MACHINE LEARNING
Earth

Part of the
MACHINE LEARNING
Series

PAPER • OPEN ACCESS

How far can we downscale? Resolution limits and physical interpretability of diffusion models for African precipitation

To cite this article: Jangho Lee and Sara Shamekh 2026 *Mach. Learn.: Earth* **2** 015018

IOP Publishing Mach. Learn.: Earth 2 (2026) 015018 https://doi.org/10.1088/3049-4753/ae7114

MACHINE LEARNING
Earth

Check for updates

PAPER

How far can we downscale? Resolution limits and physical interpretability of diffusion models for African precipitation

RECEIVED
11 March 2026

REVISED
29 April 2026

ACCEPTED FOR PUBLICATION
21 May 2026

PUBLISHED
8 June 2026

Jangho Lee* and **Sara Shamekh**

Courant Institute of Mathematics, Computing, and Data Science, New York University, New York, NY 10012, United States of America
* Author to whom any correspondence should be addressed.

E-mail: jangho.lee@nyu.edu

Keywords: diffusion model, precipitation, Africa, downscaling, AI, statistical downscaling

Supplementary material for this article is available online

Abstract

Accurate high-resolution precipitation data is critical for climate risk assessment in Sub-Saharan Africa, yet global climate models operate at coarse resolutions, failing to capture localized extremes. While deterministic deep learning models offer efficient downscaling, our evaluation shows that minimizing pointwise error inherently penalizes high-frequency features, artificially smoothing precipitation fields and suppressing extremes regardless of architectural complexity. Here, we propose a conditional denoising diffusion model for precipitation downscaling, evaluating its structural realism, physical interpretability, and resolution limits against deterministic baselines. The diffusion model successfully recovers sub-grid variability, preserving near-perfect spatial variance and heavy-tailed extreme event distributions where deterministic approaches fail. Through perturbation-based feature importance analysis, we demonstrate the generative framework acts as a context-aware emulator of atmospheric physics. It dynamically shifts predictor reliance, prioritizing low-level moisture convergence in rugged terrain and dry seasons, while favoring mid-tropospheric dynamic forcing during monsoons. Furthermore, by progressively coarsening inputs from 0.25° to 1.0°, we interrogate generative downscaling limits. As spatial constraints weaken at 1.0°, the model generates unconstrained stochastic false alarms, and the resulting ensemble smoothing directly quantifies predictive spatial uncertainty under severe information loss. A moderate 0.5° input emerges as an optimal balance, providing sufficient structural anchors to maintain both deterministic accuracy and extreme event preservation. Ultimately, this work provides a physically interpretable, resolution-aware framework for generative downscaling, offering a robust strategy for translating coarse climate projections into localized risk assessments.

1. Introduction

Accurate representation of precipitation at high spatial resolution is fundamental to climate risk assessment, water resource management, agriculture, and public health, particularly in regions where livelihoods depend heavily on rain-fed systems (Waha *et al* 2013, Grace *et al* 2015, Kisakye and Van der Bruggen 2018, Tefera *et al* 2025). In Sub-Saharan Africa, rainfall variability directly governs food security, hydrological extremes, and the transmission dynamics of vector-borne diseases (Ofori *et al* 2021, Smith *et al* 2024, Chombo *et al* 2025, Ogunbode *et al* 2025). However, reliable high-resolution meteorological observations remain sparse and unevenly distributed across the globe, creating substantial challenges for localized climate analysis and impact forecasting (Bytheway and Kummerow 2013, Wilby and Yu 2013, Lu *et al* 2020).

For future climate projections, this challenge becomes even more pronounced. Global climate models (GCMs), which form the backbone of climate change assessments, typically operate at horizontal resolutions on the order of 1° to 2°. While GCMs are indispensable for simulating large-scale atmospheric circulation and long-term climate response to greenhouse gas forcing, their coarse spatial resolution fundamentally limits their ability to resolve convective-scale precipitation processes, mesoscale organization, and terrain-induced rainfall variability (Rowell 2012, Woldemeskel *et al* 2016, Shen *et al* 2018, Tarek *et al* 2021). As a result, translating GCM outputs into actionable, location-specific climate information requires robust downscaling methodologies.

Downscaling thus plays a central role in bridging the scale gap between global climate simulations and local impact applications (Maraun *et al* 2010, Kofidou *et al* 2023). Currently, there are broadly two streams of downscaling methods: dynamical and statistical downscaling. Dynamical downscaling is done with regional climate models, and it can provide physically consistent fine-scale structure. However, it remains computationally expensive and impractical for large ensembles or multi-scenario projections (Xu *et al* 2019, Adachi and Tomita 2020). Statistical and machine learning–based downscaling approaches offer a computationally efficient alternative by learning mappings between coarse predictors and high-resolution targets (Kofidou *et al* 2023, Sun *et al* 2024). Recent regional studies have successfully applied these techniques to African domains. For instance, recent studies demonstrated the utility of convolutional neural network (CNN)-based or conditional Generative Adversarial Networks (cGANs) architectures for precipitation downscaling across specific African regions (Asfaw and Luo 2024, Antonio *et al* 2025), providing valuable regional benchmarks for model evaluation.

However, most deep learning approaches, especially CNN based models, adopt deterministic regression frameworks, optimizing pixel-wise loss functions such as mean squared error (MSE) (Tu *et al* 2021, Kim *et al* 2022, Yang *et al* 2023). This means that although such models often achieve strong performance in average-based metrics, it is fundamentally difficult to generate accurate precipitation field, which is inherently intermittent, heavy-tailed, and multiscale. Minimizing pointwise error does not necessarily guarantee preservation of spatial texture, spectral energy distribution, or extreme-event frequency. Deterministic models frequently generate smoother fields that suppress high-frequency variability and attenuate localized extremes, potentially distorting physically meaningful precipitation structures.

Recent advances in generative modeling, particularly denoising diffusion probabilistic models (DDPMs), offer a fundamentally different paradigm for high-dimensional field reconstruction (Srivastava *et al* 2024, Aich *et al* 2025, Liu *et al* 2025). Unlike deterministic CNNs that learn a single point estimate, diffusion models approximate the full conditional probability distribution of high-resolution fields given large-scale atmospheric constraints. This probabilistic formulation enables the generation of multiple physically plausible realizations, rather than a single smoothed reconstruction (Nai *et al* 2024, Liu *et al* 2025).

Diffusion models have demonstrated remarkable stability and fidelity compared to earlier generative frameworks such as GANs and Variational autoencoders (Ho *et al* 2020, Dhariwal and Nichol 2021). In particular, DDPMs avoid mode collapse and training instability commonly observed in adversarial models, while providing consistent likelihood-based optimization. Their iterative denoising formulation naturally supports multiscale structure reconstruction, progressively refining spatial detail from coarse to fine scales. This property is especially advantageous for geophysical applications, where precipitation fields exhibit hierarchical spatial organization and scale-dependent variability.

Recent applications in environmental and Earth system modeling have begun to explore diffusion-based generation of atmospheric fields, climate variables, and spatial environmental patterns (Bassetti *et al* 2024, Li *et al* 2024, Ling *et al* 2024, Mardani *et al* 2025, Tu *et al* 2025). Specifically, precipitation downscaling studies using diffusion-based models suggest that diffusion frameworks are capable of preserving fine-scale variability, spectral consistency, and heavy-tailed distributions more effectively than deterministic or adversarial approaches (Lyu *et al* 2024, Srivastava *et al* 2024, Liu *et al* 2025). Because extreme precipitation events are stochastic and exhibit non-Gaussian, heavy-tailed behavior, modeling the full conditional distribution is particularly important for risk assessment under future climate scenarios. Moreover, the stochastic sampling mechanism of diffusion models aligns naturally with ensemble-based climate projection frameworks. Rather than producing a single deterministic downscaled field, diffusion models can generate multiple conditional realizations, allowing uncertainty quantification at high resolution. This probabilistic capability is critical when downscaling GCM outputs, where structural uncertainty and internal variability must be propagated across scales.

Building on recent advancements and knowledge gaps, three fundamental questions arise when considering diffusion-based generative frameworks for climate downscaling. First, does a conditional diffusion model provide measurable advantages over deterministic convolutional architectures when reconstructing high-resolution precipitation fields? While regression-based CNNs are optimized to minimize pointwise loss, generative diffusion models are trained to approximate the conditional data distribution.

This distinction raises the possibility that diffusion models may better preserve multiscale spatial variability, intermittency, and heavy-tail behavior. A systematic comparison under identical predictor constraints is therefore necessary to assess whether diffusion-based reconstruction offers structural benefits beyond conventional regression approaches.

Second, beyond predictive performance, the scientific credibility of generative models depends on their physical interpretability. In climate applications, model acceptance requires more than visual realism; it requires consistency with physically meaningful atmospheric drivers. Diffusion models operate through iterative stochastic denoising conditioned on large-scale atmospheric and topographic variables. However, it remains unclear how such models internally prioritize these conditioning signals and whether their generative process reflects physically plausible dependencies. Evaluating feature-level sensitivity and environmental scenario–dependent behavior is therefore essential to determine whether diffusion-based downscaling preserves the governing physical mechanisms.

Third, the effectiveness of generative downscaling may depend critically on the magnitude of the resolution gap between input and target fields. In practical climate projection contexts, however, input data may originate from 1° or 2° GCM simulations. As the input resolution becomes increasingly degraded, the amount of spatial information directly available to the model decreases, and reconstruction necessarily relies more heavily on learned statistical structure. It remains an open question how diffusion-based generative reconstruction behaves under progressively coarsened inputs: whether its probabilistic framework can compensate for severe information loss, how structural realism degrades as resolution decreases, and what this implies for the practical applicability of diffusion models to future climate scenarios.

In this study, we investigate these three questions using daily precipitation over Sub-Saharan Africa as a testbed. We compare a conditional diffusion model against deterministic convolutional baselines under identical experimental settings, analyze feature importance to assess physical interpretability, and systematically examine input resolution dependency by progressively degrading coarse precipitation inputs. By doing so, we aim to clarify both the methodological advantages and the practical limitations of diffusion-based generative downscaling in the context of climate applications.

2. Data

2.1. Study domain

The study domain encompasses a vast and climatically diverse region of Sub-Saharan Africa, extending longitudinally from 18°W to 52°E and latitudinally from 7°S to 19°N (figure 1). This domain bridges the Atlantic and Indian Oceans, covering a variety of ecological zones including the semi-arid Sahelian belt, Savannah grasslands, and the perennially humid tropical rainforests. Characterized by complex topographical features, the region experiences intricate precipitation dynamics driven by the seasonal migration of the intertropical convergence zone (ITCZ) and the West African Monsoon (WAM) system. These multi-scale atmospheric processes, coupled with diverse orographic influences, result in high spatiotemporal variability in rainfall, making the region an ideal testbed for high-resolution downscaling frameworks.

Notably, this domain spans a multitude of nations across West, Central, and East Africa, including major countries such as Nigeria, Ethiopia, Sudan, Kenya, and the Democratic Republic of Congo, among dozens of others. The inclusion of many diverse countries introduces significant challenges in terms of data consistency and infrastructure, as the density and quality of ground-based meteorological networks vary across national borders. This fragmentation creates a highly heterogeneous data landscape where many nations suffer from chronic observation gaps.

Beyond these technical hurdles, the region holds immense socio-economic and public health stakes. A large majority of the population relies on rain-fed agriculture, which is increasingly vulnerable to climate-induced shifts in precipitation. Furthermore, the high variability of rainfall directly influences critical health issues, particularly the transmission dynamics of vector-borne diseases. Inconsistent or extreme rainfall patterns create stagnant water breeding sites for Anopheles mosquitoes, leading to seasonal outbreaks of malaria that pose a severe burden on the region's healthcare systems.

Consequently, developing a robust generative downscaling model for this multi-national domain is not only a technical necessity for bridging the resolution gap but also a critical step toward enhancing local climate resilience, agricultural planning, and infectious disease forecasting.

