Environmental Data Science (2026), 5: e12, 1–25
doi:10.1017/eds.2026.10039

CAMBRIDGE UNIVERSITY PRESS

APPLICATION PAPER

Correcting dry/wet classification bias in precipitation downscaling via generative adversarial networks

Shivam Singh[1], Simon Michael Papalexiou[2], Hebatallah M. Abdelmoaty[2], Tom Hartvigsen[1] and Antonios Mamalakis[1]

[1]University of Virginia, USA
[2]University of Calgary, Canada
Corresponding author: Shivam Singh; Email: wpa8me@virginia.edu

Received: 19 May 2025; **Revised:** 18 March 2026; **Accepted:** 16 April 2026

Keywords: drizzle bias correction; dry/wet classification; generative models; precipitation downscaling; Wasserstein GAN (WGAN)

Abstract

Accurately downscaling precipitation from coarse to high spatial resolutions remains a critical challenge in climate and hydrometeorological modeling. A key limitation is the frequent misclassification of dry and wet regions, which compromises the realism and reliability of high-resolution outputs. To address this, we propose a deep learning-based downscaling framework that explicitly models dry/wet classification by transforming low-resolution 6×6 precipitation inputs into high-resolution 60 × 60 binary classification fields. We evaluate two architectures, a convolutional encoder-decoder and a conditional Wasserstein generative adversarial network (WGAN), utilizing three training strategies: (1) using binary wet/dry inputs, (2) using precipitation intensity inputs, and (3) using precipitation intensity inputs and adding physical constraints. Models are trained and validated on both synthetically generated precipitation fields and real radar-estimated hourly precipitation data over the contiguous United States. Performance is assessed using metrics including the overall probability of zero (P_0) and spatial autocorrelations. Results show that incorporating intensity information improves dry/wet classification, while adding physical constraints further enhances accuracy, generalization, and physical consistency, especially for WGAN models. The convolutional encoder-decoder produces smoother outputs with stable performance regarding marginal statistics, whereas the WGAN generates sharper boundaries and more realistic dry/wet fields, improving on the spatial dependence structure. Furthermore, we demonstrate that the derived dry/wet classification fields can be used as binary masks to bias correct downscaled precipitation fields, enhancing both statistical fidelity and spatial realism. These findings highlight the value of a physically informed bias-correction strategy for improving the spatial realism of high-resolution precipitation fields from coarse-scale inputs.

Impact Statement

Underestimation of dry regions due to drizzle bias remains a persistent limitation in statistical and deep learning-based precipitation downscaling, undermining the reliability of high-resolution products for drought assessment and hydrologic applications. This study addresses the problem by training convolutional encoder-decoder and WGAN models on both synthetic and radar-based datasets to predict high-resolution dry/wet classifications from

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

coarse-resolution inputs under varying conditioning strategies. While convolutional encoder–decoders reproduced marginal dry/wet statistics reliably, WGAN with dry-region constraints yields more spatially coherent and physically consistent outputs, particularly regarding storm boundaries and spatial autocorrelation structure. We demonstrate that the predicted dry/wet masks offer an efficient correction mechanism for mitigating drizzle bias in existing downscaled outputs, enhancing their utility in water resource management, hydrologic modeling, and climate impact assessments.

1. Introduction

Precipitation is a fundamental driver of the hydrological cycle and underpins a wide range of applications in water resources, agriculture, flood risk assessment, and climate impact studies. However, global climate models and most reanalysis products represent precipitation at relatively coarse spatial resolutions, typically ranging from about ~0.25° to over 1°, which limits their ability to capture localized storm structure, the spatial intermittency of precipitation, and extremes (Maraun et al., 2010; Giorgi and Gutowski, 2015; Dueben and Bauer, 2018; Wang et al., 2021; Kumar et al., 2023; Baghanam et al., 2024; Lopez-Gomez et al., 2024). As a result, precipitation downscaling has become an essential tool for translating coarse-scale climate information into spatially refined precipitation fields suitable for hydrologic and impact-oriented applications. Dynamical downscaling based on regional climate models (RCMs) and convection-permitting models (CPMs) offers higher resolution but remains computationally expensive, making its use in long-term simulations or large ensembles prohibitive (Gao et al., 2012; Giorgi and Gutowski, 2015; Gutowski et al., 2020; Potter et al., 2020; Hobeichi et al., 2023; Nishant et al., 2023; Rahimi et al., 2024). As a result, statistical and machine learning-based downscaling methods have gained momentum due to their efficiency, scalability, reduced computational demands, and rapid deployment capabilities in translating coarse-scale climate information into finer scales suitable for impact modeling (Vrac et al., 2007; Liu et al., 2020; Tabari et al., 2021; Miralles et al., 2022; Rampal et al., 2024).

A vast majority of statistical downscaling efforts, however, focus primarily on correcting precipitation intensity and improving mean or variance estimates, often neglecting the characterization of dry and wet states (AghaKouchak et al., 2012; Mamalakis et al., 2017; Pan et al., 2021; Chen et al., 2024; Luo et al., 2025). This oversight introduces significant biases in climate risk assessments, as the misclassification of dry/wet regimes can significantly impact applications such as storm impact analysis, drought investigation, and agricultural planning, where the presence or absence of rainfall is more critical than marginal errors in precipitation amounts (Mendoza Paz and Willems, 2023; Vogel et al., 2023; Rahimi et al., 2024). Particularly in arid and semi-arid regions, or during convective extremes, traditional downscaling methods often produce unrealistic drizzle or fail to preserve extended dry spells, due to treating precipitation as a purely continuous variable optimized for mean squared error rather than its intermittent nature (Maraun, 2016; Potter et al., 2020; Suliman et al., 2020; Harder et al., 2023). Machine learning (ML) and deep learning (DL) approaches, especially generative adversarial networks (GANs), have recently emerged as promising tools for downscaling precipitation fields due to their capacity to learn complex, nonlinear patterns and replicate fine-scale spatial features (Leinonen et al., 2020; Scher and Peßenteiner, 2021; Duncan et al., 2022; González-Abad et al., 2023; Murukesh et al., 2024). GAN-based approaches have been shown to better preserve precipitation structures, spatial heterogeneity, and fine-scale variability compared to a range of conventional statistical and deterministic deep learning downscaling methods (Leinonen et al., 2020; Harris et al., 2022; Murukesh et al., 2024; Papalexiou and Mamalakis, 2025).

