Journal Name

Crossmark

RECEIVED
dd Month yyyy

REVISED
dd Month yyyy

PAPER

UPD-Diff: A Unified Precipitation Downscaling Method based on Multi-stream Elucidating Diffusion Model

Jiahan Chen[1,2], Lingzhi Kong[1], Yi Cao[3], Yuanzhi Zhong[4], Zhenfei Tang[5], Xinting Li[6] and Chengsheng Yuan[1,*]

[1]School of Computer Science, Nanjing University of Information Science and Technology, Nanjing, China
[2]Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
[3]School of Internet of Things Engineering, Wuxi University, Wuxi, China
[4]Fujian Huawang Information Technology Co., Ltd., Fujian, China
[5]Fujian Climate Center, Fujian, China
[6]National University of Defense Technology, Nanjing, China
*Author to whom any correspondence should be addressed.

E-mail: jia-han.chen@connect.polyu.hk; 202412492359@nuist.edu.cn; caoyi@cwxu.edu.cn and yuancs@nuist.edu.cn

Keywords: precipitation downscaling, super resolution, diffusion model, attention mechanism

Abstract

Precipitation downscaling plays a pivotal role in improving the resolution of coarse-grained weather and climate datasets, offering significant benefits for the issuance of localized precipitation warnings, which are indispensable for agricultural operations and economic planning in vulnerable regions. However, current downscaling approaches face several critical challenges: many deep learning methods, often adapted from computer vision, struggle to accurately model the long-tailed distribution of precipitation, leading to physically inconsistent outputs; reliance on regional a priori information severely limits model transferability, while inadequate merging of multivariate data yields ambiguous results; and prohibitive training costs and computational loads hinder real-time, large-scale deployment. To address these limitations, we propose the UPD-Diff, a unified multi-stream Elucidating Diffusion Model (EDM) framework designed for robust and efficient precipitation downscaling. Firstly, to promote physical consistency by meteorological variable conditioning and to more accurately model the long-tailed distribution of precipitation, the core EDM architecture of UPD-Diff employs refined noise scheduling and learns precipitation bias based on various physical auxiliary variables, thereby enhancing its alignment with meteorological dynamics. Secondly, to enhance model transferability and facilitate the effective fusion of multi-modal data inputs, each stream integrates a newly designed Mixed-Attention UNet (MA-UNet). This MA-UNet synergistically blends channel attention and local-importance attention, effectively capturing shared atmospheric features while preserving fine-grained local details across different geographical areas. Finally, to address the challenges associated with data and computational requirements, our method leverages its inherent architectural efficiency and bias-learning strategy to achieve superior SSIM and Corr scores compared to baseline models, even when trained on a relatively small dataset, thereby substantially reducing computational overhead. Trained in this globally sampled dataset, our method outperforms the leading SOTA baselines, showcasing competitive results with a PSNR of 29.02, a SSIM of 0.7942, a CSI of 0.867 (0.1 mm threshold) and improved event skill at higher thresholds(50/100 mm). Notably, our model supports both training and inference on a single GPU like RTX 4090, with local-region inference at about 5.1 seconds per 224×224 tile (100 steps) on the same GPU.

1 Introduction

Accurate precipitation forecasting serves as a fundamental pillar of modern meteorology, playing a pivotal role in vital applications including agricultural scheduling, water resource allocation, and disaster risk reduction [1, 2, 3]. Global climate models (GCMs) and weather prediction systems provide essential large-scale precipitation data, yet their coarse resolution—often spanning tens to hundreds of kilometers—is inadequate for capturing the fine-grained variability needed for localized

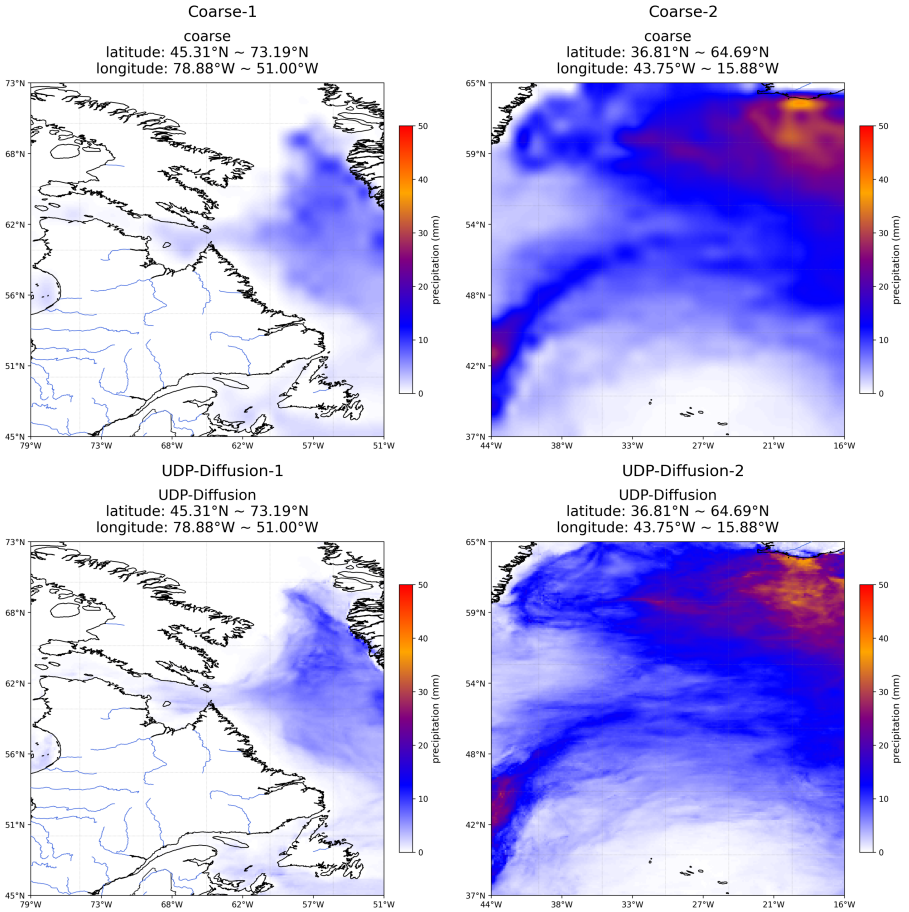

Figure 1. Visual demonstration of UPD-Diff's downscaling capabilities across different geographical regions. For each region depicted (in the top and bottom rows), the left panel displays the coarse Low-Resolution (LR) input precipitation, and the right panel presents the significantly enhanced High-Resolution (HR) output from our UPD-Diff model, highlighting its ability to reconstruct fine-scale precipitation structures.

decision-making [4]. This disparity is especially evident in regions characterized by complex topography or during extreme weather events, where precise spatial details can determine whether effective preparations are made or substantial socioeconomic losses are incurred. As climate change exacerbates the frequency and intensity of such events, the demand for high-resolution precipitation data has surged, spotlighting the need for advanced methodologies to bridge the gap between coarse model outputs and local-scale demands.

Precipitation downscaling is a crucial strategy to tackle the prevailing challenge. However, traditional approaches, including statistical downscaling and dynamical simulations, often rely heavily on region-specific prior knowledge, such as high-resolution topographical data or localized meteorological statistics. Although these techniques effectively refine the local granularity of precipitation data, they substantially limit the models' generalizability across varied geographical landscapes [1, 5, 6, 7, 8, 9]. Conversely, deep learning techniques, particularly convolutional neural networks (CNNs) like DeepSD [10] and UNet [5], have markedly elevated the accuracy of precipitation-related tasks [11]. However, due to their relatively limited adaptability, these models frequently necessitate retraining for each new geographical area [1]. Furthermore, a more fundamental challenge with many contemporary machine learning approaches is that their computational and data acquisition costs often exhibit a super-linear relationship with the target data resolution. For example, training a model at a global 0.125° resolution (yielding a grid of 1440 × 2880, or roughly 4.15 million grid points) demands substantial computational resources, making region-specific retraining unfeasible for large-scale deployment. This escalating demand for resources can render the training and deployment of models prohibitively expensive for global high-resolution applications, thus constraining their practical scalability and accessibility [1, 12]. Moreover, a pivotal issue in previous diffusion-based downscaling efforts is the inability to efficiently utilize multi-variable inputs, including low-resolution precipitation, temperature, humidity, wind components, elevation, and land-sea interface—often resulting in blurred outputs or subpar performance on extreme

precipitation events [13]. Given these limitations, there exists a pressing need for a scalable and generalizable framework capable of integrating a wide array of data sources while ensuring high-fidelity results.