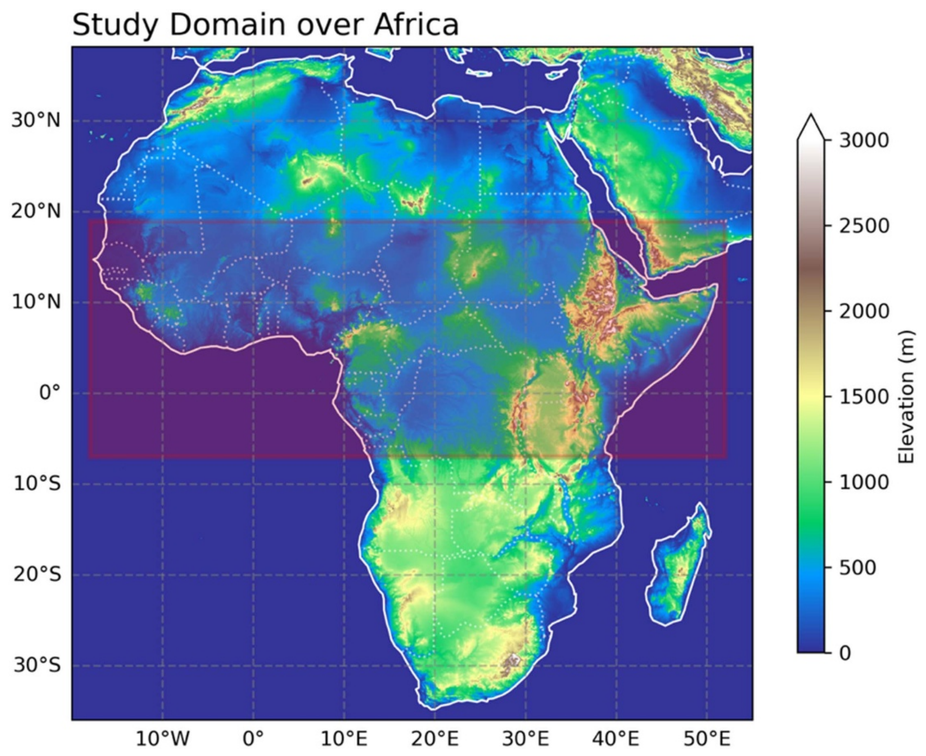


Figure 1. Study domain (red box) over the African region, with elevation overlaid as a color-shaded background.

2.2. Climate Hazards InfraRed Precipitation with Station data (CHIRPS)precipitation data

The CHIRPS is a quasi-global, high-resolution precipitation product. CHIRPS provides a unique combination of continuous satellite-based infrared estimates and rigorous *in-situ* station data, offering a 0.05° (∼5.3 km) spatial resolution across a temporal span from 1981 to the present with daily interval (Funk *et al* 2014, Dinku *et al* 2018, Du *et al* 2024), and has been evaluated in multiple regions globally (Katsanos *et al* 2016, Paredes-Trejo *et al* 2017, Dinku *et al* 2018, Rivera *et al* 2018, Du *et al* 2024).

The dataset is generated in a three-stage architectural framework, beginning with the Climate Hazards Group InfraRed Precipitation (CHIRP) foundation, which utilizes cold cloud duration data derived from Thermal Infrared observations. By calculating the duration that cloud-top temperatures remain below a convective threshold, CHIRP estimates rainfall intensity with high temporal fidelity. To ensure spatial representativeness across diverse landscapes, these infrared-derived estimates are then calibrated against a high-resolution long-term monthly climatology that incorporates satellite average fields, elevation data, and station-based averages. The final stage involves the integration of available rain-gauge data through an inverse distance weighting algorithm, blending the collection of global station records with the satellite fields to correct systematic biases (Funk *et al* 2014, 2015).

2.3. Digital elevation model

To incorporate the critical influence of complex terrain on localized precipitation patterns, this study utilizes the Copernicus Global Digital Elevation Model (GLO-30) (European Space Agency 2024, Simard *et al* 2024). GLO-30 provides a high-quality, global representation of the Earth's surface with a spatial resolution of 30 meters (1 arc-second). The integration of GLO-30 is fundamental to the downscaling framework as it provides a static yet highly detailed spatial prior. In the tropical regions of Sub-Saharan Africa, precipitation is strongly modulated by topographic forcing, where mountain ranges induce orographic lifting, leading to enhanced rainfall on windward slopes and rain-shadow effects on leeward sides. By including elevation as a primary auxiliary predictor, the Diffusion model can learn to translate coarse-scale atmospheric dynamics into physically plausible, high-resolution precipitation fields that respect the underlying terrain.

2.4. ERA5 reanalysis

Atmospheric state variables used as predictors for the downscaling model are sourced from the ERA5 reanalysis, (Hersbach *et al* 2020) the fifth-generation global atmospheric climate reanalysis produced by the European Centre for Medium-Range Weather Forecasts.

Table 1. Summary of variables used in this study.

Type	Source	Variable	Level	Unit	Resolution	Interval
Predictor	CHIRPS	Coarse-grained precipitation	Surface	mm d^{-1}	0.25°	Daily
	ERA5	Convective available potential energy	Surface	J kg^{-1}	0.25°	Daily
		2 m Temperature	Surface	K	0.25°	Daily
		2 m Dewpoint Temperature	Surface	K	0.25°	Daily
		Mean sea level pressure	Surface	Pa	0.25°	Daily
		Total column water vapor	Surface	kg m^{-2}	0.25°	Daily
		10 m U-component of the Wind	Surface	m s^{-1}	0.25°	Daily
		10 m V-component of the Wind	Surface	m s^{-1}	0.25°	Daily
		Specific humidity	850 hpa	kg kg^{-1}	0.25°	Daily
		U-component of the wind	850 hpa	m s^{-1}	0.25°	Daily
		V-component of the wind	850 hpa	m s^{-1}	0.25°	Daily
		Vertical velocity	850 hpa	Pa s^{-1}	0.25°	Daily
		Geopotential	850 hpa	m^2 s^{-2}	0.25°	Daily
		Specific humidity	500 hpa	kg kg^{-1}	0.25°	Daily
		U-component of the wind	500 hpa	m s^{-1}	0.25°	Daily
		V-component of the wind	500 hpa	m s^{-1}	0.25°	Daily
		Vertical velocity	500 hpa	Pa s^{-1}	0.25°	Daily
		Geopotential	500 hpa	m^2 s^{-2}	0.25°	Daily
	DEM	Mean elevation	Surface	m	0.05°	Static
		Max elevation	Surface	m	0.05°	Static
		Min elevation	Surface	m	0.05°	Static
		Terrain roughness (std of elevation)	Surface	m	0.05°	Static
	Time	Sine component of day of year	N/A	N/A	N/A	Daily
		Cosine component of day of year	N/A	N/A	N/A	Daily
Target	CHIRPS	Precipitation	Surface	mm d^{-1}	0.05°	Daily

The ERA5 dataset features a horizontal resolution of approximately 31 km (0.25° × 0.25° grid). For the vertical dimension, the original model operates on 137 hybrid sigma-pressure levels extending from the surface up to 0.01 hPa. In this study, we utilize data interpolated onto 37 isobaric pressure levels to ensure consistency across the surface, lower (850 hPa) and mid-troposphere (500 hPa). The temporal resolution is hourly; however, we average it to daily data to match the temporal resolution of the CHIRPS dataset.

3. Method of analysis

3.1. Experimental setup

The objective of this experiment is to downscale coarse-grained CHIRPS daily precipitation fields back to their original high-resolution structure using large-scale atmospheric predictors from ERA5 and topographic information derived from DEM. As an initial step, the original CHIRPS dataset at 0.05° spatial resolution is spatially aggregated to a 0.25° latitude–longitude grid, which matches the resolution of ERA5. The coarse graining is performed by averaging all 0.05° CHIRPS grid cells contained within each 0.25° ERA5 grid cell. In this framework, the aggregated 0.25° CHIRPS field serves as the low-resolution precipitation predictor, while the original 0.05° CHIRPS dataset is retained as the high-resolution target (ground truth).

In addition to precipitation, auxiliary predictors are derived from ERA5 and DEM. DEM, which has a resolution finer than 0.05°, is aggregated to the 0.05° grid, and four statistical descriptors are computed for each cell: mean, maximum, minimum, and standard deviation of elevation. These variables capture sub-grid orographic variability that influences convective processes. Furthermore, seasonality is represented using a cyclic encoding of the day-of-year (DOY) variable to preserve temporal continuity. Specifically, these components are calculated as $\sin\left(2\pi \times \frac{\mathrm{DOY}}{D_{\mathrm{year}}}\right)$ and $\cos\left(2\pi \times \frac{\mathrm{DOY}}{D_{\mathrm{year}}}\right)$, where D_{year} represents the total number of days in the respective year (365 or 366). The complete list of input variables is provided in table 1.

This study utilizes 25 years of daily data spanning from 2000 to 2024. The dataset is partitioned temporally to ensure robust out-of-sample evaluation. Specifically, the years 2001, 2006, 2011, 2016, and 2021 are used as validation years, while the most recent two years, 2023 and 2024, are reserved as the independent test set. This sequential chronological split (training on historical data up to 2022 and testing on the subsequent period) is specifically designed to mimic an operational climate projection scenario. This approach ensures the model's performance is evaluated on its operational readiness to generalize to unseen future climate states, naturally encompassing complex interannual variabilities without artificial temporal sampling.

The study domain covers approximately 22 million km^2, making direct full-domain training computationally prohibitive. Moreover, training a model on a single contiguous region may reduce its generalizability to other geographic domains. To address these challenges, we implement a patch-based training strategy. For each training day, the domain is subdivided into smaller spatial patches of 128 $\times$ 128 pixels, corresponding to approximately 6.4°$\times$ 6.4° at 0.05° resolution. Patches are extracted using a sliding-window approach with a stride of 100 pixels. To enhance spatial diversity and prevent overfitting to fixed grid boundaries, each patch location is randomly jittered by up to $\pm$32 pixels. This patching strategy substantially reduces computational cost while enabling the model to learn location-agnostic fine-scale relationships. As a result, the trained model is not constrained to a single geographic region and can, in principle, be extended to other parts of the world. The concept of patching and example of predictor and target variables are shown in figure 2(a). After patching, the total number of patches is approximately 160 k, where about 110 k of them are training dataset. Example of study region patch and corresponding variables are also represented in figure 2.

All variables are normalized using Z-score standardization. Specifically, each variable is centered by subtracting its mean and scaled by its standard deviation computed from the training set. For precipitation variables, including both the coarse-grained and original CHIRPS fields, a logarithmic transformation is first applied using $\log(1+x)$ to reduce skewness and stabilize variance prior to standardization.

3.2. Diffusion model

We adopt a conditional denoising diffusion model to generate high-resolution daily precipitation fields from predictor inputs. The model is designed to learn a conditional distribution over plausible fine-scale precipitation realizations, and its architecture is optimized to reconstruct spatial structure progressively from noise while remaining constrained by the conditioning information. Below, we describe the formulation, variables, network components, and why each component is necessary for the objective.

3.2.1. Conditional diffusion formulation

The model is designed to approximate the conditional probability distribution, $p_\theta(x_0|c)$, where $x_0 \in \mathbb{R}^{C_x\times H\times W}$ denotes a high-resolution precipitation patch (original CHIRPS) and $c \in \mathbb{R}^{C_c\times H\times W}$ represents the conditioning tensor constructed from the predictor variables. Here, H and W denote the spatial dimensions of each training patch (128 pixels) and C_x and C_c denote the number of channels in the target (1 channel) and conditioning tensors (24 channel). The parameter θ includes all trainable weights of the denoising neural network. Rather than directly mapping c to a deterministic estimate of x_0, the diffusion framework models the full conditional distribution, enabling stochastic generation of spatially realistic fields.

3.2.2. Forward diffusion process

The forward diffusion process defines a fixed stochastic transformation that progressively corrupts the clean target x_0 into noisy latent variables x_t, where $t \in \{1, \ldots, T\}$ indexes diffusion timesteps and T denotes the total number of diffusion steps (set to be 1000 in this study). For any timestep t, the noisy field x_t can be expressed mathematically as:

$$x_t = \sqrt{\overline{\alpha_t}}x_0 + \sqrt{1-\overline{\alpha_t}}\epsilon, \qquad \epsilon \sim N(0, I).$$

In this expression, $\epsilon \in \mathbb{R}^{C_x\times H\times W}$ is a Gaussian noise tensor with independent standard normal entries, $\mathcal{N}$ represents the Gaussian normal distribution, and I denote the identity covariance matrix. The scalar $\overline{\alpha_t} \in (0, 1]$ controls the signal preservation at timestep t and determines the relative contribution of the clean signal and injected noise. It is derived from a predefined variance schedule $\{\beta_t\}_{t=1}^{T}$, where $\beta_t \in (0, 1)$ specifies the variance increment at step t. Defining $\alpha_t = 1-\beta_t$, the cumulative coefficient is given by:

$$\overline{\alpha_t} = \prod_{s=1}^{t} \alpha_s \quad .$$

Thus, as t increases, $\overline{\alpha_t}$ will decrease, and x_t transitions smoothly from a clean field to almost purely noise field. A cosine-based variance schedule is used to ensure the gradual degradation of structure, particularly preserving resolution in low-noise regimes where fine-scale detail must be accurately learned.