However, even state-of-the-art GAN models tend to prioritize intensity reproduction, often failing to generate true dry regions and instead producing low-intensity spurious precipitation, especially near wet-dry boundaries (Harris et al., 2022; Papalexiou and Mamalakis, Papalexiou and Mamalakis, 2025). This "drizzle bias" is not merely a technical artifact, it fundamentally distorts the representation of the hydrological cycle, compromises the reliability of precipitation patterns, and can mislead water resource

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

assessments and climate impact analyses (Lazoglou et al., 2024). Papalexiou and Mamalakis (2025) systematically demonstrate this challenge through a controlled experiment using synthetic storm fields. Their evaluation of multiple neural network architectures, including a Wasserstein GAN (WGAN), highlights that while WGAN outperforms conventional models in preserving spatial structure and variability in precipitation, they fail to accurately replicate the probability of zero precipitation (P_0) and produce biased representations of extreme events. Notably, none of the evaluated models, including WGAN, were able to naturally reproduce dry regions, instead producing light precipitation (drizzle bias) across all grid cells. This behavior highlights a persistent dry/wet misclassification problem that remains a fundamental challenge in precipitation downscaling research. Recent studies reinforce this observation, pointing to a general trend where downscaling efforts, whether based on deep learning, traditional statistical approaches, or hybrid models, struggle with accurately capturing precipitation intermittency and the discrete nature of rainfall occurrence (Banõ-Medina et al., 2022; Nishant et al., 2023; Murukesh et al., 2024). For example, Chen et al. (2024) emphasize that despite improvements in spatial realism, many ML-based downscaling models underestimate dry days and overproduce light precipitation, failing to reproduce observed precipitation occurrence frequencies. Similarly, studies like those of Nishant et al. (2023) and Luo et al. (2025) highlight that statistical bias correction techniques often adjust mean values but do little to fix the structural misclassification between wet and dry states. Addressing this gap is critical, especially as climate change is expected to intensify precipitation extremes and alter dry spell dynamics (Berrang-Ford et al., 2011; Stocker, 2014; Lesnikowski et al., 2015; Magnan, 2016; Yazdandoost et al., 2021), making accurate dry/wet classification more consequential than ever. Incorporating classification-driven objectives within downscaling frameworks can substantially enhance the realism of downscaled datasets, ensuring that not only precipitation intensity but also the occurrence is accurately reproduced. Such an approach would improve hydrological model inputs, enhance the representation of precipitation occurrence and associated dry and wet conditions, and provide more actionable climate information for resource managers and policymakers.

This study proposes a generative adversarial network (GAN) framework designed to address dry/wet misclassification in precipitation downscaling. The framework incorporates explicit dry/wet classification objectives during model training to learn the spatial occurrence of precipitation, thereby reducing false wet and false dry predictions. The resulting dry/wet predictions explicitly represent the binary nature of precipitation occurrence (wet versus dry states) and can be applied as spatial masks to correct intensity-based downscaled precipitation fields. In convolutional and generative deep learning–based precipitation downscaling approaches, dry/wet corrections are often implemented using post-processing procedures based on fixed thresholds or rule-based criteria to suppress spurious low-intensity precipitation. In contrast, the proposed approach derives spatially coherent dry/wet masks directly from a data-trained generative model (WGAN), enabling the preservation of storm structure and spatial continuity. This data-driven masking strategy improves the physical consistency and interpretability of downscaled precipitation fields and reduces drizzle-related artifacts in hydrometeorological applications.

2. Data

2.1. Synthetic data

This study utilizes synthetic precipitation fields generated using the Complete Stochastic Modeling System (CoSMoS), as described in Papalexiou (2022). CoSMoS is designed to simulate hydroclimatic fields with prescribed marginal distributions, spatiotemporal autocorrelation structures, advection velocities, and anisotropy (Papalexiou et al., 2021). Its flexibility enables the creation of synthetic precipitation fields with user-defined statistical properties, making it ideal for evaluating downscaling model performance under controlled conditions. The same dataset was used by Papalexiou and Mamalakis (2025) to assess the capability of deep learning models in precipitation downscaling tasks. In this study, synthetically generated precipitation fields at a fine spatial resolution, with a grid size of 60×60, are treated as a reference dataset, enabling controlled evaluation of downscaling models against a known ground

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

truth. The corresponding low-resolution (6×6) synthetic fields are obtained by spatially aggregating the 60×60 fields using non-overlapping block averaging, consistent with the downscaling configuration adopted throughout the study. The grid dimensions describe the domain discretization of the synthetic fields and are not intended to represent a specific physical spatial resolution or to be spatially equivalent to radar-based datasets. Following aggregation, the 60×60 high-resolution precipitation fields were converted into binary dry/wet masks using the same thresholding procedure applied throughout the synthetic experiments, and these binary fields were used as the training targets. A total of 15,000 storm fields were used, comprising an equal mix of three distinct types: isotropic, anisotropic, and cyclonic storms. The dataset was randomly partitioned into 67% for training, 13% for validation, and 20% for testing, providing a balanced and robust approach for model development and assessment.

2.2. Radar data

To assess model performance on real-world precipitation, we utilized hourly accumulated precipitation fields from the NCEP Multi-Radar/Multi-Sensor (MRMS) system, accessed through the Iowa Environmental Mesonet (IEM). The MRMS dataset provides high-resolution precipitation measurements at 1 km × 1 km spatial resolution, making it well-suited for evaluating the ability of the models to capture fine-scale storm structure and dry/wet boundaries. For this study, a subset of MRMS data was extracted over 600 × 600 km spatial domains from hourly precipitation fields during a three-month winter period (November–January 2023). This period was selected to capture a wide range of cold-season precipitation characteristics and spatial intermittency typical of winter storm systems, including stratiform precipitation, frontal systems, and embedded convective features, while keeping the dataset computationally manageable for analysis (Chen et al., 2023; Akinsanola et al., 2024). To ensure geographical diversity and exposure to different climatic regimes, four distinct 600×600 km subregions were selected based on regional storm activity and hydroclimatic variability. The locations of these subregions are shown in Supplementary Figure S1, and their geographical details are summarized in Table 1.

To ensure a consistent input-output configuration with the synthetic experiments, the MRMS data were subsequently preprocessed to match the same input–output configuration used during model training, consisting of 6×6 low-resolution (LR) inputs and 60×60 high-resolution (HR) targets. This was achieved through non-overlapping block aggregation of the native 1 km precipitation fields, yielding 60×60 grids at an effective resolution of 10 km and corresponding 6×6 grids at 100 km resolution. Following aggregation, the 60×60 HR precipitation fields were converted into binary dry/wet masks using the same thresholding procedure applied in the synthetic experiments as described in Section 2.1. This preprocessing ensures that the radar-based evaluation data structurally match the synthetic input–output configuration, without implying physical spatial-scale equivalence between the two datasets.