To address these limitations, we propose a unified precipitation downscaling framework that circumvents the necessity for region-specific retraining by adeptly utilizing multi-variable inputs. Our approach leverages the Elucidating Diffusion Model (EDM) framework [14], known for its ability to capture intricate data distributions [15, 13, 16]. We enhance its generalization by infusing physical knowledge through auxiliary variables (LR precipitation, temperature, humidity, wind, elevation, and land-sea interface). To efficiently fuse these inputs, we introduce a novel Mixed-Attention UNet (MA-UNet) that integrates channel attention [17] and local-importance attention (LIA) [18]. This dual-attention mechanism prioritizes relevant meteorological variables and enhances spatial features. Our model is adept at capturing the precipitation bias—the difference between high-resolution (HR) and interpolated low-resolution (LR) precipitation—thereby enhancing performance while requiring fewer training iterations. This approach achieves state-of-the-art (SOTA) results on a limited dataset with reduced computational demands, addressing critical challenges related to scalability and cost-efficiency in precipitation modeling. The main contributions are as follows:

1. We propose UPD-Diff, a novel physics-informed diffusion framework specifically designed for achieving globally consistent precipitation downscaling. By conditioning an Elucidating Diffusion Model (EDM) on meteorological and geographical variables and training it to learn the precipitation bias, our model generates physically realistic, high-resolution outputs that adhere to the complex, long-tailed distribution of precipitation, enhancing generalization across a wide range of geographical regions.

2. We introduce a novel Mixed-Attention UNet (MA-UNet) that enhances multi-modal data fusion and model transferability. The MA-UNet integrates channel attention to weigh the influence of different physical variables with local-importance attention to accentuate critical spatial features. This dual mechanism enables the model to adaptively focus on salient information, improving generalization without relying on region-specific priors.

3. When evaluated on a globally representative dataset, our proposed UPD-Diff achieves the highest SSIM (0.7942) and Corr (0.9206) among the compared baselines, with competitive PSNR/RMSE (Table 3). Furthermore, it also shows superior skill in capturing extreme precipitation events, particularly at higher intensity thresholds.

The paper is organized as follows: Section 2 provides an overview of the related work in the fields of precipitation downscaling and diffusion models. Section 3 delves into our proposed methodology, encompassing dataset construction, the architecture of the MA-UNet, and the training strategy employed. Section 4 showcases the experimental results, including ablation studies and comparisons with baseline models. Finally, Section 5 presents the conclusions drawn from our research and outlines potential future directions.

2 Related Work

Climate downscaling aims to transform coarse-resolution model outputs into high-resolution data for local impact assessment. Methodologically, these techniques fall into two primary classes: traditional downscaling methods [19] and deep learning-based approaches [20]. While foundational, traditional methods often struggle with computational costs or questionable assumptions about climate stationarity, and many deep learning models produce overly smooth results or lack physical consistency, especially for extreme events.

2.1 Traditional Downscaling Methods

Traditional approaches include dynamical and statistical downscaling. Dynamical downscaling uses high-resolution regional climate models (RCMs), such as WRF [21], to simulate fine-scale climate by solving physical equations [19]. While physically robust, this method is computationally intensive and can inherit biases from the driving global models. In contrast, statistical downscaling establishes empirical relationships between large-scale variables and local observations. Methods range from regression-based models like SDSM [22, 23] to other transfer functions [24]. This approach is computationally efficient but assumes that historical climate relationships remain stationary, a questionable premise under climate change [25].

2.2 Deep Learning-Based Downscaling

The rise of deep learning has transformed climate downscaling, particularly for precipitation, by adapting computer vision techniques like super-resolution. These methods, including convolutional neural networks (CNNs), generative adversarial networks (GANs), and diffusion models, offer powerful tools to enhance resolution and capture complex spatial patterns.

2.2.1 CNN-Based Approaches Convolutional Neural Networks (CNNs) demonstrate remarkable prowess in extracting spatial features and learning non-linear mappings, rendering them highly suitable for downscaling tasks. The pioneering SRCNN model [26] demonstrated the potential of end-to-end CNNs for image super-resolution. However, its shallow architecture faced challenges in recovering fine details. Subsequent advances, including the use of transposed convolutions for upsampling and the incorporation of residual blocks to combat gradient vanishing, have significantly enhanced performance. In climate applications, Vandal et al. (2017) introduced DeepSD, a stacked SRCNN framework that integrates multi-scale inputs, such as topography, for statistical downscaling. This approach outperformed traditional methods like bilinear interpolation [10]. However, DeepSD's heavy reliance on a limited set of inputs—precipitation and topography—neglects other crucial variables, including temperature, wind speed, and humidity, which play pivotal roles in climate dynamics. This oversight may potentially constrain its generalizability. End-to-end architectures like UNet, characterized by their encoder-decoder design and skip connections, effectively mitigate information loss and preserve multi-scale features. In precipitation downscaling, Adewoyin et al. (2021) proposed TRU-NET, a UNet-based model with recurrent components to capture spatio-temporal dependencies, enhancing predictions of extreme precipitation events [27]. While these models are undeniably effective, they often prioritize spatial accuracy at the expense of physical consistency—a gap that our work endeavors to bridge.

2.2.2 Generative Downscaling Methods Generative models, such as GANs and diffusion models, have risen to prominence as highly effective tools for producing realistic, high-resolution outputs. By employing adversarial training strategies, GANs generate visually compelling results and have been successfully employed to downscale solar radiation, wind, and precipitation data [28, 29]. However, their training instability and susceptibility to mode collapse [30, 31] pose challenges to their reliability. Diffusion models, a more recent generative framework, model data distributions through a Markov chain that sequentially incorporates and then eliminates noise. These models demonstrate exceptional proficiency in generating high-fidelity samples and accurately quantifying uncertainty, making them highly suitable for precipitation downscaling applications [9, 13, 16]. Their probabilistic essence offers advantages over GANs, especially when it comes to predicting extreme events. Our research builds on this foundation, integrating physics-informed to enhance the realism and utility of diffusion-based downscaling.

2.2.3 Attention Mechanisms in Downscaling Attention mechanisms have revolutionized the landscape of deep learning, enabling models to selectively focus on salient features and effectively capture long-range dependencies. Within the realm of super-resolution, attention modules—such as channel attention, spatial attention, and Transformers—play a pivotal role in enhancing feature extraction and contextual modeling. Models like SwinIR and Restorme leverage the power of hierarchical attention to achieve superior global awareness [32]. In precipitation downscaling, attention can prioritize high-precipitation regions and integrate multi-modal meteorological data (e.g., wind, humidity). While their potential in generative downscaling remains underexplored, attention mechanisms offer a promising avenue for harmonizing computational efficiency with performance optimization—a direction our work actively pursues.

3 Methodology

3.1 Overview

Our research presents a concise framework for precipitation downscaling, named UPD-Diff (Unified Precipitation Downscaling leveraging Physics-Informed Diffusion and Mixed-Attention UNet). This architecture is meticulously crafted to produce high-fidelity, spatially consistent precipitation fields, demonstrating remarkable versatility across a wide range of geographical landscapes. Figure 2 offers a detailed depiction of the UPD-Diff framework, encompassing its training workflow, inference process, model architecture, and constituent submodules.

UPD-Diff integrates a physics-informed diffusion process, rooted in the Elucidating Diffusion Model (EDM), with a Mixed-Attention UNet (MA-UNet). This strategic fusion empowers the model to effectively leverage multi-modal meteorological and geographical datasets, thereby enhancing its

