Supplement of Geosci. Model Dev., 19, 7545–7567, 2026
https://doi.org/10.5194/gmd-19-7545-2026-supplement


Geoscientific Model Development
Open Access

*Supplement of*

# Comprehensive inter-comparison of generative AI models for super-resolution precipitation downscaling across hydroclimatic regimes

**Shivam Singh et al.**

*Correspondence to:* Shivam Singh (wpa8me@virginia.edu) and Antonios Mamalakis (npa4tg@virginia.edu)

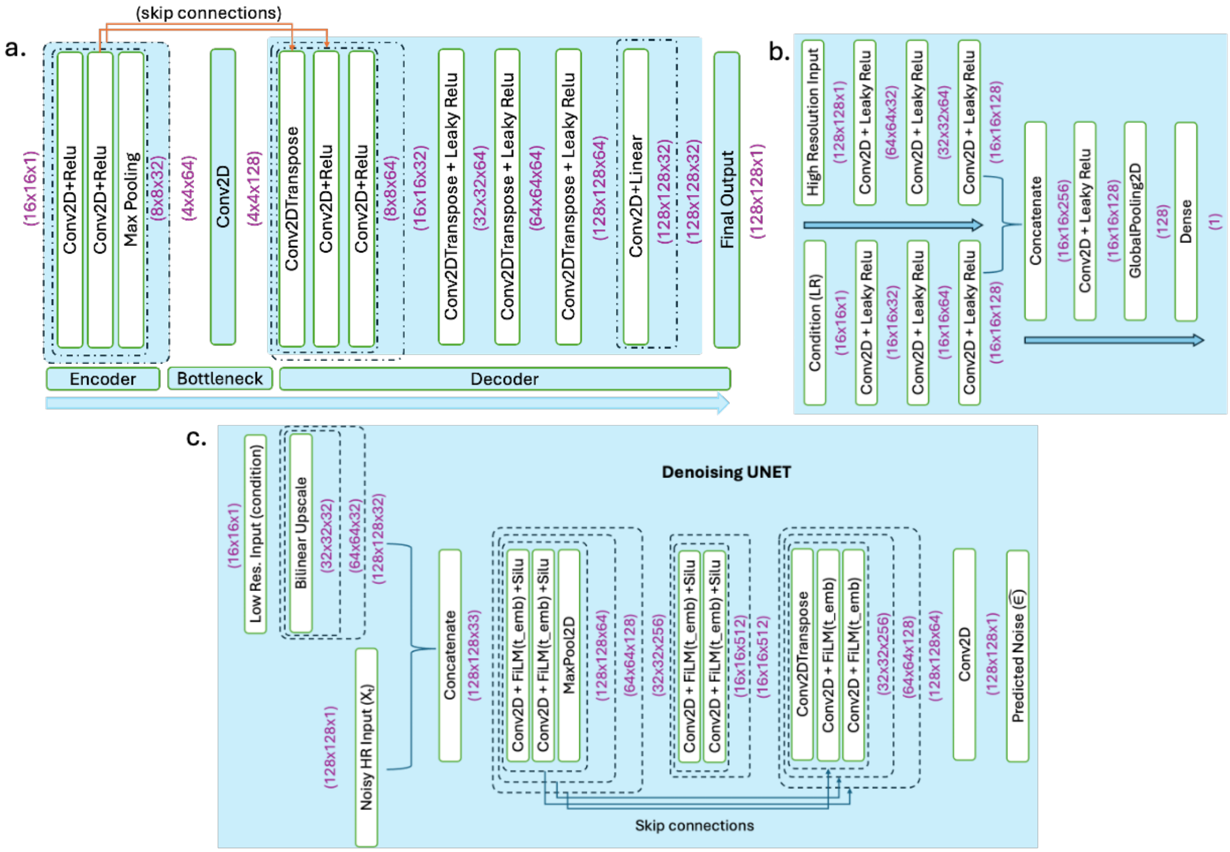


**Figure S1.** Architectures of the three downscaling models considered in this study. (a) U-Net generator used for deterministic super-resolution, mapping coarse precipitation inputs to high-resolution outputs. (b) WGAN critic, trained to distinguish between real and generated high-resolution precipitation fields under a gradient-penalty constraint. (c) Denoising U-Net used in the diffusion framework, trained to predict noise during the reverse diffusion process for probabilistic precipitation downscaling.