To focus performance assessment on nontrivial precipitation cases, samples exhibiting zero precipitation across the entire subregion at a given time step were excluded. This filtering step is necessary because one of the model configurations applies a hard constraint that preserves dry regions during training (Section 3.3). In that case, fully dry samples would therefore lead to trivial predictions and provide limited diagnostic value. Excluding these cases ensures that the analysis emphasizes situations requiring reconstruction of precipitation structure and accurate delineation of dry and wet areas. In total,

Table 1. *Geographical locations of regions selected to extract high-resolution hourly precipitation with varied spatiotemporal patterns*

S.No.	Region	Latitude range	Longitude range
1.	Western region	46 ° N, 52 ° N	–126 ° W 120 ° W
2.	Southeastern region	26 ° N, 32 ° N	–86 ° W, –80 ° W
3.	Mid-South region	30 ° N, 36 ° N	–100 ° W, –94 ° W
4.	Eastern region	40 ° N, 46 ° N	–77 ° W, –71 ° W

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

8,828 radar images were initially collected across the four regions. After filtering fully dry samples, 6,368 radar images remained. These samples were randomly shuffled using a fixed random seed to ensure reproducibility and then split into training, validation, and test sets containing 3,200, 1,248, and 1,920 samples, respectively. All splits were chosen to be exact multiples of the batch size (32) to ensure consistent batching during training and evaluation. The training set was used for model fitting, the validation set for hyperparameter tuning and early stopping, and the test set exclusively for performance assessment. This preprocessed radar dataset provides a rigorous benchmark for assessing the ability of the proposed models to reproduce fine-scale dry/wet precipitation occurrence and storm morphology in real-world conditions.

3. Methods

In this study, we present a deep learning-based framework that utilizes a convolutional encoder-decoder and WGAN to correct dry/wet classification bias and improve the representation of precipitation occurrence in downscaled fields. The methodology focuses on three main elements: (1) network architecture, which defines the core structure of the convolutional encoder-decoder and the WGAN, (2) training strategies incorporating different conditioning approaches, which guide the learning process to improve the representation of storm features, and (3) model evaluation using robust and interpretable performance metrics.

3.1. Convolutional encoder–decoder

We implement a convolutional encoder–decoder neural network to perform precipitation occurrence downscaling, with the objective of reconstructing high-resolution 60×60 dry/wet masks from coarse 6×6 low-resolution inputs. The network is designed to capture the spatial organization of precipitation and accurately delineate dry and wet regions. The model follows a standard encoder–decoder structure commonly used in image-to-image learning tasks (Badrinarayanan et al., 2017; Meghani et al., 2023; Singh and Goyal, 2023; Abdelmoaty et al., 2025). The encoder progressively extracts feature representations through convolutional layers with ReLU activation functions, capturing both local and large-scale spatial dependencies. Each convolutional layer is followed by a max-pooling operation that reduces the spatial resolution of the feature maps while retaining essential structural information, enabling the learning of hierarchical representations. The decoder reconstructs high-resolution binary precipitation occurrence fields using transposed convolutional layers that progressively increase spatial resolution. Feature reconstruction relies on the hierarchical representations learned by the encoder. The resulting network produces high-resolution dry/wet precipitation occurrence fields that capture the spatial organization of precipitation and accurately delineate dry and wet regions. To maintain architectural consistency with the WGAN-based generative framework, a Gaussian noise field is concatenated with the low-resolution input and provided to the encoder–decoder, although under binary cross-entropy training, the model behaves largely deterministically. A schematic of the network architecture is shown in Figure 2a.

3.2. Conditional Wasserstein GAN (WGAN)

While convolutional encoder–decoder networks can effectively capture the large-scale spatial organization of precipitation occurrence, training with deterministic loss functions often leads to overly smooth or repetitive binary patterns and limited representation of fine-scale spatial variability (Kalantar et al., 2021; Choubey et al., 2024; Rampal et al., 2024; Abdelmoaty et al., 2025; Papalexiou and Mamalakis, 2025). To address these limitations and improve the realism of downscaled dry/wet precipitation fields, we extend the encoder–decoder framework using a conditional WGAN. The WGAN consists of two components: (i) a generator (convolutional encoder–decoder) that performs precipitation occurrence downscaling and (ii) a critic network that evaluates the realism of the generated high-resolution binary outputs (Goodfellow, 2016). Unlike the original GAN formulation, which relies on an adversarial objective

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

implemented via binary cross-entropy loss, the WGAN optimizes the Wasserstein distance, yielding improved training stability and more informative gradient signals (Arjovsky et al., 2017; Gulrajani et al., 2017; Creswell et al., 2018; Xu et al., 2022; Han and Guan, 2023). The inclusion of a gradient penalty (WGAN-GP) further stabilizes training and mitigates mode-collapse behavior (Gulrajani et al., 2017). The WGAN is formulated as a conditional generative model to ensure consistency between the generated high-resolution dry/wet occurrence fields and the corresponding low-resolution precipitation input. Conditioning is implemented at both components (generator and critic) of the WGAN. In the generator, the low-resolution precipitation field is provided explicitly as input and propagated through the encoder–decoder network. In the critic, the low-resolution field is concatenated with intermediate feature representations derived from the high-resolution binary field, allowing the critic to assess whether the generated dry/wet patterns are statistically consistent with the coarse-scale structure of the input (Leinonen et al., 2020; Glawion et al., 2023).

As illustrated in Figure 1b, this conditioning is implemented at the feature-representation level rather than through strict pixel-wise alignment. Although these intermediate feature maps do not preserve exact spatial correspondence with either the low- or high-resolution grids, this design choice is intentional (Abdelmoaty et al., 2025; Papalexiou and Mamalakis, 2025). The aim is to enforce structural and statistical consistency across spatial scales rather than pixel-level correspondence, thereby allowing the critic to focus on large-scale coherence and the realistic spatial organization of dry/wet patterns. The generator is initialized using the pretrained encoder–decoder network (Section 3.1) and subsequently refined through adversarial training to minimize the Wasserstein distance between real and generated binary precipitation occurrence fields. This adversarial training enables the generator to learn complex spatial structures and sharper dry/wet boundaries than achievable under deterministic optimization alone. For the remainder of the study, this adversarially trained generator is referred to as the WGAN generator.