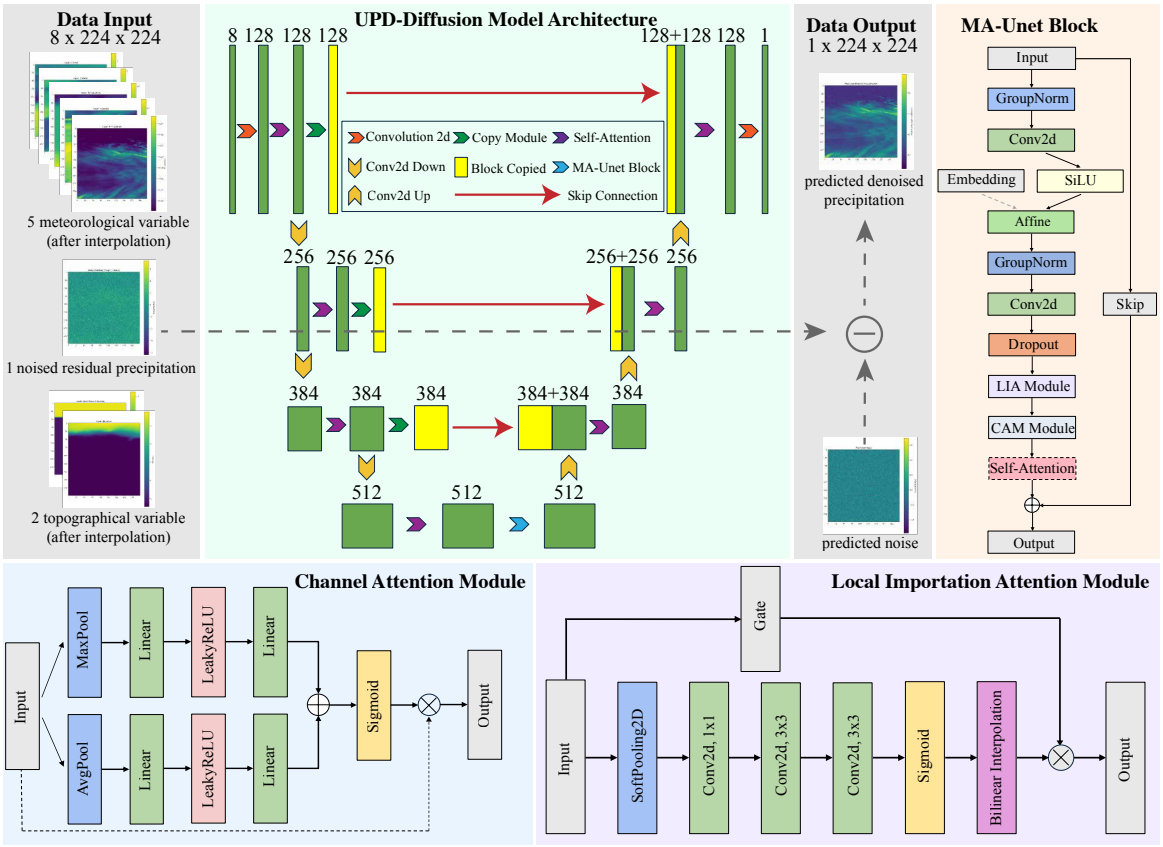

Figure 2. The flowchart of the proposed UPD-Diff model. At its core lies a U-Net-based Mixed-Attention UNet (MA-UNet), designed to process an 8-channel input that integrates meteorological, topographical, and noised residual precipitation data. The MA-UNet (central component, elaborated in the MA-UNet Block at the top right) employs skip connections alongside specialized attention modules. These modules include the Channel Attention Module (CAM, positioned at the bottom left), which dynamically recalibrates channel-wise features, and the Local Importance Attention (LIA) Module (located at the bottom right), which highlights spatially critical regions. Together, they facilitate the efficient fusion of multi-modal information and enhance feature representation for precipitation downscaling.

predictive prowess. Our methodology is meticulously structured, encompassing dataset compilation and preprocessing protocols, model architecture refinement, and a bespoke training strategy. Each component has been proposed to tackle the formidable challenges of generalization and extreme event prediction inherent in precipitation downscaling, ensuring the delivery of robust and dependable forecasts across diverse climatic conditions.

3.2 Model Training

The core of UPD-Diff lies in the integration of the Elucidating Diffusion Model (EDM) [14] framework with MA-UNet. This integration is specifically tailored to perform precipitation downscaling, guided by multi-modal inputs. Unlike traditional Denoising Diffusion Probabilistic Models (DDPM) [33, 34], EDM enhances noise scheduling and sampling techniques to better accommodate the long-tailed distribution inherent in precipitation data. These enhancements markedly improve the correlation between the generated super-resolution (SR) precipitation fields and their high-resolution (HR) ground truth counterparts. The detailed training procedure, encompassing the forward (noising) and reverse (denoising network optimization) phases, is formally presented in Algorithm 1.

Forward Diffusion Starting with the HR precipitation bias x_0, Gaussian noise is incrementally added over T timesteps according to EDM's tailored noise schedule, which prioritizes modeling extreme precipitation events. At each timestep t, the noisy version x_t is generated as:

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t}\epsilon_t$$

where α_t controls the noise schedule and $\epsilon_t \sim \mathcal{N}(0, I)$ is standard Gaussian noise. As time t progresses to T, this process incrementally transforms x_0 into isotropic noise.

Reverse Diffusion Unlike the standard Denoising Diffusion Probabilistic Models (DDPM) [33] that rely on a discrete variance schedule β_t, our framework adopts the Elucidating Diffusion Model

Algorithm 1 UPD-Diff Training Process

Require: Training dataset $\mathcal{D} = \{(x_0^{(j)}, c^{(j)})\}_{j=1}^{N}$ where x_0 is HR precipitation bias and c are conditional inputs.
Require: Log-normal distribution parameters (μ_p, σ_p), balancing parameter λ.
Initialize MA-UNet model parameters θ.
Initialize Optimizer (e.g., AdamW).
repeat
  Sample a mini-batch (x_0, c) from $\mathcal{D}$.
  Sample noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$.
  Sample noise scale $\ln(\sigma) \sim \mathcal{N}(\mu_p, \sigma_p^2)$. ▷ EDM log-normal scheduling
  Construct noisy input $x_\sigma \leftarrow x_0 + \sigma\epsilon$.
  Compute preconditioned output $\hat{x}_0 \leftarrow D_\theta(x_\sigma, \sigma, c)$. ▷ Following EDM formula
  Calculate diffusion loss $\mathcal{L}_{\text{diff}} \leftarrow \|\hat{x}_0 - x_0\|^2$.
  Calculate combined loss $\mathcal{L} \leftarrow \mathcal{L}_{\text{diff}}$.
  Update model parameters θ using optimizer and $\mathcal{L}$.
until convergence criteria met

(EDM) [14] to better handle the long-tailed nature of precipitation data. In EDM, the diffusion process is defined by a continuous noise scale σ, and the neural network D_θ is preconditioned as follows:

$$D_\theta(\mathbf{x}, \sigma, \mathbf{c}) = c_{skip}(\sigma)\mathbf{x} + c_{out}(\sigma)F_\theta(c_{in}(\sigma)\mathbf{x}, c_{noise}(\sigma), \mathbf{c}) \tag{1}$$

where c_{skip}, c_{out}, c_{in}, and c_{noise} are specific scaling factors that ensure the effective input and output of the network F_θ maintain unit variance.

This preconditioning is particularly crucial for extreme precipitation downscaling. In standard DDPM, the high-intensity peaks of extreme events often lead to large gradient fluctuations, making the model prone to underestimating peak values. EDM mitigates this by normalizing the training signal across all noise levels. Furthermore, while DDPM uses a linear or cosine noise schedule that may over-allocate capacity to uninformative high-noise regimes, EDM employs a log-normal distribution for training noise: $\ln(\sigma) \sim \mathcal{N}(\mu_p, \sigma_p^2)$. By tuning μ_p and σ_p based on the characteristic scales of precipitation bias, we enable the model to prioritize learning the spatial structures of intense rain cells.

During the training phase, instead of the traditional noise prediction objective, our MA-UNet is optimized to reconstruct the denoised precipitation bias directly through the preconditioned framework. The diffusion loss function is formulated as:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{\mathbf{x}_0, \epsilon, \sigma}\|D_\theta(\mathbf{x}_0 + \sigma\epsilon, \sigma, \mathbf{c}) - \mathbf{x}_0\|^2 \tag{2}$$

where $\mathbf{x}_0$ represents the ground-truth precipitation bias and $\mathbf{c}$ denotes the multi-modal conditional inputs. During inference, the model generates high-resolution precipitation by solving the corresponding probability flow ODE. This allows for a non-uniform sampling density that focuses on the σ ranges where the precipitation morphology evolves most rapidly, directly leading to the improved CSI and HSS scores observed for high-intensity thresholds.

As illustrated in Figure 2, the MA-UNet adopts an encoder-decoder framework enhanced with skip connections, further enhanced by two attention mechanisms. The Channel Attention Module (CAM) dynamically assigns weights to input channels based on their relevance to precipitation bias, thereby minimizing redundancy and prioritizing informative features such as humidity or wind patterns. Complementing CAM, the Local-Importance Attention (LIA) module generates a spatial importance map that highlights regions of intense precipitation, such as heavy rainfall zones, thereby enhancing local detail and improving the prediction of extreme weather events. Collectively, CAM and LIA empower the MA-UNet to efficiently integrate multi-modal data, striking a balance between global consistency and localized precision.

We utilize the AdamW optimizer with a learning rate of 1e−4 and a batch size of 16, training the model for 200 epochs. Early stopping is activated if the validation Mean Squared Error (MSE) shows no improvement over 20 consecutive epochs. Performance is evaluated using pixel-level metrics (PSNR, SSIM, RMSE) to assess spatial fidelity and event-based metrics (POD, HSS, FSS) to evaluate extreme event detection capabilities, ensuring a thorough and comprehensive assessment.