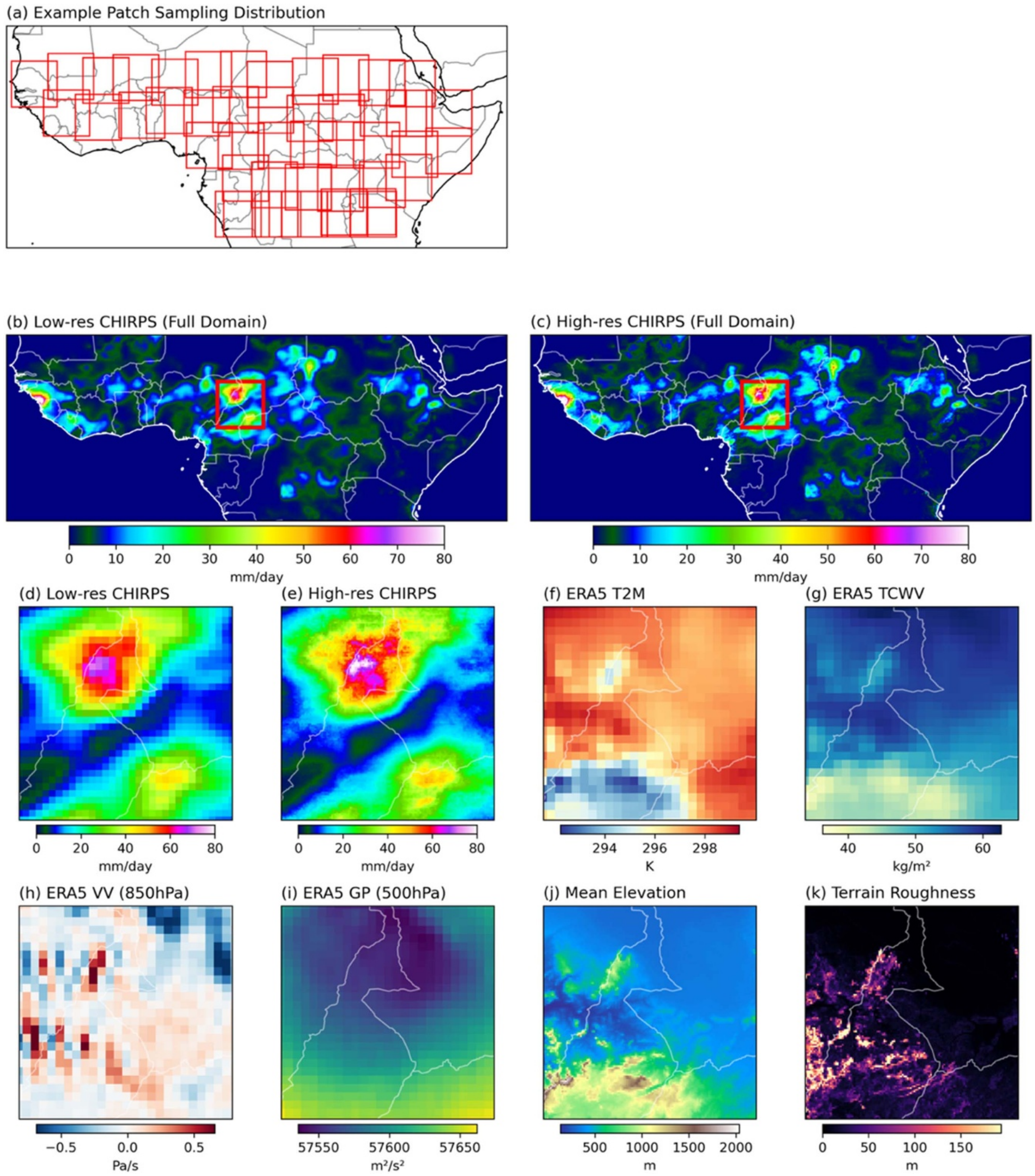


Figure 2. Overview of the patch-based training framework and input variables used for precipitation downscaling. (a) Example spatial distribution of randomly sampled 128 $\times$ 128 patches (red boxes) across the study domain. (b)–(c) Full-domain comparison between coarse-grained CHIRPS (0.25°, remapped to 0.05° grid) and original high-resolution CHIRPS (0.05°), with the highlighted red box indicating the example patch region shown below. (d)–(e) Zoomed-in example patch of low-resolution and high-resolution CHIRPS precipitation (mm d^{-1}), illustrating the recovery of fine-scale convective structure. (f)–(i) Corresponding ERA5 predictor variables within the same patch: 2 m air temperature (T2M; K), total column water vapor (TCWV; kg m^{-2}), vertical velocity at 850 hPa (VV; Pa s^{-1}) and geopotential at 500 hPa (GP; m^2 s^{-2}). (j)–(k) Static topographic predictors aggregated to 0.05° resolution: mean elevation (m) and terrain roughness (standard deviation of elevation within each grid cell; m).

This forward process contains no learnable parameters. Its purpose is to define a structured family of denoising tasks across noise levels, allowing the model to learn spatial reconstruction at multiple effective resolutions.

3.2.3. Reverse diffusion model and training objective

The reverse process seeks to recover x_0 from noisy samples x_t by iteratively removing noise. Because the exact reverse transition is intractable, it is approximated by a neural network $\in_\theta$, which predicts the Gaussian noise component present in x_t. Specifically, the network takes as input the noisy field x_t, the timestep index t, and the conditioning tensor c, and outputs a noise estimate:

$$\hat{\epsilon} = \epsilon_\theta\left(x_t, t, c\right)$$

where $\hat{\epsilon} \in \mathbb{R}^{C_x \times H \times W}$ has the same dimension as the true noise $\in$. The timestep t is provided to the network via an embedding mechanism so that the model can adapt its denoising behavior according to the noise level.

The training objective minimizes the MSE between the sampled noise $\in$ used in the forward process and the predicted noise $\hat{\in}$, which can be mathematically formulated as sum of square of L2 norm:

$$L(\theta) = E_{x_0,c,t,\epsilon}\left[\| \in - \in_\theta(x_t, t, c)\|_2^2\right].$$

The expectation is taken over the empirical data distribution of (x_0, c), random timesteps t sampled uniformly from $\{1, \ldots, T\}$, and independent Gaussian noise samples $\in$. This formulation ensures that the model learns to denoise effectively across the entire diffusion trajectory. Predicting the noise rather than directly estimating x_0 stabilizes optimization because the target distribution remains Gaussian at all timesteps, providing a consistent learning signal.

3.2.4. Conditioning mechanism

Conditioning is incorporated by concatenating the noisy input x_t and the predictor tensor c along the channel dimension prior to entering the denoising network. This produces a combined tensor in $\mathbb{R}^{(C_x+C_c)\times H \times W}$, allowing the model to process both latent precipitation information and conditioning variables jointly at every diffusion step. Because denoising is iterative, providing conditioning information at each step is essential to ensure that the reconstructed field remains consistent with predictor constraints throughout the reverse trajectory.

3.2.5. Denoising network architecture

The denoising function $\in_\theta$ is implemented using a U-Net architecture composed of an encoder–decoder structure with residual blocks and skip connections. The encoder progressively reduces spatial resolution through downsampling operations, enabling the network to capture long-range spatial dependencies and global structure. The decoder restores spatial resolution using upsampling layers, while skip connections transfer high-resolution feature information from corresponding encoder stages to prevent loss of localized detail.

Within each resolution stage, residual convolutional blocks are augmented with Squeeze-and-Excitation (SE) modules to perform adaptive channel-wise recalibration. The SE mechanism first aggregates spatial information through global pooling to obtain a compact channel descriptor and subsequently applies learned gating to reweight feature channels. This channel-wise attention mechanism allows the network to dynamically emphasize physically informative representations while suppressing less relevant activations.

The timestep index t is embedded using a sinusoidal positional encoding followed by a learned projection, and this embedding is injected into intermediate residual blocks through feature-wise affine transformations. This mechanism allows the same network parameters to represent distinct denoising operators depending on the noise level. Without explicit timestep conditioning, the model would be unable to adapt its reconstruction behavior across varying corruption regimes.

Residual connections within blocks facilitate gradient flow and stabilize deep network training, while channel-wise modulation mechanisms enhance representational flexibility by allowing dynamic feature reweighting. The multiscale nature of the U-Net backbone is critical for the diffusion objective, as it enables simultaneous modeling of large-scale spatial coherence and fine-scale texture during iterative denoising.

The U-Net in this study uses a base channel width of 48. Channel dimensionality is scaled across resolution levels using multipliers of (1, 2, 4, 4), resulting in feature widths of 48, 96, 192, and 192 channels from shallow to deep layers. This progressive channel expansion increases representational capacity at lower spatial resolutions, where global context is aggregated. Dropout with a rate of 0.1 is applied within residual blocks, and Group Normalization with eight groups is used for stable training across batch sizes.

3.2.6. Sampling procedure

At inference time, generation begins with a random Gaussian field $x_T \sim \mathcal{N}(0, I)$ of the same dimensionality as the target field. The model then iteratively applies the learned reverse transition from timestep T down to 1, each time using the predicted noise $\in_\theta(x_t, t, c)$ to compute x_{t-1}. The reverse update at each step is given by:

$$x_{t-1} = \tfrac{1}{\sqrt{\alpha_t}}\left(x_t - \tfrac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \in_\theta(x_t, t, c)\right) + \sigma_t \omega, \quad \omega \sim N(0, I)\ .$$

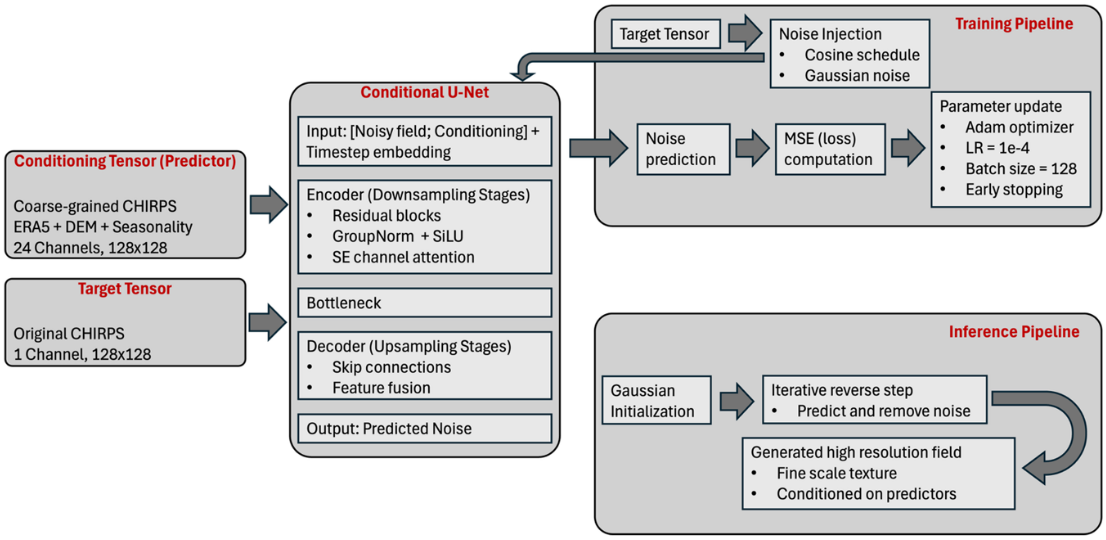


Figure 3. Conceptual schematic of diffusion model used in this study.

Here, ω is an independent standard Gaussian noise tensor injected at each reverse sampling step. Under the standard Denoising Diffusion Probabilistic Model (DDPM) formulation, the reverse transition variance is set to $\sigma_t^2 = \beta_t$, which corresponds to the original variance-preserving parameterization of Ho *et al* (2020).

The final output x_0 represents a generated high-resolution precipitation field conditioned on c. Because sampling is stochastic, multiple realizations can be generated from the same conditioning input, allowing the model to represent variability in fine-scale spatial structure while preserving consistency with predictor information.

3.2.7. Model summary

The proposed diffusion architecture integrates three key components: a cosine-scheduled forward corruption process with $T = 1000$ steps, a conditional U-Net denoiser with timestep-aware modulation, and iterative reverse sampling constrained by predictor inputs. The forward–reverse formulation enables progressive reconstruction of spatial structure across noise scales, while multiscale skip connections and channel-wise modulation ensure stable representation of both large-scale coherence and fine-scale texture. Together, these elements provide a structured probabilistic framework for conditional high-resolution field generation.

The model is trained with a batch size of 256 using the Adam optimizer with an initial learning rate of 0.0001. Training is conducted for up to 100 epochs, with early stopping applied if the validation loss does not improve for 10 consecutive epochs. The schematic of the conditional diffusion model used in this study is depicted in figure 3.

3.3. Simple fully convolutional network (Simple FCN)

As a deterministic deep learning benchmark, we implement a Simple FCN to provide a direct regression benchmark against the proposed diffusion framework. This Simple FCN baseline represents a lightweight deterministic convolutional model that performs direct pixel-wise regression without multiscale feature aggregation or probabilistic sampling, thereby providing a controlled benchmark for evaluating the benefits of the proposed diffusion architecture.

The architecture consists of a straightforward stack of five convolutional layers without encoder–decoder structure, skip connections, or attention mechanisms. The network begins with three consecutive convolutional layers with 64 feature maps and kernel size of 3 × 3, followed by a fourth layer reducing the channel dimension to 32, and a final 3 × 3 convolution producing the single-channel output field. ReLU activation functions are applied after each intermediate convolution, and zero-padding is used to preserve spatial resolution throughout the network.

