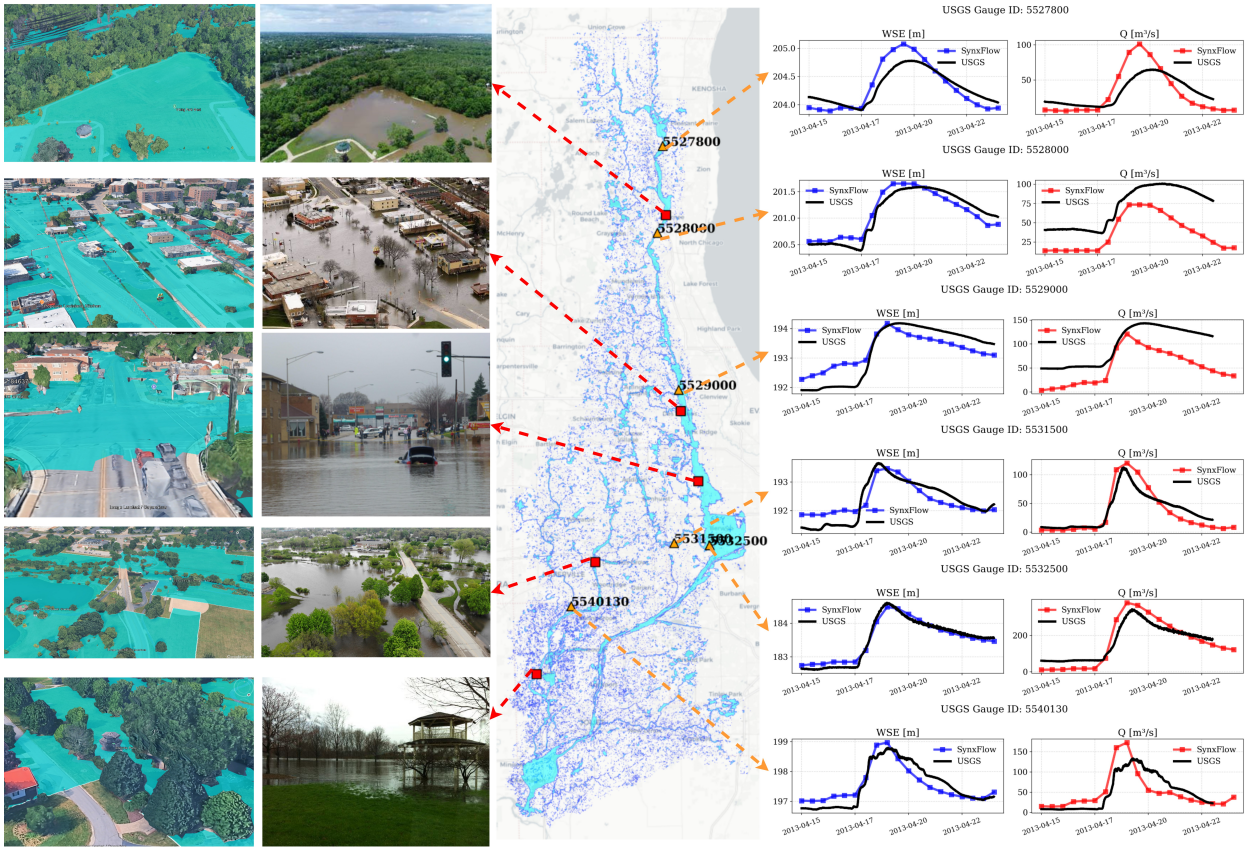


图3：2013年4月芝加哥历史洪水事件期间，针对德斯普兰斯河流域的SynxFlow水动力模型验证。洪水淹没范围，以及用于验证的美国地质调查局河流水位站位置和可用的社区社交媒体照片（中）。在Google Earth中对模拟洪水淹没多边形进行三维可视化，并与洪水事件期间收集的相应社区照片进行对比（左）。在选定的美国地质调查局测站位置，对模拟与观测的水面高程和流量过程线进行比较（右）。

时间窗口涵盖标称事件日期前一天至后三天，形成持续四天的风暴序列。

为构建德斯普兰斯河流域的降水强迫，提取每个历史事件在与德斯普兰斯流域几何范围相同的空间域内的数据，然后将其投影到流域栅格网格上。此过程使模型能够接触到多样的风暴模式、强度和空间降雨分布，同时保持计算域一致。需要指出的是，该设计有意将美国中西部其他地区观测到的风暴转移至德斯普兰斯计算域；因此，所得事件集在物理上合理，但并非对该流域洪水历史的重建。

对于每个事件，在四天的历史降水序列前添加了24小时预热期。模拟从干河床状态开始，并在预热期内施加三种不同的均匀降水强度：0.1、0.5和1.0 毫米/小时。这样，每个历史事件对应三种前期湿润情景，使水动力模拟总数达到 $113 \times 3 = 339$（每种空间分辨率）。因此，每次模拟总时长为121小时（包括24小时预热和97小时风暴窗口），模型状态每小时记录一次。

图4 给出了根据NOAA Atlas 14对伊利诺伊州乔利埃特的降水频率估算值得到的强度–历时–频率（IDF）曲线和深度–历时–频率（DDF）曲线 [82]。这些曲线涵盖的降雨历时范围为5分钟至4天，重现期介于1年至200年之间。这些频率曲线为刻画所生成洪水模拟数据集中降水事件的严重程度与罕见程度提供了参考框架。

图4 中的浅色散点表示用于训练扩散模型的历史降水事件。降水统计数据根据在德斯普兰斯河流域的投影降水栅格内同一空间位置提取的降雨样本计算得到。训练事件的分布表明，该数据集涵盖了广泛的降水严重程度与重现期，其中事件更集中地对应于约2年至25年的重现期范围。

实心散点对应于六个留出的测试事件，记为事件A–F。这些事件的选取旨在覆盖目录中所代表的风暴强度范围，从流域平均累积量最小的事件A，到小时峰值强度远高于其余事件的事件F。表2 总结了六个测试事件的特征。需要强调的是，风暴强度并非单一标量：事件A–F的排序在流域平均累积量和小时峰值强度上均非单调。事件D产生了最大的流域平均累积量和流域最大累积量，而事件F在短历时强度方面最为极端。因此，标签A–F应被理解为六种不同的严重程度等级，而非严格递增的强度轴，第4 节中指标随事件变化的图示亦应据此解读。

训练数据集和测试数据集的 IDF 与 DDF 统计量均通过对降水时间序列进行滑动窗口分析计算得出。窗口历时根据目标历时进行调整，涵盖 5 分钟、1 小时和 1 天等短历时及长历时累计降水量。对于每种历时，保留滑动窗口分析得到的最大累计降水量或降雨强度，并用于频率表征。

数据集生成通过运行 SynxFlow 完成，涵盖 339 种事件–预热组合，空间分辨率分别为 30 m、100 m 和 150 m。30米模拟用作高分辨率目标数据集，而 100 m 和 150米模拟则用作扩散模型的低分辨率条件输入，详见第 3 节。所有模拟均在阿贡国家实验室实验室计算资源中心（LCRC）的 SWING集群上，使用八个 NVIDIA A100 GPU 执行。整个计算任务在 30 m、100 m 和 150米分辨率下的实际运行时间分别约为 50.7、7.2 和 4.1 GPU节点小时。

图块提取为150 m条件约束配对保留了113个事件中的111个，并为100 m配对保留了全部113个事件，原因是存在两个不完整的粗分辨率输出。六个测试事件在两种配对中均存在，因此所有报告的模型比较均在相同的测试集上评估；训练集相差两个事件（100 m和150 m配置的训练事件分别为107个和105个）。

表2：六场保留测试降水事件的特征，基于投影到德斯普兰斯河流域范围后的 Stage IV 场计算。累积量统计时段为 97 小时风暴时段。

事件索引		Date	起源	流域平均值累积量 [毫米]	流域最大值累积量 [毫米]	1小时峰值强度 [毫米/小时]
A	40	2021年6月8日	北达科他州	22.2	138.5	67.8
B	85	2016年8月10日	明尼苏达州	69.0	181.5	56.8
C	95	2015年5月6日	内布拉斯加州	99.3	227.7	52.0
D	15	2021年6月24日	堪萨斯州	131.7	354.5	63.9
E	5	2017年4月28日	印第安纳州	120.4	275.0	59.6
F	115	2018年8月20日	威斯康星州	80.6	322.0	145.4

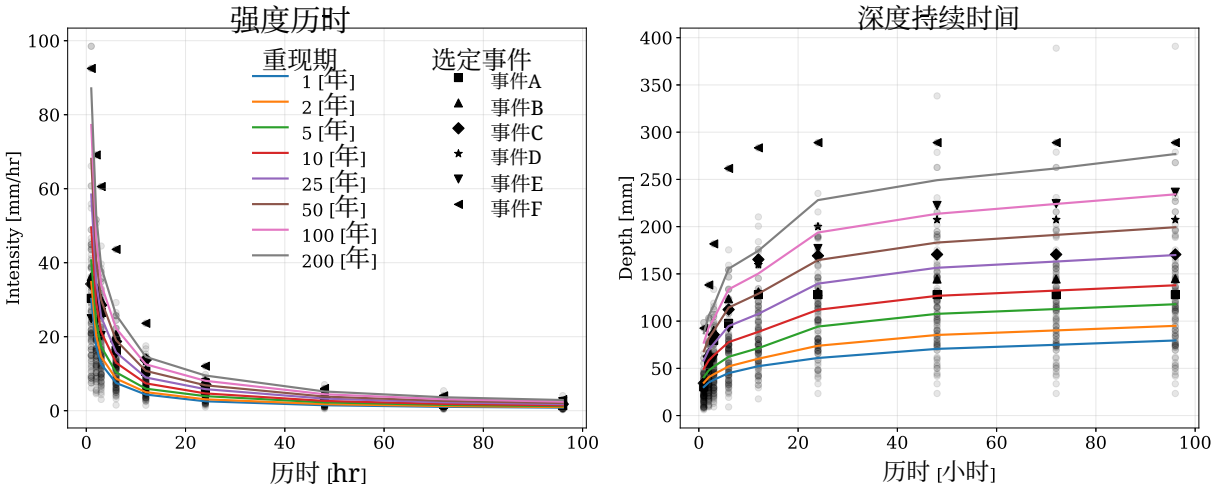


图4：根据 NOAA Atlas 14 的降水频率估算值绘制的伊利诺伊州乔利埃特（位于德斯普兰斯河流域）的强度–历时–频率（IDF）曲线（左）和深度–历时–频率（DDF）曲线（右）。浅色散点表示用于训练数据集的历史降水事件，实心散点则表示用于模型评估的测试降水事件。

3. 方法论

3.1. 条件去噪扩散概率模型

我们采用去噪扩散概率模型（DDPM），在静态地形、观测降水、初始水动力状态以及可选的低分辨率（LR）淹没场条件下，预报高分辨率（HR）洪水淹没深度。DDPM 提供了一种生成框架，通过逆转随机扩散过程来学习合成高分辨率场 [65]。如图5a所示，扩散模型采用双阶段机制：（i）一个前向过程，将数据逐步退化为高斯噪声；（ii）一个可训练的反向过程，模型通过迭代去除所添加的噪声来学习重建数据。

前向扩散过程 $q(x_{1:T} \mid x_0)$ 是一个不可训练的马尔可夫链，它通过在 T 个时间步中注入高斯噪声，逐步扰动一个 HR 样本 x_0。在任意噪声时间步 t，均可直接从 x_0 采样得到噪声状态 x_t，公式如下：

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{3}$$

其中 $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$，$\bar{\alpha}_t$ 由噪声调度 β_t 定义，使得 $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$。对于适当选择的噪声调度，$\bar{\alpha}_T \approx 0$，因此 x_T近似服从标准正态分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$。本文采用 Sigmoid噪声调度，将调度的分辨率集中在中等噪声水平。

反向过程 $p_\theta(x_{t-1} \mid x_t, \zeta)$ 在 ζ条件下，近似于难以处理的反向转移$q(x_{t-1} \mid x_t)$。本研究使用 v-预测参数化[83], $v_\theta(x_t, t, \zeta) \approx v_t$训练卷积神经网络，其中目标 v_t 定义如下：

$$v_t \equiv \sqrt{\bar{\alpha}_t}\epsilon - \sqrt{1-\bar{\alpha}_t} x_0. \tag{4}$$

该目标有时称为扩散“速度”，与第 v 2 节中的流速无关；它平衡信号和噪声成分，并提升训练稳定性和样本质量，尤其适用于高分辨率合成。根据预测的 v_θ，可将干净样本估计为 $\hat{x}_0 = \sqrt{\bar{\alpha}_t} x_t - \sqrt{1-\bar{\alpha}_t} v_\theta$。随后对反向转移进行采样：

$$p_\theta(x_{t-1} \mid x_t, \zeta) = \mathcal{N}\big(x_{t-1}; \mu_\theta(x_t, t, \zeta), \sigma_t^2 \mathbf{I}\big), \tag{5}$$

其中，均值 $\mu_\theta(x_t, t, \zeta)$ 是使用估计得到的 $\hat{x}_0$ 计算出的解析后验均值 $\tilde{\mu}(x_t, \hat{x}_0)$，方差 σ_t^2 则设为后验方差 $\tilde{\beta}_t$。对于高斯前向过程，这两个量均可通过闭式公式求得 [65]:

$$\tilde{\mu}(x_t, \hat{x}_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{1-\bar{\alpha}_t}\,\hat{x}_0 + \frac{\sqrt{\alpha_t}\,(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\,x_t, \qquad \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\,\beta_t, \tag{6}$$

其中 $\alpha_t = 1 - \beta_t$。

采样采用涵盖所有 T 个反向步骤的完整随机祖先采样器：链从 $x_T \sim \mathcal{N}(0, \mathbf{I})$ 初始化，并在每一步 σ_t 注入幅度为 $t > 0$ 的独立高斯扰动。不使用确定性采样器或蒸馏采样器。因此，对于固定的条件约束 ζ，训练好的模型每次评估都会从学习到的预测分布 $p_\theta(x_0 \mid \zeta)$中返回一个独立样本，而非单一的确定性预报。第 3.6 节直接利用这一特性来构建预报集合。

为引导生成过程得到符合物理规律的洪水淹没图，该过程以 $\zeta = \{\zeta_{\text{static}}, \zeta_{\text{dynamic}}\}$为条件约束，代表环境和水动力约束。关于这些条件变量的详细说明见第 3.4 节和 3.5节。

3.2. 预报目标与基于分块的 *formulation*

模型预报的量是在四个每日预见期 h 对应的高分辨率水深场。为区分物理时间与扩散步 t，预见期记为 τ。设 $\tau = 0$ 为结束时刻

24小时预热期的结束时刻（即风暴开始时刻），模型预报 h 在 $\tau = 24$，48，72 和 96 小时的水深，对应于 SynxFlow 模拟中的第 48、72、96 和 120 小时。这些是模型在相应时刻的瞬时状态，而非日平均值或日最大值。四个预见期由一次网络评估的四个输出通道表示，因此整个四天预报一次即可生成，而无需进行自回归滚动预测。水深是唯一的预报变量；动量分量仅作为条件约束信息输入模型。

训练和推理均基于空间图块进行，而非整个流域。候选图块大小为 512×512 个网格单元（在 30 m 分辨率下为 15.36×15.36 km^2，即每个图块覆盖 235.9 km^2）。这些图块沿规则网格提取，步长为 256 个网格单元，对应于两个方向上均有 50% 的重叠。在 95 个候选位置中，保留了 60 个数字高程模型中无数据单元占比低于 50% 的图块。保留图块的几何并集覆盖了流域面积的 97.5%。这 60 个重叠分区用于训练。对于流域尺度的推理，计算域被划分为 26 个大小相同且不重叠的斑块，因此完成一次预报需要进行 26 次分块评估。

所有条件约束变量和目标变量均通过基于最小-最大缩放的方法归一化到单位区间，所用全局统计量在整个训练数据集上计算一次，以确保每个图块和每个事件采用相同的归一化方式。在计算任何指标或绘制任何地图之前，均使用相同的全局统计量对水深进行反归一化。

3.3. 网络架构

反向扩散过程由条件 U-Net 架构参数化 [60]，，该架构整合了卷积残差块 [84]，、跳跃连接、下采样/上采样阶段以及自注意力机制 [85]，，如图5b所示。

在每个反向步骤中，网络根据噪声状态 x_t 预测 v_θ，并在静态和动态条件集 $\zeta = \{\zeta_{\text{static}}, \zeta_{\text{dynamic}}\}$ 的引导下，由此获得四帧水深序列 $\hat{x}_0$ 和下一状态 x_{t-1}。

噪声状态 x_t 按通道维度与条件张量拼接，经初始卷积层处理后，依次输入由编码器、瓶颈层和解码器组成的 U-Net，其通道乘数为 (1，2、4、8)。编码器在每个下采样阶段都将特征/通道深度逐步增加一倍。相反，解码器在每个上采样阶段逐步将通道数减少一半，直至恢复原始空间分辨率，随后通过最终投影得到四个输出通道。基础通道维度和扩散步数通过第3.8节所述的敏感性分析进行选择。

条件张量的构成总结见表3。除了这两个二维空间条件张量外，网络还通过专用嵌入网络将四个标量条件变量与扩散时间步一并纳入 t。扩散时间步向模型指示扩散过程中的当前噪声水平，使去噪网络能够相应地调整预测结果。这四个标量分别是归一化高程、土地覆盖、坡度和大阈值HAND图层的斑块均值，为网络提供全局斑块级地形上下文，以补充卷积编码器在最粗空间尺度上的纯局部感受野。需要指出的是，土地覆盖图层的斑块均值是对分类变量的汇总，应将其理解为斑块内城市化程度的粗略代理指标，而非物理量。

四个标量斑块描述符和扩散步 t分别由独立的多层感知机（MLP）处理。随后，将得到的嵌入注入 U-Net 的每个残差块（图5b），以确保全局上下文在所有空间尺度上得以保留。

3.4. 静态条件

静态条件变量，以 ζ_{static} 表示，源自流域地形和陆面特征，用于提供控制径流的空间信息。采用的静态条件图层如图 6 所示，共包含 13 个通道：高程、土地覆盖、坡度、采用 D8 和 LDD 编码的局部流向、河流级别、汇流累积、距最近排水高度（HAND）以及地貌洪泛平原图。HAND 和地貌洪泛平原栅格

表3：条件 U-Net 对应一个 512×512 分区的输入张量和输出张量组成。低分辨率块仅存在于 LR 条件配置中。

Role	块	Ch.	描述
目标	x_0	4	预见期为 24、48、72、96 小时的水深 h
条件约束	ζ_{static}	13	地形和地表层（图6）
	ζ_{init}	3	$[h, hu, hv]$ 在预热结束时（$\tau = 0$）
	ζ_{precip}	1	Stage IV 时空降水矩阵
	ζ_{LR}	15	5 个时间层级 $\times [h, hu, hv]$，分辨率为 100 或 150 米
	总计（含低分辨率）	32	
	总计（不含低分辨率）	17	
标量	t，分区均值	1+4	扩散步；高程、土地的分区均值 覆盖、坡度与HAND

数据集分别采用三种不同的上游面积阈值 $\{0.5, 5.0, 50.0\}$ km^2 生成，以捕捉多个空间尺度上的排水特征，共计占 13 个通道中的 6 个。