UPD-Diff sets itself apart by integrating physics-informed EDM diffusion with mixed attention mechanisms. By explicitly targeting the prediction of precipitation bias and harnessing multi-modal data, it produces high-resolution outputs that adhere to physical principles. This residual-learning

Algorithm 2 UPD-Diff Inference Process

Require: Conditional inputs c (normalized).
Require: Low-resolution interpolated precipitation $Y_{LR,\text{interp}}$.
Require: Trained MA-UNet model parameters θ.
Require: Number of diffusion steps T_{inf} (e.g., 100 from your paper).
Require: Mean μ_{bias} and std σ_{bias} of precipitation bias from training set.
Sample initial noise $x_{T_{inf}} \sim \mathcal{N}(0, \mathbf{I})$.
for $t = T_{inf}, \ldots, 1$ **do**
    Predict mean $\mu_\theta(x_t, c, t)$ using MA-UNet. ▷ Based on noise prediction ϵ_θ
    Sample $z \sim \mathcal{N}(0, \mathbf{I})$ if $t > 1$, else $z \leftarrow 0$.
    $x_{t-1} \leftarrow \mu_\theta(x_t, c, t) + \Sigma_\theta(t)^{1/2} z$. ▷ EDM reverse step, $\Sigma_\theta(t)$ is variance
end for
Predicted normalized bias $\hat{y}_{\text{bias,norm}} \leftarrow x_0$.
Inverse normalize bias: $\hat{y}_{\text{bias}} \leftarrow \hat{y}_{\text{bias,norm}} \cdot \sigma_{\text{bias}} + \mu_{\text{bias}}$.
Reconstruct HR precipitation: $Y_{SR} \leftarrow Y_{LR,\text{interp}} + \hat{y}_{\text{bias}}$.
Post-process: $Y_{SR,\text{final}} \leftarrow \text{Max}(Y_{SR}, 0)$.
Return $Y_{SR,\text{final}}$.

strategy allows the model to excel at capturing extreme precipitation events, thereby improving structural fidelity and event-skill metrics over CNN/GAN baselines.

3.3 Model Inference

During the inference phase, our Unified Precipitation Downscaling with Physics-Informed Diffusion and Mixed-Attention Unet (UPD-Diff) framework generates high-resolution (HR) precipitation fields by predicting the precipitation bias—defined as the residual between HR precipitation and bicubic-interpolated low-resolution (LR) precipitation. By adopting this bias-centric strategy, the model can focus on capturing the fine-scale details that are pivotal for effective downscaling, instead of endeavoring to reconstruct the entire HR precipitation field directly.

The inference pipeline seamlessly combines several crucial steps: data normalization, iterative denoising via the diffusion process, inverse normalization, bias addition, and post-processing. The inference pipeline seamlessly combines several crucial steps: data normalization, iterative denoising via the diffusion process, inverse normalization, bias addition, and post-processing. A detailed description of this integrated pipeline is presented in Algorithm 2. This integrated approach guarantees both the accuracy and physical coherence of the outputs. Below, we delineate each step in detail to offer a transparent and reproducible methodology.

Data Normalization To ensure consistency with the training phase, all input meteorological variables, encompassing low-resolution (LR) precipitation, temperature, humidity, u- and v-wind components, elevation, and the land-sea interface, undergo normalization via the Z - score method. The normalization transformation for each variable x is carried out as follows:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma}$$

where μ and σ are the mean and standard deviation, respectively, derived from the training data set. This normalization process standardizes the inputs, maintaining their relative contributions, and promoting stable conditioning within the model.

Iterative Denoising via Diffusion The inference process leverages the reverse diffusion mechanism of the Elucidating Diffusion Model (EDM), where the model iteratively denoises a noisy sample of the precipitation bias, conditioned on the normalized multi-modal inputs. We configure the denoising process with 100 steps, striking a balance between computational efficiency and output fidelity. Initiating from a randomly generated noise sample $x_T \sim \mathcal{N}(0, I)$, the model applies the learned reverse transition at each timestep t:

$$x_{t-1} = \mu_\theta(x_t, t, c) + \Sigma_\theta(t)^{1/2} z, \quad z \sim \mathcal{N}(0, I)$$

where $\mu_\theta(x_t, t, c)$ is the mean predicted by the Mixed-Attention UNet (MA-UNet), conditioned on c (the normalized LR precipitation and auxiliary variables), and $\Sigma_\theta(t)$ is the optimized time-dependent variance from EDM's noise schedule. After 100 iterations, the model outputs the predicted precipitation bias in its normalized form, denoted $\hat{y}_{\text{bias, norm}}$.

Table 1. Inference metrics on a single RTX 4090 (tile size 224×224, batch size 8, 100 diffusion steps).

Tiles	Time	Throughput	Per-tile (s)	VRAM (GB)
400	34 min	0.20	5.1	6.8

Inverse Normalization and Bias Addition Since the diffusion process operates on normalized data, the predicted bias $\hat{y}_{\text{bias, norm}}$ is first transformed back to its original scale using inverse Z-score normalization:

$$\hat{y}_{\text{bias}} = \hat{y}_{\text{bias, norm}} \cdot \sigma_{\text{bias}} + \mu_{\text{bias}}$$

where μ_{bias} and σ_{bias} are the mean and standard deviation of the bias computed from the training set. The super-resolution (SR) precipitation is then reconstructed by adding this bias to the bicubic-interpolated LR precipitation:

$$\hat{y}_{\text{SR}} = \hat{y}_{\text{LR, interp}} + \hat{y}_{\text{bias}}$$

This step ensures that the model's output builds upon the coarse-scale information while enhancing it with fine-scale details captured in the bias.

Post-Processing To enforce physical plausibility, we implement a thresholding correction on the super-resolved (SR) precipitation data:

$$\hat{y}_{\text{SR}} = \max(\hat{y}_{\text{SR}}, 0)$$

This eliminates negative precipitation values, which are physically unrealistic, thereby enhancing the output's suitability for practical applications like weather forecasting and climate modeling. The generated super-resolution (SR) precipitation field combines high spatial fidelity with adherence to meteorological constraints.

This inference pipeline leverages the strengths of physics-informed diffusion and the MA-UNet architecture, delivering SR precipitation fields that are rich in detail and operationally feasible. By addressing inherent biases and integrating multi-modal conditioning, UPD-Diff ensures robust scalability and generalizability across a wide range of global regions.

4 Experiment and Analysis

4.1 Experiment Setup

In this section, we evaluate the performance of our proposed UPD-Diff framework by conducting ablation studies and benchmarking it against state-of-the-art (SOTA) baselines. To ensure a rigorous and fair comparison, strictly identical experimental protocols were applied across all models. Specifically, all baselines utilized the exact same set of 7-channel input variables (encompassing low-resolution precipitation, temperature, humidity, wind components, elevation, and land-sea mask), underwent uniform Z-score normalization, and were trained and evaluated on consistent dataset splits. Our evaluation metrics encompass pixel-level fidelity (e.g., PSNR, SSIM) and meteorological proficiency (e.g., CSI, HSS), with a specific focus on accurately capturing extreme precipitation events. All experiments are performed on a globally sampled test set sourced from the CMA-GFS and ERA5 datasets.

In addition, all experiments are performed on a single Nvidia RTX 4090 GPU. The models are trained for 200 epochs, with early stopping implemented based on a patience threshold of 20 epochs. Our implementation is based on the PyTorch framework. The complete codebase is publicly available at https://anonymous.4open.science/r/UPD-Diffusion-2BDA, facilitating reproducibility and further research.

For runtime characteristics, with 100 diffusion steps we process 400 tiles in 34 minutes (batch size 8) on an RTX 4090. The per-tile time is ≈5.1 s (0.20 tiles/s; peak VRAM: 6.8G); see Table 1.

4.2 Datasets and Evaluation Metrics

Our study utilizes a dataset integrating high-resolution (0.125°) meteorological variables from CMA-GFS with elevation and land-sea data from ERA5. This multi-modal dataset, covering the entire globe, captures both dynamic weather patterns and static geographical features. Preprocessing includes Z-score normalization and extracting 224×224 patches. Our model learns to predict the bias between high-resolution (HR) and bicubic-interpolated low-resolution (LR) precipitation. We use a balanced training set of 4000 images, with 400 images each for validation and testing. Detailed information on dataset construction and preprocessing is available in the Supplementary Materials.

Table 2. Pixel-level metrics for ablation study configurations. UPD-Diff's superior results validate its mixed attention and multi-stream design for characterizing precipitation patterns.