## Training behavior of UNET, WGAN and DDPM

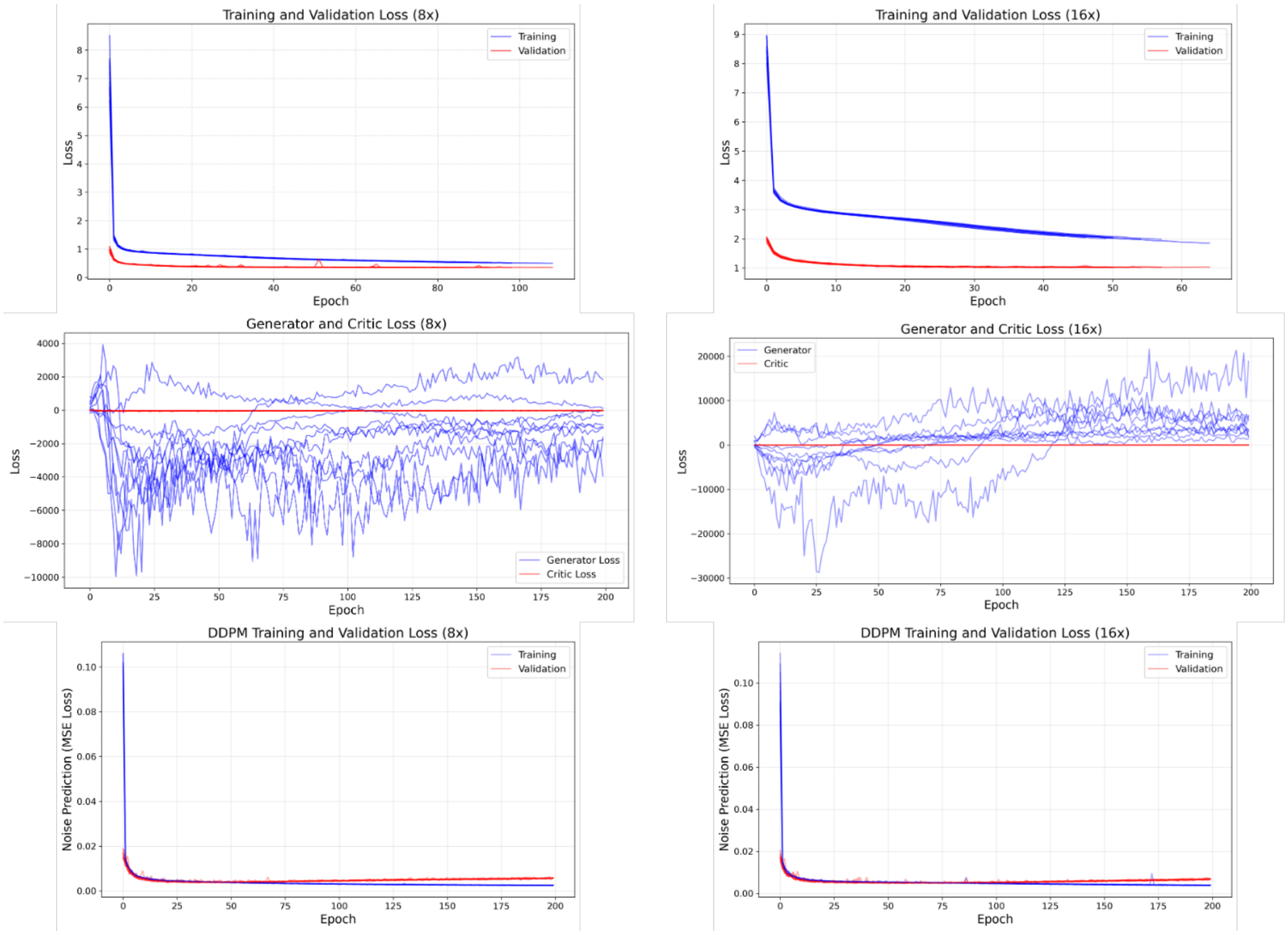


**Figure S2.** Training and validation curves for the U-Net, WGAN, and DDPM models under 8× and 16× downscaling configurations. (top) Training and validation loss for U-Net; (middle) Generator and critic losses for WGAN; (bottom) Noise-prediction MSE loss for DDPM. Each subplot represents one of ten independent (random) initialization seeds, shown separately for 8× (left column) and 16× (right column). Blue line represents training loss behavior (generator loss in case of WGAN) whereas red line represents validation loss behavior with epochs (critic loss in case of WGAN).