高程数据取自美国地质调查局三维高程计划（3DEP）数字高程模型，空间分辨率为30 m；土地覆盖信息则取自国家土地覆盖数据库（NLCD）2021数据集，空间分辨率相同。在推导地形变量之前，通过洼地填充程序去除地表洼地，以确保水文上一致的排水网络。pyflwdir软件包中实现的洼地填充算法遵循 [86] 中提出的高效洼地填充方法。

若干由地形导出的水文变量使用开源Python软件包pyflwdir[87]计算。坡度由数字高程模型计算得出，采用在移动 3×3 像元窗口内通过有限差分估计的一阶偏导数，表示局地地形梯度，即以单位水平距离上的高程变化表示（m/m）。

局部流向采用 D8 和 PCRaster 局部排水方向（LDD）两种规则进行编码；二者使用不同的数值代码表示同一个单向排水网络。这些变量描述了相邻网格单元之间的下游连通性，并确定地表水在地形上的运动方向。下游流路采用最陡下降法确定，即每个网格单元都向下坡梯度最陡的相邻网格单元排水。流向网络基于采用 [86] 方法获得的洼地填平数字高程模型（DEM）推导而来。由于 D8 和 LDD 是同一排水网络的不同编码方式，这两个通道相互冗余；此处保留二者是为了与归档数据集保持一致，并将这种冗余视为轻微的低效之处，而非额外的信息来源。

基于推导得到的流向网络，计算了河流级别和汇流累积图。河流级别采用斯特拉勒河流级序分类系统 [88]，计算，其中一级河流代表源头河道。两条同级河流汇合时，下游河段的级别增加一级；不同级别的河流汇流时，下游河段则继承较高的河流级别。汇流累积表示流经每个网格单元的上游汇水面积总量，常用作河道起始、排水密度和流域连通性的代理指标。

HAND图层是在 pyflwdir 中根据 [89]，所述的方法论生成的，该方法计算地形网格单元与其最近下游排水河道之间的垂直距离。通过将地形相对于局部排水网络进行归一化，HAND 可提供具有物理意义的洪水易发性和潜在淹没范围指标。HAND值较低的网格单元通常洪水脆弱性较高，因为相对于局部排水网络，它们处于较低的高程。

此外，我们采用 [90]中提出的 GFPLAIN 方法生成了地貌洪泛平原图。该方法利用上游汇水面积与水深之间的幂律关系（水力尺度关系）划定洪泛平原边界，并以此定义被视为洪泛平原一部分的最大 HAND值

。所得洪泛平原图层提供了在地貌学上易发洪水区域的信息，并已证明可作为数据驱动水文应用中预测洪水范围的有效指标。

这些静态条件变量共同提供了互补信息，用于描述流域地形、排水结构和洪水易发性，从而使去噪扩散概率模型框架能够学习控制洪水淹没动态的空间因素。

3.5. 动态条件

动态条件集合 $\zeta_{\text{dynamic}} = \{\zeta_{\text{init}}, \zeta_{\text{precip}}, \zeta_{\text{LR}}\}$ 由三个部分组成：

- **初始水动力状态（ζ_{init}）：** 水深和两个单位宽流量分量 $[h, hu, hv]^T$取自30米模拟在24小时预热期结束时的数据，即风暴开始的时刻。该部分定义了分区的前期湿润状态，也是三个预热强度情景进入模型的途径。

- **降水（ζ_{precip}）：** 将流域上97小时的逐小时 Stage IV降水序列表示为单通道数据。逐小时降水块的形状为（97小时 $\times\ n_y \times\ n_x$），单位为毫米/小时；其两个空间维度被展平，并重新采样至一个 512×512 矩阵，该矩阵的第一轴对应时间，第二轴对应展平后的空间索引，如图5b所示。该编码通过一个与其他条件层形状相同的通道，为网络提供完整的暴雨雨量过程线及其空间异质性。需要明确说明两点。第一，由于 Stage IV网格约为4千米，而目标网格为30米，降水通道仅保留风暴结构的原生中尺度分辨率。第二，该通道按事件构建一次，因此对每个分区都相同：降水作为全域时空描述符输入模型，而非局部补丁降雨场。局部补丁强迫信息则通过ζ_{init} 以及（若可用）$\zeta_{\text{低分辨率}}$传递。

- **低分辨率水动力解（$\zeta_{\text{低分辨率}}$）：** SynxFlow以100 m或150 m分辨率生成的粗网格水动力解，包含三个水动力变量 $[h, hu, hv]^T$在五个时间层级上的数据（预热结束状态加上四个预报提前期），共15个通道。对于每个分区，均将粗分辨率场裁剪至相应的地理窗口，并双线性上采样至 512×512 目标网格。低分辨率帧与初始状态及四个目标帧在时间上相对应，因此无需进行时间插值或偏移。值得注意的是，$\zeta_{\text{低分辨率}}$是一种可选模态：在无法获得低分辨率模拟时，该架构仍可仅利用静态、初始状态和降水条件输入进行预报；本文始终将这种配置称为“无LR”。

3.6. 集成构建

确定性机器学习洪水替代模型仅返回单个场，无法提供预报置信度信息，而这恰恰是应急管理人员在决定是否疏散某个街区时所需要的信息。在气象学 [74] 和业务化洪水预报中，集合预报都是成熟的解决方案 [9]。

此处采用的扩散公式无需任何额外机制，即可提供这一能力。由于反向过程是上述随机马尔可夫链，训练好的网络定义的并非从条件约束到预报的函数，而是一个条件分布 $p_\theta(x_0 \mid \zeta)$，可以从中抽取相互独立的样本。使用相同的条件张量，对同一个训练好的模型进行 N 次评估ζ，即可得到 N 种各不相同且同样合理的高分辨率洪水演变结果。这是模型类别本身的结构属性，而非对模型的扩展：第 4 节所报告的集合无需重新训练、无需扰动输入、无需辅助方差网络，也无需更改架构或训练好的权重。这也是扩散模型相较于第 1 节讨论的生成对抗网络的主要实际优势；后者容易出现模式崩溃，因此往往无法充分表现

合理结果的多样性。近期研究表明，在全球天气预报方面，基于扩散的集合已可与业务物理集合相媲美 [75]。

实际操作中，通过沿批次维度将条件张量 $N = 20$ 复制多次，并对祖先采样器进行一次批量调用来生成集合。各成员之间的独立性源于采样器对每个批次元素分别独立抽取初始状态 x_T 和每个中间增量。

集合所体现的不确定性来源

明确所采样的不确定性来源十分重要，因为这些集合并不构成完整的预测不确定性预算。

由于所有成员的条件张量 ζ 都保持不变，因此其离散程度体现了学习到的生成映射本身的不确定性：给定特定的地形、前期状态、风暴以及（如有使用）粗分辨率水动力学解，它量化了模型在精细尺度水体分布上保留的自由度。这种不确定性结合了两部分：重建条件约束无法确定的30米淹没细节时存在的不可约歧义，以及训练好的网络所具有的认知不确定性。这是报告代理模型步骤本身不确定性的恰当指标，也是确定性代理模型悄然舍弃的那部分不确定性。

另外三种不确定性来源被有意排除在抽样之外。由于模型由观测到的 Stage IV降水驱动，而非降水预报驱动，因此不考虑气象不确定性。由于曼宁糙率根据土地覆盖确定性地指定，因此不考虑水动力参数不确定性。前期条件不确定性仅通过三个离散的预热情景进行粗略表征。

从同一框架自然可以构建完整的运行不确定性预算，而其结构正是该方法具有吸引力的原因。给定 K 个成员的数值天气预报降水集合，每个成员 k 都定义自己的条件张量 ζ_k，据此可以抽取 M 个扩散成员，从而形成嵌套 $K \times M$ 集合，联合传播气象不确定性和生成不确定性。其成本主要来自 K 粗分辨率水动力模拟，而非 KM 扩散样本，具体量化见第4.6节。要验证此类系统的预报技巧，需要使用预报降水，这将留待未来工作。

3.7. 两种运行模式的解读

ζ_{LR} 的有无界定了两种概念上截然不同的运行模式。明确说明每种模式的含义十分重要，因为它们回答的是不同的问题。

在低分辨率条件模式下，模型会在与待预测量相同的预见期接收粗分辨率水动力学解。因此，该模型并非独立预报洪水演变，而是充当一个学习得到的空间降尺度算子，将粗分辨率水动力预报映射到30米网格，同时补充粗分辨率解无法解析的精细尺度淹没结构。在实际运行中，这对应于一个两阶段流程：先由预报降水驱动计算成本较低的粗分辨率浅水方程模型，再由扩散模型对其输出进行细化。此模式下报告的技能应理解为该细化步骤的保真度，前提是粗分辨率预报可用；这并不能证明网络学会了随时间向前积分浅水方程。

相比之下，在无条件配置（无长波辐射）下，模型仅接收静态地形描述符、风暴开始时的前期水动力状态以及降水序列。模型不会获得有关预报时段的任何水动力信息。因此，这种模式是真正的端到端代理模型，可将强迫和初始状态直接映射为四天的淹没演变；如果完全不运行水动力模型，这便是可部署的配置。其误差显著较大，详见第4节；这反映的是该任务本身更具难度，而非架构失效。

全文均报告这两种模式，而它们之间的对比构成了本文提出物理引导条件化论点的定量依据。

3.8. 超参数选择与训练

针对图 5 所示的去噪扩散概率模型架构，我们进行了超参数敏感性分析，以确定 U-Net架构中的扩散去噪时间步数和基础通道数。

训练了六种配置，涵盖 $\{100, 250, 500\}$ 扩散步数和 $\{32, 64\}$ 基础通道数的所有组合。

为在同等条件下比较各项配置，所有模型均使用相同的图形处理器硬件、学习率和批次大小，固定实际运行时间为四小时进行训练。

采用水深均方误差（MSE）和临界成功指数（CSI）对各项配置进行了比较，包括 100m、150m 和无条件配置（图7）。

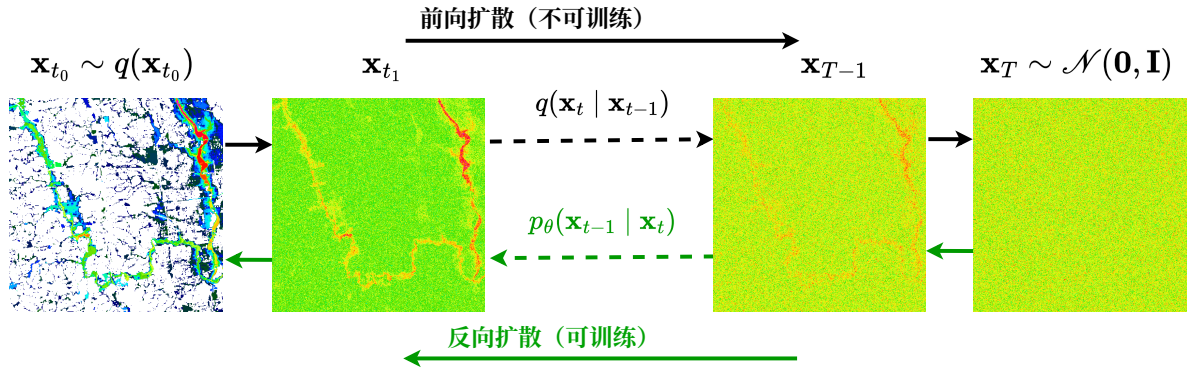
结果表明，对于 U-Net 基础通道维度较大的 64 配置，将扩散去噪时间步从 100 增加到 500 通常能够提升预报性能，使 MSE 值降低、CSI 值升高。相比之下，对于基础通道维度较小的 32 配置，增加扩散去噪时间步往往会使 MSE 升高并导致 CSI 值降低。在此采用的等实际运行时间训练方案下，较大的配置在规定的四小时内完成的优化器更新次数较少，因此在解读图 7时应注意这一混杂因素：该比较反映的是单位训练预算下可达到的准确度，而不是收敛时的准确度。

具体而言，采用 32 个基础通道和 100 个扩散步的配置，其表现与采用 64 个基础通道和 500 个扩散步的配置相当。正如表4所总结的，在相同的 GPU 硬件上，无论增加扩散去噪时间步数还是 U-Net 通道数，都会显著延长四天流域尺度预报的推理时间。因此，本研究最终采用的去噪扩散概率模型（DDPM）配置在训练和推理时均使用 100 个扩散去噪时间步，U-Net 的基础通道维度为 32。

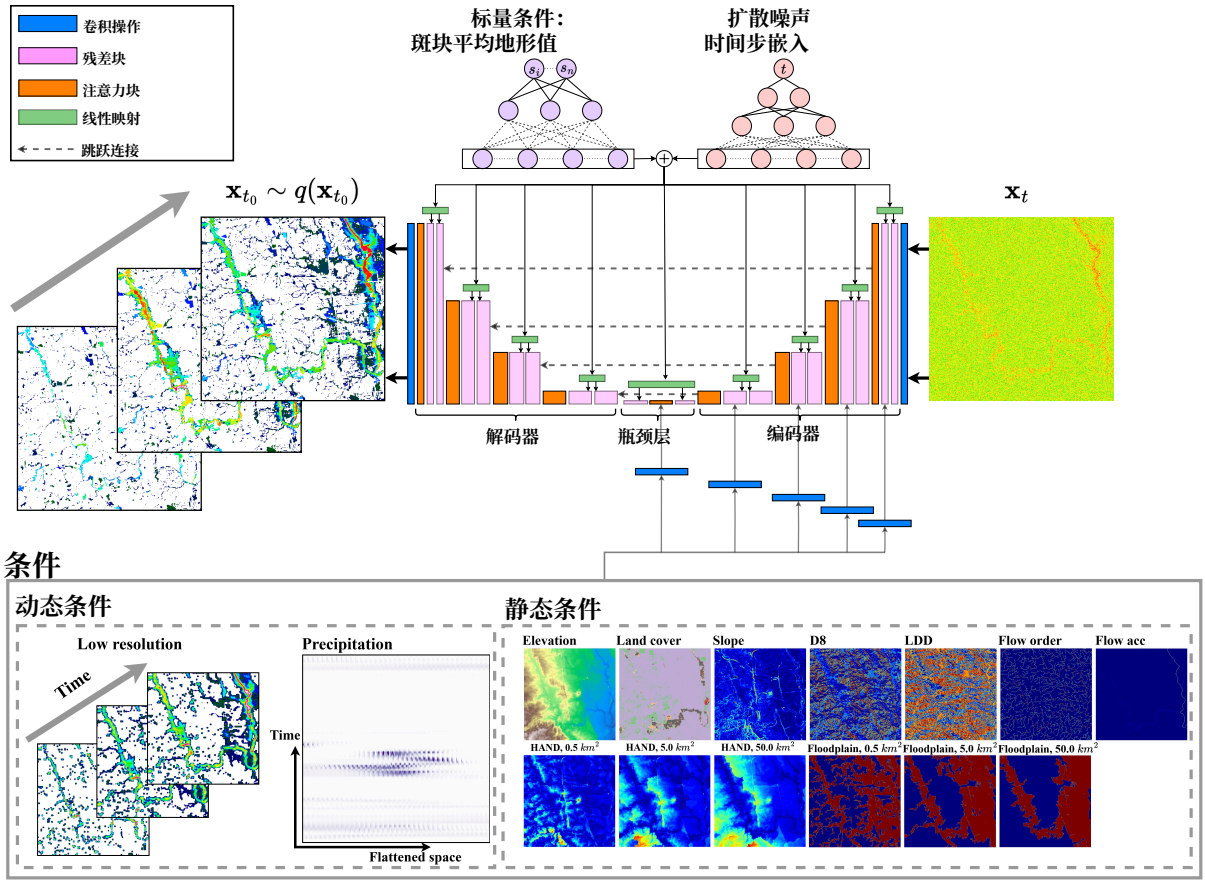
对于第4节展示的洪水预测结果，我们针对三种条件配置（100 m LR、150 m LR 和无 LR）分别训练了一个模型。每个模型均使用 Adam优化器 [91]，并采用恒定学习率 5×10^{-5}、批次大小为 4 个图像块以及混合精度（FP16）运算；扩散去噪时间步数为 100，U-Net 的基础通道维度为 32。100 m 配置训练了 49 轮，150 m 和无长波辐射配置训练了 50 轮，此时训练损失已趋于平稳，接近 1×10^{-4}。各配置所报告的模型均为训练损失按轮次取平均后最低的检查点。数值实验在阿贡国家实验室（Argonne National Laboratory）实验室计算资源中心（LCRC）的 SWING GPU集群上进行，每次训练使用一块 NVIDIA A100图形处理器。

表4：在单个 NVIDIA A100图形处理器上，对完整的四天流域尺度洪水预报进行 DDPM推理所需的时间（由测得的每个分区推理时间计算得出），其中包含 26 个不重叠斑块，每个斑块包含 512×512 个网格单元。该时间针对 U-Net 基础通道维度和扩散去噪时间步数的不同组合。数值为三种条件配置下的平均值，各配置的计时结果相差不到 2%。

基础通道	扩散步数	推理时间 [秒]
32	100	123
32	250	292
32	500	578
64	100	163
64	250	390
64	500	743

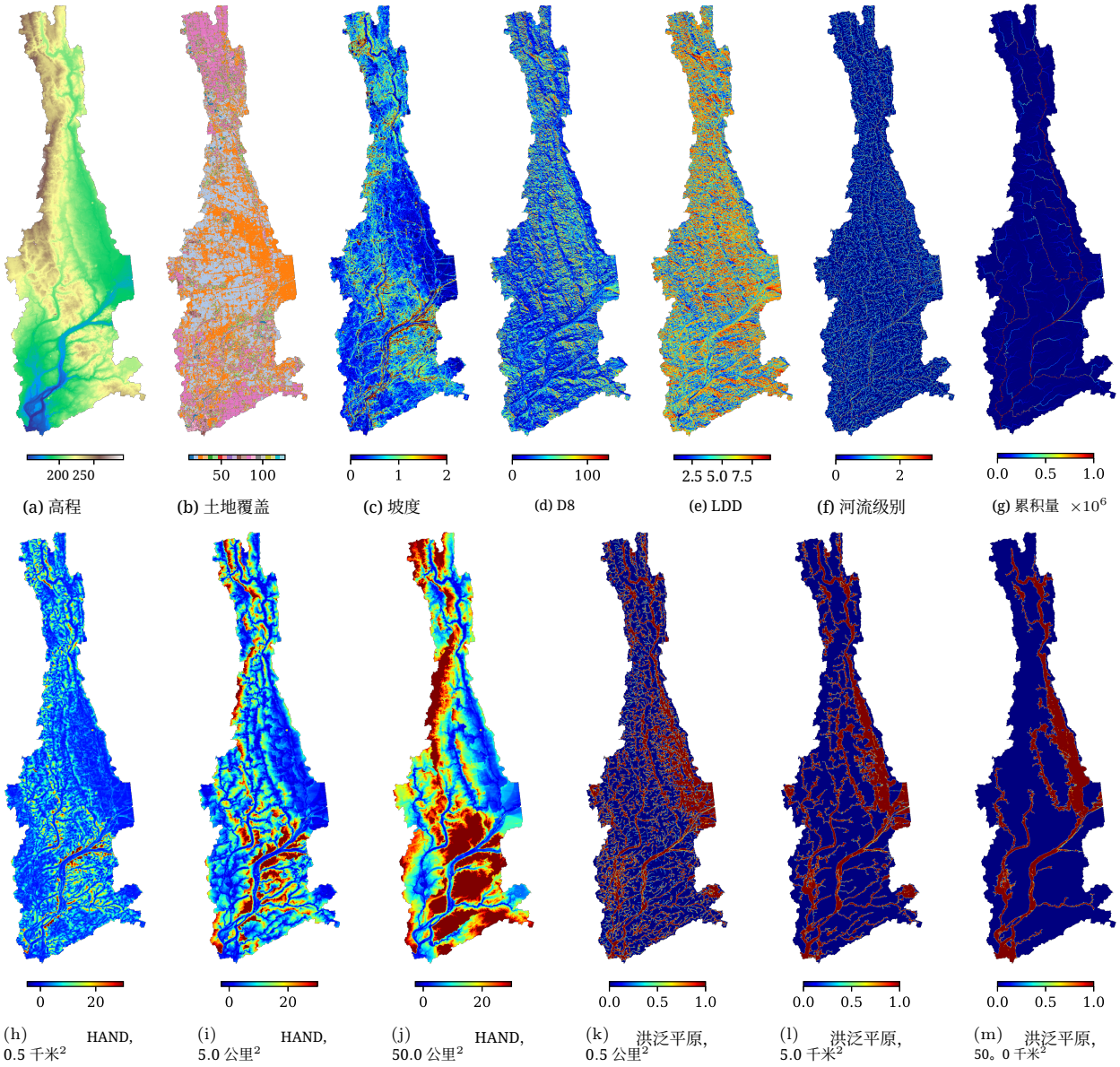


(a) 去噪扩散概率模型（DDPM）的前向扩散和反向去噪过程。前向扩散过程被构建为一个马尔可夫链，逐步向输入数据添加高斯噪声，直至其接近标准高斯分布。反向去噪过程是一种可训练的生成过程，它从含噪潜在分布中迭代重建干净的数据样本。