Configuration	PSNR	SSIM	RMSE	Corr
EDM-SingleInput	27.81	0.6672	4.1497	0.9164
EDM-MultiInput	27.91	0.7776	4.4500	0.9006
EDM-ChannelAttn	28.08	0.7561	4.5106	0.9071
EDM-LocalAttn	27.92	0.7770	4.8263	0.8999
EDM-CrossAttn	27.98	0.7207	4.5351	0.9066
EDM-ChannelSpatialAttn	27.78	0.7642	4.5709	0.9017
UPD-Diff(Ours)	**29.02**	**0.7941**	**3.9816**	**0.9206**

To assess performance, we use a comprehensive set of metrics. Pixel-level fidelity is measured using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Root Mean Square Error (RMSE), and the Correlation Coefficient (Corr). Event detection skill, particularly for extreme events, is evaluated using the Probability of Detection (POD), Critical Success Index (CSI), and Heidke Skill Score (HSS) at various precipitation thresholds (0.1, 1, 10, 20, 50, and 100 mm). Detailed definitions of these metrics are provided in the Supplementary Materials.

4.3 Quantitative and qualitative results

4.3.1 Ablation Studies To assess the contributions of different components in our Elucidating Diffusion Model (EDM), we conducted experiments using the following configurations, each tailored to emphasize a distinct facet of the framework:

- **EDM-SingleInput:** Building upon the baseline, this variant integrates LR precipitation with auxiliary meteorological factors (e.g., wind speed, humidity) and geographical features (e.g., elevation).
- **EDM-MultiInput:** Incorporates LR precipitation plus auxiliary meteorological (e.g., wind, humidity) and geographical (e.g., elevation) variables.
- **EDM-ChannelAttn:** By augmenting the UNet architecture with channel attention mechanisms, this setup prioritizes the most relevant feature channels for improved performance.
- **EDM-LocalAttn:** This configuration incorporates local importance attention (LIA) into the UNet, directing focus towards spatially critical regions within the data
- **UPD-Diff:** Representing the full model, this architecture combines channel attention and LIA within a Mixed-Attention UNet framework, offering a comprehensive solution.
- **EDM-ChannelSpatialAttn:** Leveraging both channel and spatial attention, this variant provides a broader emphasis on features, enhancing the model's ability to capture complex patterns.
- **EDM-CrossAttn:** Utilizing cross-attention mechanisms, this setup models the intricate relationships between precipitation and auxiliary variables, drawing inspiration from [35].

These names are thoughtfully designed to be highly intuitive: "SingleInput" and "MultiInput" clearly convey the varying degrees of input complexity, whereas "ChannelAttn," "LocalAttn," and "CrossAttn" succinctly denote the distinct attention mechanisms employed. Additionally, "UPD-Diff" encapsulates the entire, meticulously optimized framework.

As demonstrated in Table 2, the outcomes of our ablation studies offer pivotal insights into the respective contributions of different components within the UPD-Diff framework.

Beginning with the input configuration, the shift from **EDM-SingleInput** (which relies solely on LR precipitation data) to **EDM-MultiInput** (which integrates auxiliary meteorological and geographical variables) leads to a notable enhancement in SSIM, increasing from 0.6672 to 0.7776. This highlights the significance of multi-modal information in capturing the intricate spatial patterns inherent in precipitation fields. However, while PSNR exhibits a slight improvement and RMSE shows a minor deterioration, it suggests that although multi-input integration is advantageous for structural fidelity, further refinement is necessary to achieve optimal pixel-wise accuracy.

An analysis of the impact exerted by different attention mechanisms integrated into the multi-input baseline model reveals distinct outcomes. **EDM-ChannelAttn** exhibits an enhancement in both PSNR (28.08) and Corr (0.9071) compared to EDM-MultiInput, underscoring that

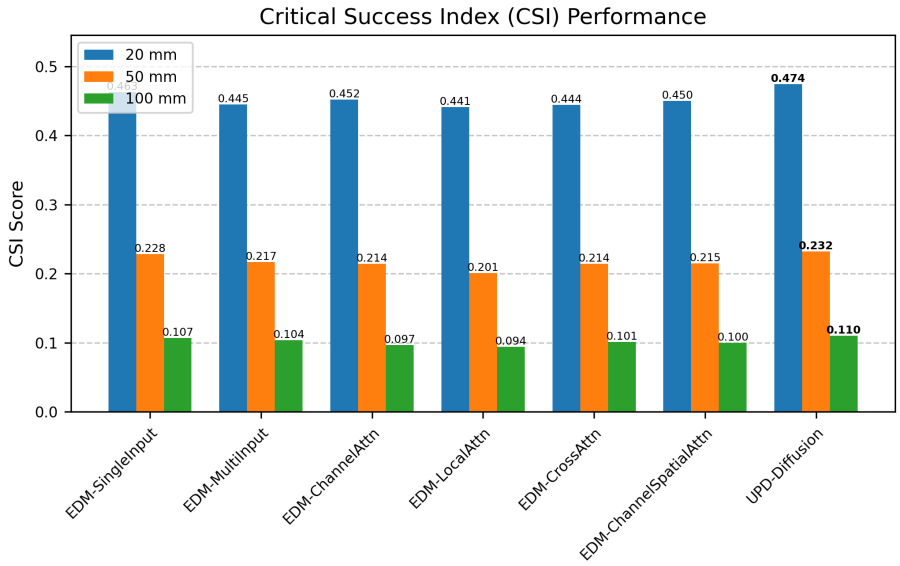

Figure 3. Critical Success Index (CSI) for Extreme Precipitation. The bar chart displays the CSI scores for different Elucidating Diffusion Model (EDM) configurations (namely EDM-SingleInput, EDM-MultiInput, EDM-ChannelAttn, EDM-LocalAttn, EDM-CrossAttn, EDM-ChannelSpatialAttn) and the proposed UPD-Diff framework. Scores are presented across three distinct precipitation thresholds: 20 mm (blue bars), 50 mm (orange bars), and 100 mm (green bars), highlighting UPD-Diff's consistent lead in accurately identifying significant precipitation events.

prioritizing pertinent feature channels elevates pixel-level precision and linear correlation. Nevertheless, a marginal decline in SSIM is observed. In contrast, **EDM-LocalAttn**, designed to concentrate on spatially critical regions, results in a significant rise in RMSE (4.8263) and a slight downturn in SSIM and Corr relative to EDM-MultiInput. This suggests that an exclusive focus on local significance, devoid of global context or channel-specific feature weighting, might undermine overall structural coherence and inflate pixel errors. The **EDM-CrossAttn** configuration, aimed at modeling intricate relationships between precipitation and auxiliary variables, showed a slight PSNR and Corr improvement but a significant drop in SSIM (0.7207). This aligns with our subsequent discussion on the challenges of aligning heterogeneous data distributions, where the complexity of cross-attention might not be fully realized without proper distribution alignment, potentially disrupting structural information. The **EDM-ChannelSpatialAttn**, combining standard channel and spatial attention, fails to outperform simpler configurations, highlighting that a naive combination of attention mechanisms is not guaranteed to improve performance.

Ultimately, the proposed **UPD-Diff** framework, which integrates a Mixed-Attention UNet (MA-UNet) combining channel attention and Local-Importance Attention (LIA), demonstrates strong performance, with the highest SSIM (0.7942) and Corr (0.9206) among the compared baselines, and competitive PSNR (29.02) and RMSE (3.9816). This indicates complementary benefits of the MA-UNet design. Channel attention effectively weighs the contribution of diverse input features, while LIA refines the focus on spatially significant precipitation patterns. This combined approach allows UPD-Diff to maintain competitive pixel-wise fidelity (PSNR, RMSE), preserve structural information (SSIM), and achieve high overall correlation (Corr), validating the effectiveness of our mixed attention strategy in capturing both the fine-grained details and the broader contextual characteristics of precipitation.

Notably, the proposed UPD-Diff demonstrates robust performance not only in pixel-level reconstruction (high SSIM and Corr) but also in detecting meteorological events across varying intensities.

As detailed in the subsequent analysis, Figures 3 and 4 visually compare the Critical Success Index (CSI) and Heidke Skill Score (HSS) for various EDM configurations against UPD-Diff. These comparisons, made at precipitation thresholds of 20 mm, 50 mm, and 100 mm, illustrate UPD-Diff's superior capability in identifying moderate to extreme precipitation events compared to single-input or limited-attention baselines.

4.3.2 Comparison with SOTA Baselines We evaluate the performance of UPD-Diff by benchmarking it against several SOTA methods, including CNN-based approaches like DeepSD [10] and FSRCNN-ESM [6], the GAN-based Encoded_CGAN [36], and a standard UNet baseline. The UNet baseline shares our model's architecture but is trained with a pixel-wise L1 loss. This selection provides a robust foundation for a comprehensive comparison with our diffusion modeling framework.