The model is trained to minimize MSE between predictions and ground-truth precipitation. Optimization is performed using the Adam optimizer with a learning rate of 0.0001 and a batch size of 256. Training proceeds for up to 100 epochs, with early stopping applied if validation loss does not improve for 10 consecutive epochs. Gradient norm clipping with a threshold of 1.0 is used to stabilize training. Furthermore, to guarantee a strictly fair architectural comparison across all experiments, the SE modules utilized in the U-Net were identically integrated into the convolutional layers.

3.4. Super resolution CNN

To provide additional benchmark for performance comparison, we implement a SRCNN as second baseline model. SRCNN is one of the earliest and most widely used convolutional architectures for single-image super-resolution and serves as a lightweight yet effective reference model.

The baseline model follows the three-layer SRCNN architecture, consisting of convolutional layers with nonlinear activation. The network takes same input as the diffusion model and outputs a single-channel high-resolution precipitation estimate. The architecture consists of A first convolutional layer with 64 filters of size 9×9, a second convolutional layer with 32 filters of size 5×5, A final reconstruction layer with a single 5×5 convolution producing the output field. Padding is applied so that spatial dimensions are preserved across layers.

The SRCNN model is trained to minimize the MSE between the predicted high-resolution precipitation field and the ground-truth field. Optimization is performed using the Adam optimizer with a learning rate of 0.0001. The model is trained with a batch size of 256 for up to 100 epochs. Early stopping is applied based on validation loss, with training terminated if no improvement is observed for 10 consecutive epochs. The best-performing model (based on validation MSE) is retained for evaluation. Similar to Simple FCN, SE modules utilized in the U-Net were identically integrated into the convolutional layers.

3.5. Deterministic U-Net

To rigorously isolate the performance gains attributable to the probabilistic diffusion formulation from the structural benefits of the U-Net architecture, we implemented a deterministic U-Net as a direct baseline (Sha *et al* 2020, Mishra Sharma and Mitra 2024, Watt and Mansfield 2024). This model shares an identical backbone architecture with the denoising network of our diffusion model, including the multiscale encoder-decoder structure, skip connections, and SE modules for adaptive channel-wise recalibration.

However, instead of learning to reverse a noise process, the Deterministic U-Net, denoted as f_ϕ, learns a direct mapping from the conditioning predictor tensor c to the high-resolution precipitation field x_0. The network parameters ϕ are optimized by directly minimizing the pointwise MSE objective:

$$\mathcal{L}_{\mathrm{MSE}}\left(\phi\right) = \mathrm{E}_{x_0,c}[||x_0 - f_\phi\left(c\right)||_2^2].$$

Here, $f_\phi\left(c\right)$ represents the networks deterministically predicted high-resolution precipitation field, and $||x_0 - f_\phi\left(c\right)||_2^2$ denotes the squared L2 norm, which computes the pixel-wise intensity differences between the prediction and the ground truth x_0. The expectation E is taken over all pairs of predictor and target patches in the training dataset. Because this deterministic objective strictly penalizes any pixel-wise deviation from the target, it inherently forces the model to predict the conditional mean of the possible precipitation distribution. This mathematical behavior naturally results in the spatial smoothing of high-frequency textures and the underestimation of localized extreme events. By ensuring that this baseline utilizes the exact same architectural refinements and identical hyperparameter settings as the diffusion model, any disparity in the preservation of spatial variance and extreme values can be confidently attributed to the generative paradigm itself, rather than structural discrepancies.

The SE modules were identically integrated, and deterministic U-Net model was trained using a unified hyperparameter configuration: the Adam optimizer with a learning rate of 0.0001 and a batch size of 256.

3.6. Statistical metrics

To evaluate the downscaling performance, we employ a comprehensive set of statistical metrics that quantify magnitude accuracy, spatial structure, spectral fidelity, distributional consistency, and extreme-event representation. For the magnitude accuracy, we utilize root MSE (RMSE) and mean absolute error (MAE). For the spatial alignment, we use spatial correlation (Corr). Below, we explain some of more advanced statistics that are used to evaluate the model's performance.

3.6.1. Structural similarity index (SSIM)

For spatial structure, SSIM is utilized. For two local patches x and $\hat{x}$, SSIM is defined as equation below:

$$\mathrm{SSIM}\left(x,\hat{x}\right) = \frac{\left(2\mu_x\mu_{\hat{x}} + C_1\right)\left(2\sigma_{x\hat{x}} + C_2\right)}{\left(\mu_x^2 + \mu_{\hat{x}}^2 + C_1\right)\left(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2\right)}$$

where μ denotes mean intensity, σ denotes standard deviation, and constants C_1 and C_2 represents stabilize division. $\sigma_{x\hat{x}}$ denotes the cross-covariance between x and $\hat{x}$. SSIM simultaneously measures the structural agreement, making it particularly suitable for evaluating precipitation texture preservation. The SSIM value ranges from -1 to 1, where an ideal value of 1 indicates perfect structural similarity.

3.6.2. Fractional standard deviation (FSD)

To assess variability reproduction, we compute the FSD:

$$\mathrm{FSD} = \frac{\sigma_{\hat{x}}}{\sigma_x}.$$

where σ denotes spatial standard deviation. Values close to 1 indicate accurate representation of spatial variability, whereas values below 1 suggest excessive smoothing.

3.6.3. Power spectral density (PSD)

Spectral fidelity is evaluated using the PSD. Let $\mathcal{F}(x)$ denote the two-dimensional (2D) Fourier transform of the field. The power spectrum is computed as:

$$P(k) = |\mathcal{F}(x)|^2$$

and averaged radially over spatial frequencies k. Spectral agreement is quantified by computing the Pearson correlation between the logarithmic spectra of predicted and reference fields. This metric evaluates the model's ability to reproduce multiscale spatial energy distribution and fine-scale texture. The Pearson correlation for the PSD ranges from -1 to 1, with an ideal value of 1 representing perfect spectral agreement.

3.6.4. Kullback–Leibler divergence (KLD)

Distributional similarity is assessed using the KLD between empirical precipitation intensity distributions:

$$\mathrm{KLD}(p||q) = \sum_b p(b) \log \frac{p(b)}{q(b)}$$

where $p(b)$ and $q(b)$ denote histogram-based probability densities of reference and predicted intensities across bins (b). KLD ranges between 0 to infinity, while lower values indicate better distributional agreement.

3.6.5. WaveSim

To quantitatively evaluate the structural realism and multi-scale physical consistency of the generated precipitation fields, we utilize the WaveSim metric proposed by Accarino *et al* (2025). WaveSim utilizes a 2D discrete wavelet transform to decompose spatial fields into scale-specific detail coefficients. For each decomposition level s, the metric computes the three orthogonal components based on the energy (E) of the detail coefficients: magnitude (M_s), Displacement (D_s), and structure (S_s).

The Magnitude component assesses the overall intensity similarity:

$$M_s = 1 - \frac{\left|\mu\left(E_{\mathrm{gt}}\right) - \mu\left(E_{\mathrm{pred}}\right)\right|}{\mu\left(E_{\mathrm{gt}}\right) + \mu\left(E_{\mathrm{pred}}\right)}$$

where $\mu(E)$ denote the mean energy of the coefficients, and the subscripts gt and pred represent the ground truth (high-resolution reference) and the predicted precipitation fields, respectively.

The Displacement component measures spatial shifts using the Jensen-Shannon divergence (JSD) between the marginal energy distributions (ρ) along the latitudinal and longitudinal axes:

$$D_s = \max\left(0, 1 - \mathrm{JSD}\left(\rho_{\mathrm{gt}}^{\mathrm{lat}}, \rho_{\mathrm{pred}}^{\mathrm{lat}}\right)\right) \times \max\left(0, 1 - \mathrm{JSD}\left(\rho_{\mathrm{gt}}^{\mathrm{lon}}, \rho_{\mathrm{pred}}^{\mathrm{lon}}\right)\right).$$

Here, the superscripts lat and lon specifically denote the marginal energy distributions computed along the latitudinal and longitudinal axes, respectively.

The Structural component evaluates morphological consistency using the cosine similarity $(\cos(\theta))$ of the flattened and sorted detail coefficient x, penalized by average magnitude variations:

$$S_s = \frac{1 + \cos(\theta)}{2} \times \max\left(0, 1 - \frac{\frac{||x_{\mathrm{gt}} - x_{\mathrm{pred}}||}{N}}{\mu\left(|x_{\mathrm{gt}}|\right) + \mu\left(|x_{\mathrm{pred}}|\right)}\right).$$

The final WaveSim score is formulated as the weighted sum of these components across all L decomposition levels:

$$\text{WaveSim} = \sum_{s=1}^{L} w_s M_s D_s S_s$$

where w_s represents the weight at scale s. The composite WaveSim score, as well as its individual components (M_s, D_s, and S_s), ranges from 0 to 1, where an ideal value of 1 indicates perfect structural and physical multi-scale consistency.

3.6.6. Extreme bias

Lastly, the representation of extreme precipitation dynamics is assessed using percentile-based exceedance ratios, specifically denoted as Extreme Bias (EB95 and EB99). By calculating the ratio of pixels exceeding the 95th and 99th percentile thresholds in the generated outputs compared to the target data, these metrics systematically evaluate the structural realism of the models. An optimal ratio of 1.0 indicates a perfect preservation of extreme event frequencies, ensuring that the downscaled fields do not systematically underestimate heavy precipitation.

4. Results and discussions

4.1. Model performance

We first evaluate the performance of the diffusion model for precipitation downscaling. Rather than assessing the entire test dataset in aggregate, we focus on representative rainfall events for two critical reasons. First, generating a multi-member ensemble for every daily timestep across a multi-year test dataset is computationally prohibitive, given that each diffusion realization requires 1000 iterative reverse sampling steps over an expansive spatial domain. Second, long-term aggregate evaluations tend to smooth out localized performance nuances; focusing on specific events allows for a more rigorous examination of the model's structural realism and physical consistency under distinct hydrometeorological and topographic conditions. Specifically, seven case studies are selected. The first four cases correspond to high-intensity precipitation events, defined by the 99th, 95th, 90th, and 75th percentiles of daily rainfall within the test dataset. These cases allow us to assess the model's ability to reconstruct extreme and near-extreme precipitation structures, which are typically characterized by strong spatial intermittency and sharp gradients.

In addition to intensity-based selection, we further examine three terrain-conditioned cases. These include: (i) high-precipitation events over high-elevation regions, defined by mean DEM values above the 90th percentile; (ii) high-precipitation events exceeding the 90th percentile occurring over rough terrain, defined as regions where the standard deviation of elevation (DEM std) exceeds the 90th percentile; and (iii) high-precipitation events over relatively flat terrain, defined by DEM std below the 10th percentile. This terrain-stratified evaluation assesses whether the model can consistently capture fine-scale spatial structures under diverse physical and topographic constraints.

Unlike deterministic models (e.g. SimpleFCN, SRCNN, and Deterministic U-Net) that map a specific input to a single fixed output, the diffusion model is inherently stochastic. It learns the conditional probability distribution of the high-resolution precipitation field and generates samples by iteratively removing noise. This stochasticity allows us to generate an ensemble of multiple distinct plausible realizations for the same input. To comprehensively evaluate the diffusion framework, we analyze three variants: Diffusion Single (the individual realization), Diffusion Ens (5) (the pixel-wise mean of 5 realizations), and Diffusion Ens (25) (the mean of 25 realizations).

Figure 4 illustrates spatial comparisons between the low-resolution input, high-resolution reference, and the downscaling models for a representative 95th-percentile extreme precipitation event. Visualizations for the remaining case studies are provided in the supplementary material. The deterministic CNN baselines (SimpleFCN, SRCNN, and Deterministic U-Net) invariably generate excessively smoothed precipitation fields, suppressing localized extreme rainfall cores. In sharp contrast, the Diffusion Single model successfully reconstructs the sharpest spatial gradients and fine-scale structures, visually preserving the true morphological characteristics of the Ground Truth. The ensemble averages, Diffusion Ens (5) and Ens (25), demonstrate a progressive spatial smoothing effect as stochastic noise is averaged out; however, the structural integrity and geographic placement of their rainfall bands remain vastly superior to the blurred outputs of the deterministic baselines.