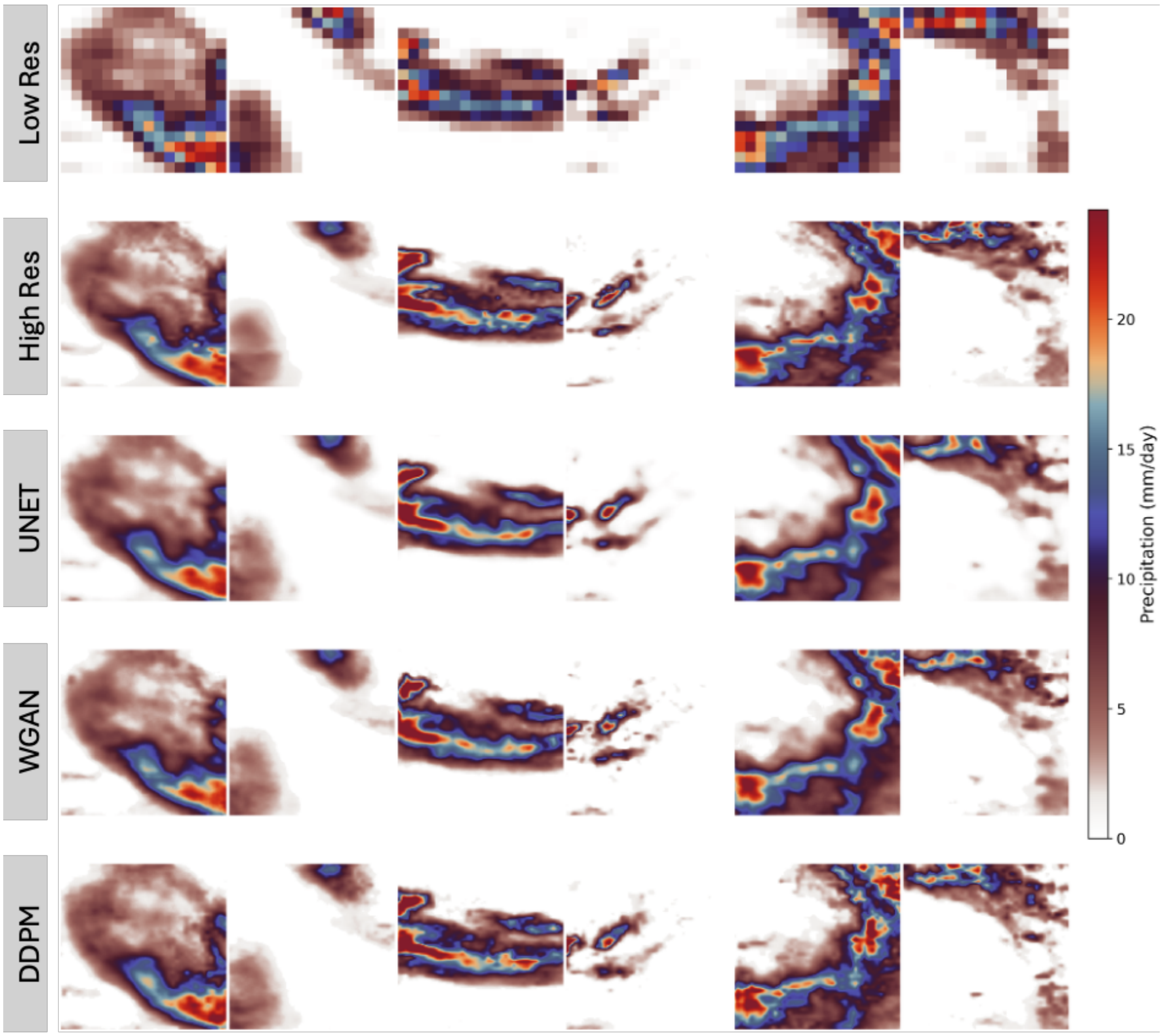


**Figure S3.** Representative examples of 16× precipitation downscaling from the Northeast test region. Columns correspond to four randomly selected precipitation events, while rows show the low-resolution input (8 × 8), the high-resolution target (ERA5-Land), and the corresponding reconstructions generated by U-Net, WGAN, and DDPM. All models successfully recover the large-scale precipitation structures present in the target fields; however, differences emerge in the representation of fine-scale spatial variability.

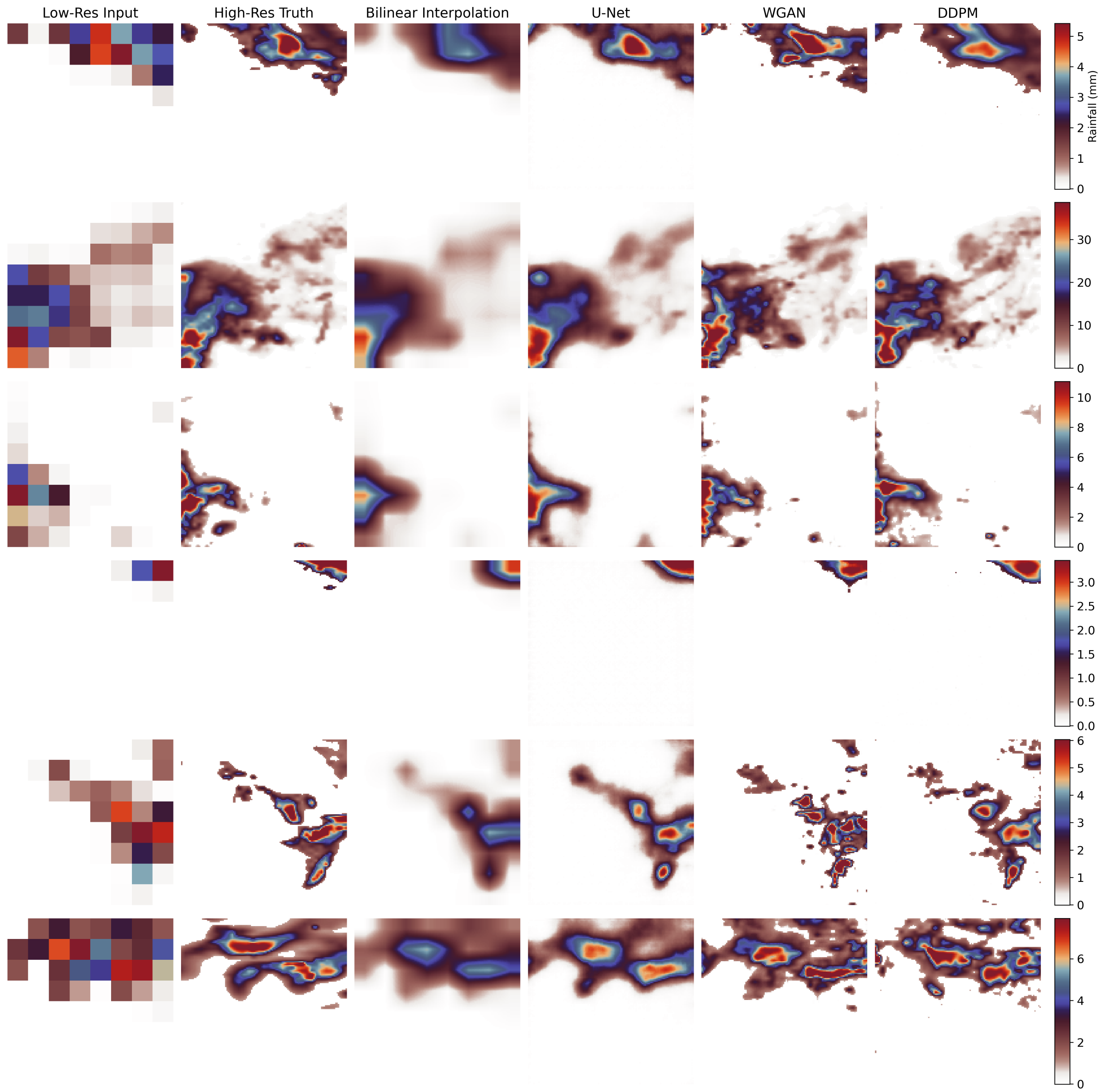


**Figure S4.** Representative qualitative comparison of precipitation downscaling results for five test samples from the Northeast evaluation region in the 16× super-resolution experiment. Columns show the low-resolution input (8×8), high-resolution ERA5-Land target, bilinear interpolation baseline, and predictions from the U-Net, WGAN, and DDPM models. Rows correspond to different precipitation events spanning a range of spatial structures and intensities. **Bilinear interpolation produces overly smooth fields and fails to recover fine-scale precipitation features.**