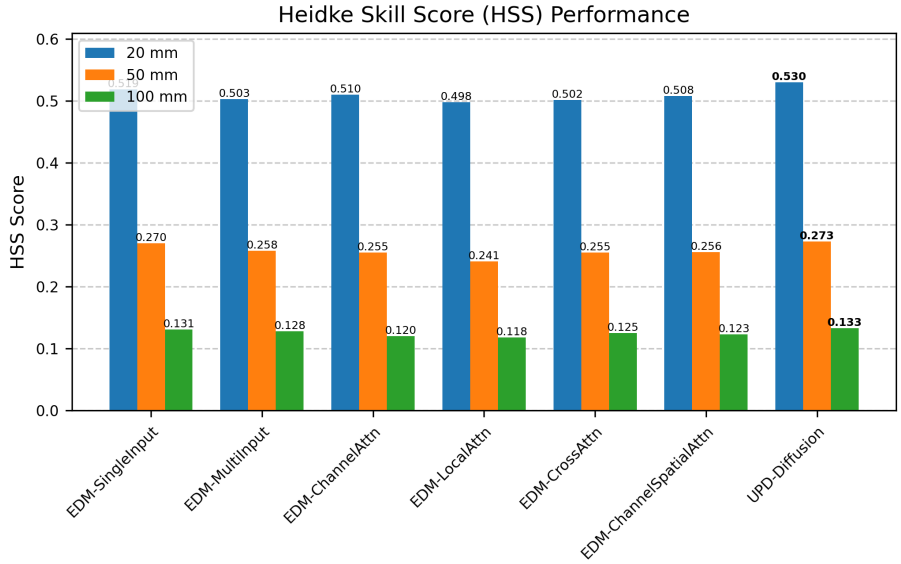

Figure 4. Heidke Skill Score (HSS) for Extreme Precipitation. This bar chart illustrates the HSS for the same set of EDM configurations (EDM-SingleInput, EDM-MultiInput, EDM-ChannelAttn, EDM-LocalAttn, EDM-CrossAttn, EDM-ChannelSpatialAttn) and the UPD-Diff model. The HSS values are shown for precipitation thresholds of 20 mm (blue bars), 50 mm (orange bars), and 100 mm (green bars), further demonstrating UPD-Diff's enhanced predictive skill for moderate to extreme precipitation relative to other configurations.

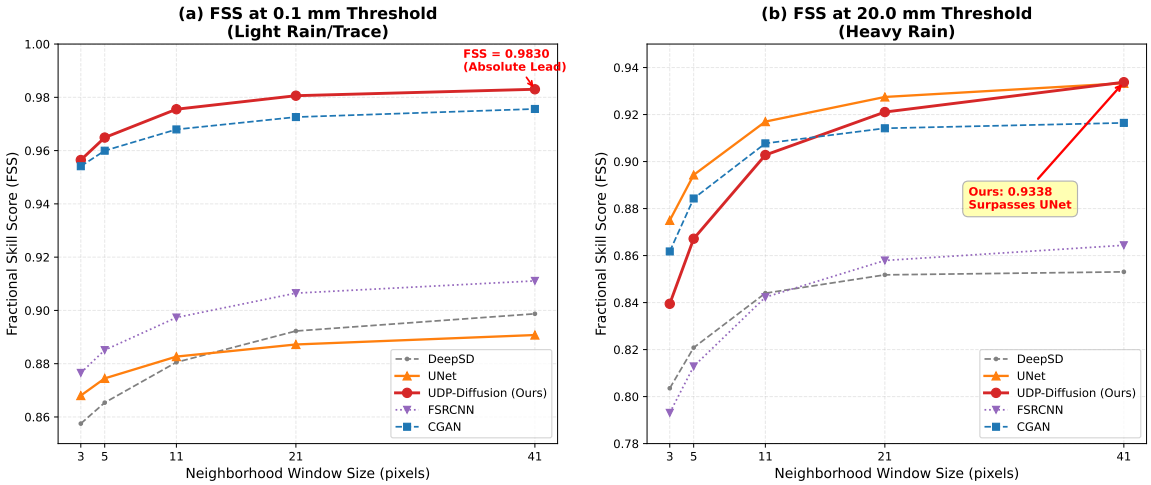

Figure 5. Fractional Skill Score (FSS) as a function of neighborhood window size. (a) At the 0.1 mm threshold (representing rain area delineation), UPD-Diff consistently outperforms all baselines across spatial scales, demonstrating superior capability in capturing the correct extent of precipitation systems. **(b)** At the 20 mm threshold (representing heavy precipitation), while pixel-wise optimized models like UNet show high initial scores due to smoothing, UPD-Diff demonstrates a rapid performance gain with increasing window size, eventually matching or surpassing UNet at meso-scales (e.g., 41×41). This indicates that UPD-Diff correctly positions extreme weather events while preserving high-frequency textures that are often lost in regression-based methods.

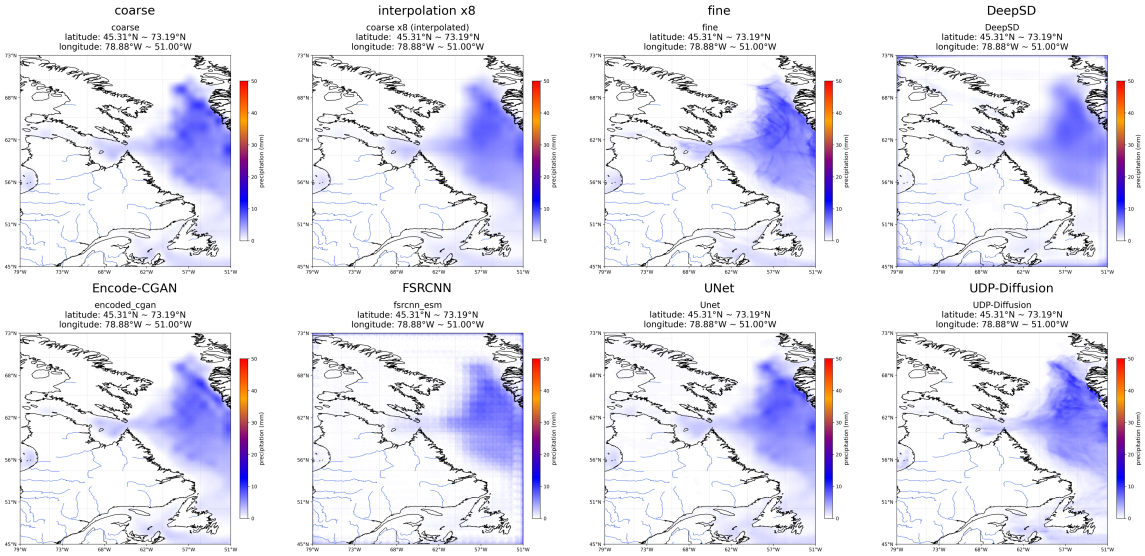

Figure 6. Qualitative comparison of high-resolution precipitation downscaling results from various state-of-the-art (SOTA) models for a representative precipitation event. The first row, from left to right, displays: the Low-Resolution (LR) input, the High-Resolution (HR) observation (Ground Truth), the output from Bilinear Interpolation, and the output from DeepSD. The second row, from left to right, shows results from: FSRCNN-ESM, Encoded-Conditional GAN, UNet, and our proposed UPD-Diff model. This visual comparison shows that UPD-Diff recovers fine-scale precipitation patterns and intensities more faithfully in this example, consistent with its higher SSIM/Corr in Table 3. All panels share the same colorbar range, so identical hues indicate identical intensities across methods.

Table 3. Pixel-level metrics vs. SOTA baselines. UPD-Diff shows top SSIM/Corr (superior structural/distributional modeling). Slightly lower PSNR/RMSE reflects its focus on overall data distribution learning over pixel-level optimization.

Model	PSNR	SSIM	RMSE	Corr
DeepSD [10]	27.28	0.7281	3.5723	0.8732
FSRCNN-ESM [6]	25.36	0.6651	4.3064	0.8262
Encoded_CGAN [36]	**30.43**	0.7907	3.4539	0.907
Unet	30.51	0.7817	**3.151**	0.901
UPD-Diff(Ours)	29.02	**0.7942**	3.9816	**0.9206**

Detailed descriptions of these baseline models and their configurations are provided in the Supplementary Materials.

An analysis of the pixel-level results in Table 3 reveals that while UPD-Diff outperforms all baselines in SSIM (0.7942) and Correlation (0.9206), it exhibits slightly lower PSNR and RMSE compared to L1-optimized models like UNet. This discrepancy highlights an inherent bias in PSNR and RMSE, which typically reward smoother, more blurred images. Because these metrics calculate error strictly at the individual pixel level, they favor models that produce a "statistical average" of all possible precipitation patterns. While such over-smoothing numerically reduces pixel-wise variance, it inevitably erases the sharp, localized intensity peaks that are vital for capturing extreme weather.