(b) 用于时空洪水淹没预报的去噪扩散概率模型（DDPM）的U-Net架构。该网络在反向去噪过程中根据含噪潜在输入场预测高分辨率时空水深淹没图。DDPM同时使用静态和动态物理输入进行条件约束。静态条件变量来源于数字高程模型（DEM），而动态条件变量包括时空降水强迫、初始水动力状态，以及由水动力模拟生成的可选低分辨率时空洪水淹没图。

图5：用于时空洪水水深淹没预报的条件去噪扩散概率模型架构。



(a) 高程 (b) 土地覆盖 (c) 坡度 (d) D8 (e) LDD (f) 河流级别 (g) 累积量 $\times 10^6$

(h) HAND, 0.5 千米^2 (i) HAND, 5.0 公里^2 (j) HAND, 50.0 公里^2 (k) 洪泛平原, 0.5 公里^2 (l) 洪泛平原, 5.0 千米^2 (m) 洪泛平原, 50。0 千米^2

图6：条件去噪扩散概率模型框架所使用的静态条件变量。数字高程模型（DEM）取自美国地质调查局三维高程计划（3DEP），空间分辨率为30 m；土地覆盖数据取自国家土地覆盖数据库（NLCD）2021，分辨率为30 m。其余地貌与水文条件变量，包括坡度、流向、河流级别、汇流累积、HAND和地貌洪泛平原图层，其推导基于数字高程模型，并使用`pyflwdir` 软件包 [87]。

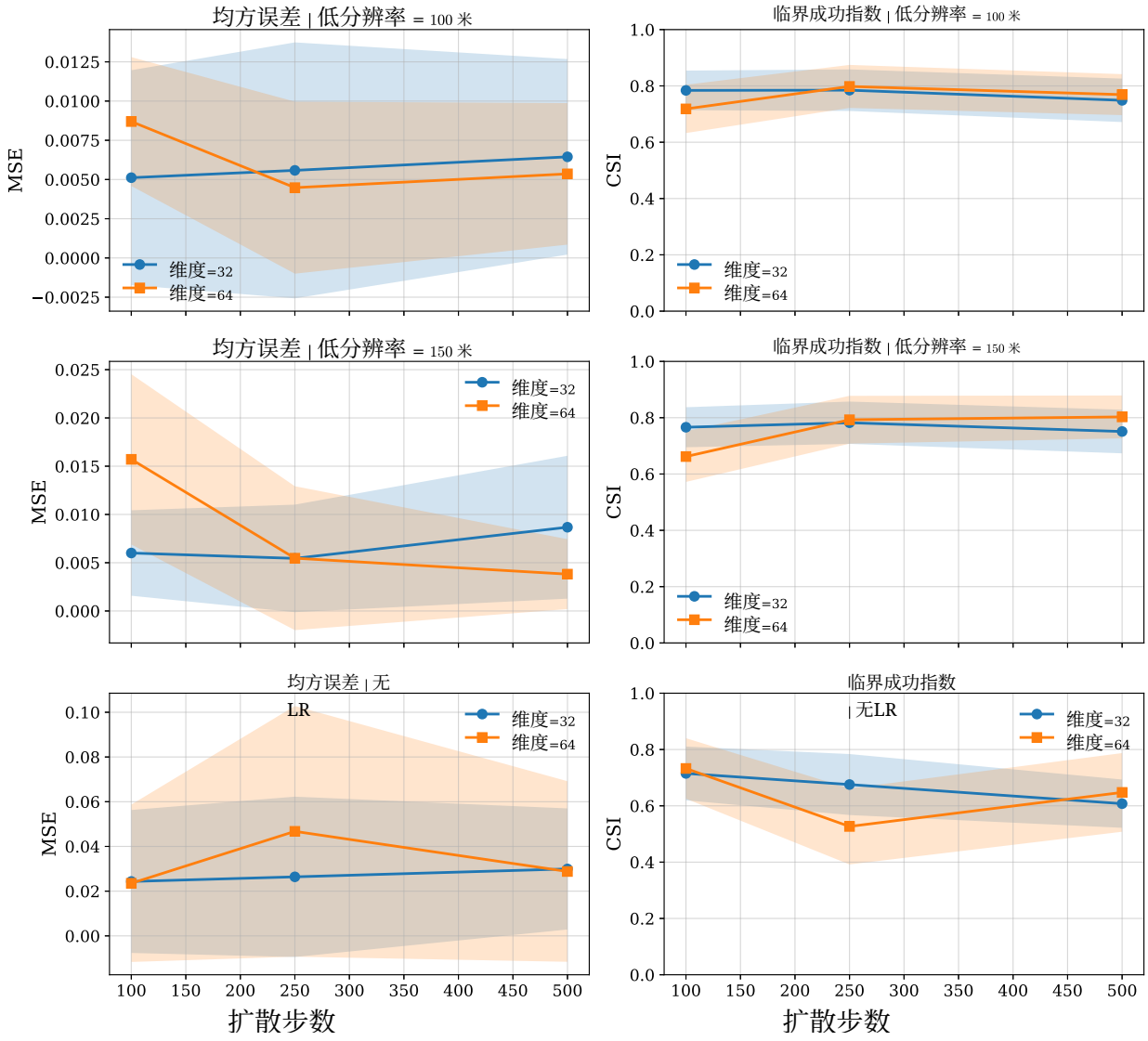


图 7：水深　　均方误差（MSE）（左）和临界成功指数（CSI）（右），对应不同的扩散去噪组合　DDPM 架构中使用的时间步数和 U-Net 基础通道维度，如图5所示。

4. 结果与讨论

本节报告的所有预报均为集合预报。如第3.6节所述，反向扩散过程是随机的，因此对训练好的模型进行一次评估，只会从已学习的预测分布中得到一次任意抽样，其本身并不能构成定义明确的预报产品。因此，我们始终报告20成员集合：需要单一场时给出集合平均，以不确定性为关注量时给出集合分布。单次实现的结果不作为预报产品呈现。

比较了三种条件配置：分别以100米和150米低分辨率SynxFlow模拟为条件的去噪扩散概率模型，以及仅接收静态地形、前期水动力状态和降水的无条件配置（“无LR”）。所有结果均对照作为参考的30 米 SynxFlow 模拟进行评估。

评估范围。该集合实验涵盖表2中的六个留出的测试事件、图8中突出显示的六个分析区块，以及0.1毫米/小时前期湿润情景，共计 $3\times6\times6\times20$ =2160 次四天分区预报。这六个图像块因其较高的人口密度和城市暴露度而被选取，并共同覆盖流域的北部、中部和南部河段。它们是用于训练的60个重叠分区的子集，因此下文的汇总评分具有流域代表性，而非覆盖整个流域；第4.4 节表明，它们能够密切再现全流域汇总。除非另有说明，当某网格单元的水深超过0.15米时，即将其判定为淹没，该标准对预报和参考同样适用。

4.1. 预报淹没的时空演变

图 9 至 14 展示了事件F（峰值小时降雨强度最大，145.4 毫米/小时，表 2）下六个分析区块的集合平均水深场。每幅图均对比了30米参考结果与三种条件配置在24、48、72和96小时预见期的集合平均结果。

条件集合在淹没范围和时序上均与参考结果高度吻合。干流沿线的河道淹没出现在正确的预见期，洪泛平原按照正确顺序被淹没，且72至96小时之间的退水过程也得到了再现。图中标注的各面板均方根误差值，在所有预见期内、两种条件配置下均保持在约0.09米以下。

无条件集合的表现则在定性上截然不同，且偏差方向始终一致：洪水来得过早，淹没范围也过大。对于区块8（图 9），在参考结果几乎完全干燥的24小时预见期，它就已显示出大范围的河道和洪泛平原淹没；到72小时时，其均方根误差达到0.26米，而100米和150米条件约束集合的误差分别为0.076米和0.029米。其他区块也反复出现同样的过早、过度淹没模式。由于无条件模型未接收有关预报时段的水动力信息，因此必须仅凭降水描述符和前期状态推断洪水波的时序；显然，它消解这种不确定性的方式，是提早淹湿地貌洪泛平原，而非传播洪水波。

这是核心定性结果：低分辨率水动力条件提供了生成模型无法仅凭强迫恢复的时间结构，而扩散模型则提供了粗分辨率模拟无法分辨的精细尺度空间细节。

4.2. 洪水范围技巧

图15 至20将预报洪水范围分解为命中、漏报和虚警，分别对应按流域平均累积量衡量最弱的事件（事件A）和峰值强度最大的事件（事件F），时间为参考淹没最大值所在日。列联指标定义为

$$HR=\frac{TP}{TP+FN},\qquad MR=\frac{FN}{TP+FN}=1-HR,\qquad FAR=\frac{FP}{TP+FP},\qquad CSI=\frac{TP}{TP+FP+FN}, \tag{7}$$

其中， TP、 FP和 FN分别表示预报与参考中均被淹没的网格单元数、仅在预报中被淹没的网格单元数，以及仅在参考中被淹没的网格单元数。计数范围为两个场均有效的网格单元。每个面板均标注了

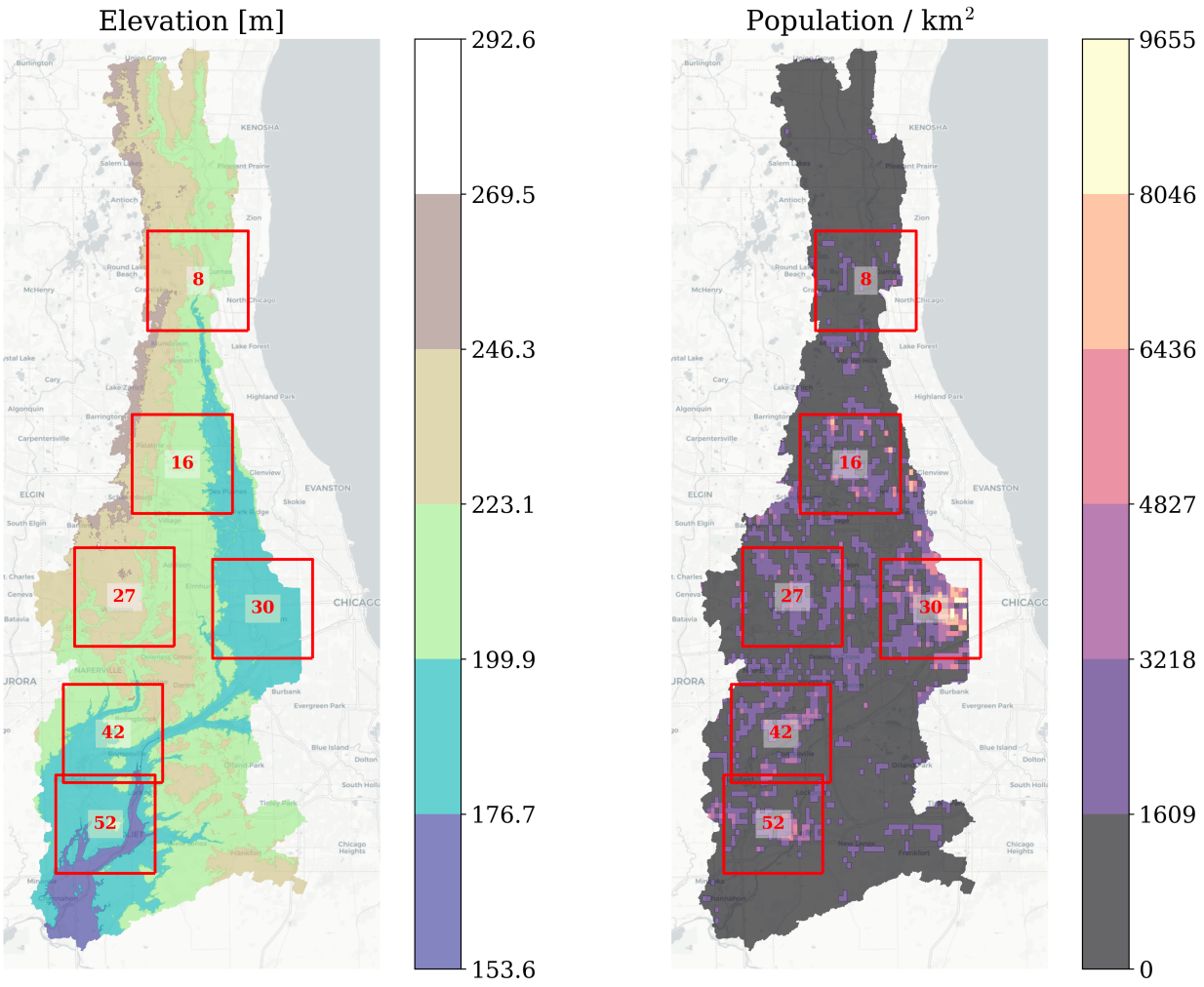


图8：本节通篇使用的六个分析区块，分别显示在流域地形高程（左）和每km^2 人口密度（右）上。这些图像块因高人口密度和城市暴露度而被选取，并覆盖德斯普兰斯河流域的北部、中部和南部河段。

集合平均的评分；对于临界成功指数（CSI），还标注了该评分在20个单个成员中的标准差，该标准差反映了任一实现的技巧水平变化幅度。

两种情形显而易见。对于较弱的事件A，计算域的淹没比例较小，三种配置均高估了淹没范围，但程度差异很大：在区块8，无条件集合达到 $FAR = 0.96$ 和 $CSI = 0.04$，相比 $FAR = 0.55$、 $CSI = 0.41$（100 m条件约束）以及 $FAR = 0.73$、 $CSI = 0.25$ （150 m条件约束）。对所有配置而言，弱事件都是最困难的情形，这与它们在训练目录中代表性不足相一致（图4）。

对于严重的事件F，情况恰好相反，评分变得较高且差异明显：所有配置的命中率都接近1，漏报很少，而配置之间的差异几乎完全体现在虚警上。在区块8，150米条件约束集合达到 $CSI = 0.90$ ，其 $FAR = 0.09$，而100米配置为 $CSI = 0.82$， $FAR = 0.18$，无条件配置为 $CSI = 0.74$， $FAR = 0.26$。整个过程中，CSI的成员间标准差较小，介于0.02和0.06之间，这证实了配置之间的技巧差异远大于集合内部的抽样变异。

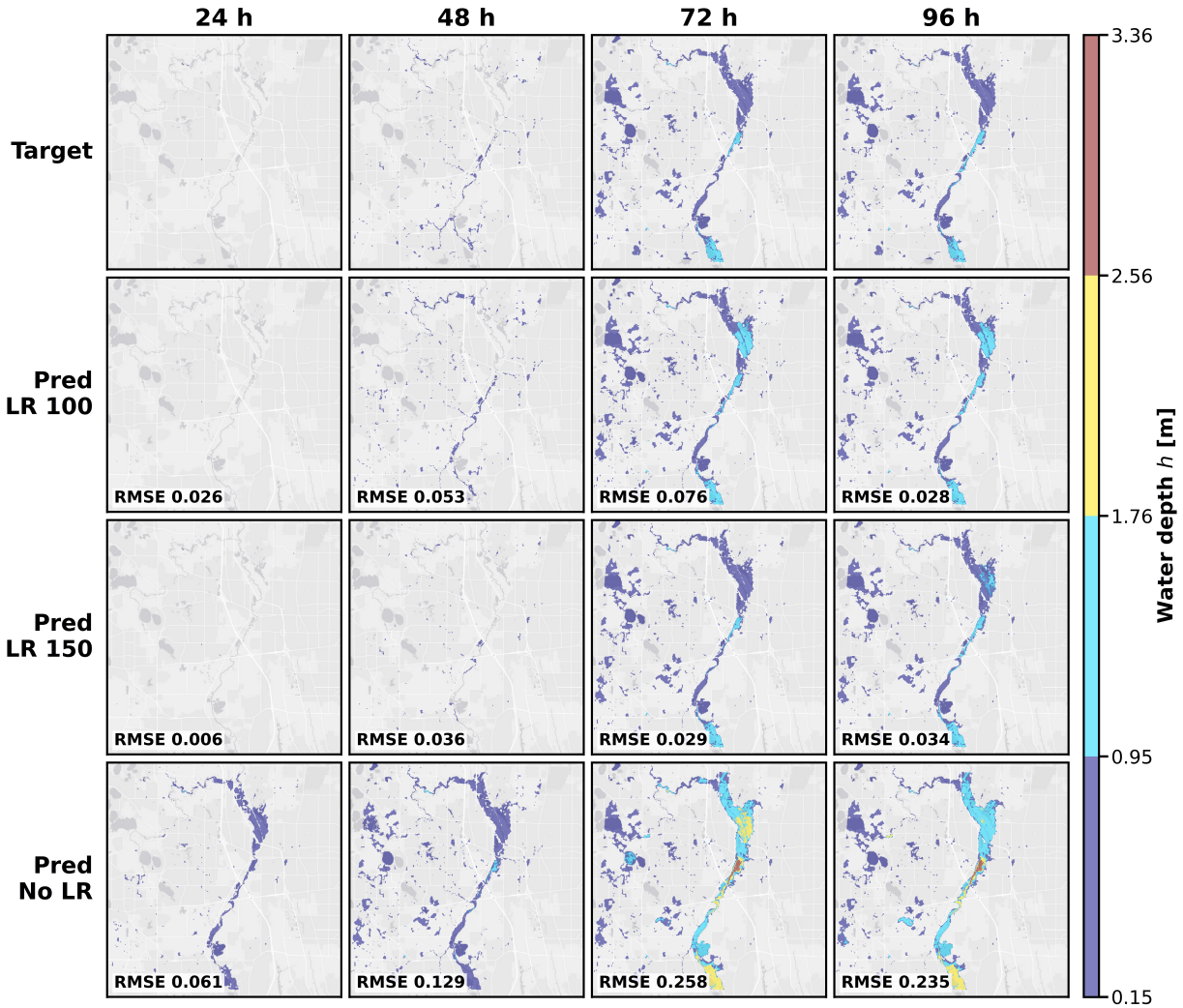


图9：事件F在区块8的集合平均水深演变。各行依次为：30 m SynxFlow 参考结果，以及 100 m 条件约束、150 m 条件约束和无条件配置的 DDPM 集合平均值（$N = 20$）。各列依次为预见期 24、48、72 和 96 h。低于 0.15 m 淹没阈值的水深以透明显示；标注的均方根误差为集合平均值相对于参考结果的误差。