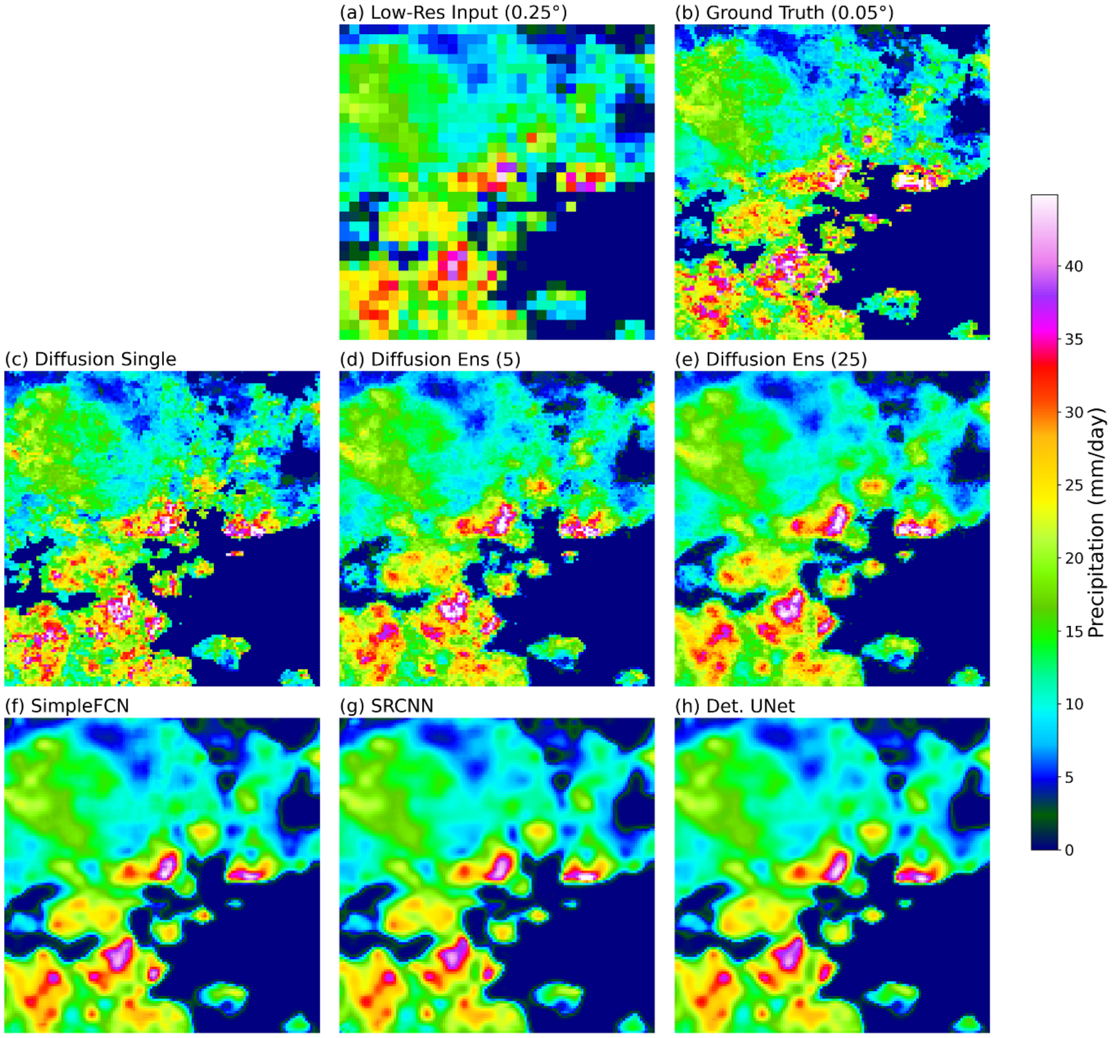


Figure 4. Spatial comparison of precipitation downscaling results for the 95th percentile rainfall events. Columns show the coarse-resolution input, high-resolution reference (ground truth), Diffusion Single, Diffusion Ens (5), Diffusion Ens (25), SimpleFCN, SRCNN, and Deterministic U-Net, respectively. Rows correspond to percentile-based extreme precipitation cases and terrain-conditioned cases.

Quantitative evaluations (summarized in table 2) explicitly reveal the fundamental tradeoffs between pointwise regression accuracy and physical realism. In terms of pixel-wise magnitude error, the Diffusion Ens (25) achieves the best overall performance across all models, recording the lowest RMSE (4.0907 mm d^{-1}) and the highest Pearson correlation (0.9123). Among the baselines, the Deterministic U-Net achieves the lowest error (RMSE: 4.2611 mm d^{-1}, MAE: 2.4268 mm d^{-1}). However, the individual Diffusion Single realization inherently yields a higher pointwise error (RMSE: 5.4891 mm d^{-1}) due to the 'double-penalty' effect common in high-resolution spatial verification, where slightly displaced stochastic high-frequency peaks are heavily penalized by grid-to-grid metrics. Yet, aggregating these stochastic samples (Diffusion Ens 25) effectively minimizes this deterministic error, surpassing even the most advanced deterministic U-Net architecture.

More importantly, the analysis of spatial variance and extreme values reveals a critical limitation of deterministic modeling. The Diffusion Single model demonstrates near-perfect preservation of natural spatial variability with a FSD of 1.0008 and captures the frequency of heavy rainfall with exceptional precision (EB95: 0.9248, EB99: 0.9509). Furthermore, it attains the lowest (KLD: 0.0125), proving its robust distributional similarity to the observational truth. In contrast, all deterministic models suffer from severe variance suppression (FSD < 0.90). A profound insight emerges from the Deterministic U-Net results: while it successfully minimizes the overall MAE better than other baselines, it exhibits the worst extreme event preservation, capturing only 23.31% of the 99th-percentile extremes (EB99: 0.2331)—a performance inferior even to the much simpler SimpleFCN (0.3067). This powerfully illustrates that the

Table 2. Quantitative performance metrics of the downscaling models averaged across the seven representative cases.

	Diff Single	Diff Ens (5)	Diff Ens (25)	SimpleFCN	SRCNN	Det. UNet
RMSE	5.4891	4.4921	4.0907	4.4045	4.4502	4.2611
MAE	3.1935	2.6484	2.4448	2.5084	2.5479	2.4268
Corr	0.8489	0.8945	0.9123	0.9012	0.9000	0.9084
SSIM	0.5713	0.6505	0.6833	0.6663	0.6623	0.6748
FSD	1.0008	0.9466	0.9280	0.8967	0.8853	0.8894
PSD_Corr	0.9932	0.9938	0.9901	0.9863	0.9851	0.9860
KLD	0.0125	0.0421	0.0732	0.1353	0.1255	0.1252
EB95	0.9248	0.8323	0.7454	0.5777	0.5373	0.5496
EB99	0.9509	0.6626	0.5951	0.3067	0.2699	0.2331
WaveSim	0.9456	0.6629	0.5050	0.3549	0.3444	0.3646
WaveSim_M	0.9731	0.6797	0.5210	0.3758	0.3682	0.3869
WaveSim_D	0.9742	0.9783	0.9730	0.9271	0.8922	0.9160
WaveSim_S	0.9975	0.9945	0.9902	0.9749	0.9608	0.9685
Max	88.4934	56.0117	54.7387	44.9884	53.0644	45.2572
Mean	11.2378	11.2460	11.2409	10.4073	10.3093	10.4009

inability to resolve convective extremes is not a consequence of architectural simplicity, but an unavoidable mathematical artifact of the MSE optimization objective, which aggressively attenuates outliers to predict the conditional mean.

Lastly, the multiscale physical consistency is confirmed by the WaveSim and PSD analyses. The Diffusion Single model achieves an overall WaveSim score of 0.9456, significantly outperforming the SimpleFCN (0.3549), SRCNN (0.3444), and Deterministic U-Net (0.3646). Decomposing this metric provides a precise diagnosis of the structural limitations inherent in deterministic downscaling. While all models demonstrate a high capability in capturing the overall morphological structure (WaveSim_$S > 0.96$) and general spatial displacement patterns (WaveSim_$D > 0.89$), they diverge drastically in the magnitude component. The Diffusion Single model successfully retains the correct energy magnitudes across all wavelet decomposition scales (WaveSim_M: 0.9731). In contrast, the deterministic baselines suffer a severe collapse in this specific metric (WaveSim_M dropping to approximately 0.37–0.39). This indicates that while regression models can place general rainfall shapes in roughly correct locations, they fail completely to mathematically sustain the required intensity of high-resolution microstructures.

This scale-dependent limitation is further visually corroborated by the spatial PSD curves presented in figure 5. The Ground Truth reference exhibits a gradual decay of spectral energy with increasing spatial frequency, which reflects the natural cascade of atmospheric energy from broad synoptic scales down to localized, highly intense convective cores. The Diffusion Single model tracks this reference spectrum almost perfectly across the entire frequency domain (PSD_Corr: 0.9932), confirming its ability to inject physically appropriate high-frequency details. Conversely, all deterministic models display a sharp, premature drop-off in energy density at higher spatial frequencies. This visible loss of high-frequency energy explicitly demonstrates the smoothing effect of the MSE objective, proving that deterministic approaches inherently filter out the essential sub-grid textures and extreme peak intensities required for realistic precipitation modeling.

4.2. Feature importance analysis

To evaluate whether the diffusion model performs a physically consistent reconstruction rather than mere statistical pattern completion, feature importance was analyzed across eight distinct environmental scenarios. These scenarios include: (1) an overall baseline evaluation, the four meteorological seasons (2–5) DJF, MAM, JJA, SON, (6) extreme heavy precipitation events (>90th percentile), (7) high-altitude regions (upper decile of mean elevation), and (8) rugged terrain (upper decile of elevation standard deviation).

Feature importance was quantified using a perturbation-based sensitivity analysis, measuring the degradation in the model's denoising accuracy when specific input channels were masked. The resulting ranks (figure 6) illuminate the internal hierarchy of the model's generative logic, revealing a dynamic interplay between static constraints and transient atmospheric drivers.

Across every single scenario, the model exhibits a rigid structural hierarchy at the top: the coarse-grained precipitation (chirps_low) invariably occupies Rank 1, immediately followed by the primary topographic constraints (dem_mean, dem_max, and dem_min) which dominate Ranks 2 through 4. This