To further investigate the models' ability to reproduce realistic spatial structures across various scales, we analyze the Radially Averaged Power Spectral Density (RAPSD), as shown in Figure 7. The RAPSD plot reveals how well each model captures the variance at different spatial frequencies. Our UPD-Diff model's spectrum closely matches that of the high-resolution reference across a wide range of frequencies, particularly demonstrating superior performance at higher frequencies, which correspond to finer-scale details. Many baseline models often fall short in accurately reproducing these small-scale features. This indicates that UPD-Diff not only preserves large-scale patterns but also generates more realistic fine-scale textures compared to other DL models. This spectral fidelity is crucial for capturing the hierarchical nature of precipitation systems.

Finally, we present a comprehensive evaluation of meteorological skill against SOTA baselines. Table 4 details the POD, CSI, and HSS scores across three representative intensity thresholds: Light Rain (> 0.1 mm), Heavy Rain (> 50 mm), and Torrential Rain (> 100 mm).

Performance on Light Precipitation: At the low threshold (0.1 mm), regression-based models like UNet and Encoded-CGAN achieve high POD (0.9804) and HSS (0.608) scores. This is expected, as L1/L2 loss functions tend to produce smoother fields that cover wider areas, essentially hedging against "misses." However, our UPD-Diff still maintains a highly competitive CSI of 0.867,

Table 4. Comparison of Meteorological Skill Scores. The table reports Probability of Detection (POD), Critical Success Index (CSI), and Heidke Skill Score (HSS) at three distinct precipitation thresholds.

Model	POD	CSI	HSS
Threshold = 0.1 mm (Light Rain)			
DeepSD [1]	0.9500	0.7780	0.3290
FSRCNN-ESM [2]	0.9367	0.7690	0.3000
Encoded-CGAN [3]	0.9163	0.8630	**0.6080**
UNet	**0.9804**	0.7930	0.3800
UPD-Diff (Ours)	0.9019	**0.8670**	0.5300
Threshold = 50 mm (Heavy Rain)			
DeepSD [1]	0.2430	0.2170	0.2470
FSRCNN-ESM [2]	0.2244	0.1980	0.2280
Encoded-CGAN [3]	0.2578	0.2260	0.2710
UNet	0.2704	0.2220	0.2710
UPD-Diff (Ours)	**0.2879**	**0.2320**	**0.2730**
Threshold = 100 mm (Torrential Rain)			
DeepSD [1]	0.1172	0.1020	0.1330
FSRCNN-ESM [2]	0.1067	0.0920	0.2280
Encoded-CGAN [3]	0.1123	0.1030	0.2710
UNet	0.1216	0.1020	0.2710
UPD-Diff (Ours)	**0.1384**	**0.1100**	**0.2730**

outperforming DeepSD and FSRCNN, which indicates a precise delineation of rain/no-rain boundaries.

Superiority in Extreme Precipitation: The advantages of UPD-Diff become pronounced at higher intensities, where baseline models falter. As shown in Table 4, for Heavy Rain (> 50 mm), UPD-Diff surpasses all baselines, achieving the highest POD (0.2879), CSI (0.232), and HSS (0.273). This trend continues to the Torrential threshold (> 100 mm), where UPD-Diff achieves a POD of 0.1384 and CSI of 0.110, significantly outperforming DeepSD (CSI 0.102) and UNet (CSI 0.102).

To further assess the spatial structural consistency and tolerance to minor spatial displacements, we evaluated the Fractional Skill Score (FSS) across different neighborhood scales. As shown in Figure 5, UPD-Diff exhibits distinct advantages: (1) For rain area delineation (> 0.1 mm), our model consistently achieves the highest FSS across all spatial scales, indicating superior structural fidelity. (2) For heavy precipitation (> 20 mm), while pixel-wise optimized baselines (e.g., UNet) show high scores at small scales due to smoothing, UPD-Diff demonstrates a rapid performance gain with increasing window size, matching or surpassing baselines at meso-scales (e.g., 41×41). This confirms that our physics-informed diffusion model correctly positions extreme weather systems.

Unlike regression baselines that often blur extreme values towards the mean (leading to underestimation of extremes), the generative nature of UPD-Diff allows it to reconstruct the long-tailed distribution of precipitation. By learning the precipitation bias explicitly, our model successfully recovers high-intensity structures that are critical for hazardous weather warnings, validating its practical utility for meteorological applications.

4.3.3 Visualization of Attention Mechanism To further validate the interpretability of the proposed method and understand how the Local-Importance Attention (LIA) module contributes to performance gains, we visualized the learned spatial importance maps. By extracting the intermediate outputs of the LIA module, we analyzed the correlation between the attention weights and the precipitation intensity.

Figure 8 presents the visualization results across multiple samples with varying precipitation intensities. The results reveal a clear mechanism:

- **High-Frequency Capture:** The LIA module demonstrates a strong capability to locate high-frequency spatial details. As observed in Sample 307 and Sample 304, the regions characterized by heavy rainfall (torrential events) correspond to the highest activation values in the LIA importance map (indicated by dark red zones).
- **Spatial Adaptability:** The overlay visualization confirms that the attention mechanism is

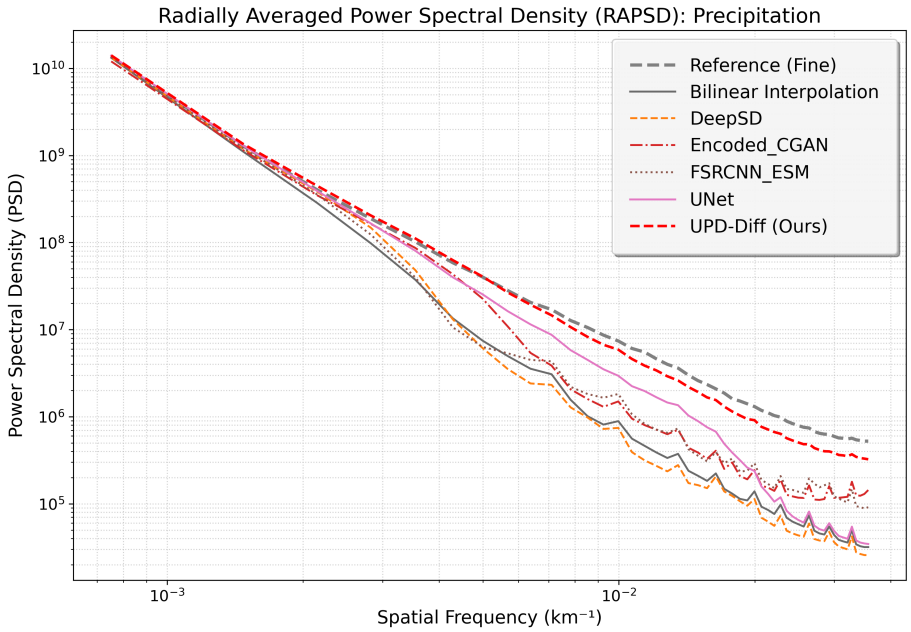

Figure 7. Radially Averaged Power Spectral Density (RAPSD) of precipitation fields for different downscaling models compared against High-Resolution (Reference) data. The plot illustrates the spatial variance distribution across different spatial frequencies. The closer a model's spectrum is to the Reference, particularly at higher frequencies (smaller scales), the better its ability to reproduce fine-scale details and textures.