4.3. 概率洪水图

集合与确定性预报之间的区别在于超越概率。图21 展示了 $P(h > 0.15$ 米)，即20个成员中超过淹没阈值的比例，覆盖事件F的全部六个分析区块，并叠加了参考洪水边界。

概率场在空间上连贯，与参考轮廓高度吻合：河道和内侧洪泛平原内的网格单元概率接近 1，谷坡上的网格单元概率接近 0。具有参考价值的结构是两者之间的过渡带；在此区域，条件约束无法确定湿润锋的位置，也是实际应用人员会重点关注的区域。

不过，这一带很窄，通常只有几个网格单元宽，因此概率场接近二元。第4.5节量化的离散度不足，直观地体现为这种锐利特征：集合预报很有把握，而且其把握程度超出了自身误差所能证明的范围。因此，这些地图有助于勾勒洪水的不确定边缘，但不应将其概率值解读为校准后的超越频率。

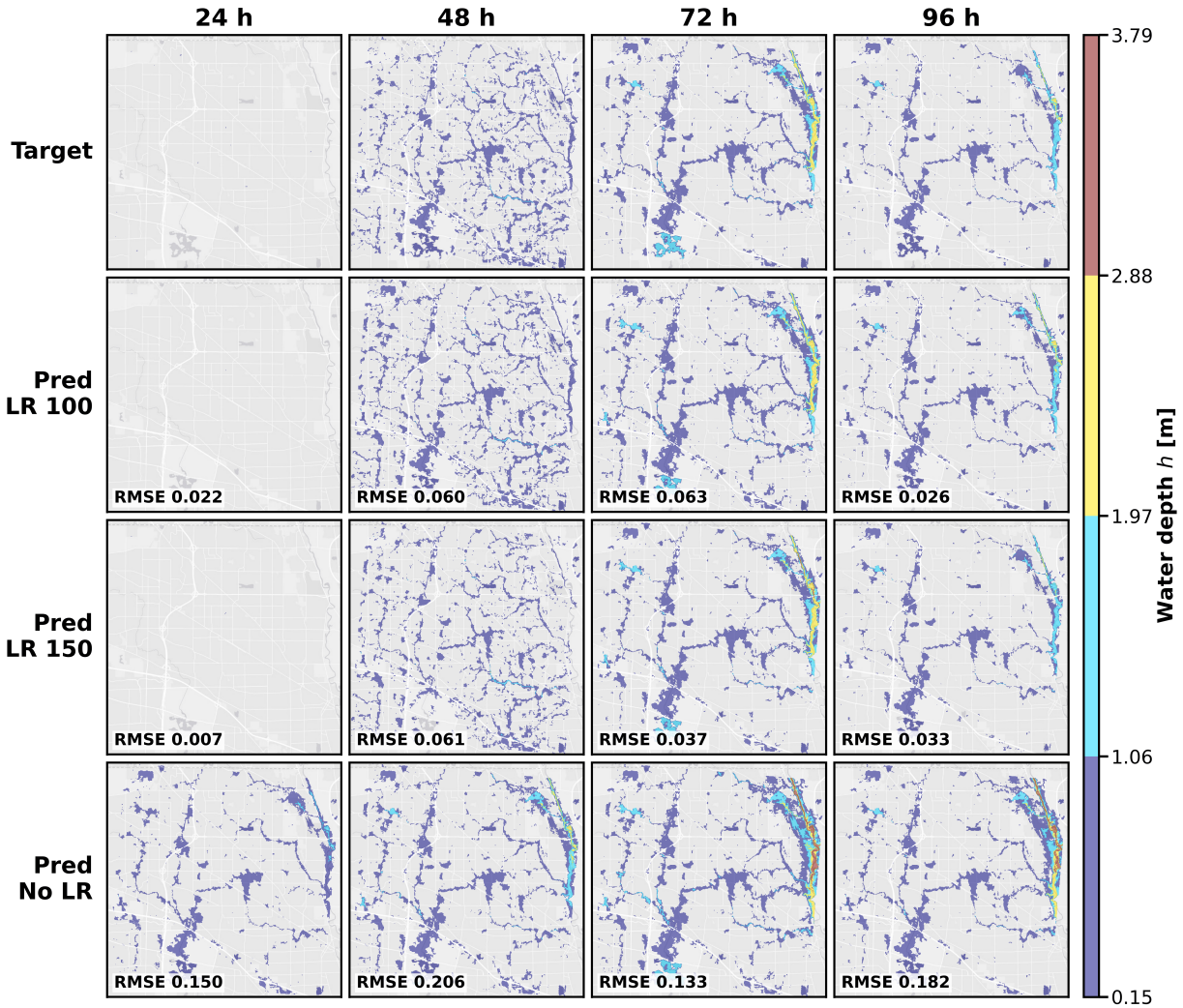


图 10：事件F在第 16 个区域的集合平均水深演变，绘制方式与图 9相同。

4.4. 汇总性能

图22 汇总了六个分析区块和六个测试事件中集合平均预报的结果。计算评分前，先对各图像块的平方误差和列联计数进行汇总，因此，淹没区域较大的图像块会按其范围占比产生相应影响。表5 列出了相应数值。

左图显示集合平均的汇总均方根误差随预见期的变化。两个条件集合的误差都从第 1 天的较小值逐渐增大，在第 3 天达到峰值，之后回落；150 m 条件约束下的峰值为 0.067 m，100 m 条件约束下则为 0.086 m。无条件集合在每个预见期的误差都是最大的：其第 1 天的均方根误差（0.078 m）已经超过 150 m 配置在第 3 天的峰值，并在第 3 天达到 0.250 m。其阴影带宽度是其他集合的三至五倍，表明误差很大程度上取决于所预报的风暴。

右侧面板和表 5汇总了峰值日的列联评分。无条件集合的 CSI（0.678）和 FAR（0.271）表现最差，其四日平均 RMSE 为 0.170 m，是任一条件配置的三倍以上。两种条件配置的表现模式一致，这也是本研究的主要定量发现：100 m 条件约束的命中率更高（0.957，对比 0.913），但 150 m 条件约束在其余三项评分上更优，误差更低

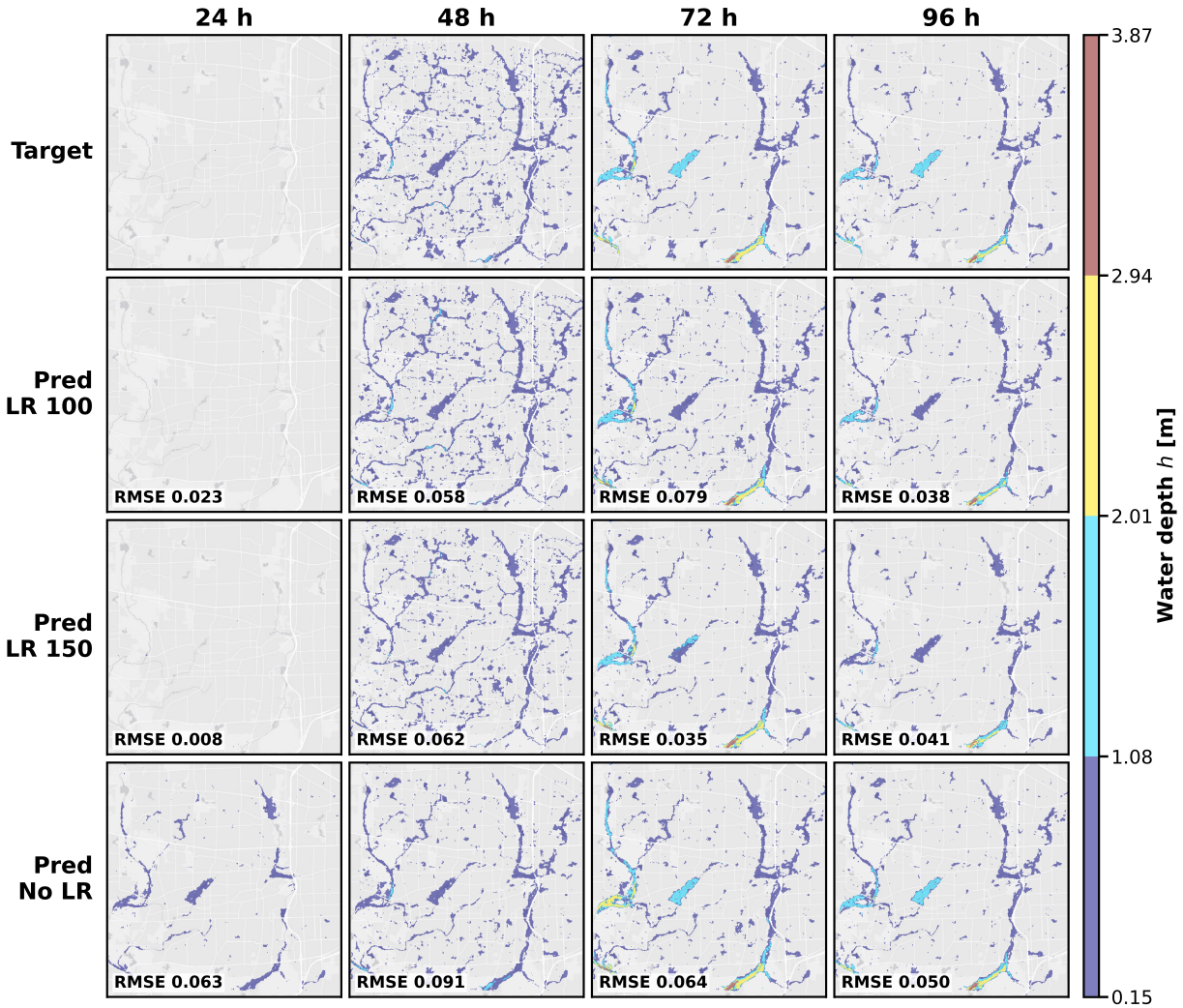


图11：事件F中分块27的集合平均水深演变，绘制方式与图9相同。

RMSE 更低（0.048，对比 0.054 m）、CSI 更高（0.839，对比 0.785），且虚警比显著更低（0.089，对比 0.179）。

淹没范围图展示了这一机制：100 m 条件约束集合产生了系统性偏湿的预报，使命中和虚警同时增加，最终导致基于重叠度的临界成功指数（CSI）下降。第4.5 节的秩直方图也独立证实了这一偏差。由于 150 m 条件约束的计算成本也较低（第4.6节），在本文考察的范围内，更精细的条件化分辨率并未提升洪水制图效果，因此较粗的选项具有更好的精度—成本权衡。

为检验这六个分析区块是否具有代表性，我们将这些汇总值与单次实现预报进行了比较；这些预报覆盖全部60个图像块和全部三种前期情景，仅用于此项检验。无条件配置、150 m配置和100 m配置的预报四日平均均方根误差（RMSE）分别为0.190、0.050和0.058 m，临界成功指数（CSI）分别为0.689、0.832和0.787。CSI差异约在0.02以内、RMSE差异约在0.02 m以内，这表明这六个分析区块能够代表整个流域，且三种配置的排名并非评估范围缩小所致的假象。

图23r 通过测试事件得出相同的四项评分。各配置的排序保持稳定
在各事件中均如此 临界成功指数和虚警率。无条件集合在较弱的事件A中表现明显最差

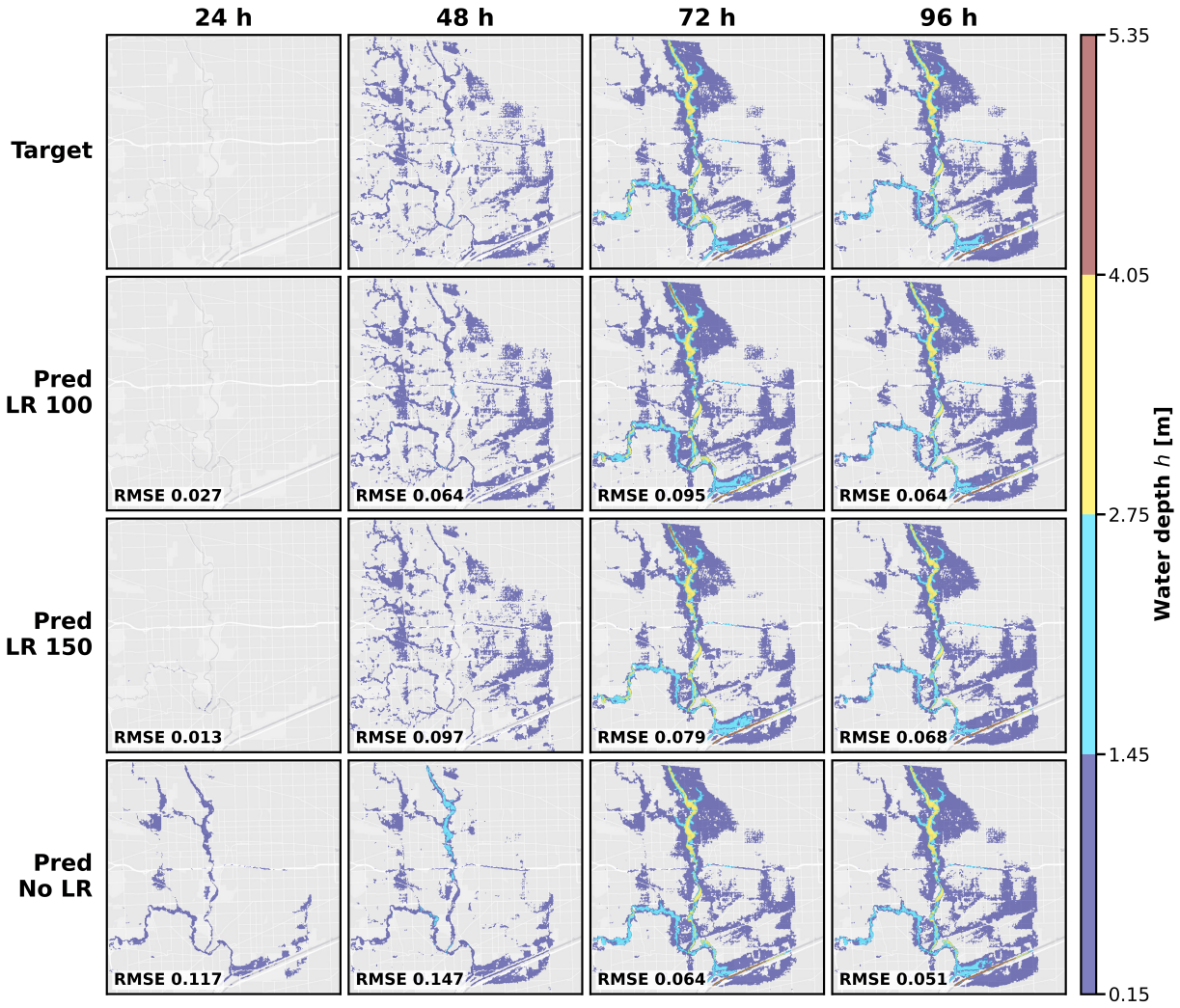


图12：事件F在区块30处的集合平均水深演变，绘制方式同图9。

而在严重程度更高的事件中差距缩小，这与淹没范围图中观察到的模式一致：较大的风暴会产生空间上更连片的淹没，并与作为静态条件输入提供的地貌洪泛平原和 HAND图层相吻合；在没有水动力引导的情况下，这类模式也更容易重建。正如第 2.4 节所述，事件A–F代表六种不同的严重程度等级，而非单调变化的严重程度轴，因此横轴应视为分类变量。

表5：在0.1毫米/小时前期情景下，六个分析区块和六个测试事件的汇总集合平均性能。RMSE为四个预报日的合并平均值；HR、CSI和FAR为参考淹没最大值当日的合并值。各指标的最佳值以粗体标出。

配置	RMSE [米] ↓	命中率 ↑	临界成功指数 ↑	虚警率 ↓
DDPM + 低分辨率（100 米）	0.054	**0.957**	0.785	0.179
去噪扩散概率模型 + 低分辨率（150 m）	**0.048**	0.913	**0.839**	**0.089**
去噪扩散概率模型（无 LR）	0.170	0.928	0.678	0.271

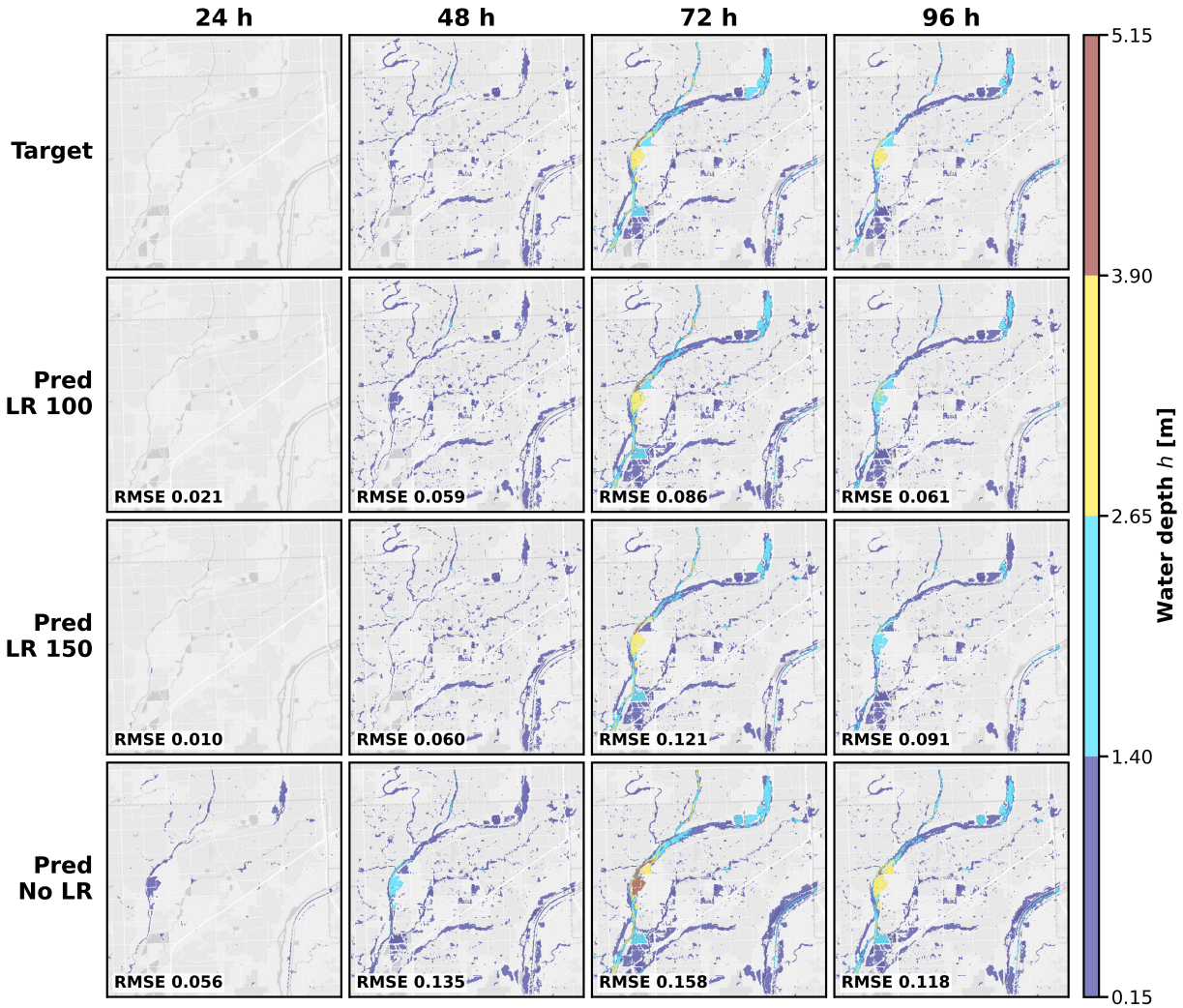


图13：事件F在斑块42处的集合平均水深演变，绘制方式同图9。

4.5. 概率检验与校准

上述结果使用集合平均作为点预测。本小节评估集合分布，以确定第3.6节中的概率产品是否可信。我们采用四种标准诊断指标，均针对四个预见期中的每一个，在有效网格单元上计算。