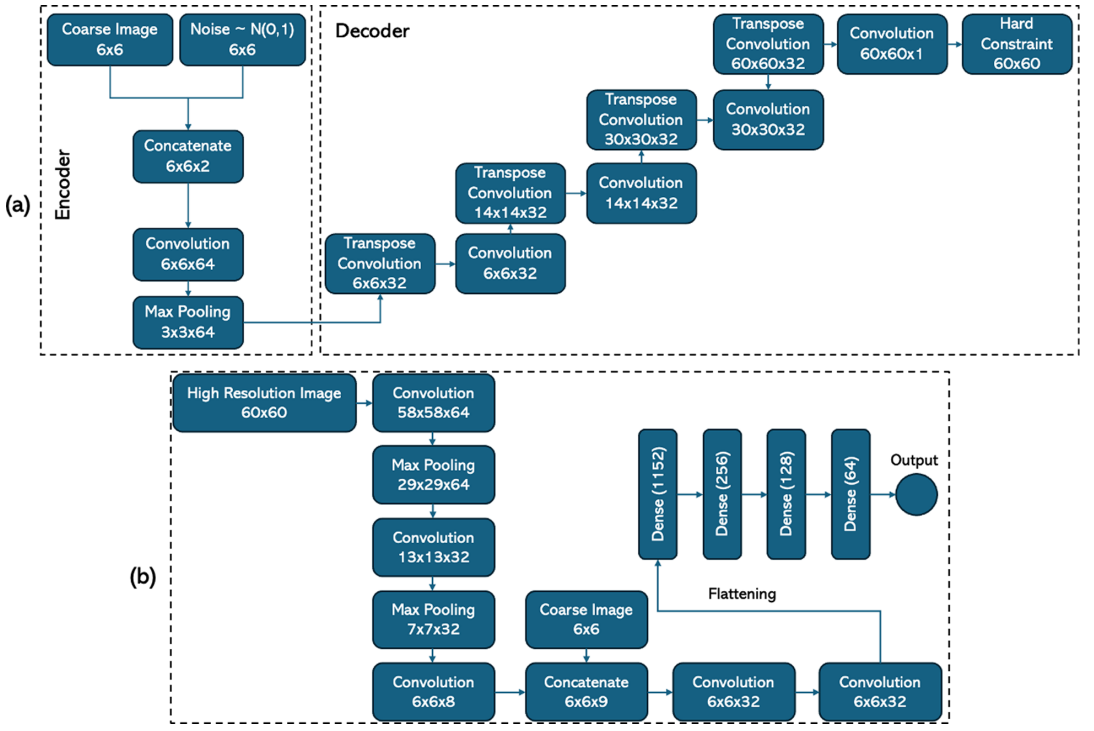


Figure 1. *The architectures of (a) convolutional encoder–decoder (Generator) and (b) critic used in this study as components of WGAN.*

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

The encoder–decoder generator and critic contain 121,633 and 382,729 trainable parameters, respectively (Figure 1).

3.3. Training strategy

Models were trained using the Adam optimizer (Kingma and Ba, 2014) with a batch size of 32. Training was performed independently for the synthetic CoSMoS dataset and the MRMS radar dataset. For each dataset, samples were shuffled with a fixed random seed and divided into training, validation, and test sets. The validation set was used for hyperparameter selection, while the test set was reserved exclusively for performance evaluation.

To introduce stochastic variability in the generated outputs, the generator receives a stochastic noise input during both training and inference. Specifically, a noise field defined on the low-resolution grid is sampled from a standard normal distribution and concatenated channel-wise with the low-resolution precipitation input. Although the noise is sampled independently at each grid cell, the convolutional architecture of the generator transforms this input into spatially coherent variability, avoiding unphysical pixel-scale randomness and promoting realistic spatial structure. This stochastic input provides the primary source of variability in the generated dry/wet occurrence fields. When trained using a binary cross-entropy loss alone, the encoder–decoder network tends to suppress the influence of the noise input and converge toward a largely deterministic solution. In contrast, adversarial training within the WGAN framework encourages the generator to utilize the noise input to produce diverse yet realistic dry/wet spatial configurations. The critic does not generate variability itself, but instead constrains how stochastic variability is expressed by penalizing unrealistic spatial organization. As a result, the WGAN produces multiple plausible high-resolution dry/wet realizations conditioned on the same low-resolution input. For evaluation purposes, a fixed random seed is employed during inference to ensure reproducibility and consistent comparison across experiments. These models are trained and evaluated independently on the synthetic and MRMS datasets, and no transfer learning or cross-dataset generalization is assumed in this study.

To examine how different forms of low-resolution information influence precipitation occurrence reconstruction, three conditioning strategies were evaluated:

a) Binary input: The low-resolution input is provided as a binary dry/wet mask.
b) Intensity input: The low-resolution input consists of continuous precipitation intensity values.
c) Constrained input: In addition to intensity input, a hard dry constraint is imposed during training, which enforces that high-resolution pixels corresponding to dry low-resolution pixels remain dry.

For each conditioning strategy and model type (the convolutional encoder–decoder and the WGAN), models were trained using five independent training runs with different random initializations (seeds = [8889, 42, 100, 1025, 61]) to assess the robustness of the results with respect to stochastic training variability. This procedure is not intended as formal uncertainty quantification, but rather to evaluate the stability of model behavior across training realizations. The model output layer produces pixel-wise probabilities of wet occurrence. For evaluation, these probabilities were converted to binary dry/wet classifications using a threshold of 0.5, where values below 0.5 were classified as dry and values equal to or greater than 0.5 as wet.

3.3.1. Convolutional encoder–decoder network training

The convolutional encoder–decoder network was first trained independently as a binary precipitation occurrence downscaling model, with the objective of classifying each high-resolution pixel as dry or wet. Training was conducted for up to 500 epochs with early stopping based on validation loss, using a patience of 30 epochs to prevent overfitting. The network was optimized using the binary cross-entropy (BCE) loss, defined as

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(\widehat{y}_i) + (1-y_i)\, log(1-\widehat{y}_i)] \quad \text{Eq (1)}$$

where N is the number of pixels in a batch, $y_i \in \{0,1\}$ is the true dry/wet label derived from precipitation values using a threshold of precipitation >0, and $\widehat{y}_i \in [0,1]$ is the predicted probability of a wet pixel. The final binary prediction is obtained by applying a threshold of 0.5 to $\widehat{y}_i$. This stage produces a deterministic encoder–decoder model trained using the binary cross-entropy loss for pixel-wise dry/wet classification.