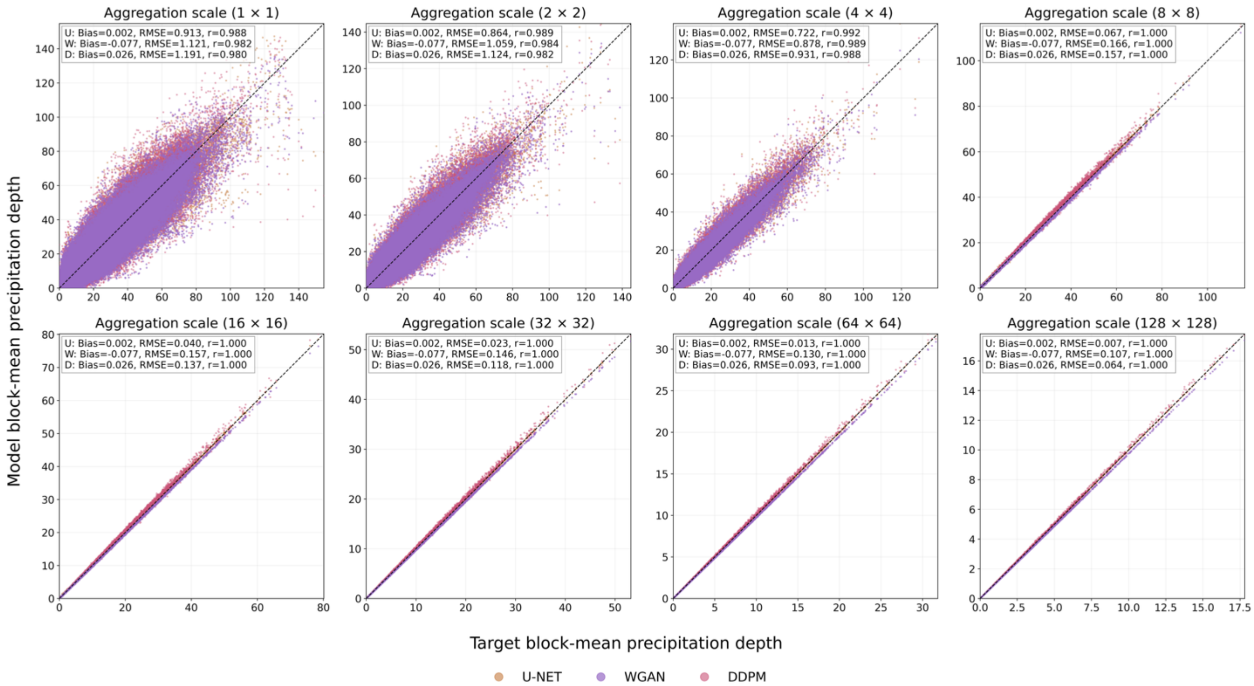


**Figure S5**. Mass (rainfall depth) consistency between observed and predicted precipitation across spatial aggregation scales for the 8× downscaling experiment. For all three models, discrepancies evident at fine scales progressively diminish with increasing spatial aggregation, demonstrating robust preservation of large-scale rainfall mass despite differences in small-scale variability.