连续排序概率评分 [92, 93] 是一种严格适当的评分规则 [94]，将平均绝对误差推广到概率预报。这里采用公平（无偏）的集合估计量，

$$\mathrm{CRPS} = \frac{1}{N}\sum_{i=1}^{N}|x_i - y| - \frac{1}{2N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}|x_i - x_j|, \tag{8}$$

其中 x_i 为成员，y 为参考值。**离散度—技巧一致性** 比较集合方差均方根与集合平均的均方根误差；对于统计可靠的集合，两者应当一致 [95]。**秩直方图** [96, 97] 统计参考值在排序后的集合中的秩，仅针对参考值为湿的网格单元；平坦直方图表示校准良好，U形表示离散度不足，非对称尖峰则表示存在系统偏差。**布里尔评分** [98] 和可靠性图用于评估超越概率。

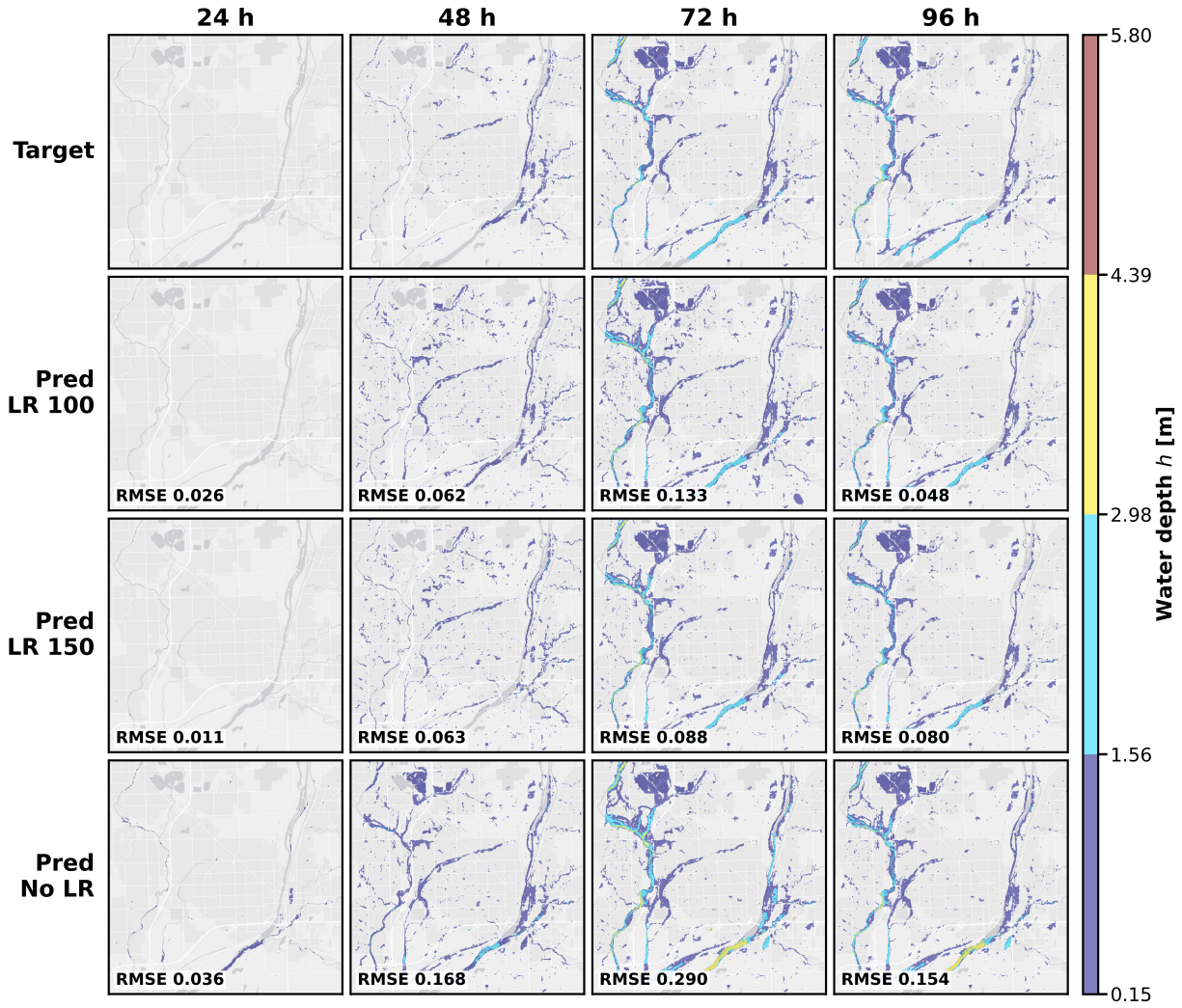


图14：事件F在斑块52处的集合平均水深演变，绘制方式同图 9。

表 6和图 25 报告了结果。与第 4.4节的汇总评分不同，此处的每项评分均针对每个分区、事件和预见期分别计算，再以相同权重取平均，因此集合平均均方根误差的数值与表 5略有不同（例如，对于 150 m 配置，分别为 0.045 m 和 0.048 m），但排名不变。由此得出三项发现。

首先，对成员进行平均是有益的，而且条件约束越弱，收益越大。与单个成员的平均值相比，集合平均使100 m、150 m和无条件配置的水深均方根误差分别降低了4.9%、8.2%和10.7%。这一排序符合预期，因为无条件模型的自由度最高，因此可通过平均消除的单次实现误差占比也最大。

其次，配置的概率排名重现了点预测排名，而且差异更加显著。150米条件约束集合在所有概率评分中均取得最佳值：连续排序概率得分（CRPS）为0.019 m，而100 m配置为0.036 m、无条件配置为0.040 m；布里尔评分为0.014，而另外两者分别为0.019和0.045。两种条件配置之间的CRPS差距，按比例远大于其集合平均均方根误差的差距，这表明，较粗分辨率条件约束的优势在于其预测分布的形状，而不仅仅在于中心估计的准确性。

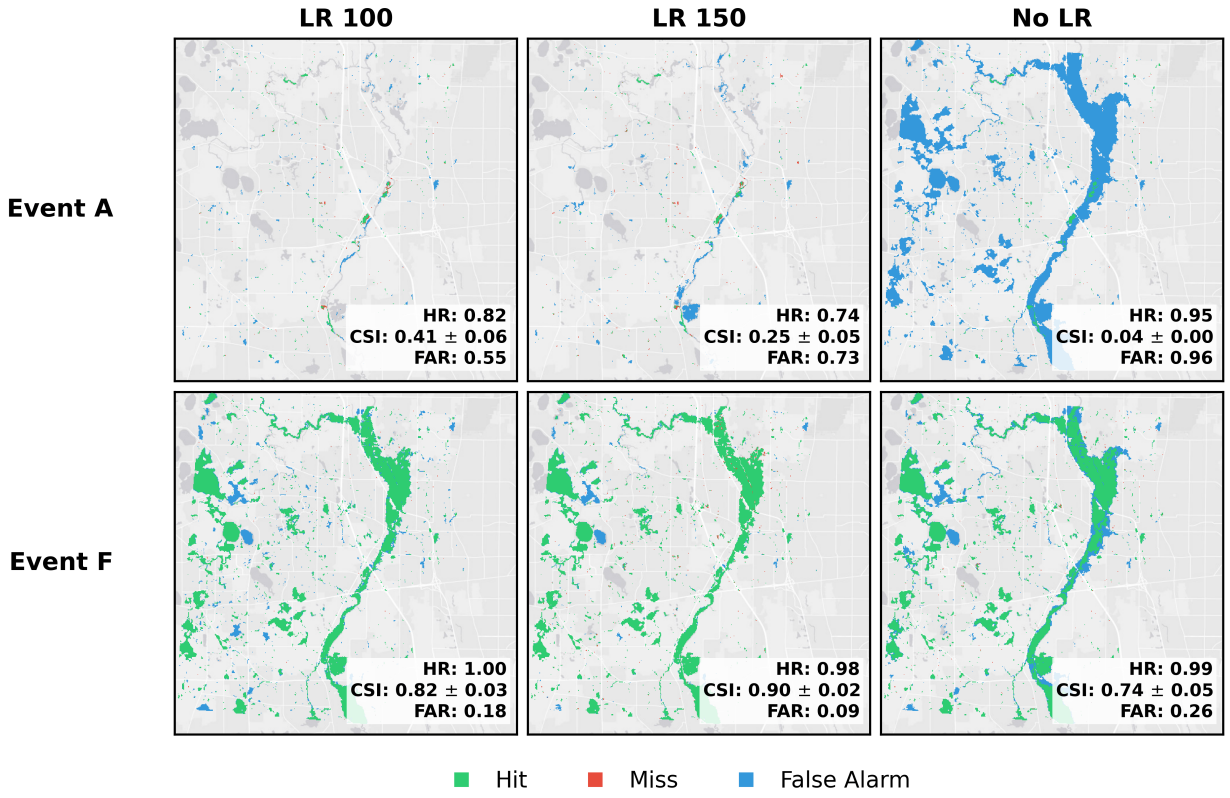


图15：参考淹没最大值当天，区块8处集合平均洪水范围分解为命中、漏报和虚警。行：事件A和F。列：三种条件配置。标注给出集合平均的命中率、临界成功指数（以及20个成员间临界成功指数的标准差）和虚警比。

第三，对于解读概率图至关重要的是，三个集合的离散度都明显**不足**。100 m、150 m和无条件配置的集合离散度与集合平均误差之比分别为0.35、0.45和0.73，而统计一致集合的该比值应为1。这些集合都过度自信，其中条件约束信息最丰富的配置尤为如此。图25d中的秩直方图在所有情况下都呈明显的U形，并表现出诊断性不对称：100 m配置中有33%的参考值落在最低秩，这意味着参考水深低于所有成员的预测水深，且预报偏湿；而150 m配置中有31%的参考值落在最高秩，表现为偏干。这两种特征分别证实了第4.4节中对比鲜明的命中率和虚警表现，这些结论是从预报的分布结构而非阈值化范围中得出的。图25c中的可靠性图与此一致：在大部分概率范围内，100 m配置和无条件配置的曲线位于对角线下方，而150 m配置的曲线最接近对角线。

图25b 得出了一个实际结论。由于生成式离散度较窄，连续排序概率得分和集合平均均方根误差在成员数约为 3 至 5 个时趋于稳定，此后基本保持不变；将 N 从 5 增至 20，对所有配置的连续排序概率得分影响均不到 2%。因此，对于按当前方式训练的该系统，无需使用大型集合，这显著降低了概率产品的成本。更深层的启示是，重采样扩散噪声并不能构成完整的不确定性预算：其产生的离散度是真实存在的，但范围较窄；业务系统还需要扰动气象强迫，采用第3.6节所述的嵌套设计。直接校准生成组件，例如通过事后离散度膨胀，或采用离散度感知目标函数进行训练，是一种自然的扩展，但本研究未对此进行尝试。

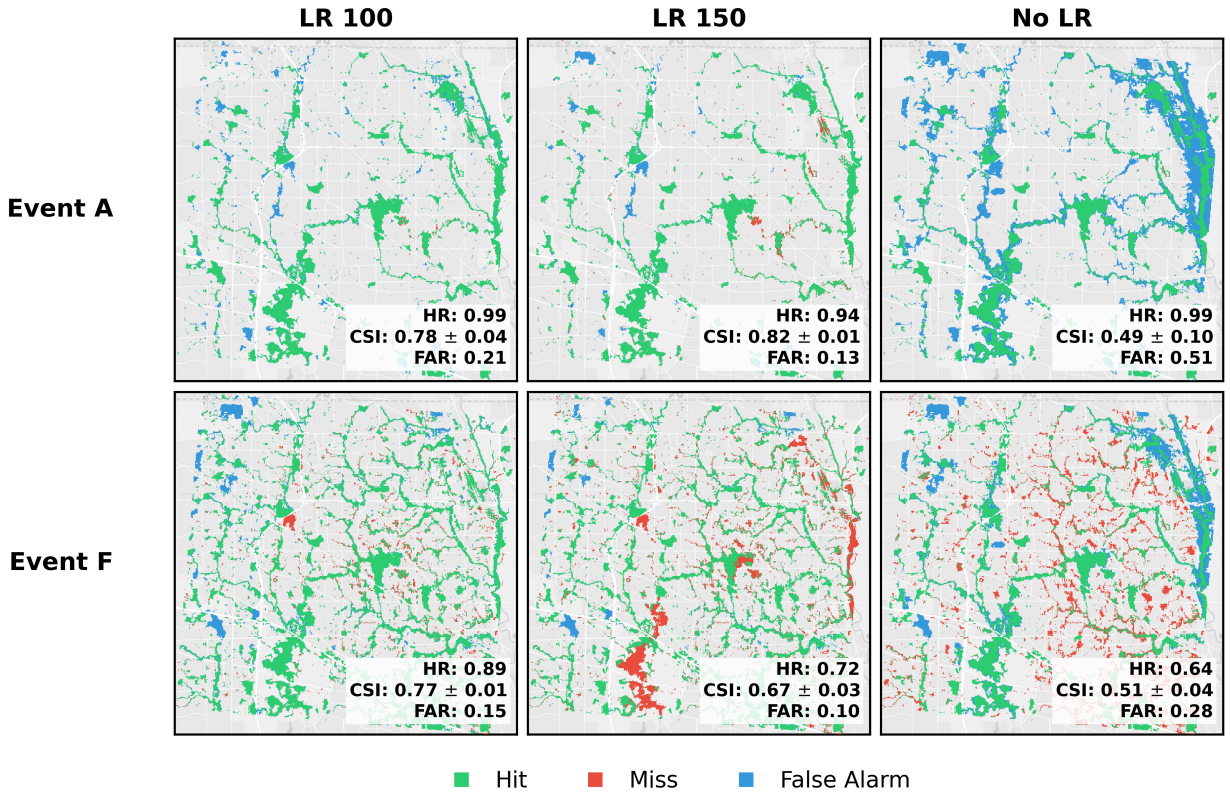


图16：集　　合平均洪水范围分解为第 16 个区域中的命中、漏报和虚警，如图　　15.

表6：20成员集合在六个分析区块和六个测试事件上的概率检验，按区域、事件和预见期分别计算，并以相同权重取平均（因此与表5中的汇总均方根误差略有差异）。水深单位为米。离散度误差比低于 1 表示集合离散度不足（过度自信）。每列最佳值以粗体显示。

配置	RMSE 单成员	RMSE 集合平均 ↓	连续排序概率得分 ↓	离散度/误差比	Brier ↓
去噪扩散概率模型 + 低分辨率 (100 m)	0.056	0.053	0.036	0.35	0.019
去噪扩散概率模型 + 低分辨率 (150 m)	**0.049**	**0.045**	**0.019**	0.45	**0.014**
去噪扩散概率模型（无 LR）	0.176	0.158	0.040	**0.73**	0.045

4.6. 计算成本

框架的计算成本通过对德斯普兰斯河流域进行完整的四天预报来评估，结果汇总于表7。所有计时均在同一集群的 NVIDIA A100 GPU 上测得。进行业务推理时，将流域划分为 26 个互不重叠的 512 × 512分区，DDPM 的成本则根据实测的单分区推理时间计算。

用于构建训练数据集的水动力模拟在八块GPU上执行；对339次存档模拟取平均后，一次121小时的运行所需实际运行时间在30米分辨率为538秒，在100米为76秒，在150米为43.6秒。由于去噪扩散概率模型在单GPU上进行推理，因此额外针对测试事件F在最干旱前期情景下对求解器开展了专门的单GPU基准测试，从而给出表7中同等条件下的单设备成本。一次单成员流域尺度四天预报包含100个去噪步下的26次分块评估，在一块GPU上条件配置耗时120秒，无条件配置耗时122秒，并且基本上与条件配置无关（表4）。

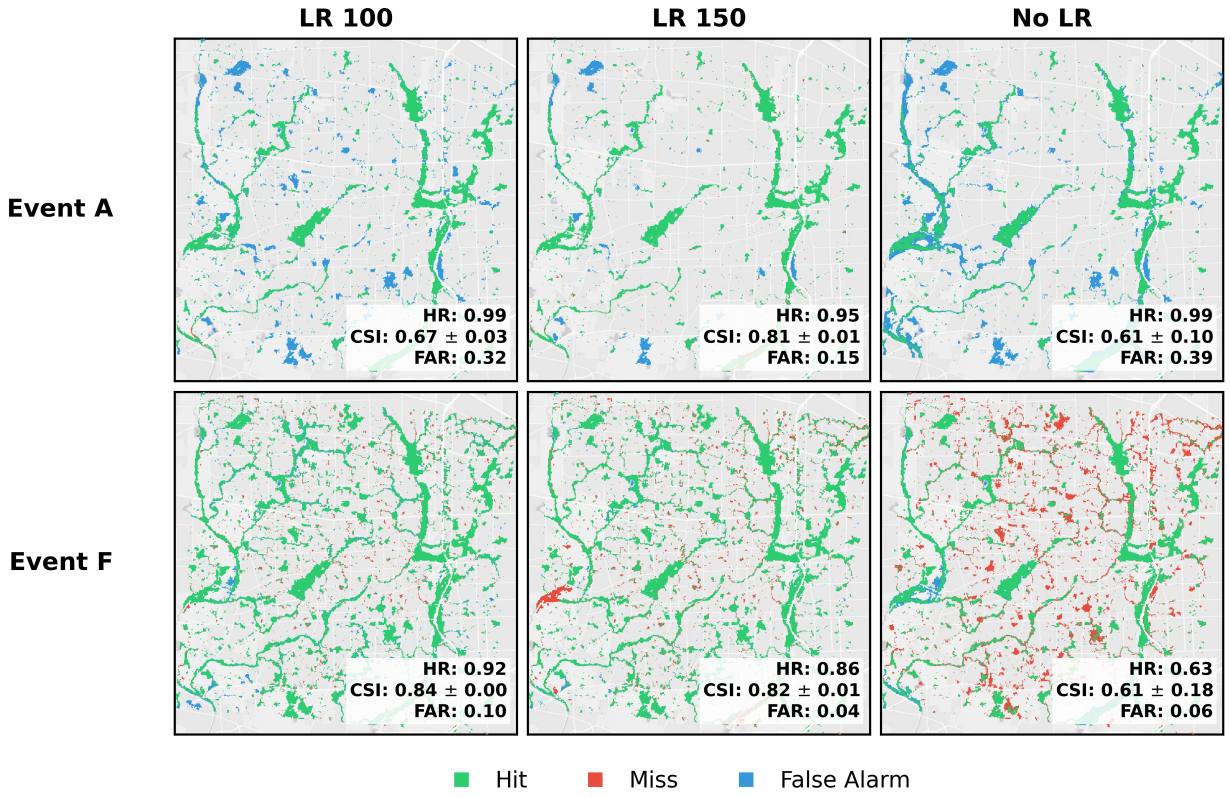


图17：集　　合平均洪水范围在分块27处拆分为命中、漏报和虚警，绘制方式同图　　15.