3.3.2. Conditional WGAN training

To improve the realism and spatial variability of the reconstructed dry/wet fields, the pretrained encoder–decoder network was subsequently used to initialize the generator in a conditional Wasserstein GAN with gradient penalty (WGAN-GP) framework. Under this conditional WGAN-GP formulation, the critic and generator are optimized using the following objectives. The critic loss is defined as

$$\mathcal{L}_{\mathrm{C}} = \mathbb{E}_{\tilde{x}\sim P_{\mathrm{g}}}[C(\tilde{x}, x_{\mathrm{LR}})] - \mathbb{E}_{x\sim P_{\mathrm{r}}}[C(x, x_{\mathrm{LR}})] + \lambda \mathbb{E}_{\widehat{x}}\left[\left(\|\nabla_{\widehat{x}} C(\widehat{x}, x_{\mathrm{LR}})\|_2 - 1\right)^2\right] \quad \text{Eq (2)}$$

where

$x \sim P_{\mathrm{r}}$ denotes real high-resolution binary precipitation fields drawn from the real data distribution P_r,
$\tilde{x} = G(z, x_{\mathrm{LR}})$ denotes generated high-resolution samples produced by the generator G,
$\widehat{x} = \varepsilon x + (1-\varepsilon)\tilde{x}$ denotes samples interpolated between real and generated data, with $\varepsilon \sim \mathrm{Uniform}(0,1)$,
x_{LR} denotes the corresponding low-resolution precipitation input used as conditioning information,
P_{g} denotes the distribution of generated samples,

$C(\cdot)$ *denotes the critic that assigns a scalar score to each sample conditioned on* x_{LR}

λ is the gradient penalty coefficient enforcing the Lipschitz constraint.

The generator loss is given by

$$\mathcal{L}_{\mathrm{G}} = -\mathbb{E}_{\tilde{x}\sim P_{\mathrm{g}}}[C(\tilde{x}, x_{\mathrm{LR}})] \quad \text{Eq (3)}$$

The critic was updated three times per generator update, following established practice (Gulrajani et al., 2017; Abdelmoaty et al., 2025; Papalexiou and Mamalakis, 2025). The Adam optimizer was used for both the generator and the critic, with momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.9$. The gradient penalty coefficient was fixed at $\lambda = 10$, following Gulrajani et al. (2017). Models were trained up to 50 epochs and learning rates were selected empirically based on validation performance and training stability, with values of 1×10^{-4} for the critic and 2×10^{-4} for the generator.

4. Evaluation for dry/wet storm prediction

Evaluating the performance of deep learning models for dry/wet classification requires robust and interpretable metrics that can effectively capture both classification accuracy and spatial consistency. In this work, we focus on two key metrics: probability of zero (P_0) and spatial autocorrelation, both of which are particularly relevant in precipitation downscaling (Papalexiou and Mamalakis, 2025).

4.1. Probability of zero

The probability of zero **(P_0)** is used to quantify the proportion of dry pixels (zero precipitation) in a given image. This metric is especially important in assessing the model's ability to reproduce the percentage of dry/wet regions correctly (Baño-Medina et al., 2020; Papalexiou, 2022; Papalexiou and Mamalakis, 2025). Mathematically, P_0 is defined as the fraction of pixels with zero precipitation (dry) within an image, expressed as:

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

$$P_0 = \frac{\sum_{i=1}^{N} 1(X_i = 0)}{N} \quad \text{Eq (4)}$$

Here, N is the total number of pixels in the image, X_i is the binary value (0 for dry,1 for wet) at pixel *i.*

By comparing P_0 values between the ground truth (e.g., radar observations) and model predictions (e.g., from a convolutional encoder-decoder or WGAN), we can assess whether the model tends to overestimate wet areas or fails to capture dry zones. This is particularly critical in storm boundary detection. In addition to the mean P_0, we also evaluate the mean bias and mean root mean squared error (RMSE) between predicted and observed P_0 values to assess systematic deviations in dryness representation.

4.2. Spatial autocorrelation

To complement the evaluation of model performance in predicting dry and wet regions, we also use lagged spatial autocorrelation to assess how well these models preserve the spatial structure of storm fields (Papalexiou and Mamalakis, 2025). This metric computes the Pearson autocorrelation coefficient between the original binary storm field and its shifted version in both the horizontal (longitudinal) and vertical (latitudinal) directions, for spatial lags up to six pixels. The autocorrelation is expressed as:

$$r = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Z_i - \overline{Z})}{\sqrt{\sum_{i=1}^{N} (X_i - \overline{X})^2 \sum_{i=1}^{N} (Z_i - \overline{Z})^2}} \quad \text{Eq (5)}$$

Here, N denotes the total number of pixels in the field. X_i and Z_i represent the binary values (0 for dry, 1 for wet) at the *i*th pixel in the original and spatially shifted (lagged) fields, respectively. $\overline{X}$ and $\overline{Z}$ are the mean values of the original and shifted fields across all pixels. Higher lagged spatial autocorrelations indicate stronger structural dependence as a function of spatial lags. Together, the probability of zero and lagged spatial autocorrelation metrics provide a comprehensive evaluation framework for dry–wet storm prediction models, combining pixel-level accuracy with structural realism across multiple spatial scales.

5. Results

5.1. Synthetic storm data

We evaluated the performance of the convolutional encoder–decoder and WGAN models on synthetic data using three conditioning settings: Binary input, intensity input, and constrained. For the convolutional encoder–decoder model, training and validation loss curves demonstrate distinct behaviors across the settings (Figure 2a). When training using a binary input setting, the training loss decreases steadily, but the validation loss begins to diverge after approximately 20 epochs, indicating signs of overfitting. In contrast, the intensity input setting yields better generalization, with flatter validation loss curves and reaching lower loss values, suggesting that the added precipitation intensity information helps the model learn more spatial context (Figure 2a). The most stable and lowest loss values are observed in the constrained input setting, where the inclusion of dry-region constraint leads to faster convergence and improved generalization, highlighting the benefit of incorporating physically meaningful constraints for bias correction.

For the WGAN model, the generator and critic losses exhibit the oscillatory behavior characteristic of adversarial training (Figure 2b). When conditioned on binary inputs, loss fluctuations remain within a moderate range, indicating relatively stable generator–critic dynamics. In contrast, intensity-only conditioning leads to pronounced instability, with large-amplitude oscillations in the generator loss, reflecting difficulty in maintaining adversarial balance when relying solely on intensity information. Notably, incorporating an explicit dry-region constraint substantially stabilizes training, reducing both generator and critic loss variability after the initial training phase. This behavior suggests that dry constraints act as an effective regularizer, anchoring the adversarial learning process toward physically plausible dry/wet spatial configurations and improving training stability. While WGAN-GP loss magnitudes are not