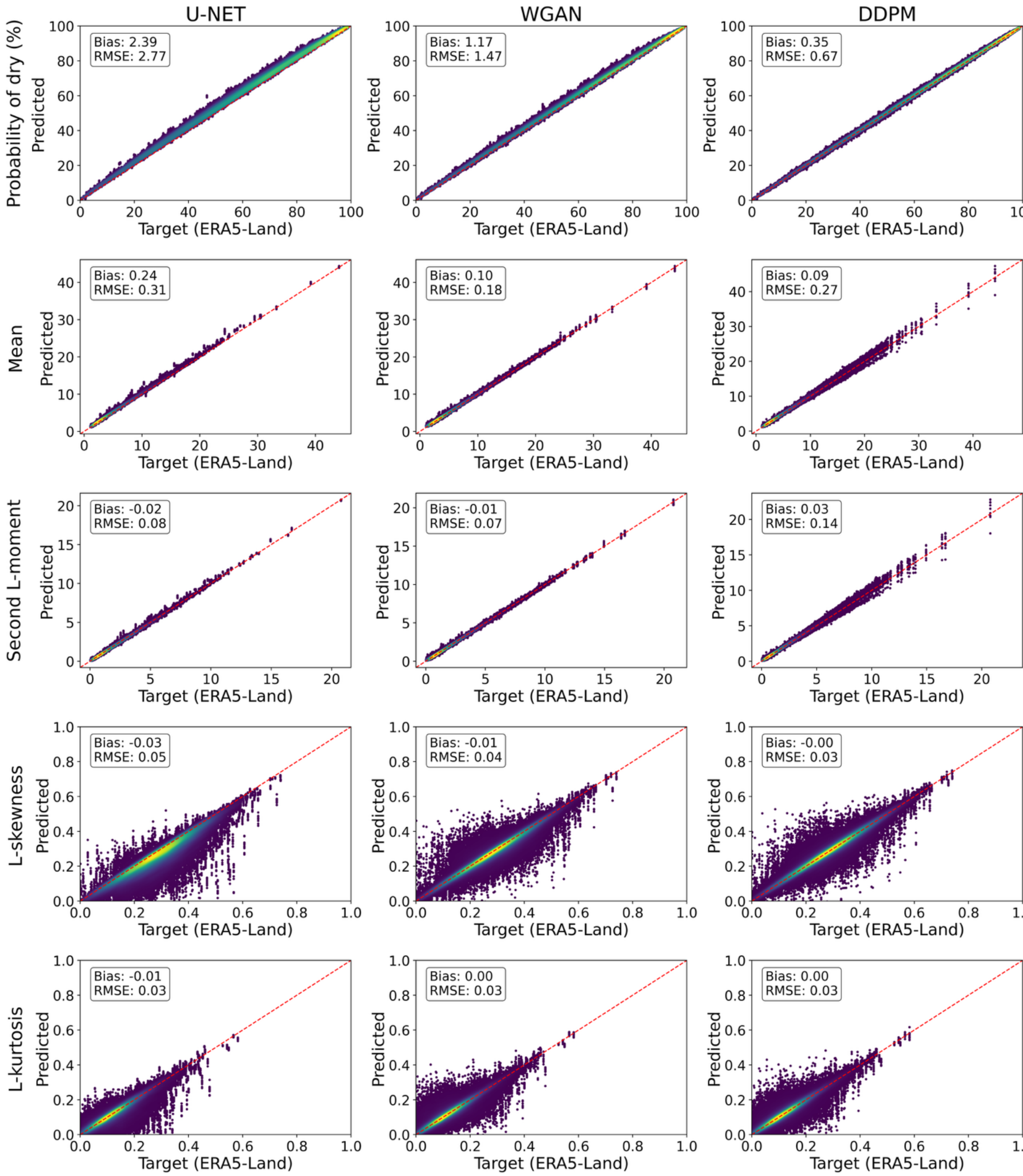


**Figure S6.** Pixel-wise comparison of precipitation occurrence and higher-order statistical moments between model predictions and the target (ERA5-Land) for the 8× downscaling experiment. Columns correspond to U-Net, WGAN, and DDPM, while rows show the probability of dry pixels ($P_0$), mean precipitation, second L-moment, L-skewness, and L-kurtosis. Each point represents a grid cell in the Northeast test region, and the dashed red line denotes the 1:1 relationship. Bias and RMSE values are shown in each panel.

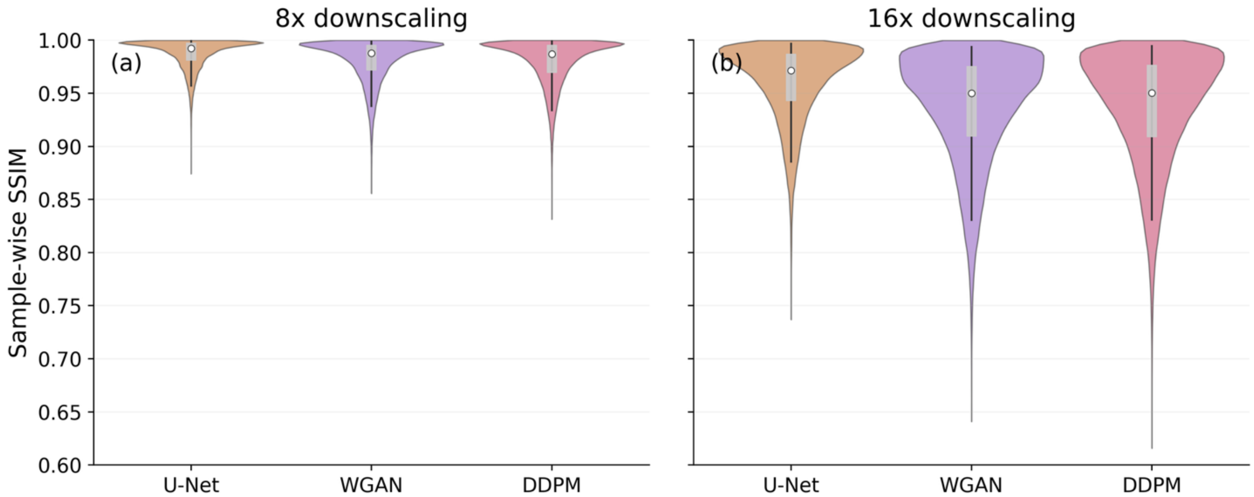


**Figure S7.** Distribution of sample-wise Structural Similarity Index Measure (SSIM) values for U-Net, WGAN, and DDPM across (a) 8× and (b) 16× precipitation downscaling experiments. SSIM was computed between each predicted precipitation field and the corresponding Target (ERA5-Land) field for all test samples and all 10 independently initialized seeds. Violin widths represent the density of SSIM values, while the internal vertical bar and marker denote the interquartile range (25th–75th percentile) and median, respectively. All models achieve high structural similarity at 8× downscaling, with distributions concentrated near SSIM ≈ 1.0. At 16× downscaling, the distributions broaden substantially, indicating increased difficulty in reconstructing fine-scale precipitation structures. U-Net exhibits the highest median SSIM and the narrowest distribution, whereas WGAN and DDPM display greater variability and longer lower tails, reflecting increased diversity in generated spatial patterns at higher downscaling factors.