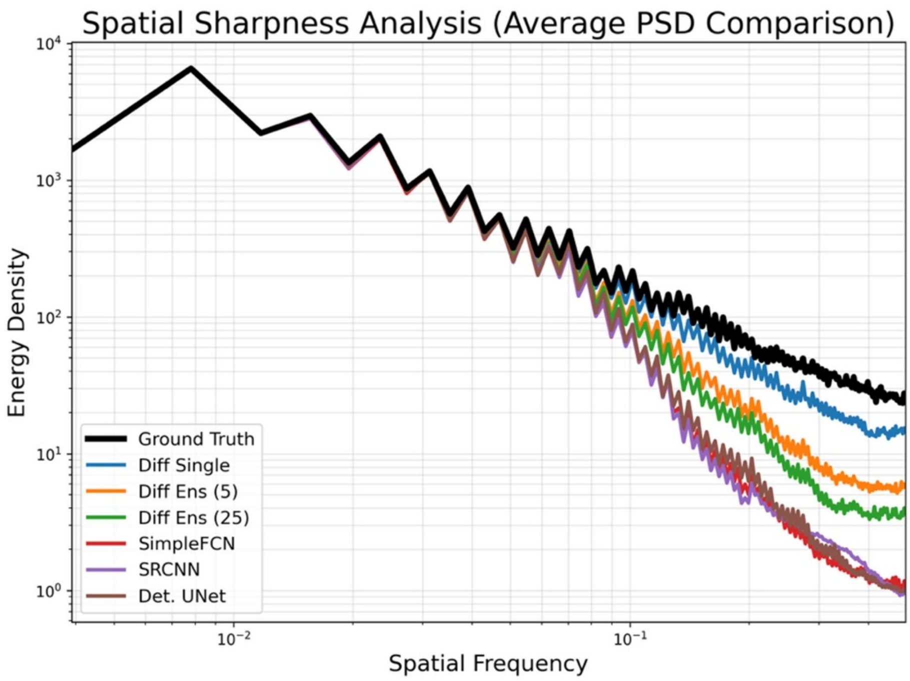


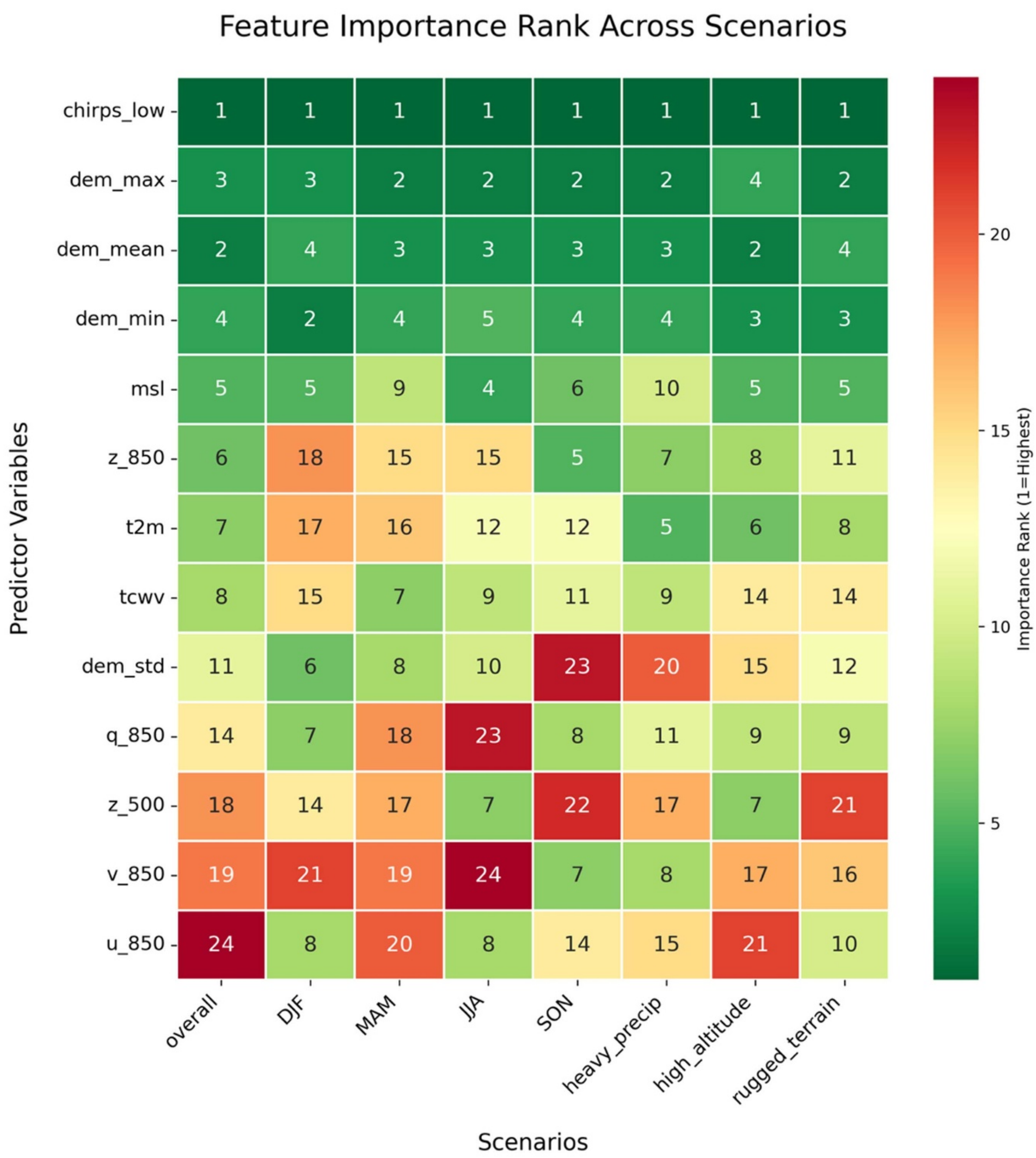
Figure 5. Average spatial power spectral density (PSD) of the high-resolution reference and precipitation reconstructions from each model across the seven representative cases.

unwavering stability provides a profound insight into the diffusion framework's mechanism. The model relies on the coarse input to dictate the absolute volumetric mass and general placement of the precipitation, while the high-resolution DEM provides the static spatial skeleton required to distribute that mass realistically. The diffusion process does not attempt to generate precipitation out of context; rather, it firmly anchors its generation to known mass boundaries and topographical realities before consulting any secondary atmospheric variables.

Beyond these primary anchors, the model's prioritization of auxiliary atmospheric variables shifts, proving its capacity to adapt to Sub-Saharan Africa's complex seasonal dynamics. In the DJF (winter) scenario, the model exhibits a sharp reliance on low-level moisture and wind vectors. Specific humidity at 850 hPa (q_850) climbs to Rank 7 (up from an overall baseline of 14), and zonal wind (u_850) climbs drastically from Rank 24 to Rank 8. Conversely, variables like lower-level geopotential height (z_850) and surface temperature (t2m) are heavily de-prioritized to Ranks 18 and 17, respectively. This indicates that during the dry season, when large-scale synoptic forcing is generally weak, the model correctly identifies specific low-level moisture convergence and local wind transport as the precise physical triggers needed to resolve isolated convective cells, closely aligning with the physically plausible behavior of dry-season convection in the region.

In contrast, the JJA (summer) scenario forces the model to adopt a different strategy that demonstrates high physical plausibility regarding the ITCZ and the WAM. During the monsoon, lower-level moisture (q_850) loses its discriminative value, dropping to Rank 23 as widespread humidity is a given. Instead, the model turns to mid-to-upper atmospheric dynamics: upper-level geopotential (z_500) escalates to Rank 7 (from an overall Rank 18), and low-level zonal wind (u_850) reaches Rank 8. This effectively captures the influence of the African Easterly Jet and associated African Easterly Waves. This directly aligns with known regional mechanisms where the interaction of mid-level wind shear and atmospheric instability, rather than mere surface moisture availability, dictates where extreme, fine-scale squall lines and rainfall cores will actually materialize.

During heavy precipitation extremes (>90th percentile), the model restructures its logic to focus heavily on moisture transport and surface thermodynamics. Meridional wind (v_850), which often dictates the transport of deep tropical moisture from the equator or oceans, surges to Rank 8 (up from an overall Rank 19), while surface temperature (t2m) rises to Rank 5. To accommodate these convective drivers, the model suppresses the influence of mean sea level pressure, which drops from its usual Rank 5 to Rank 10. In the context of localized extremes, the diffusion model correctly recognizes that broad-scale atmospheric stability (pressure) is less relevant than the sheer availability of moisture transport mechanisms delivering energy to the grid cell.



	overall	DJF	MAM	JJA	SON	heavy_precip	high_altitude	rugged_terrain
chirps_low	1	1	1	1	1	1	1	1
dem_max	3	3	2	2	2	2	4	2
dem_mean	2	4	3	3	3	3	2	4
dem_min	4	2	4	5	4	4	3	3
msl	5	5	9	4	6	10	5	5
z_850	6	18	15	15	5	7	8	11
t2m	7	17	16	12	12	5	6	8
tcwv	8	15	7	9	11	9	14	14
dem_std	11	6	8	10	23	20	15	12
q_850	14	7	18	23	8	11	9	9
z_500	18	14	17	7	22	17	7	21
v_850	19	21	19	24	7	8	17	16
u_850	24	8	20	8	14	15	21	10

Figure 6. Comparison of feature importance rankings across different environmental scenarios. The values represent the rank of each variable (1–24) based on its sensitivity score within the specific scenario.

The model's behavior bifurcates when dealing with different types of complex terrain. In high-altitude environments, the model relies on upper-level dynamics, with z_500 rising to Rank 7, while de-emphasizing total columnar moisture, as tcwv drops to Rank 14. This suggests the model understands that precipitation at high elevations is primarily driven by synoptic-scale flow intercepting massive mountain barriers, requiring information about mid-tropospheric steering currents.

Conversely, in highly rugged terrain, broad synoptic signals become useless for resolving fine-scale spatial variations. Consequently, z_500 plummets to Rank 21. Instead, the model strongly prioritizes the immediate surface environment: low-level moisture (q_850) reaches Rank 9, and low-level zonal wind (u_850) climbs to Rank 10. This demonstrates that in rugged landscapes, the model is actively simulating localized orographic lifting, using low-level wind crashing into complex, high-resolution topography to trigger small-scale convective generation.

Ultimately, this sensitivity analysis confirms that the diffusion model is not operating as a black-box statistical interpolator. By dynamically shifting its attention between upper-level forcing during monsoons, horizontal moisture transport during extremes, and low-level wind convergence in rugged terrain, the generative framework successfully encodes and exploits the underlying physical laws governing atmospheric dynamics.

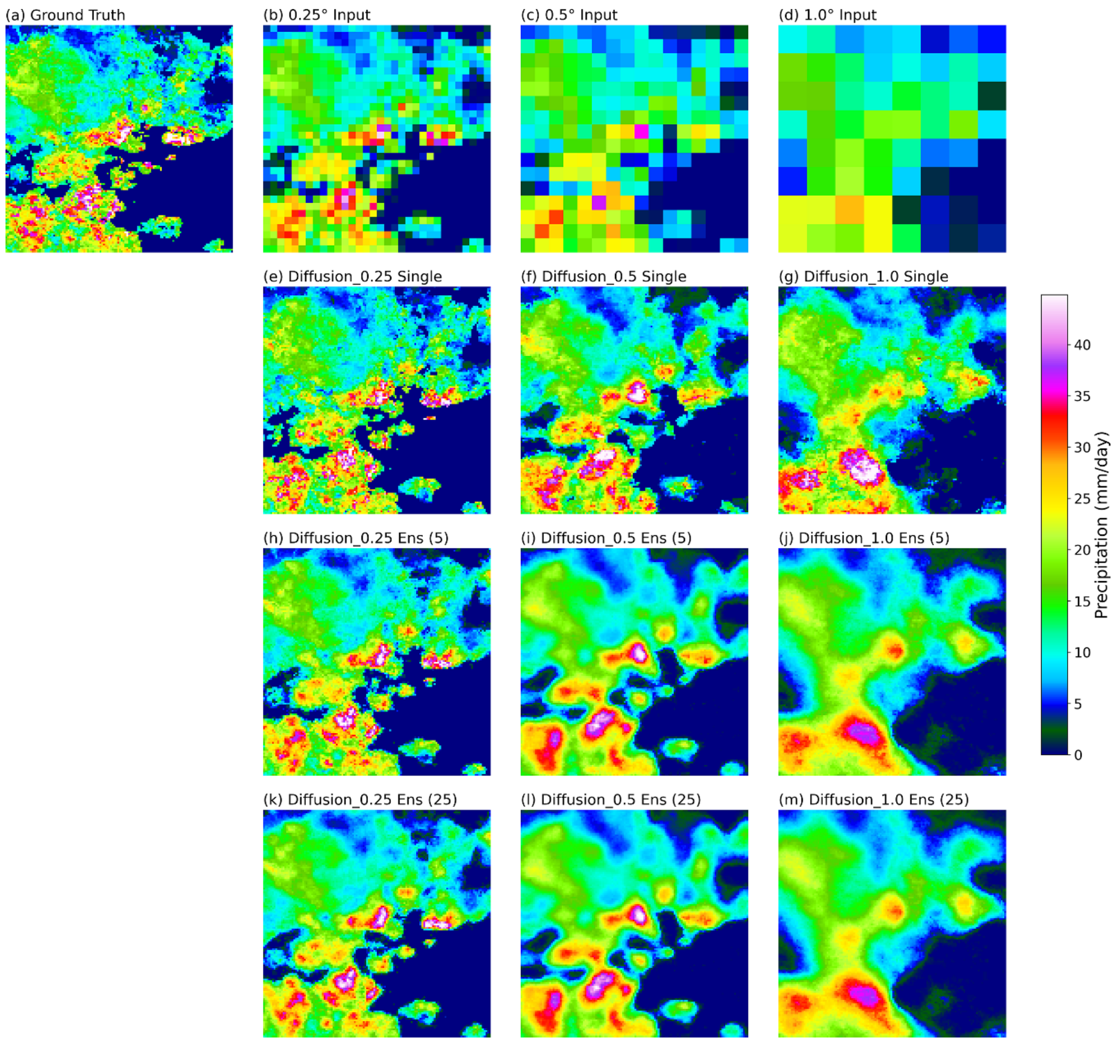


Figure 7. Spatial comparison of precipitation downscaling across different input resolutions for a representative 99th-percentile extreme event. The top row (a)–(d) displays the high-resolution ground truth alongside 0.25°, 0.5°, and 1.0° coarse-grained inputs. Subsequent rows present the downscaled reconstructions from the Diffusion_0.25, Diffusion_0.5, and Diffusion_1.0 models, illustrating the single realization (e)–(g) and the ensemble means of the 5 (h)–(j) and 25 (k)–(m) realizations.

4.3. Input resolution dependency

In practical climate applications, statistical downscaling frameworks are fundamentally tasked with bridging the scale gap between coarse GCM outputs and local impact scales. However, operating GCMs or dynamical regional models at highly resolved scales (e.g. 0.25°) imposes exponentially higher computational costs compared to running them at standard macroscopic resolutions (1.0° or 2.0°). This computational burden necessitates a critical scientific inquiry: To what extent can the probabilistic generative capacity of a diffusion model compensate for the severe information loss inherent in computationally inexpensive, coarse-resolution inputs? To systematically evaluate the boundaries of generative downscaling, we progressively degraded the high-resolution precipitation target into 0.5° and 1.0° coarse-grained inputs, training dedicated diffusion models for each spatial constraint. This experimental design allows us to isolate the model's dependence on structural input and analyze the trade-offs between computational efficiency and downscaling fidelity.

Figure 7 visually illustrates the model's response to progressive information degradation for a 99th-percentile extreme precipitation event. In the top row, the transition from the 0.25° to the 1.0° coarse-grained input reveals a profound loss of spatial constraints; at 1.0°, entire mesoscale convective structures are homogenized into uniform, low-intensity blocks. Despite this severe lack of structural guidance, the single stochastic realizations (Diffusion Single, second row) generated across all resolutions successfully reconstruct remarkably sharp spatial gradients and localized convective cores. Unlike deterministic CNNs that typically respond to coarse inputs with artificial blurring, the diffusion model leverages its learned internal priors to stochastically generate physically plausible sub-grid variability.