相对于单GPU的30米参考模拟，单成员预报在以100米和150米模拟为条件约束时的加速比分别为 $7.8\times$ 和 $11.0\times$ ，因为仍须先运行粗分辨率模拟；在无水动力条件时则为 $16.3\times$ 。

本文报告的集合预报优势更大，因为条件输入（包括任何粗分辨率水动力模拟）只需计算一次，便可供每个成员重复使用。测得每增加一个成员的边际成本为每分区3.95秒，因此流域尺度的 N成员集合需要 $103N$ 秒。设 T_{30} 表示一次30米模拟的耗时， T_{LR} 表示一次粗分辨率模拟的耗时， $T_{\mathrm{DM}} = 103$ 表示每个成员的耗时，则 N成员集合相对于 N成员基于物理的集合的加速比为

$$S(N) = \frac{N\,T_{30}}{T_{\mathrm{LR}} + N\,T_{\mathrm{DM}}}, \tag{9}$$

该加速比随 N 增大至 $T_{30}/T_{\mathrm{DM}} = 19.3\times$而单调递增。表8 对此进行了评估。由于 T_{DM} 是以每批十个成员的方式采样时，单个成员的边际成本，因此在9处的式 $N = 1$ 所得数值（$8.4\times$、 $12.2\times$ 和 $19.3\times$）略高于表7 中未分批的单成员加速比（$7.8\times$、 $11.0\times$ 和 $16.3\times$）；后者适用于单次预报。流域尺度的20成员集合耗时约为0.6 GPU小时，而等效的30米模拟则需11.0 GPU小时。加速采样器或蒸馏采样器可减少每个成员的反向扩散步数，是进一步降低成本的直接途径。

除了这些实测比值外，代理模型成本也呈现出良好的扩展性：由于预见期是作为输出通道生成的，而不是在CFL约束下按顺序积分，因此推理成本随预报时域的增长幅度较小；通过独立的分块评估，成本会随预报区域扩大，而随着计算域增大，30米雪水当量解的成本则会迅速上升；并且，如公式（9）所示，集合规模越大，其效率越高。

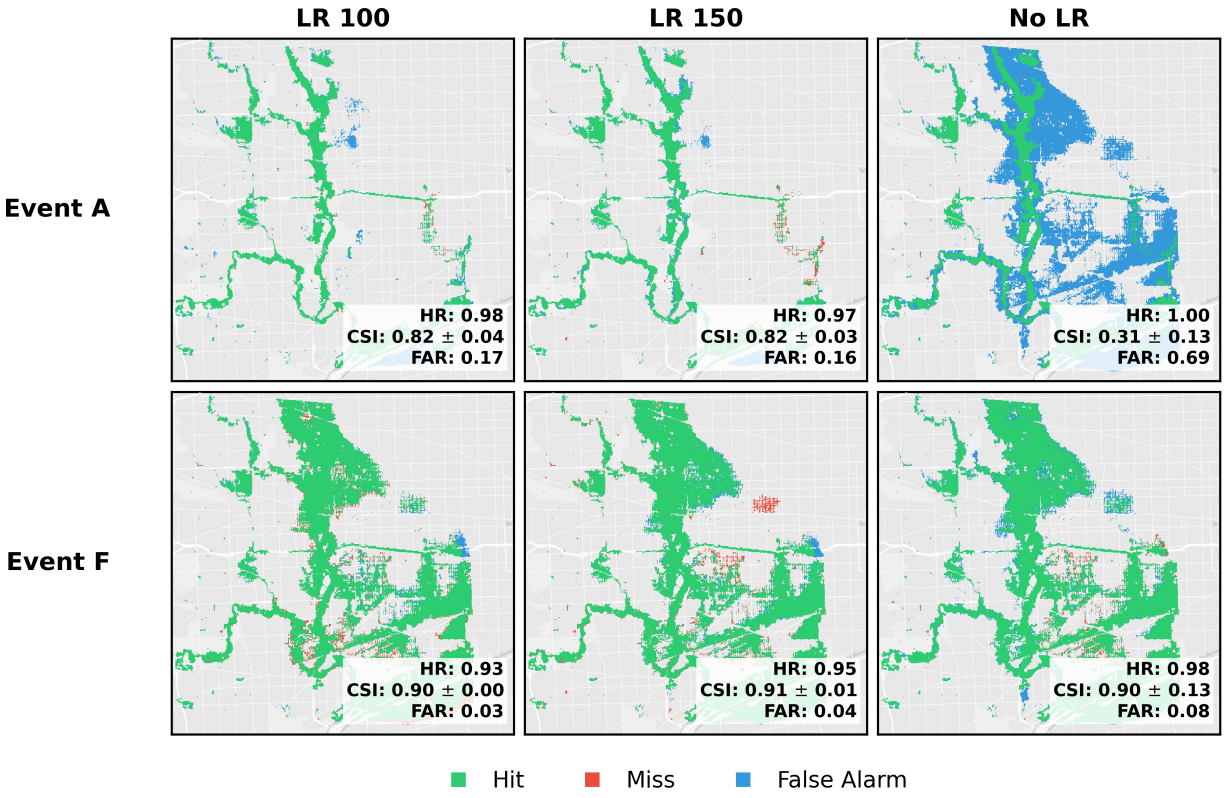


图18：集合 在区块30处将集合平均洪水范围分解为命中、漏报和虚警，绘图方式同图 15.

预计加速比.. 通过增加输出通道，可将预报时域延长至14天，因此单次网络推理遍历仍可生成完整预报，而SWE运行时间则随模拟时段成比例增长（增长 $361/121 \approx 3$倍）。表 9 报告了相对于单GPU 30米模拟的加速比；这些是根据表 7中的实测时间推算得出的。

表7：德斯普兰斯河流域四天洪水预报的计算成本。水动力学计时对应一次121小时的SynxFlow运行；八图形处理器列为339次归档数据集模拟的平均值，单图形处理器列则是针对0.1毫米/小时前期情景下事件F的专门基准测试。对于该事件，30、100和150米分辨率下匹配的八图形处理器运行时间分别为487、67和38秒，因此单图形处理器基准测试对应于在八个设备上的并行加速比 $4.1\times$、$2.0\times$ 和 $1.6\times$ 。DDPM推理在一块图形处理器上，对一个集合成员的26个不重叠斑块进行100步去噪。加速比以单图形处理器30米模拟为参照。

配置	8 个 GPU [s]	1GPU [s]	相对于30米的加速比（1 个图形处理器）
SynxFlow 30 米（参考）	538	1986	–
SynxFlow 100 米	76	133	–
SynxFlow 150 米	44	60	–
DDPM，一个成员	–	120–122	–
去噪扩散概率模型 + 低分辨率（100 m）	–	253	7.8×
去噪扩散概率模型 + 低分辨率（150 m）	–	180	11.0×
去噪扩散概率模型（无 LR）	–	122	16.3×

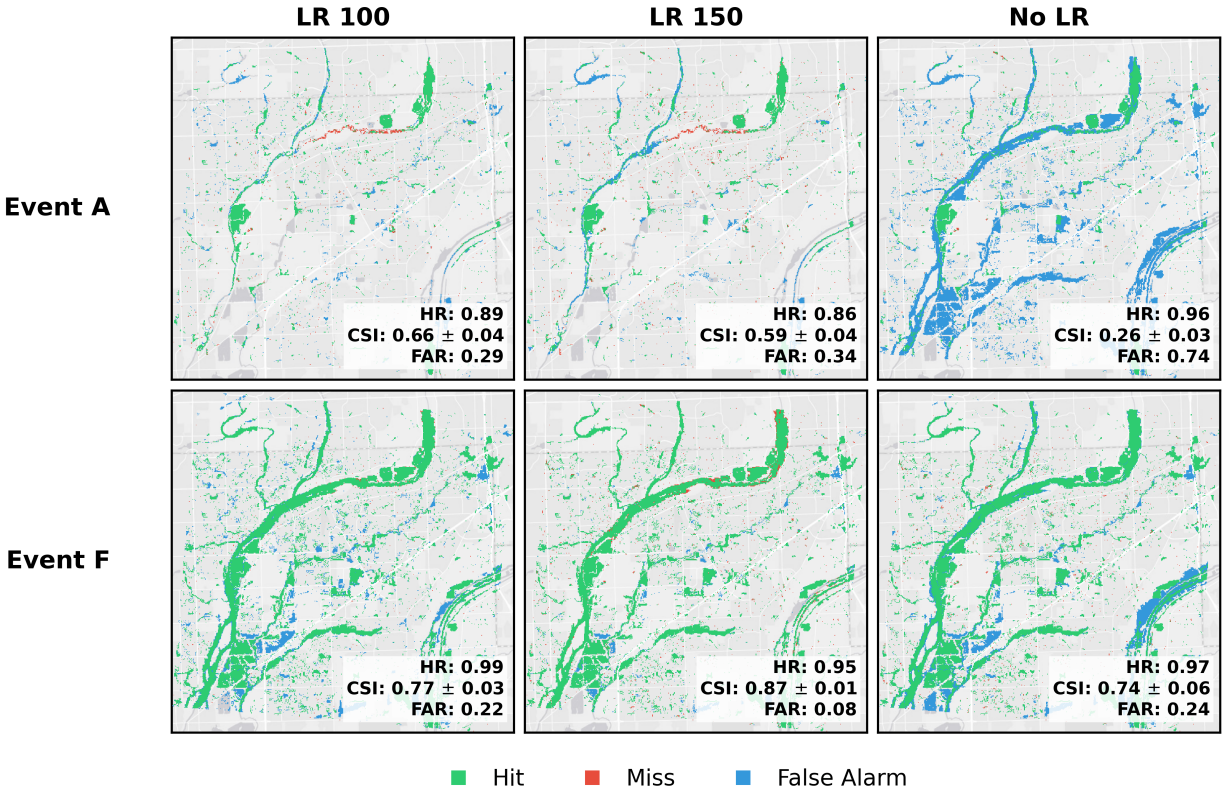


图19：集　　合平均洪水范围在斑块42处分解为命中、漏报和虚警，如图所示　　15.

表8：在单个图形处理器上运行 N成员流域尺度集合的加速比，根据公式（9），采用表7 中的计时结果以及$T_{\mathrm{DM}} = 103$ 秒/成员，即以每批十个成员采样时单个成员的边际成本。基于物理的参考生成 N 个相互独立的30米模拟。 $N = 1$ 列为批处理成本极限；未批处理的单成员加速比见表7。

配置	$N = 1$	$N = 5$	$N = 20$
去噪扩散概率模型 + 低分辨率（100米）	8.4×	15.4×	18.2×
去噪扩散概率模型 + 低分辨率（150米）	12.2×	17.3×	18.8×
去噪扩散概率模型（无 LR）	19.3×	19.3×	19.3×

表9：在包含26个不重叠斑块的14天预报中，相对于单GPU 30 m SynxFlow模拟的预计加速比。

配置	1个成员	20个成员
去噪扩散概率模型 + 低分辨率（100 m）	11.5×	48.4×
去噪扩散概率模型 + 低分辨率（150 m）	19.8×	53.1×
去噪扩散概率模型（无 LR）	48.7×	57.7×

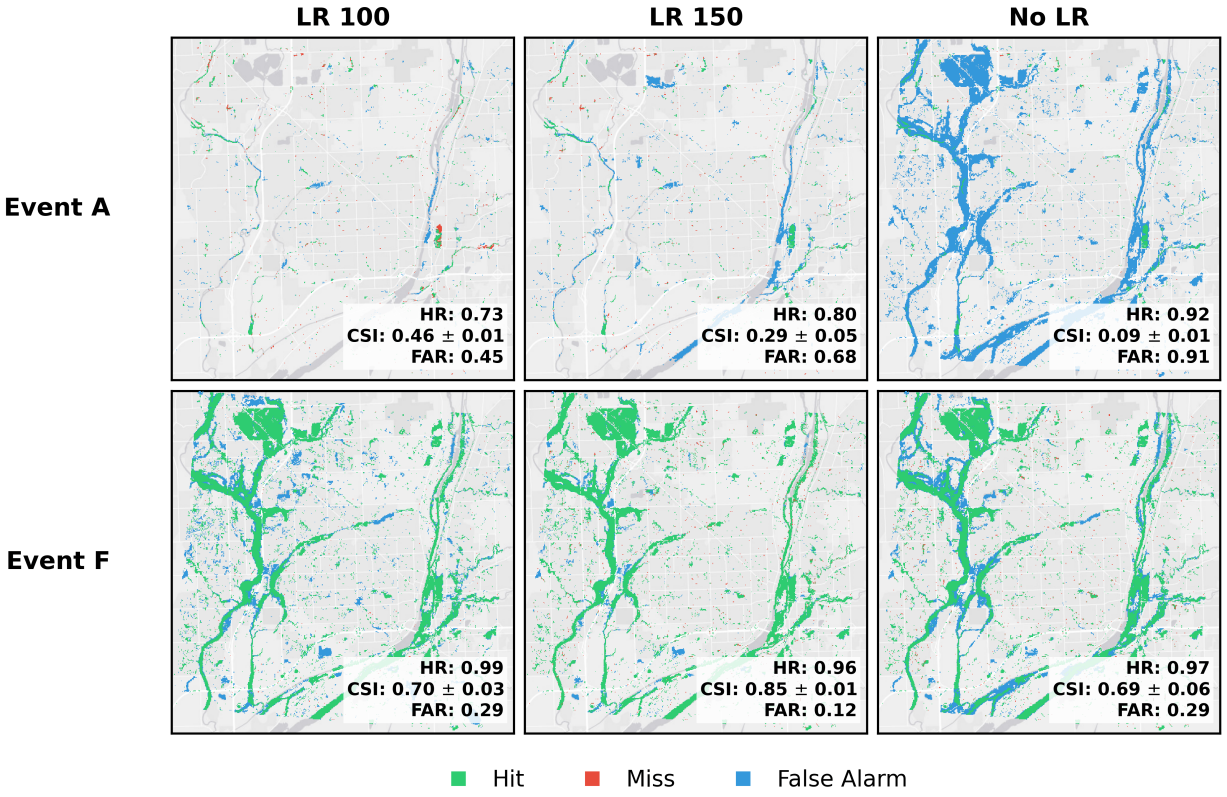


图20：E　　在斑块52处，将集合平均洪水范围分解为命中、漏报和虚警，绘图方式如图所示 15.

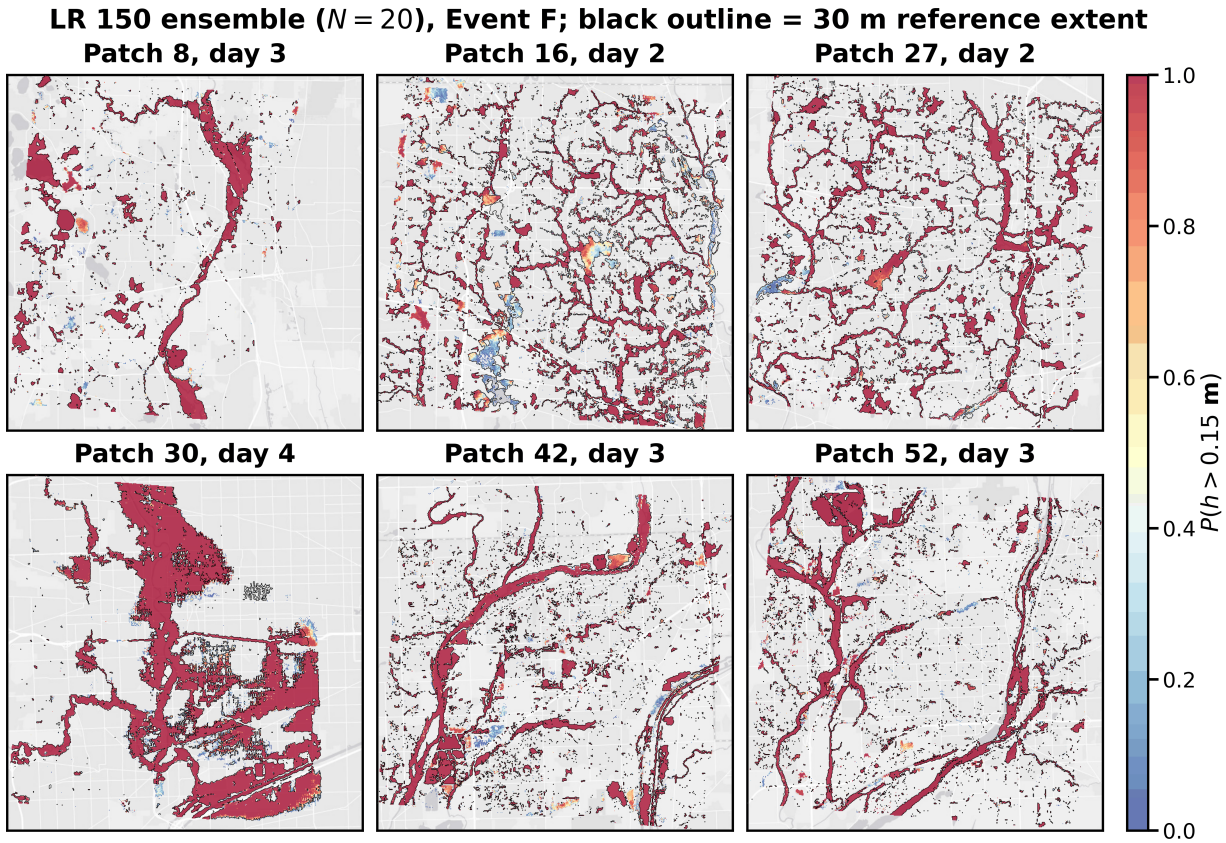


图21：事件F中，150 m条件化的20成员集合在参考淹没最大值当天、六个分析区块上的超越概率 $P(h > 0.15\ \text{m})$。黑色轮廓为30 m参考洪水范围。低于0.02的概率设为透明。