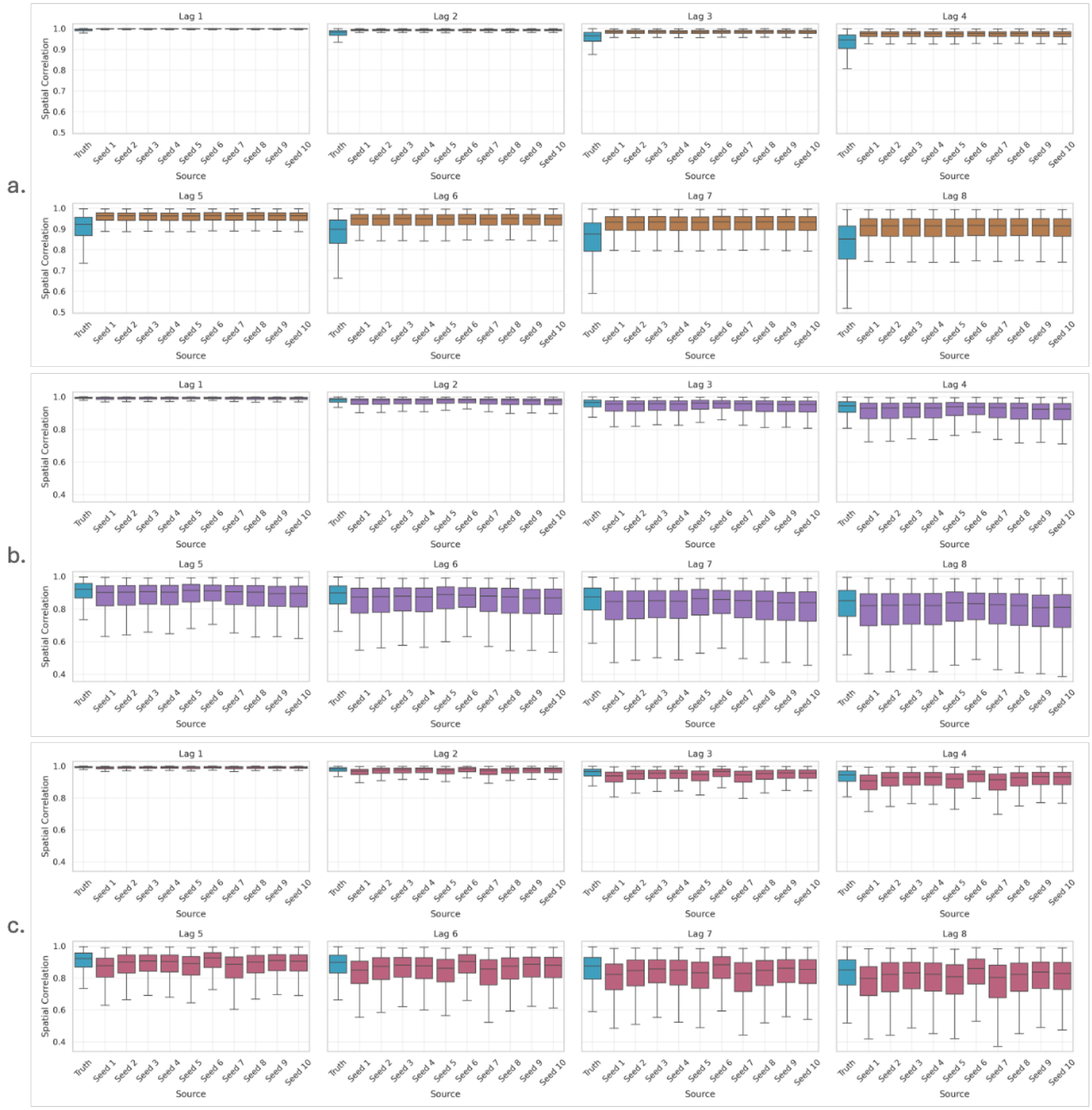


**Figure S8.** Lagged spatial correlation of daily precipitation fields for 16× downscaling. Boxplots summarize spatial correlation coefficients computed at increasing horizontal lag distances (lags 1–8, left to right) for (a) U-Net, (b) WGAN, and (c) DDPM. For each lag, correlations are evaluated over the full test set. The leftmost box in each panel corresponds to the reference (observed) precipitation field, while subsequent boxes represent individual model realizations from ten random seeds. Colored boxes denote model predictions, with each seed treated independently. The decay of spatial correlation with increasing lag provides a measure of how well each model reproduces multi-scale spatial dependence. Compared to the deterministic U-Net, the generative models (WGAN and DDPM) exhibit larger inter-seed variability, particularly at larger lags, reflecting stochastic representation of fine-scale spatial structure and uncertainty in spatial coherence.