However, generating plausible textures does not equate to deterministic spatial accuracy. Because the diffusion model is probabilistic, it generates high-resolution details by sampling from a conditional distribution. At 0.25°, the conditioning signal acts as a tight physical anchor, restricting the generated storm cells to their true geographical locations. At 1.0°, the spatial degrees of freedom expand dramatically. The model correctly interprets the coarse input as indicative of a high-volume convective environment, but it lacks the deterministic constraints required to precisely locate the storm's epicenter.

This inherent spatial uncertainty is explicitly shown by the ensemble averages (Rows 3 and 4 in figure 7). Averaging multiple stochastic realizations mathematically functions as a spatial low-pass filter. For the 0.25° model, the 25-member ensemble (Ens 25) remains structurally coherent with the ground truth, demonstrating that the individual members are tightly clustered with low spatial uncertainty. Conversely, for the 1.0° model, the Ens (25) field is heavily smoothed. Because the individual stochastic members place the unconstrained fine-scale storm cores in vastly different locations across the 1.0° grid, their arithmetic mean smears the precipitation across the domain. This reveals a fundamental property of generative downscaling: the degree of ensemble smoothing serves as a direct, quantifiable proxy for the model's predictive spatial uncertainty under weak physical constraints.

The quantitative metrics (table 3) provide a rigorous breakdown of this resolution-dependent divergence between pointwise accuracy and structural realism. As the input coarsens from 0.25° to 1.0°, deterministic metrics for the Diffusion Single model predictably deteriorate. RMSE increases from 5.4891 mm d^{-1} (0.25°) to 7.5725 mm d^{-1} (1.0°), and Corr drops from 0.8489 to 0.7025.

Despite the degradation in spatial alignment, the generative structural metrics exhibit remarkable resilience up to a certain threshold. The FSD for the Single realizations stays near unity at both 0.25° (1.0008) and 0.5° (0.9838), indicating that the diffusion process flawlessly preserves the natural variance of the precipitation field. However, a critical boundary is crossed at 1.0°. At this resolution, the KLD increases significantly from 0.0125 to 0.1239. More importantly, the model's ability to accurately reconstruct the absolute volumetric peaks of extreme events becomes highly unconstrained. The EB99 shifts from a highly accurate 0.9509 at 0.25° to an overestimation of 1.0859 at 1.0°. This highlights an absolute limit of generative capability: while the model can reconstruct generic storm textures from 1.0° inputs, the input signal is too diluted to provide the physical constraints required to place peak extreme intensities accurately, resulting in stochastic false alarms across the unconstrained domain.

Consistently, based on the single sampling models, as the input data resolution coarsens from 0.25 ° to 0.5° and 1.0 °, the overall WaveSim score drops precipitously from 0.9456 to 0.5710 and 0.3140, respectively. Decomposing this metric into its sub-components provides a clearer diagnosis of this failure. While the single models manage to reasonably maintain spatial displacement patterns (WaveSim_D > 0.8946) and structural similarity (WaveSim_*S* > 0.9880) even at the coarsest 1.0-degree input, the multiscale energy magnitude (WaveSim_M) collapses entirely, falling from 0.9731 (at 0.25°) to just 0.3552 (at 1.0°). This indicates that the loss of fine-scale spatial conditioning information directly translates to an inability to mathematically sustain the correct intensity of high-resolution microstructures. Simultaneously, across all resolution conditions, a consistent downward trend in WaveSim scores is observed as the number of ensemble members increases to 5 and 25 (e.g. dropping from 0.3140 to 0.0655 and 0.0501 at the 1.0-degree resolution). Consequently, these results quantitatively demonstrate that coarser input resolutions accelerate the initial loss of multiscale energy, and the spatial smoothing induced by ensemble averaging acts as a critical detriment to preserving the physical texture of precipitation fields, regardless of the baseline input resolution level.

The preservation of multiscale variability is further corroborated by the PSD analysis (figure 8). Across all three subplots, the Diffusion Single realizations (blue lines) track the Ground Truth spectrum (black line) with near-perfect fidelity across the frequency domain, maintaining PSD correlations of 0.9888 or higher regardless of input resolution. This confirms that the diffusion framework consistently injects physically appropriate high-frequency energy.

The PSD plots also visualize the mechanics of spatial uncertainty discussed earlier. As explicitly highlighted by the annotations and zoom-in panels in figure 8, the vertical variance gap between the Single spectrum and the Ens (25) spectrum represents the loss of high-frequency energy due to destructive interference among ensemble members. In figure 8(a) (Diffusion_0.25), this gap is relatively narrow, reflecting a tightly constrained downscaling regime where high-resolution inputs anchor the stochastic generation. Conversely, in figure 8(c) (Diffusion_1.0), the gap drastically widens. The 1.0° input permits a vast latent space of plausible high-resolution configurations; consequently, averaging them neutralizes almost all high-frequency spectral energy. This resolution-dependent divergence is explicitly quantified in figure 8(d). The log energy gap between the single stochastic realization and the 25-member ensemble mean increases sharply as the input resolution coarsens, rising from 0.57 at 0.25° to 1.03 at 0.5°, and eventually reaching 1.72 at 1.0°. These rising values provide a transparent, quantifiable measure of the

Table 3. Quantitative performance metrics of the diffusion downscaling models across varying input resolutions (0.25°, 0.5°, and 1.0°) and ensemble sizes (Single, 5-member, and 25-member means), averaged across the seven representative cases.

	Diffusion_0.25 Single	Diffusion_0.25 Ens (5)	Diffusion_0.25 Ens (25)	Diffusion_0.5 Single	Diffusion_0.5 Ens (5)	Diffusion_0.5 Ens (25)	Diffusion_1.0 Single	Diffusion_1.0 Ens (5)	Diffusion_1.0 Ens (25)
RMSE	5.4891	4.4921	4.0907	5.8044	5.2028	5.2033	7.5725	6.8866	6.9194
MAE	3.1935	2.6484	2.4448	3.5844	3.2040	3.1981	4.8834	4.4867	4.4970
Corr	0.8489	0.8945	0.9123	0.8290	0.8579	0.8575	0.7025	0.7407	0.7371
SSIM	0.5713	0.6505	0.6833	0.4951	0.5538	0.5539	0.3261	0.3847	0.3837
FSD	1.0008	0.9466	0.9280	0.9838	0.9390	0.9355	0.9613	0.8929	0.8870
PSD_Corr	0.9932	0.9938	0.9901	0.9941	0.9858	0.9853	0.9888	0.9848	0.9840
KLD	0.0125	0.0421	0.0732	0.0585	0.1259	0.1221	0.1239	0.2924	0.3201
EB95	0.9248	0.8323	0.7454	1.0294	0.8849	0.8813	0.9241	0.6659	0.6034
EB99	0.9509	0.6626	0.5951	0.8650	0.4540	0.4110	1.0859	0.3452	0.3923
WaveSim	0.9456	0.6629	0.5050	0.5710	0.2675	0.2516	0.3140	0.0655	0.0501
WaveSim_M	0.9731	0.6797	0.5210	0.6060	0.2794	0.2631	0.3552	0.0735	0.0570
WaveSim_D	0.9742	0.9783	0.9730	0.9458	0.9659	0.9659	0.8946	0.9147	0.9086
WaveSim_S	0.9975	0.9945	0.9902	0.9937	0.9868	0.9861	0.9880	0.9762	0.9743
Max	88.4934	56.0117	54.7387	61.6647	50.9585	49.9406	55.0168	39.9277	38.5378
Mean	11.2378	11.2460	11.2409	11.6125	11.5785	11.5664	11.7033	11.6398	11.6267

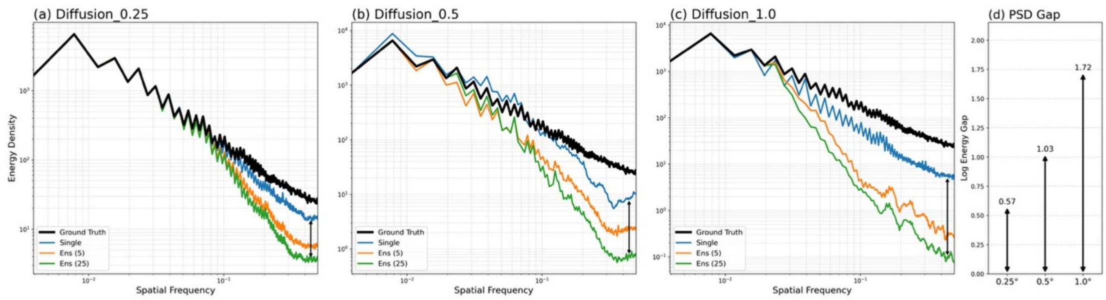


Figure 8. Average spatial power spectral density (PSD) across the seven representative cases at varying input resolutions: (a) Diffusion_0.25, (b) Diffusion_0.5, and (c) Diffusion_1.0. Each panel compares the high-resolution Ground Truth (black line) with the model's single realization (blue), 5-member ensemble mean (orange), and 25-member ensemble mean (green). Vertical arrows in (a)–(c) indicate the energy gap at high spatial frequencies. Panel (d) shows the quantitative log energy gap between the single realization and the 25-member ensemble mean for each resolution level.

model's transition from a tightly constrained structural recovery at high input resolutions to a highly unconstrained, purely stochastic generation regime under severe information loss.

These findings provide crucial strategic guidance for the design of future climate projection frameworks. Generative downscaling is an exceptionally powerful tool, but it cannot manufacture deterministic certainty out of a vacuum. If a GCM input is excessively coarse (e.g. 1.0°), the diffusion model will convert deterministic structural errors into massive spatial uncertainty, ultimately failing to capture the true geographic magnitude of extreme convective risks.

Therefore, a deliberate balance must be struck. While running GCMs at 0.25° is optimal for downscaling, it is often computationally prohibitive for large multi-decadal ensembles. Our quantitative results suggest that a moderate resolution of 0.5° serves as a computational optimum in this case. At 0.5°, the diffusion framework mathematically aligns to provide both solid generative fidelity (maintaining an EB99 of 0.8650 and FSD of 0.9838 for the single realization) and reliable deterministic accuracy when ensembled (RMSE of 5.2033 for Ens 25). Ultimately, this demonstrates that while diffusion models drastically enhance downscaling fidelity by recovering sub-grid variability, providing a sufficient baseline structural anchor remains an indispensable prerequisite for reliable local climate risk assessments.

5. Conclusion

Accurate high-resolution precipitation data is fundamentally required for climate risk assessment, hydrological modeling, and water resource management, particularly in regions with complex hydroclimatic regimes like Sub-Saharan Africa. In this study, we systematically evaluated a conditional denoising diffusion model for precipitation downscaling. By comparing its performance against deterministic convolutional baselines (SimpleFCN, SRCNN, and a Deterministic U-Net) across diverse intensity thresholds and topographic conditions, we investigated the model's capacity to recover sub-grid variability, its physical interpretability, and its sensitivity to progressive input resolution degradation.

Our quantitative evaluation highlights an inherent methodological trade-off between pointwise accuracy and structural realism in statistical downscaling. Deterministic architectures, optimized to minimize MSE, inherently penalize the misplacement of high-frequency features. By utilizing a Deterministic U-Net baseline identical in architecture to our diffusion model, we explicitly demonstrated that this failure to capture extremes is not a structural limitation, but an unavoidable mathematical artifact of the MSE objective. Consequently, deterministic models produce artificially smoothed precipitation fields that yield lower RMSE but systematically underestimate heavy rainfall events and suppress natural spatial variance. This loss of multiscale structural integrity is quantitatively reflected in their notably low WaveSim scores. In contrast, individual realizations generated by the diffusion model successfully preserve the complex, heavy-tailed distributions of precipitation. Across all tested scenarios, the single diffusion model maintained a FSD near unity, accurately reproduced extreme event frequencies, and achieved superior WaveSim scores, demonstrating its robust capacity to restore physical structures across multiple scales. Furthermore, the generative framework provides operational flexibility through ensembling. While individual stochastic realizations are optimal for applications requiring precise structural textures and extreme values, the ensemble mean effectively filters out high-frequency noise. This filtering mechanism achieves competitive pointwise accuracy, though the accompanying sharp decline in WaveSim scores explicitly quantifies the spatial smoothing and degradation of fine-scale physical textures induced by ensemble averaging.

Beyond statistical performance, our feature importance analysis provides critical insights into the internal mechanisms of the diffusion model, demonstrating that its generative process is physically interpretable rather than purely correlative. The model establishes a robust structural hierarchy: coarse precipitation and topographic variables consistently serve as the primary spatial anchors for volumetric mass and placement. More importantly, the model dynamically shifts its reliance on auxiliary atmospheric predictors in response to varying environmental contexts. For example, during dry seasons or over rugged terrain, the model actively prioritizes low-level moisture and wind convergence to resolve localized convective cells. Conversely, during the summer or in high-altitude environments, its sensitivity shifts toward mid-tropospheric dynamic forcing and upper-level geopotential. This context-dependent prioritization strongly indicates that the diffusion framework successfully encodes and exploits relevant meteorological principles, such as moisture transport and orographic lifting, to guide its sub-grid reconstruction.