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

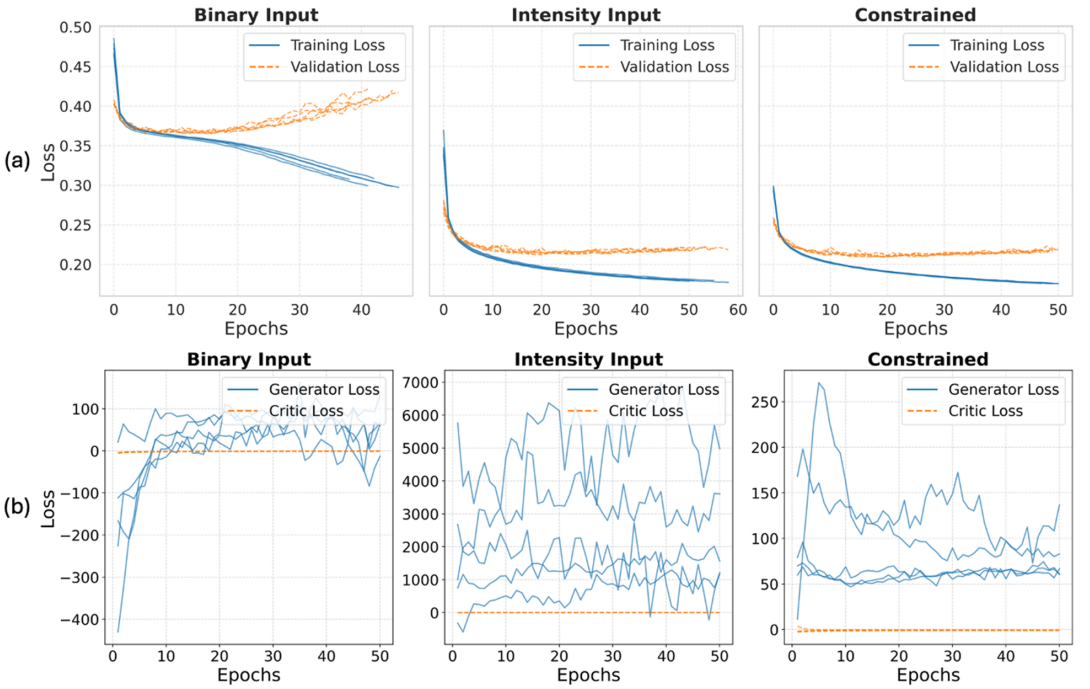


Figure 2. *Training and validation loss curves for convolutional encoder-decoder (top row, [a]) and WGAN (bottom row, [b]) models under three conditioning settings: binary input, intensity input, and constrained.*

absolute indicators of convergence, their relative behavior across conditioning strategies provides insight into adversarial balance and robustness. Training stability was further assessed using complementary diagnostics, including monitoring the gradient penalty term and visually inspecting the spatial coherence and diversity of generated dry/wet fields throughout training. The model did not collapse to trivial solutions (e.g., uniformly dry or wet fields), and the Constrained input configuration consistently preserved realistic dry/wet spatial structures.

The predictions by the convolutional encoder-decoder are provided in Figure 3. We observe that predictions based on the binary input setting tend to produce coarse and overly smoothed wet areas, with poor alignment to fine-scale features in the high-resolution ground truth. The model struggles to reconstruct intricate structures like spirals or narrow wet regions due to limited spatial information in the binary input setting.

The intensity input setting significantly improves the spatial accuracy and wet-region detection by the model. Fine-scale structures begin to emerge, although some noise and discontinuities persist. Similar results are produced when trained with a dry constraint. The WGAN model also exhibits similar behavior regarding the three different training settings, but the generated fields show sharper dry/wet boundaries compared to the smoother outputs produced by the convolutional encoder–decoder. The wet/dry fields show improved sharpness and are much more realistic (Figure 4). Using the binary input setting, WGAN predictions lack spatial detail, often failing to recover sub-grid storm patterns.

The intensity input setting improves the resolution and extent of wet regions, but the output contains fragmented structures and some overprediction in dry zones. The constrained input setting yields similar results to the Intensity input setting, but is more physically consistent (Figure 4). The constraint makes wet regions more continuous, and dry regions better preserved. This reflects the regularizing effect of the dry constraint, which helps the generator maintain realistic spatial distributions. Overall, these visual results confirm the effectiveness of using intensity and dry-region information as conditioning inputs, with the