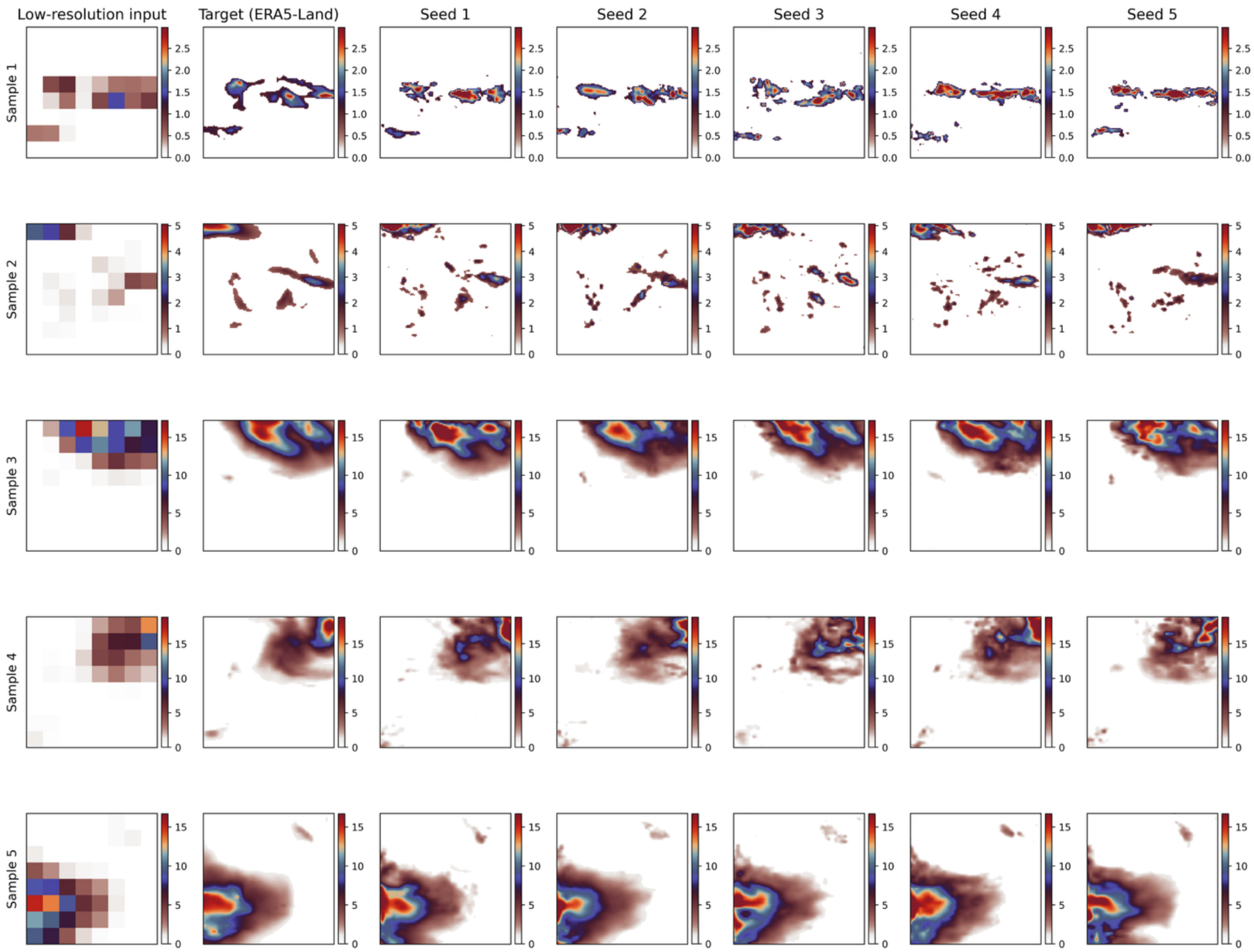


**Figure S9.** DDPM reconstructions (16x) for the T = 100 diffusion configuration. Each row shows a different precipitation event, with the low-resolution input (8×8), corresponding ground truth (128×128), and five independent DDPM realizations (seeds 1–5).