By systematically coarsening the input constraints from 0.25° to 1.0°, we also delineated the fundamental limits of generative downscaling under severe information loss. As the input resolution decreases, the diffusion model continues to generate structurally realistic fine-scale textures, effectively avoiding the blurring effect seen in CNNs. However, this visual realism is accompanied by a natural decline in deterministic spatial accuracy. Because a 1.0° coarse input provides highly generalized spatial constraints, the model experiences a vast expansion in its spatial degrees of freedom. This unconstrained environment causes individual stochastic members to misplace intense convective cores, occasionally leading to overestimations or false alarms in localized extremes. Consequently, when these varied realizations are averaged, the resulting ensemble mean exhibits substantial spatial smoothing, heavily neutralizing the high-frequency spectral energy. This behavior provides a critical scientific insight: the degree of smoothing across a diffusion ensemble is a quantifiable measure of the predictive spatial uncertainty inherently caused by the resolution limits of the input data. Ultimately, our findings suggest that a moderate input resolution of 0.5° serves as a practical computational optimum, providing sufficient structural anchors to mathematically align solid deterministic accuracy with near-perfect extreme event preservation.

Despite the demonstrated advantages of generative downscaling, several limitations must be addressed in future research. First, the iterative reverse sampling process inherent to diffusion models introduces significantly higher computational costs during inference compared to single-step deterministic models. Scaling this framework to generate massive, multi-decadal ensembles for global climate projections will require the integration of advanced sampling acceleration techniques. Second, like all statistical downscaling methodologies, this framework relies on the stationarity assumption, the premise that the learned relationships between large-scale atmospheric predictors and local precipitation will remain constant under future climate conditions. The robustness of these generative priors under out-of-distribution greenhouse gas forcing scenarios requires rigorous validation. Finally, the physical realism of the model is fundamentally bounded by the observational uncertainties of the target dataset (CHIRPS), which can be particularly pronounced in gauge-sparse regions of Africa. Future studies should explore the integration of multi-source observational constraints and physically guided loss functions to further ensure hydrometeorological consistency across generated climate scenarios.

Acknowledgment

This work was carried out thanks to the support by Wellcome (Grant n. 309864/Z/23/Z). The views expressed herein do not necessarily represent those of UNESCO, ICTP or Wellcome.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://data.chc.ucsb.edu/products/CHIRPS/v3.0/.

Supplementary Information available at https://doi.org/10.1088/3049-4753/ae7114/data1.

ORCID iD

Jangho Lee 0000-0002-8942-1092

References

Accarino G, Acquaviva V, Shamekh S, Watson-Parris D and Lawrence D 2025. WaveSim: a wavelet-based multi-scale similarity metric for weather and climate fields (arXiv:2512.14656)

Adachi S A and Tomita H 2020 Methodology of the constraint condition in dynamical downscaling for regional climate evaluation: a review *J. Geophys. Res. Atmos.* **125** e2019JD032166

Aich M, Hess P, Pan B, Bathiany S, Huang Y and Boers N 2025 Conditional diffusion models for downscaling & bias correction of Earth system model precipitation *EGUsphere* **2025** 1–21

Antonio B, McRae A T, MacLeod D, Cooper F C, Marsham J, Aitchison L, Palmer T N and Watson P A 2025 Postprocessing East African rainfall forecasts using a generative machine learning model *J. Adv. Model. Earth Syst.* **17** e2024MS004796

Asfaw T G and Luo J-J 2024 Downscaling seasonal precipitation forecasts over East Africa with deep convolutional neural networks *Adv. Atmos. Sci.* **41** 449–64

Bassetti S, Hutchinson B, Tebaldi C and Kravitz B 2024 Diffesm: conditional emulation of temperature and precipitation in earth system models with 3D diffusion models *J. Adv. Model. Earth Syst.* **16** e2023MS004194

Bytheway J L and Kummerow C D 2013 Inferring the uncertainty of satellite precipitation estimates in data-sparse regions over land *J. Geophys. Res. Atmos.* **118** 9524–33

Chombo S, Bachwenkizi J, Omary H, Paulo H A, Luoga P, Zacharia A, Mbishi J V, Andrew J D and Addo I Y 2025 Associations of temperature and precipitation with malaria in children under 5: a multi-country study in Sub-Saharan Africa *PLoS One* **20** e0335031

Dhariwal P and Nichol A 2021 Diffusion models beat gans on image synthesis *Advances in Neural Information Processing Systems* vol 34 pp 8780–94

Dinku T, Funk C, Peterson P, Maidment R, Tadesse T, Gadain H and Ceccato P 2018 Validation of the CHIRPS satellite rainfall estimates over eastern Africa *Q. J. R. Meteorol. Soc.* **144** 292–312

Du H, Tan M L, Zhang F, Chun K P, Li L and Kabir M H 2024 Evaluating the effectiveness of CHIRPS data for hydroclimatic studies *Theor. Appl. Climatol.* **155** 1519–39

European Space Agency 2024 *Copernicus Global Digital Elevation Model* (https://doi.org/10.5069/G9028PQB)

Funk C C, Peterson P J, Landsfeld M F, Pedreros D H, Verdin J P, Rowland J D, Romero B E, Husak G J, Michaelsen J C and Verdin A P 2014 A quasi-global precipitation time series for drought monitoring

Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L and Hoell A 2015 The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes *Sci. Data* **2** 1–21

Grace K, Davenport F, Hanson H, Funk C and Shukla S 2015 Linking climate change and health outcomes: examining the relationship between temperature, precipitation and birth weight in Africa *Global Environ. Change* **35** 125–37

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R and Schepers D 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049

Ho J, Jain A and Abbeel P 2020 Denoising diffusion probabilistic models *Advances in Neural Information Processing Systems* vol 33 pp 6840–51

Katsanos D, Retalis A and Michaelides S 2016 Validation of a high-resolution precipitation database (CHIRPS) over Cyprus for a 30-year period *Atmos. Res.* **169** 459–64

Kim J, Lee M, Han H, Kim D, Bae Y and Kim H S 2022 Case study: development of the CNN model considering teleconnection for spatial downscaling of precipitation in a climate change scenario *Sustainability* **14** 4719

Kisakye V and Van der Bruggen B 2018 Effects of climate change on water savings and water security from rainwater harvesting systems *Resour. Conserv. Recycl.* **138** 49–63

Kofidou M, Stathopoulos S and Gemitzi A 2023 Review on spatial downscaling of satellite derived precipitation estimates *Environ. Earth Sci.* **82** 424

Li L, Carver R, Lopez-Gomez I, Sha F and Anderson J 2024 Generative emulation of weather forecast ensembles with diffusion models *Sci. Adv.* **10** eadk4489

Ling F, Lu Z, Luo J-J, Bai L, Behera S K, Jin D, Pan B, Jiang H and Yamagata T 2024 Diffusion model-based probabilistic downscaling for 180-year East Asian climate reconstruction *npj Clim. Atmos. Sci.* **7** 131

Liu Y, Doss-Gollin J, Dai Q, Veeraraghavan A and Balakrishnan G 2025 Downscaling extreme precipitation with wasserstein regularized diffusion *IEEE Trans. Geosci. Remote Sens.* **63** 5011018

Lu S, Veldhuis M-C T and van De Giesen N 2020 A methodology for multiobjective evaluation of precipitation products for extreme weather (In a data-scarce environment) *J. Hydrometeorol.* **21** 1223–44

Lyu R, Wang L, Sun Y, Bai H and Lu C-T 2024 Downscaling precipitation with bias-informed conditional diffusion model *2024 IEEE Int. Conf. on Big Data (Bigdata)*

Maraun D, Wetterhall F, Ireson A M, Chandler R E, Kendon E J, Widmann M, Brienen S, Rust H W, Sauter T and Themeßl M 2010 Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user *Rev. Geophys.* **48** RG3003

Mardani M, Brenowitz N, Cohen Y, Pathak J, Chen C-Y, Liu C-C, Vahdat A, Nabian M A, Ge T and Subramaniam A 2025 Residual corrective diffusion modeling for km-scale atmospheric downscaling *Commun. Earth Environ.* **6** 124

Mishra Sharma S C and Mitra A 2024 MAUNet: a max-average neural network architecture for precipitation downscaling *Neural Comput. Appl.* **36** 16465–84

Nai C, Pan B, Chen X, Tang Q, Ni G, Duan Q, Lu B, Xiao Z and Liu X 2024 Reliable precipitation nowcasting using probabilistic diffusion models *Environ. Res. Lett.* **19** 034039

Ofori S A, Cobbina S J and Obiri S 2021 Climate change, land, water, and food security: perspectives from Sub-Saharan Africa *Front. Sustain. Food Syst.* **5** 680924

Ogunbode T O, Oyebamiji V O, Oyelami A A, Akinkuolie T A, Adekiya A O and Ifabiyi P I 2025 Assessing water security in Sub-Saharan Africa in the context of climate change threats: a case study *Int. J. Clim. Change Strateg. Manage.* **17** 805–26

Paredes-Trejo F J, Barbosa H and Kumar T L 2017 Validating CHIRPS-based satellite precipitation estimates in Northeast Brazil *J. Arid Env.* **139** 26–40

Rivera J A, Marianetti G and Hinrichs S 2018 Validation of CHIRPS precipitation dataset along the Central Andes of Argentina *Atmos. Res.* **213** 437–49

Rowell D P 2012 Sources of uncertainty in future changes in local precipitation *Clim. Dyn.* **39** 1929–50

Sha Y, Gagne II D J, West G and Stull R 2020 Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part II: daily precipitation *J. Appl. Meteorol. Climatol.* **59** 2075–92

Shen M, Chen J, Zhuan M, Chen H, Xu C-Y and Xiong L 2018 Estimating uncertainty and its temporal variation related to global climate models in quantifying climate change impacts on hydrology *J. Hydrol.* **556** 10–24

Simard M, Denbina M, Marshak C and Neumann M 2024 A global evaluation of radar-derived digital elevation models: SRTM, NASADEM, and GLO-30 *J. Geophys. Res.-Biogeosci.* **129** e2023JG007672

Smith M W, Willis T, Mroz E, James W H, Klaar M J, Gosling S N and Thomas C J 2024 Future malaria environmental suitability in Africa is sensitive to hydrology *Science* **384** 697–703

Srivastava P, Yang R, Kerrigan G, Dresdner G, McGibbon J, Bretherton C S and Mandt S 2024 Precipitation downscaling with spatiotemporal video diffusion *Advances in Neural Information Processing Systems* vol 37 pp 56374–400

Sun Y, Deng K, Ren K, Liu J, Deng C and Jin Y 2024 Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: a systematic review *ISPRS J. Photogramm. Remote Sens.* **208** 14–38

Tarek M, Brissette F and Arsenault R 2021 Uncertainty of gridded precipitation and temperature reference datasets in climate change impact studies *Hydrol. Earth Syst. Sci.* **25** 3331–50

Tefera M L, Seddaiu G, Carletti A and Awada H 2025 Rainfall variability and drought in West Africa: challenges and implications for rainfed agriculture *Theor. Appl. Climatol.* **156** 1–24

Tu S, Fei B, Yang W, Ling F, Chen H, Liu Z, Chen K, Fan H, Ouyang W and Bai L 2025 Satellite observations guided diffusion model for accurate meteorological states at arbitrary resolution *Proc. Computer Vision and Pattern Recognition Conf.*

Tu T, Ishida K, Ercan A, Kiyama M, Amagasaki M and Zhao T 2021 Hybrid precipitation downscaling over coastal watersheds in Japan using WRF and CNN *J. Hydrol.* **37** 100921

Waha K, Müller C and Rolinski S 2013 Separate and combined effects of temperature and precipitation change on maize yields in Sub-Saharan Africa for mid-to late-21st century *Glob. Planet. Change* **106** 1–12

Watt R A and Mansfield L A 2024 Generative diffusion-based downscaling for climate (arXiv:2404.17752)

Wilby R and Yu D 2013 Rainfall and temperature estimation for a data sparse region *Hydrol. Earth Syst. Sci.* **17** 3937–55

Woldemeskel F, Sharma A, Sivakumar B and Mehrotra R 2016 Quantification of precipitation and temperature uncertainties simulated by CMIP3 and CMIP5 models *J. Geophys. Res. Atmos.* **121** 3–17

Xu Z, Han Y and Yang Z 2019 Dynamical downscaling of regional climate: a review of methods and limitations *Sci. China Earth Sci.* **62** 365–75

Yang D, Liu S, Hu Y, Liu X, Xie J and Zhao L 2023 Predictor selection for CNN-based statistical downscaling of monthly precipitation *Adv. Atmos. Sci.* **40** 1117–31