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

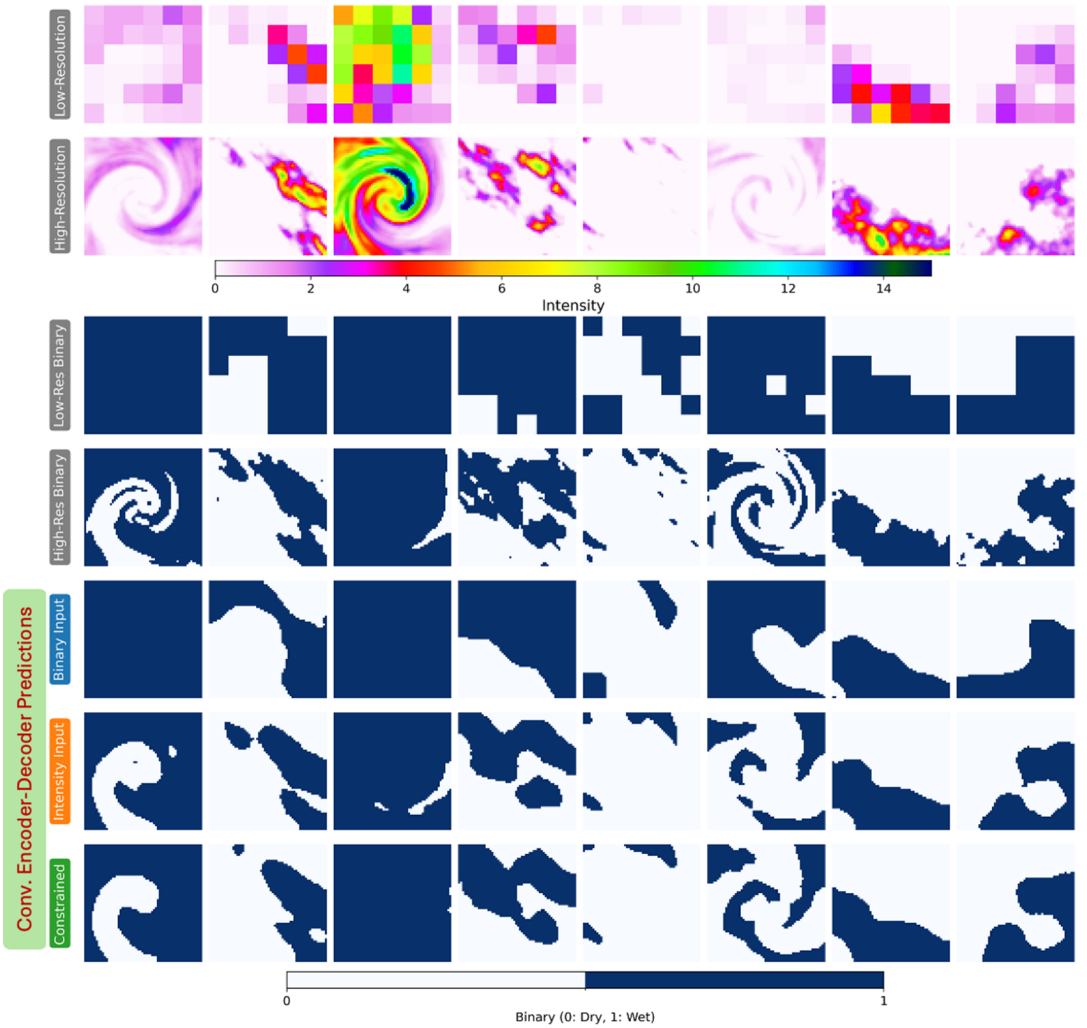


Figure 3. *Spatial representation of storm structures and dry/wet classification from synthetic test data. The first and second row displays the low-resolution and high-resolution storm fields whereas the third and fourth rows present their binary equivalents, respectively. The fifth to seventh rows illustrate the predicted high-resolution storm fields generated by a convolutional encoder-decoder model trained under the three distinct settings. The results presented here are obtained after classifying the predicted probabilities into dry (0) and wet (1) using a threshold of 0.5.*

constrained input setting consistently yielding better predictions in terms of both structure and realism. This pattern holds across both model architectures, although the convolutional encoder-decoder tends to produce smoother and more interpretable outputs, while WGAN results exhibit more pixel-level noise but sometimes sharper boundaries.

The performance of the trained models was evaluated in terms of their ability to correctly classify dry regions, quantified by the predicted probability of zero (P_0). With the binary input setting, the convolutional encoder–decoder model displays the most pronounced bias (Bias: −8.37%, RMSE: 13.31%), with consistent underprediction of P_0 (i.e., overprediction of wet areas). This behavior stems from the limited information in the binary input, which lacks precipitation intensity information and makes the model prone to misclassifying uncertain regions as wet. A slight reversal occurs only in very dry scenes

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

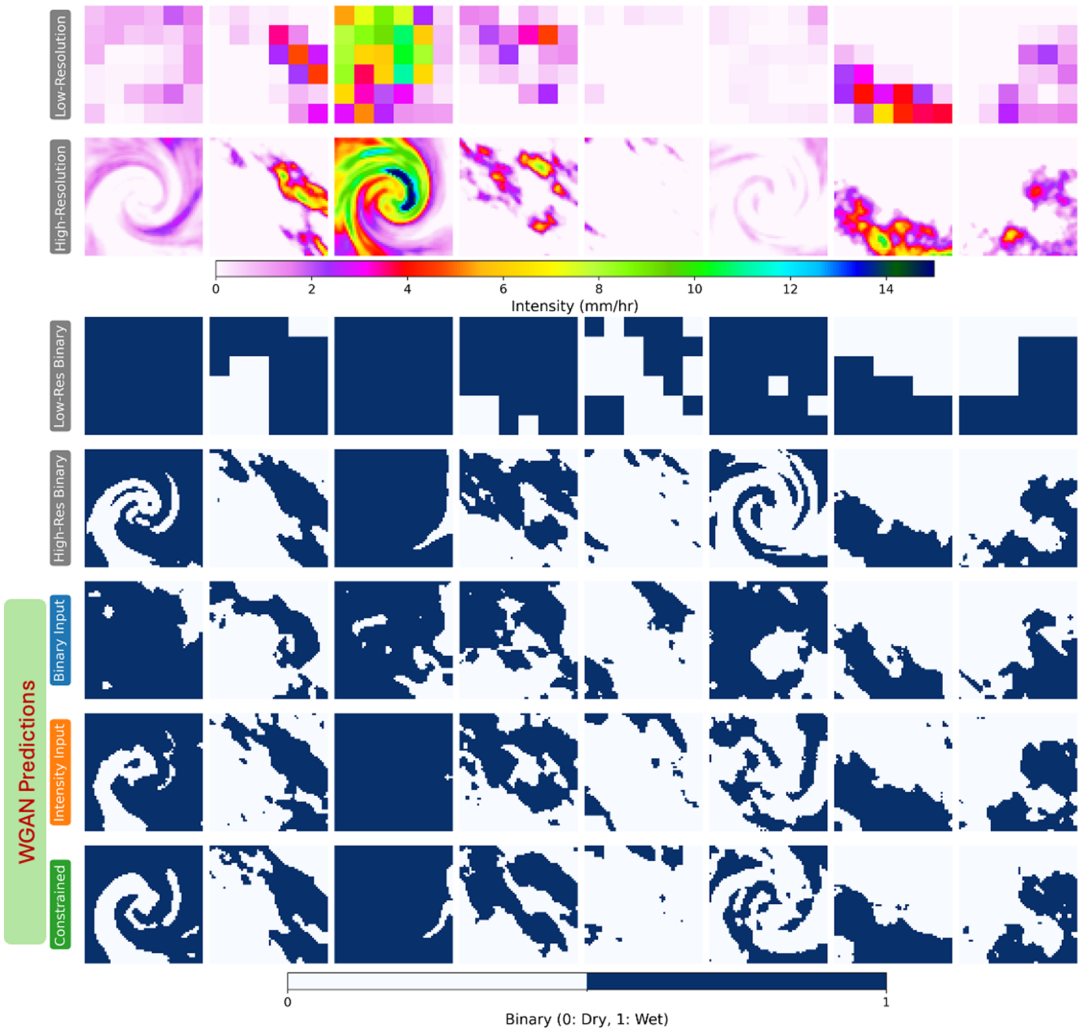


Figure 4. *Spatial representation of storm structures and dry/wet classification from synthetic test data. The first and second rows display the low-resolution and high-resolution storm fields, whereas the third and fourth rows present their binary equivalents, respectively. The fifth to seventh rows illustrate the predicted high-resolution binary storm fields generated by the WGAN model trained under the three distinct settings. (The results presented here are obtained after classifying the predicted probabilities into dry (0) and wet (1) with a threshold of 0.5).*

($P_0 > 90\%$), where the model slightly overpredicts dry conditions. With the intensity input setting, performance improves substantially (Bias: 0.56%, RMSE: 2.99%). Predictions are more tightly aligned with the 1:1 line, but a consistent trend is observed: the model underpredicts dry conditions in wetter scenes ($P_0 < 50\%$), while in drier scenes ($P_0 > 50\%$), it slightly overpredicts dry conditions, as reflected by the shift of points above the diagonal (Figure 5a). This behavior indicates that while Intensity input setting helps anchor storm structure, the model still tends to slightly exaggerate coarse-resolution conditions—dry scenes are predicted drier and wet scenes are predicted wetter. When adding the dry constraint, the model predictions are similar to the ones with the intensity input (Bias: 0.66%, RMSE: 3.06%), but with a slightly more pronounced overprediction of dryness in the mid to high P_0 range. This shift likely results from the explicit enforcement of dry-region preservation, which encourages