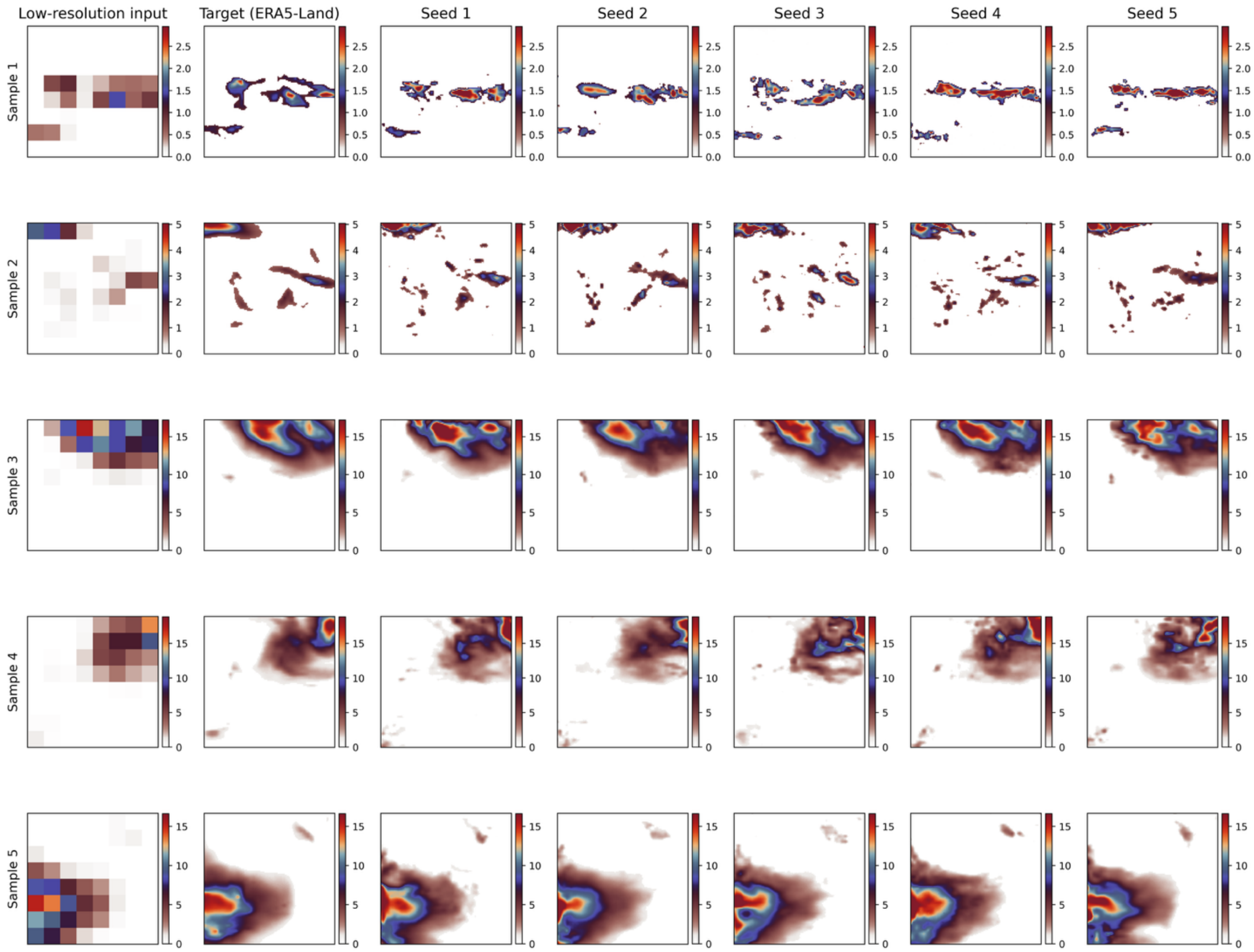


**Figure S10.** DDPM reconstructions (16x) for the T = 50 diffusion configuration. Each row shows a different precipitation event, with the low-resolution input (8×8), corresponding ground truth (128×128), and five independent DDPM realizations (seeds 1–5).

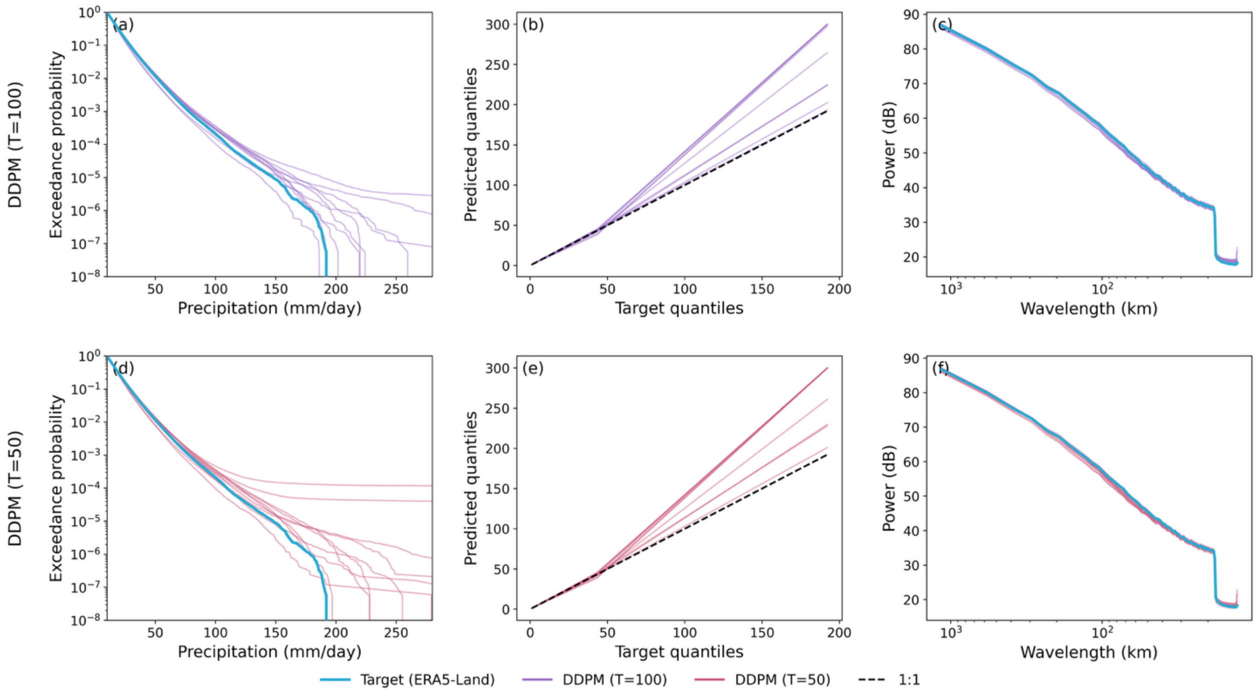


**Figure S11.** Sensitivity of DDPM precipitation downscaling performance to further reductions in diffusion inference steps from T=100 to T=50 for the 16× downscaling experiment. Panels (a, d) show exceedance probability distributions for precipitation intensities exceeding 10 mm/day, panels (b, e) show quantile–quantile (Q–Q) relationships relative to Target (ERA5-Land), and panels (c, f) show radial averaged power spectral density (RAPSD) as a function of wavelength. Results are shown for ten independently initialized DDPM seeds. The top row compares Target (ERA5-Land) with DDPM predictions generated using T=100 inference steps, while the bottom row shows the corresponding comparison for T=50. Cyan lines denote Target (ERA5-Land), purple lines denote DDPM (T=100), pink lines denote DDPM (T=50), and the dashed line in the Q–Q panels represents the one-to-one relationship.

**Table S1.** Number of trainable parameters for the architectures evaluated in this study, including the U-Net baseline, the WGAN generator and critic, and the DDPM denoising U-Net.

| Model | Trainable Parameters |
| --- | --- |
| U-Net | 497,121 |
| WGAN Generator | 497,121 |
| WGAN Critic | 552,417 |
| DDPM Denoising U-Net | 14,567,649 |