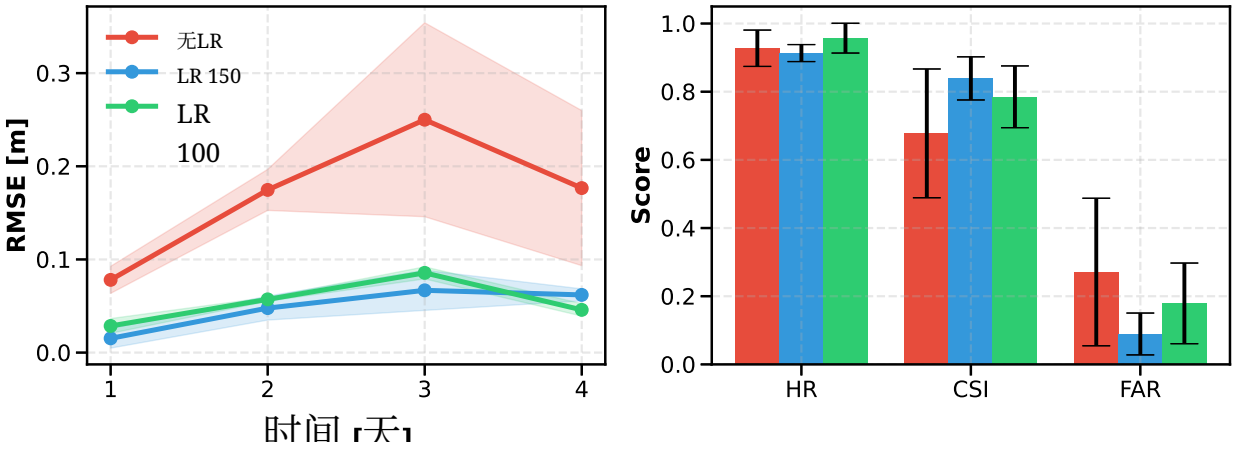


图22：六个分析区块和六个测试事件上的集合总体表现。左：集合平均相对于预见期的汇总均方根误差，阴影带表示跨事件的一个标准差。右：参考淹没最大值当天集合平均的汇总命中率、临界成功指数和虚警比，误差条表示跨事件的一个标准差。

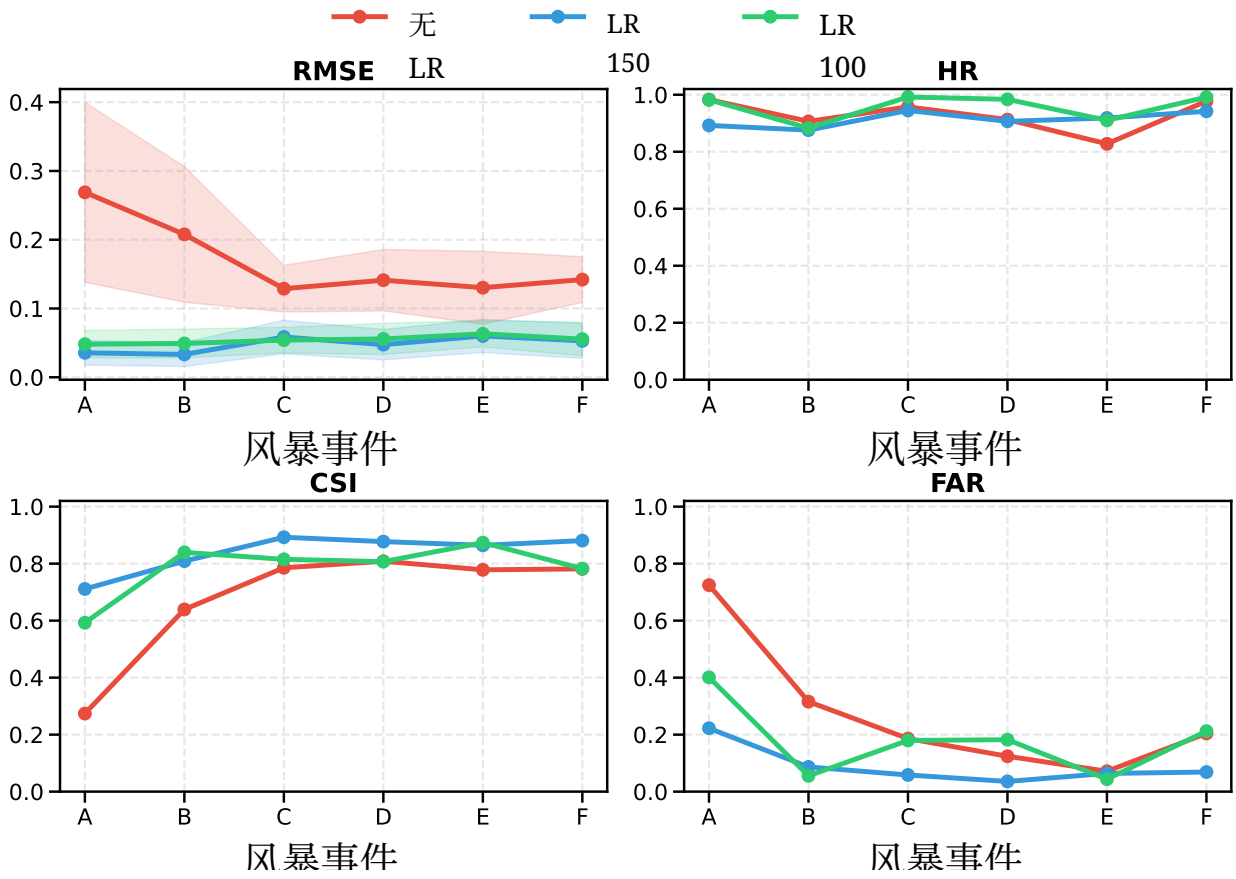


图23：按测试事件解析的集合总体表现。RMSE 为四个预见期的平均值，阴影带表示各预见期之间一个标准差的范围；HR、临界成功指数和虚警率在参考淹没最大值当天进行评估。横轴为类别轴；各事件的严重程度特征见表 2。

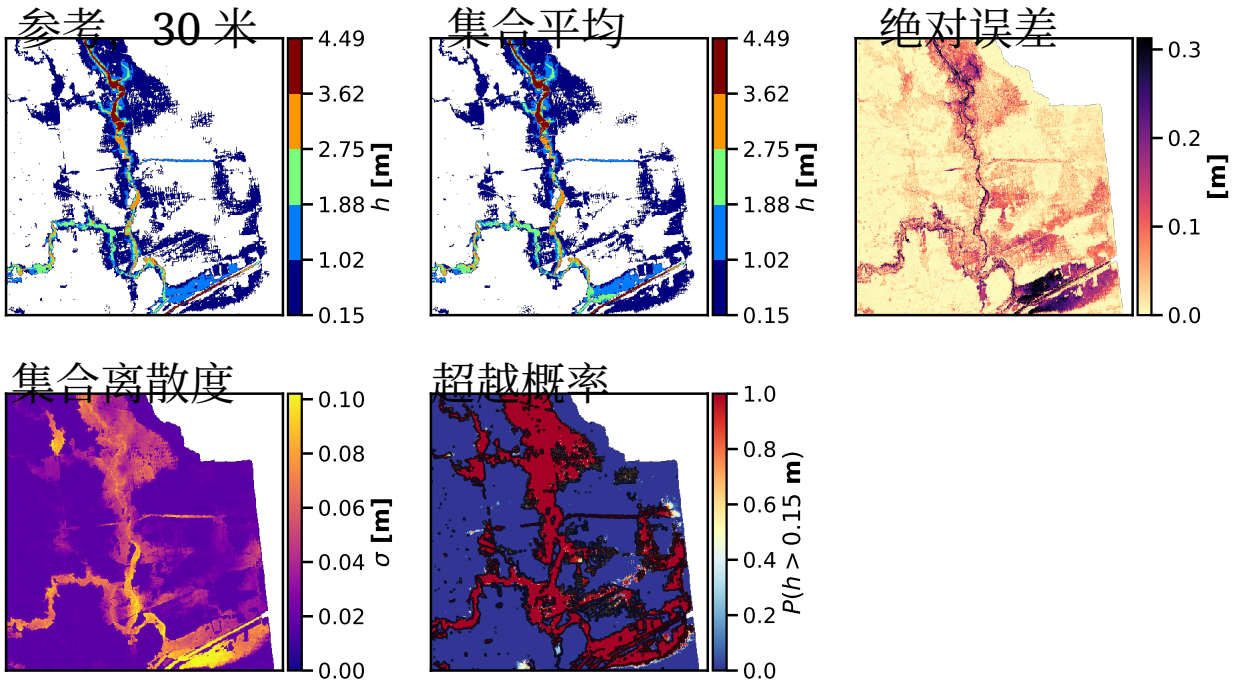


图24：单个集合预报的构成：事件F、区块30、三天预见期、150 m 条件约束，$N = 20$。从左到右依次为：30 m SynxFlow 参考水深；集合平均；集合平均的绝对误差；集合标准差；以及超越概率 $P(h > 0.15\ \text{m})$，其中参考洪水边界以黑色显示。离散度集中在淹没区域的边缘，因为条件约束无法确定湿润锋的位置。

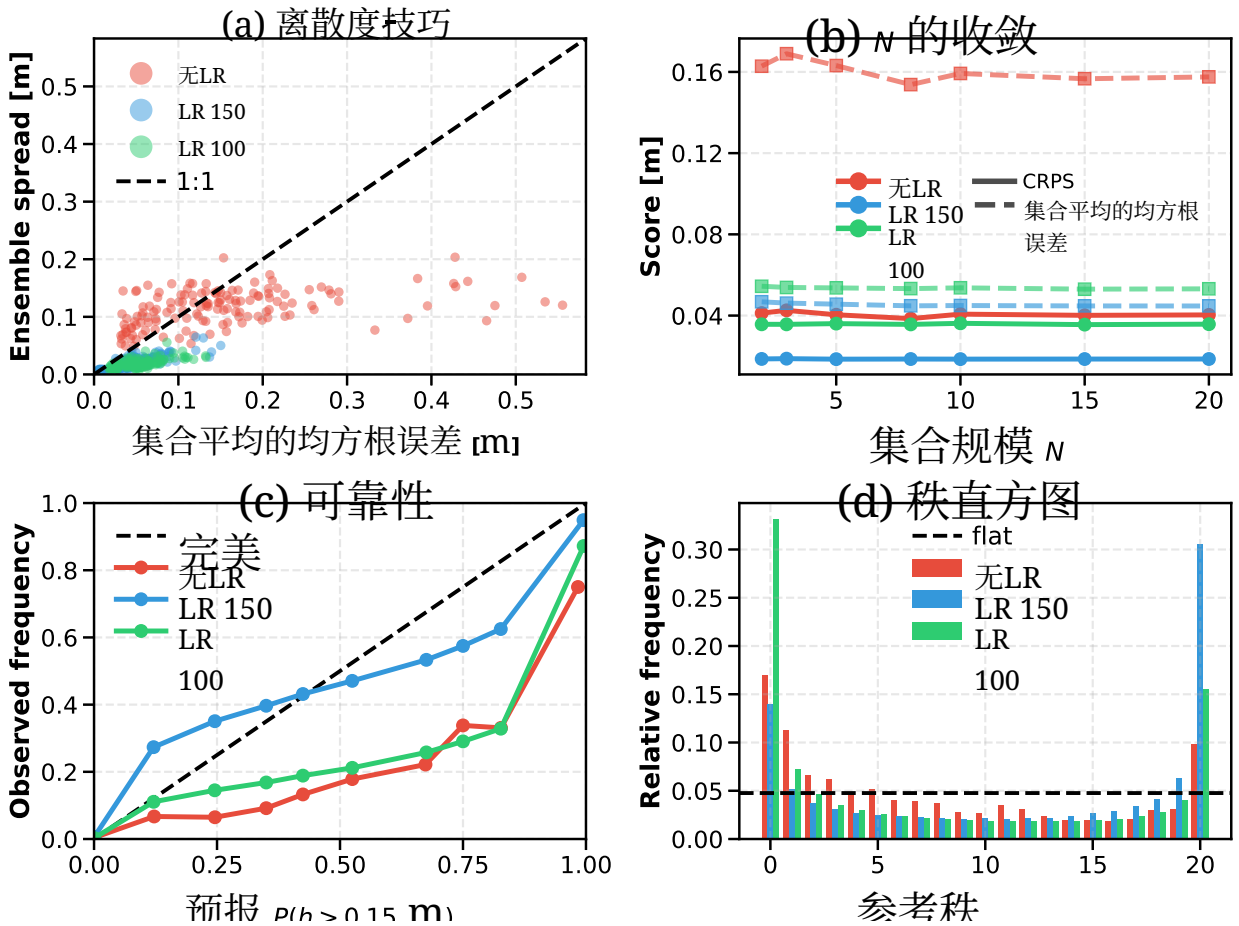


图25：DDPM集合的概率检验。(a) 离散度–技巧关系，每个个例和预见期对应一个点，1:1线表示统计一致性；所有配置均位于其下方。(b) 连续排序概率得分（实线）和集合均值均方根误差（虚线）随集合规模 N 的收敛情况。(c) 超越概率 $P(h > 0.15\ \mathrm{m})$ 的可靠性图。(d) 参考值在参考场中为湿的网格单元内的集合中的秩直方图；虚线表示校准集合预期呈现的平坦直方图。

5. 结论

本研究开发了一种条件去噪扩散概率模型（DDPM），用于伊利诺伊州东北部德斯普兰斯河流域的二维时空洪水淹没预报。该模型通过一次网络评估即可预报四个逐日预见期内30 m水深的演变，而非仅预测二元淹没范围或洪峰快照，因此能够捕捉洪水的传播和退水过程。模型采用基于物理的数据集进行训练；该数据集由 SynxFlow 多GPU浅水方程求解器生成，驱动数据为2003年至2024年间在美国中西部十二个州观测到的113场历史降水事件，并在三种前期湿润情景下进行模拟。每种网格分辨率对应339次水动力模拟，涵盖了范围广泛的风暴强度、降水格局和前期条件。

以低分辨率水动力模拟作为条件约束，显著提升了预报的准确性和时间一致性。与无条件模型相比，这一方法将集合平均的四日平均水深均方根误差从0.170 m降至150 m和100 m条件配置下的0.048 m和0.054 m，并将临界成功指数从0.678提高至0.839和0.785。条件模型对六场留出的测试风暴展现出良好的泛化能力，其中包括降雨强度最极端且历时最短的事件。这表明，粗分辨率水动力学解提供了洪水的时间结构，而扩散模型则补足了粗分辨率模拟无法解析的精细尺度空间细节。

一个具有实际意义的发现是，较粗的150 m条件约束在水深均方根误差、临界成功指数、虚警比、连续排序概率得分和布里尔评分上均优于较细的100 m条件约束，同时也是成本更低的方案。100 m条件约束模型取得了更高的命中率，但这是通过生成系统性偏湿的预报实现的，同时也提高了其虚警比。因此，在所考察的范围内，更细的条件化分辨率并不能保证更好的洪水制图能力，而较粗的方案提供了更好的精度—成本权衡。

由于反向扩散过程是一条随机马尔可夫链，训练好的模型就其构造而言是概率性的，本研究中的每一次预报都是一个无需重新训练、输入扰动或架构修改即可获得的20成员集合。对成员进行平均后，相对于单次实现，100 m、150 m和无条件配置的水深均方根误差分别降低了4.9%、8.2%和10.7%，而集合离散度则提供了空间显式的洪水超越概率。150 m配置再次表现最佳，取得了最低的连续排序概率得分和布里尔评分。概率检验进一步表明，这些集合离散度不足，离散度-误差比为0.35–0.73，其秩直方图识别出100 m处的偏湿和150 m处的偏干，这与洪水范围分析中的命中率和虚警表现相一致。由于生成式离散度较窄，得分在三到五个成员时即趋于收敛，因此紧凑集合已足够，而将生成式集合与扰动后的气象强迫相结合，是获得完整业务不确定性估计的自然途径。

从计算视角来看，在单个图形处理器上进行单成员四天预报时，对于 100 m、150 m 和无条件配置，其速度分别比 30 m 水动力模拟快 $7.8\times$、$11.0\times$ 和 $16.3\times$ 倍。随着集合规模增大，这一优势也会增加，因为条件输入（包括任何粗分辨率水动力模拟）只需计算一次，即可供每个成员重复使用；随着预报时域延长，这一优势同样会增加，因为额外的预报时效是作为输出通道生成的，而非通过时间积分获得。对于 14天预报，单成员的预计加速比分别达到 $11.5\times$、$19.8\times$ 和 $48.7\times$ 倍，20成员集合的预计加速比分别达到 $48.4\times$、$53.1\times$ 和 $57.7\times$ 倍。

总体而言，本研究表明，物理引导扩散模型可作为高效、准确的区域尺度洪水预报概率代理模型。该框架结合了水动力模拟的物理一致性与扩散模型的生成能力，在降低计算成本的同时保留高分辨率淹没结构，并能原生提供业务洪水预警所需的集合信息；其基于分块的 formulation 也为扩展至更多流域提供了可扩展的途径。未来工作将把该框架扩展至伊利诺伊州的区域尺度洪水预报，以分块局地强迫取代全域降水描述符，利用数值天气预报集合驱动模型，使气象不确定性与生成不确定性

能够联合传播，纳入河流边界条件，并探索加速采样器，以进一步降低大规模集合的成本。

CRediT 作者贡献声明

Omar Sallam: 概念化、方法论、软件、形式分析、调查、数据整理、可视化、撰写初稿。**David Coe:** 审阅与编辑。**Jeremy Feinstein:**审阅与编辑。**Hugo Torres:** 审阅与编辑。**Sofia Sahagun Covarrubias:**审阅与编辑。**Eugene Yan:** 监督、经费获取、项目管理、撰写、审阅与编辑。

利益冲突声明

作者声明，他们不存在已知的可能影响本文所报告工作的竞争性财务利益或个人关系。

致谢

本研究使用了阿贡国家实验室实验室计算资源中心（LCRC）的资源。阿贡国家实验室由 UChicago Argonne, LLC 代表美国能源部运营，依据合同编号 DE-AC02-06CH11357。

数据可用性

降水强迫数据取自公开可用的 NCEP/NWS Stage IV 多传感器降水分析。地形和土地覆盖数据来自美国地质调查局 3DEP、美国地质调查局 NHD 和 NLCD 2021 的公开档案。SynxFlow 求解器为开源软件。生成的水动力模拟数据集、训练后的模型权重和分析脚本可在合理请求后向通讯作者获取。