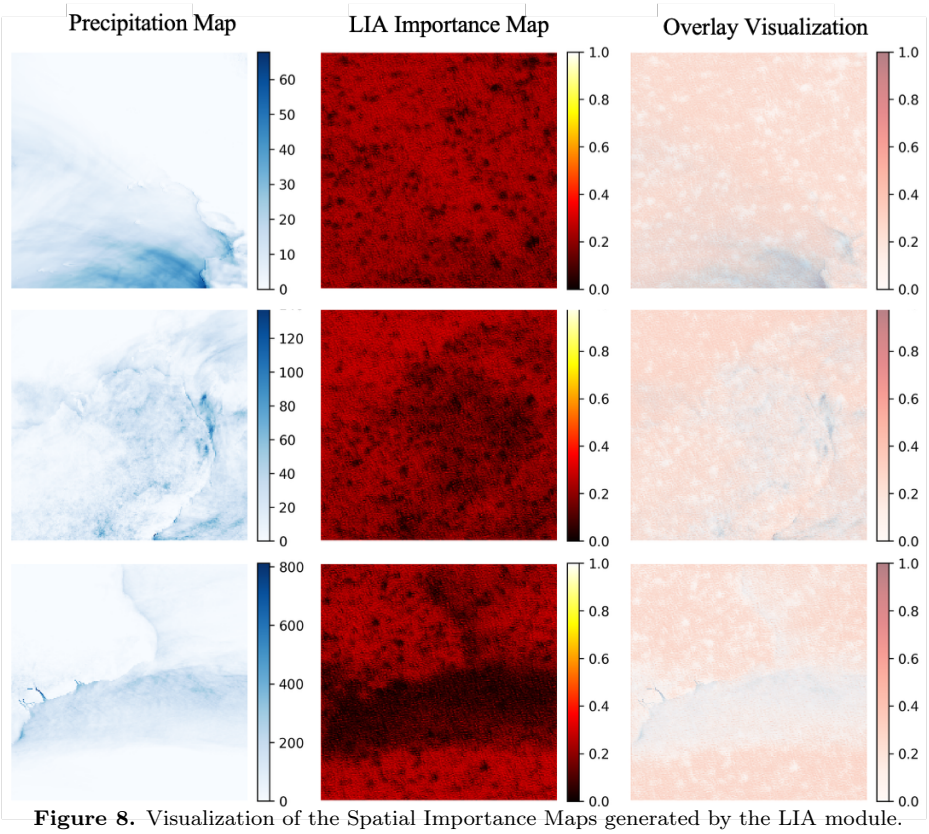

Figure 8. Visualization of the Spatial Importance Maps generated by the LIA module.

spatially adaptive. Unlike static weights, the LIA dynamically adjusts its focus based on the input meteorological conditions, effectively acting as a "soft gate" that emphasizes physically significant features while suppressing background noise.

This visual evidence supports the ablation study results in Table 2, explaining why the inclusion of LIA significantly improves the SSIM and event-based metrics (CSI/HSS) for extreme precipitation.

5 Conclusions and Future work

In this study, we propose UPD-Diff, a unified precipitation downscaling framework that introduces a novel Mixed-Attention UNet (MA-UNet) and integrates Elucidating Diffusion Models (EDMs) with meteorological and geographic variables. This approach addresses the limitations of existing models that rely on region-specific knowledge and require extensive retraining, thereby optimizing the use of multivariate inputs and improving spatial detail. Results on global datasets demonstrate that UPD-Diff significantly outperforms state-of-the-art benchmarks in both pixel-level metrics and meteorological skill scores. Looking ahead, our future work will focus on developing domain-adaptive attention mechanisms, incorporating more physics-informed principles, and extending the framework to simulate temporal dynamics or simultaneously downscale multiple climatic variables.

References

[1] Morteza Mardani et al. "Residual corrective diffusion modeling for km-scale atmospheric downscaling". In: *Communications Earth & Environment* 6.1 (2025), p. 124.

[2] Nan Yang, Chong Wang, and Xiaofeng Li. "Improving tropical cyclone precipitation forecasting with deep learning and satellite image sequencing". In: *Journal of Geophysical Research: Machine Learning and Computation* 1.2 (2024), e2024JH000175.

[3] Nan Yang and Xiaofeng Li. "An individual motion-driven artificial intelligence method for precipitation forecasting using radar image sequencing". In: *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024), pp. 1–16.

[4] Jiahan Chen et al. "A Review of Super Resolution Technology for Extended-Range Precipitation Forecasting". In: *2024 International Conference on Computer and Applications (ICCA)*. IEEE. 2024, pp. 1–7.

[5] Takeyoshi Nagasato et al. "Extension of convolutional neural network along temporal and vertical directions for precipitation downscaling". In: *arXiv preprint arXiv:2112.06571* (2021).

[6] Linsey S Passarella et al. "Reconstructing high resolution ESM data through a novel fast super resolution convolutional neural network (FSRCNN)". In: *Geophysical Research Letters* 49.4 (2022), e2021GL097571.

[7] Luca Glawion et al. "Global spatio-temporal downscaling of ERA5 precipitation through generative AI". In: *arXiv preprint arXiv:2411.16098* (2024).

[8] Seth Bassetti et al. "DiffESM: Conditional emulation of temperature and precipitation in Earth system models with 3D diffusion models". In: *Journal of Advances in Modeling Earth Systems* 16.10 (2024), e2023MS004194.

[9] Ting-Yu Dai and Hayato Ushijima-Mwesigwa. "PrecipDiff: Leveraging image diffusion models to enhance satellite-based precipitation observations". In: *arXiv preprint arXiv:2501.07447* (2025).

[10] Thomas Vandal et al. "Deepsd: Generating high resolution climate change projections through single image super-resolution". In: *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*. 2017, pp. 1663–1672.

[11] Nan Yang, Chong Wang, and Xiaofeng Li. "Evaluation of precipitation forecasting methods and an advanced lightweight model". In: *environmental research letters* 19.9 (2024), p. 094006.

[12] Nan Yang and Xiaofeng Li. "Lightweight AI-powered precipitation nowcasting". In: *The Innovation Geoscience* 2.2 (2024), p. 100066.

[13] Naufal Shidqi et al. "Generating High-Resolution Regional Precipitation Using Conditional Diffusion Model". In: *arXiv preprint arXiv:2312.07112* (2023).

[14] Tero Karras et al. "Elucidating the design space of diffusion-based generative models". In: *Advances in neural information processing systems* 35 (2022), pp. 26565–26577.

[15] Michael Aich et al. "Conditional diffusion models for downscaling & bias correction of Earth system model precipitation". In: *arXiv preprint arXiv:2404.14416* (2024).

[16] Robbie A Watt and Laura A Mansfield. "Generative diffusion-based downscaling for climate". In: *arXiv preprint arXiv:2404.17752* (2024).

[17] Sanghyun Woo et al. "Cbam: Convolutional block attention module". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 3–19.

[18] Yan Wang et al. "PlainUSR: Chasing Faster ConvNet for Efficient Super-Resolution". In: *Proceedings of the Asian Conference on Computer Vision*. 2024, pp. 4262–4279.

[19] Nidhi Nishant et al. "Comparison of a novel machine learning approach with dynamical downscaling for Australian precipitation". In: *Environmental Research Letters* 18.9 (2023), p. 094006.

[20] Chris Huntingford et al. "Machine learning and artificial intelligence to aid climate change research and preparedness". In: *Environmental Research Letters* 14.12 (2019), p. 124007.

[21] Jeff Chun-Fung Lo, Zong-Liang Yang, and Roger A Pielke Sr. "Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) model". In: *Journal of Geophysical Research: Atmospheres* 113.D9 (2008).

[22] Robert L Wilby et al. "Statistical downscaling of general circulation model output: A comparison of methods". In: *Water resources research* 34.11 (1998), pp. 2995–3008.

[23] Robert L Wilby and Christian W Dawson. "The Statistical Downscaling Model: insights from one decade of application." In: *International Journal of Climatology* 33.7 (2013).

[24] Robert L Wilby et al. "The statistical downscaling model-decision centric (SDSM-DC): conceptual basis and applications". In: *Climate Research* 61.3 (2014), pp. 259–276.

[25] Alfonso Hernanz et al. "A critical view on the suitability of machine learning techniques to downscale climate change projections: Illustration for temperature with a toy experiment". In: *Atmospheric Science Letters* 23.6 (2022), e1087.

[26] Chao Dong et al. "Learning a deep convolutional network for image super-resolution". In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*. Springer. 2014, pp. 184–199.

[27] Rilwan A Adewoyin et al. "TRU-NET: a deep learning approach to high resolution prediction of rainfall". In: *Machine Learning* 110 (2021), pp. 2035–2062.

[28] Jussi Leinonen, Daniele Nerini, and Alexis Berne. "Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network". In: *IEEE Transactions on Geoscience and Remote Sensing* 59.9 (2020), pp. 7211–7223.

[29] Karen Stengel et al. "Adversarial super-resolution of climatological wind and solar data". In: *Proceedings of the National Academy of Sciences* 117.29 (2020), pp. 16805–16815.

[30] Luke Metz et al. "Unrolled generative adversarial networks". In: *arXiv preprint arXiv:1611.02163* (2016).

[31] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. "Which training methods for GANs do actually converge?" In: *International conference on machine learning*. PMLR. 2018, pp. 3481–3490.

[32] Ze Liu et al. "Swin transformer: Hierarchical vision transformer using shifted windows". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 10012–10022.

[33] Jonathan Ho, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models". In: *Advances in neural information processing systems* 33 (2020), pp. 6840–6851.

[34] Prafulla Dhariwal and Alexander Nichol. "Diffusion models beat gans on image synthesis". In: *Advances in neural information processing systems* 34 (2021), pp. 8780–8794.

[35] Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[36] Jiali Wang et al. "Fast and accurate learned multiresolution dynamical downscaling for precipitation". In: *Geoscientific Model Development Discussions* 2021 (2021), pp. 1–24.