参考文献

[1] 灾害流行病学研究中心（CRED），《2025年灾害数据：大地、风暴与火灾》，技术报告，灾害流行病学研究中心（CRED），访问日期：2026-05-29（2026年4月）。网址：https://reliefweb.int/report/world/2025-disasters-numbers[2] S. I. Seneviratne、X. Zhang、M. Adnan、W. Badi、C. Dereczynski、A. D. Luca、S. Ghosh、I. Iskandar、J. Kossin、S. Lewis 等，《气候变化背景下的天气与气候极端事件》，《2021年气候变化：物理科学基础：政府间气候变化专门委员会第六次评估报告第一工作组贡献》（2021）1513–1766。[3] J. S. Rogers、M. P. Maneta、S. R. Sain、L. E. Madaus、J. P. Hacker，《气候与人口变化在全球洪水暴露度和脆弱性中的作用》，《自然·通讯》16（1）（2025年）1287。doi：10.1038/s41467-025-56654-8。[4] W. Chen、Y. Zheng、Y. Zhou、T. Zhang、Y. Zhang、D. Z. Zhu、X. Zhou、L. Zhang、X. Zhang、Z. Huang，《变暖气候下河流—潮汐相互作用加剧复合洪水风险的全球趋势》，npj《自然灾害》3（1）（2026）13。doi:10.1038/s44304-026-00179-7。[5] IPCC，《2022年气候变化：影响、适应与脆弱性》，技术报告，政府间气候变化专门委员会（2022年）。[6] 世界气象组织，《全民早期预警：2023–2027年执行行动计划》（2022年）。网址：https://public.wmo.int/en/resources/bulletin/early-warnings-all[7] 世界气象组织，《洪水预报与预警手册》（2011年）。[8] J. Teng、A. J. Jakeman、J. Vaze、B. F. Croke、D. Dutta、S. Kim，《洪水淹没建模：方法、最新进展与不确定性分析综述》，《环境建模与软件》90（2017年）201–216。doi:10.1016/j.envsoft.2017.01.006。[9] H. L. Cloke、F. Pappenberger，《集合洪水预报：综述》，《水文学杂志》375（3-4）（2009年）613–626。[10] M. Asif、M. M. Kuglitsch、I. Pelivan、R. Albano，《短期洪水预报机器学习应用的综述与比较》，《水资源管理》39（5）（2025年）1971–1991。doi:10.1007/s11269-025-04093-x。[11] B. Liu、Y. Li、M. Ma、B. Mao，《洪水深度估计机器学习方法综述》，《国际灾害风险科学杂志》16（3）（2025年）433–445。doi:10.1007/s13753-025-00639-0。[12] F. T. Zahura、J. L. Goodall、J. M. Sadler、Y. Shen、M. M. Morsy、M. Behl，《基于高保真物理模型训练机器学习代理模型：在城市沿海社区实时街道尺度洪水预测中的应用》，《水资源研究》56（10）（2020年）e2019WR027038。[13] M. Ma、G. Zhao、B. He、Q. Li、H. Dong、S. Wang、Z. Wang，《基于XGBoost的方法用于山洪风险评估》，《水文学杂志》598（2021）126382。[14] C.-C. Lee、L. Huang、F. Antolini、M. Garcia、A. Juan、S. D. Brody、A. Mostafavi，《利用物理信息机器学习模型预测峰值淹没深度》，《科学报告》14（1）（2024年）14826。doi:10.1038/s41598-024-65570-8。[15] W. Liu、Q. Wu、Z. Liu、X. Zheng、T. Liu、H. Gao，《用于城市洪水深度预测的PINN框架：融合数据驱动洞见与物理约束》，《水文学杂志》665（2026）134693。doi:10.1016/j.jhydrol.2025.134693。

[16] A. Y. Sun、Z. Li、W. Lee、Q. Huang、B. R. Scanlon、C. Dawson，基于傅里叶神经算子的快速洪水淹没预报，载于：IEEE/CVF计算机视觉国际会议论文集，2023年，第3733–3739页。[17] Q. Xu、Y. Shi、J. L. Bamber、C. Ouyang、X. X. Zhu，基于floodcast的大规模洪水建模与预报，《水研究》264（2024年）122162。[18] Q. Xu、Y. Shi、J. Zhao、X. X. Zhu，Floodcastbench：用于洪水建模与预报的大规模数据集与基础模型，《科学数据》12（1）（2025年）431。[19] C. Kumar、D. McSpadden、M. Schram、H. Richter、Y. Wang、B. Roy、J. L. Goodall，人工智能驱动的时间超分辨率方法用于弗吉尼亚州诺福克的洪水预测，发表于NeurIPS 2025机器学习应对气候变化研讨会，气候变化人工智能（CCAI）研讨会，美国加利福尼亚州圣迭戈（2025年）。URL https://s3.us-east-1.amazonaws.com/climate-change-ai/papers/neurips2025/26/ slides.pdf[20] P. Si、Y. Qiu、O. Sallam、J. Feinstein、Z. He、E. Yan、P. Chen，迈向人工智能驱动的都市洪水数字孪生：浅水方程的条件潜在动力学网络代理模型，《水文学杂志》680（2026年）136461。doi:10.1016/j.jhydrol.2026.136461。[21] R. Bentivoglio、E. Isufi、S. N. Jonkman、R. Taormina，基于水力学的图神经网络实现快速时空洪水建模，《水文学与地球系统科学》27（23）（2023年）4227–4246。[22] A. Kazadi、J. Doss-Gollin、A. Sebastian、A. Silva，Floodgnn-gru：用于洪水预测的时空图神经网络，《环境数据科学》3（2024年）e21。doi:10.1017/eds.2024.19。[23] A. Kazadi、J. Doss-Gollin、A. L. Da Silva，基于水力学信息引导的消息传递的暴雨洪水模拟，载于：第四十一届国际机器学习大会，2024年。[24] R. Bentivoglio、E. Isufi、S. N. Jonkman、R. Taormina，用于洪水建模的多尺度水力图神经网络，《自然灾害与地球系统科学》25（1）（2025年）335–351。[25] M. Taghizadeh、Z. Zandsalimi、M. A. Nabian、M. Shafiee-Jood、N. Alemazkoor，用于洪水预报的可解释物理信息图神经网络，《计算机辅助土木与基础设施工程》40（18）（2025年）2629–2649。doi:10.1111/mice.13484。[26] C. M. Acosta、H. M. V. V. Herath、J. Y. Lim、A. Saha、S. Rasnayaka、L. Marshall，DUALFloodGNN：用于业务洪水建模的物理信息图神经网络，载于：第三十五届国际人工智能联合会议（IJCAI-26）论文集，2026年，第6408–6416页。doi:10.24963/ijcai.2026/713。[27] Z. Guo、V. Moosavi、J. P. Leitão，具有集水区泛化能力的数据驱动快速洪水预测制图，《水文学杂志》609（2022年）127726。[28] B. Pradhan、S. Lee、A. Dikshit、H. Kim，使用可解释人工智能（XAI）模型进行洪水易发性空间制图，《地球科学前沿》14（6）（2023年）101625。[29] H. Choi、H. Woo、M. Kim、H. Ryu、J.-H. Lee、S. Lee、S. J. Noh，Flo-sr：基于深度学习的城市洪水超分辨率模型，《水文学杂志》661（2025年）133529。doi:10.1016/j.jhydrol.2025.133529。[30] R. Löwe、J. Böhm、D. G. Jensen、J. Leandro、S. H. Rasmussen，U-flood——用于预测城市暴雨洪水水深的地形深度学习，《水文学杂志》603（2021）126898。

[31] Z. Wang、H. Lyu、G. Fu、C. Zhang，用于时空城市洪水模拟的时间引导卷积神经网络，《水文学杂志》645（2024年）132250。doi:10.1016/j.jhydrol.2024.132250。[32] H. M. V. V. Herath、L. Marshall、A. Saha、S. Rasnayaka、S. Seneviratne，用于高分辨率洪水制图的子网格知情神经网络，《水文学杂志》660（2025年）133329。doi:10.1016/j.jhydrol.2025.133329。[33] J. He、L. Zhang、T. Xiao、H. Wang、H. Luo，深度学习实现时空变化暴雨条件下超分辨率水动力洪水过程建模，《水研究》239（2023年）120057。[34] Y. Zhou、W. Wu、R. Nathan、Q. J. Wang，一种结合深度学习与空间降维与重构的快速洪水淹没建模框架，《环境建模与软件》143（2021）105112。doi:10.1016/j.envsoft.2021.105112。[35] G. Nearing、D. Cohen、V. Dube、M. Gauch、O. Gilon、S. Harrigan、A. Hassidim、D. Klotz、F. Kratzert、A. Metzger等，无测站流域极端洪水的全球预测，《自然》627（8004）（2024年）559–563。[36] D. Cohen、R. Amira、R. Aschner、Y. Carny、B. Feinstein、H. Fester、S. Fronman、M. Gauch、O. Gilon、R. Green等，拓展中期全球洪水预报：谷歌全球洪水预报模型第2版，EGUsphere 2026（2026年）1–31。doi:10.5194/egusphere-2026-2283。[37] J. Hofmann、H. Schüttrumpf，FloodGAN：利用深度对抗学习实时预测暴雨洪水，《Water》13（16）（2021）2255。[38] C. A. Do Lago、M. H. Giacomoni、R. Bentivoglio、R. Taormina、M. N. G. Junior、E. M. Mendiondo，利用条件生成对抗网络将快速洪水预测泛化至未见过的城市汇水区，《水文学杂志》618（2023年）129276。[39]L. Huang、F. Antolini、A. Mostafavi、R. Blessing、M. Garcia、S. D. Brody，利用生成式机器学习结合大规模合成降水与淹没数据进行高分辨率洪水概率制图，《计算机辅助土木与基础设施工程》40（19）（2025年）2859–2875。doi:10.1111/mice.13490。[40] Z. Yin、Y. Saadati、B. Hu、A. S. Leon、M. H. Amini、D. McDaniel，利用超分辨率技术快速生成高保真洪水淹没图，《水文信息学杂志》26（1）（2024年）319–336。doi:10.2166/hydro.2024.228。[41] H. Shen、H. Hu、C. Xin、Y. Yang、Y. Wu、Z. Liang、J. Wen，用于沿海特大城市高精度城市暴雨洪水预报的基于注意力的超分辨率模型，《水文学杂志》（2026年）135330。[42] P. Shao、J. Feng、J. Lu、P. Zhang、C. Zou，用于洪水预报的数据驱动与知识引导去噪扩散模型，《专家系统及其应用》244（2024年）122908。doi:10.1016/j.eswa.2023.122908。[43] Z. Ou、C. Nai、B. Pan、Y. Zheng、C. Shen、P. Jiang、X. Liu、Q. Tang、W. Li、M. Pan，概率性的扩散模型推动极端洪水预报，《地球物理研究快报》52（15）（2025年）e2025GL115705。doi:10.1029/2025GL115705。[44] S. H. Neo、S. Seneviratne、H. M. V. V. H. Herath、A. Saha、S. Rasnayaka、L. A. Marshall，Flood-LDM：用于快速准确零样本高分辨率洪水制图的可泛化潜在扩散模型，载于：IEEE/CVF计算机视觉应用冬季会议论文集，2026年，第8063–8072页。[45] W. Yang、H. Ji、L. Lonzarich、Y. Song、C. Shen，基于扩散的概率性建模用于小时尺度径流预测与同化，arXiv预印本 arXiv:2510.08488（2025年）。

[46] J. Chen, Y. Li, S. Zhang，基于 CNN-LSTM 的城市洪水水深快速预测，《Water》15 (7) (2023年) 1397。[47] F. Yang, W. Ding, J. Zhao, L. Song, D. Yang, X. Li，基于物理信息深度学习方法的城市洪水淹没快速预报，《水文学杂志》643 (2024) 131998。doi:10.1016/j.jhydrol.2024.131998。[48] Y. Liao, Z. Wang, H. Yu, W. Gao, Z. Zeng, X. Li, C. Lai，利用 ConvLSTM 深度学习模型加速时空变化暴雨条件下的城市洪水淹没模拟，《水资源研究》61 (8) (2025年) e2025WR040433。doi:10.1029/2025WR040433。[49] Y. Wang, J. L. Goodall, C. Kumar, D. McSpadden, S. A. Barbosa, B. Roy, A. Shahabi, N. Tahvildari，一种用于弗吉尼亚州诺福克超高分辨率时空洪水预报的混合 CNN-LSTM 代理模型，《水文学杂志：区域研究》64 (2026) 103234。doi:10.1016/j.ejrh.2026.103234。[50] L. Breiman，随机森林，《机器学习》45 (1) (2001年) 5–32。[51] J. H. Friedman，贪心函数逼近：梯度提升机，《统计学年刊》(2001年) 1189–1232。[52] T. Chen, C. Guestrin，XGBoost：一种可扩展的树提升系统，载于：第22届ACM SIGKDD知识发现与数据挖掘国际会议论文集，2016年，第785–794页。[53] M. Raissi, P. Perdikaris, G. E. Karniadakis，物理信息神经网络：用于求解涉及非线性偏微分方程的正问题与逆问题的深度学习框架，《计算物理学期刊》378 (2019年) 686–707。[54] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar，用于参数偏微分方程的傅里叶神经算子，arXiv预印本 arXiv:2010.08895 (2020年)。[55] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini，图神经网络模型，《IEEE神经网络汇刊》20 (1) (2009年) 61–80。[56] T. N. Kipf, M. Welling，利用图卷积网络进行半监督分类，arXiv预印本 arXiv:1609.02907 (2016年)。[57] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner 等，关系归纳偏置、深度学习与图网络，arXiv预印本 arXiv:1806.01261 (2018年)。[58] J. L. Elman，在时间中发现结构，《认知科学》14 (2) (1990年) 179–211。[59] S. Hochreiter, J. Schmidhuber，长短期记忆，《神经计算》9 (8) (1997年) 1735–1780。[60] O. Ronneberger, P. Fischer, T. Brox，U-Net：用于生物医学图像分割的卷积网络，载于：医学图像计算与计算机辅助介入国际会议，施普林格，2015年，第234–241页。[61] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio，生成对抗网络，《神经信息处理系统进展》27 (2014年)。[62] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen，改进生成对抗网络训练的技术，《神经信息处理系统进展》29 (2016年)。[63] L. Metz, B. Poole, D. Pfau, J. Sohl-Dickstein，展开生成对抗网络，arXiv预印本 arXiv:1611.02163 (2016年)。

[64] M. Arjovsky, S. Chintala, L. Bottou，Wasserstein生成对抗网络，载于：国际机器学习会议，PMLR，2017年，第214–223页。[65] J. Ho, A. Jain, P. Abbeel，去噪扩散概率模型，《神经信息处理系统进展》33 (2020年) 6840–6851。[66] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer，基于潜在扩散模型的高分辨率图像合成，载于：IEEE/CVF计算机视觉与模式识别会议论文集，2022年，第10684–10695页。[67] Y. Song, P. Dhariwal, M. Chen, I. Sutskever，一致性模型，载于：第40届国际机器学习会议论文集，《机器学习研究论文集》第202卷，PMLR，2023年，第32211–32252页。[68] K. Fukami, K. Fukagata, K. Taira，利用机器学习对湍流进行超分辨率重建，《流体力学杂志》870 (2019年) 106–120。[69] K. Fukami, K. Fukagata, K. Taira，基于机器学习的湍流时空超分辨率重建，《流体力学杂志》909 (2021) A9。[70] G. Buster, B. N. Benton, D. Rastogi, S.-C. Kao, G. Castelao, J. Eisenman，利用生成式机器学习生成的第二代降尺度地球系统模型数据，Data in Brief 61 (2025年) 111774。doi:10.1016/j.dib.2025.111774。[71] B. N. Benton, G. Buster, P. Pinchuk, A. Glaws, R. N. King, G. Maclaurin, I. Chernyakhovskiy，利用再分析数据中的风数据对可再生能源资源数据进行超分辨率处理及其在乌克兰的应用，《Energies》18 (14) (2025年) 3769。[72] K. Stengel, A. Glaws, D. Hettinger, R. N. King，气候风能和太阳能数据的对抗式超分辨率，《美国国家科学院院刊》117 (29) (2020年) 16805–16815。[73] O. Sallam, A. Glaws, M. Sanchez-Gomez, G. Deskos, A. Cortiella, R. King, M. Fürth，利用条件扩散去噪概率模型实现热带气旋超分辨率：从中尺度模拟到大涡模拟，Environmental Research Communications 7 (12) (2025年) 125025。 doi:10.1088/2515-7620/ae2a96。[74] M. Leutbecher, T. N. Palmer，集合预报，《计算物理学期刊》227 (7) (2008年) 3515–3539。[75] I. Price, A. Sanchez-Gonzalez, F. Alet, T. R. Andersson, A. El-Kadi, D. Masters, T. Ewalds, J. Stott, S. Mohamed, P. Battaglia, R. Lam, M. Willson，基于机器学习的概率天气预报，《自然》637 (8044) (2025年) 84–90。[76] X. Xia, Q. Liang, X. Ming，基于高性能集成水动力建模系统（HiPIMS）的全尺度河流洪水建模框架，《水资源进展》132 (2019年) 103392。[77] X. Xia, X. Ming，SynxFlow：用于多灾害模拟的GPU加速Python软件包，《开源软件杂志》10 (111) (2025年) 7586。 doi:10.21105/joss.07586。[78] X. Xia, Q. Liang, X. Ming, J. Hou，一种具有新型源项离散格式的高效稳定水动力模型，用于地表径流与洪水模拟，《水资源研究》53 (5) (2017年) 3730–3759。[79] X. Xia, Q. Liang，一种用于离散浅水方程中刚性摩擦项的高效新型隐式格式，《水资源进展》117 (2018年) 87–97。[80] E. Campos, J. Wang, 2013年4月芝加哥洪水的数值模拟与分析，《水文学杂志》531 (2015年) 454–474。

[81] 美国国家气象局，《美国洪水损失报告——2013水文年》，技术报告，美国国家海洋和大气管理局，访问日期：2026-05-24（2013年）。网址：https://www.weather.gov/media/water/WY13%20Flood%20Loss%20Summary.pdf[82] G. M. Bonnin、D. Martin、B. Lin、T. Parzybok、M. Yekta、D. Riley，《NOAA Atlas 14：美国降水频率图集》第2卷第3.0版：俄亥俄河流域及周边州，技术报告，NOAA，美国国家气象局，马里兰州银泉（2006年）。[83] T. Salimans、J. Ho，渐进蒸馏：加速扩散模型采样，CoRR abs/2202.00512（2022年）。arXiv:2202.00512。网址：https://arxiv.org/abs/2202.00512[84] K. He、X. Zhang、S. Ren、J. Sun，面向图像识别的深度残差学习，载于：电气与电子工程师学会计算机视觉与模式识别会议论文集，2016年，第770–778页。[85] H. Zhang、I. Goodfellow、D. Metaxas、A. Odena，自注意力生成对抗网络，载于：国际机器学习会议，PMLR，2019年，第7354–7363页。[86] L. Wang、H. Liu，一种在数字高程模型中识别并填充地表洼地、用于水文分析与建模的高效方法，《国际地理信息科学杂志》20（2）（2006年）193–213。[87] D. Eilander、M. Visser，Deltares/pyflwdir：v0.5（2021）。doi:10.5281/zenodo.5534756。网址：https://doi.org/10.5281/zenodo.5534756[88] A. N. Strahler，流域地貌学的定量分析，《Eos：美国地球物理联合会汇刊》38（6）（1957年）913–920。[89] A. D. Nobre、L. A. Cuartas、M. R. Momo、D. L. Severo、A. Pinheiro、C. A. Nobre，手形轮廓：一种预测淹没范围的新型代理指标，《水文过程》30（2）（2016年）320–333。[90] F. Nardi、A. Annis、G. Di Baldassarre、E. Vivoni、S. Grimaldi，Gfplain250m：全球高分辨率地球洪泛区数据集，《科学数据》6（2019年）180309。doi:10.1038/sdata.2018.309。[91] D. P. Kingma、J. Ba，Adam：一种随机优化方法，版本号：9（2014年）。doi:10.48550/ARXIV.1412.6980。网址：https://arxiv.org/abs/1412.6980[92] J. E. Matheson、R. L. Winkler，连续概率分布的评分规则，《管理科学》22（10）（1976年）1087–1096。[93] H. Hersbach，集合预报系统连续排序概率评分的分解，《天气与预报》15（5）（2000年）559–570。[94] T. Gneiting、A. E. Raftery，严格适当的评分规则、预测与估计，《美国统计学会杂志》102（477）（2007年）359–378。[95] V. Fortin、M. Abaza、F. Anctil、R. Turcotte，为什么集合离散度应与集合平均的均方根误差相匹配？《水文气象学杂志》15（4）（2014年）1708–1713。[96] O. Talagrand、R. Vautard、B. Strauss，概率预报系统的评估，欧洲中期天气预报中心可预报性研讨会论文集（1997年）1–25。[97] T. M. Hamill，秩直方图在检验集合预报中的解读，《月刊天气评论》129（3）（2001年）550–560。[98] G. W. Brier，以概率形式表达的预报检验，《月刊天气评论》78（1）（1950年）1–3。