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

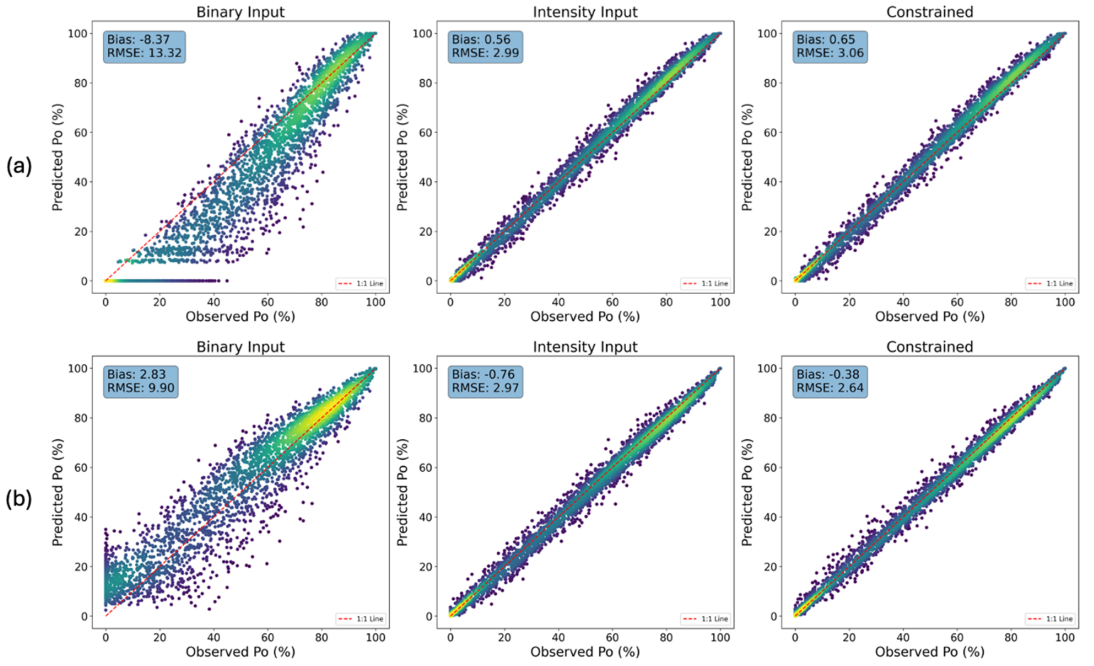


Figure 5. *Comprehensive evaluation of model performance in predicting the Probability of Zero (P_0), defined as the percentage of dry pixels (i.e., pixels with zero precipitation) in each high-resolution output. The evaluation was conducted across five independently trained model runs for each of the three conditioning settings: Binary, Intensity, and Dry Constraint. (a) Convolutional encoder–decoder model. (b) WGAN model. Each scatter plot compares the predicted P_0 against the observed P_0 across the entire test set, providing insight into model calibration, bias, and consistency in preserving dry regions.*

extrapolation of dry areas and reduces false positives in wet prediction. The result is a modest dry bias that enhances physical consistency, particularly in scenes with large-scale dry conditions.

For WGAN, predictions based on the binary input setting show reduced bias compared to the convolutional encoder-decoder (Bias: 2.83%, RMSE: 9.90%), but exhibit greater spread across samples. Using intensity input setting improves alignment with the 1:1 line (Bias: −0.76%, RMSE: 2.97%) and reduces spread, but the model still underpredicts dry conditions. The constrained input setting provides the most balanced performance for WGAN (Bias: −0.38%, RMSE: 2.64%), with predictions closely following the 1:1 line across the full range. The dry constraint appears to shift the model's tendency by reducing a negative bias from −0.76% (in the Intensity input setting) to −0.38%, suggesting that the constraint helps correct underestimation of dry conditions and promotes more accurate dry-region reconstruction (Figure 5b).

Next, we evaluate the spatial structure of predicted high-resolution dry/wet fields using lagged spatial autocorrelation across pixel lags from 1 to 6 in both horizontal and vertical directions. For the convolutional encoder-decoder model (Supplementary Figure S2), the predicted lagged spatial autocorrelations using binary input setting are consistently higher than the ground truth across all lags and directions, indicating over-smoothing and exaggerated spatial persistence. This pattern is consistent in both horizontal and vertical directions and is observed across all five training runs. As the spatial lag increases, the interquartile spread widens, particularly beyond lag 3, reflecting greater variability and instability in preserving long-range spatial structure. Conditioning on intensity input improves both alignment with the ground truth and model consistency. Although the autocorrelations still slightly exceed those of the ground truth, the interquartile range becomes narrower, indicating reduced variability and improved

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

generalization. The constrained convolutional encoder-decoder shows similar results to the Intensity input setting, showing improved structural stability and only marginal differences in lagged autocorrelation magnitude and spread across seeds. Both conditioning strategies help regularize spatial continuity while limiting excessive smoothing.

For the WGAN (Figure 6), the interquartile spread is generally narrower than in the convolutional encoder-decoder across all three training settings, indicating greater consistency across different seeds, even when using binary inputs. The WGAN trained in Intensity input setting most consistently follow the autocorrelation trend, with medians closely aligned to ground truth across all lags and directions. This suggests that intensity conditioning helps preserve both short- and long-range spatial structure. Predictions from the Binary input setting show wider variability across seeds and slightly lower correlations, particularly at higher lag (beyond lag 3), indicating limited spatial coherence when only binary dry/wet information is provided. The Constrained input setting, while maintaining high fidelity as of the intensity input setting and enhanced spatial realism, exhibits somewhat more inter-seed variability, possibly due to stricter regularization suppressing fine-scale features. Notably, in all three cases, WGAN predictions do not over-smooth the fields, as reflected by their ability to retain spatial correlation beyond short lags.

Overall, the integration of intensity and dry-region constraint substantially improves the performance of both convolutional encoder-decoder and WGAN models in predicting high-resolution dry/wet fields. While intensity input settings enhance the spatial localization of wet regions, the dry constraint setting provides an additional regularizing effect, reinforcing physical consistency and enhanced spatial realism. The convolutional encoder–decoder exhibits similar or slightly better performance to the WGAN regarding the marginal statistics of the dry/wet fields (see probability of zero in Figure 5 and more statistics in the Supplementary Figures S5 and S6), but they tend to produce overly smooth outputs with inflated spatial coherence. In contrast, WGAN fields, though sometimes noisier, better align with targeted spatial structures and maintain stronger consistency across different initialization seeds.

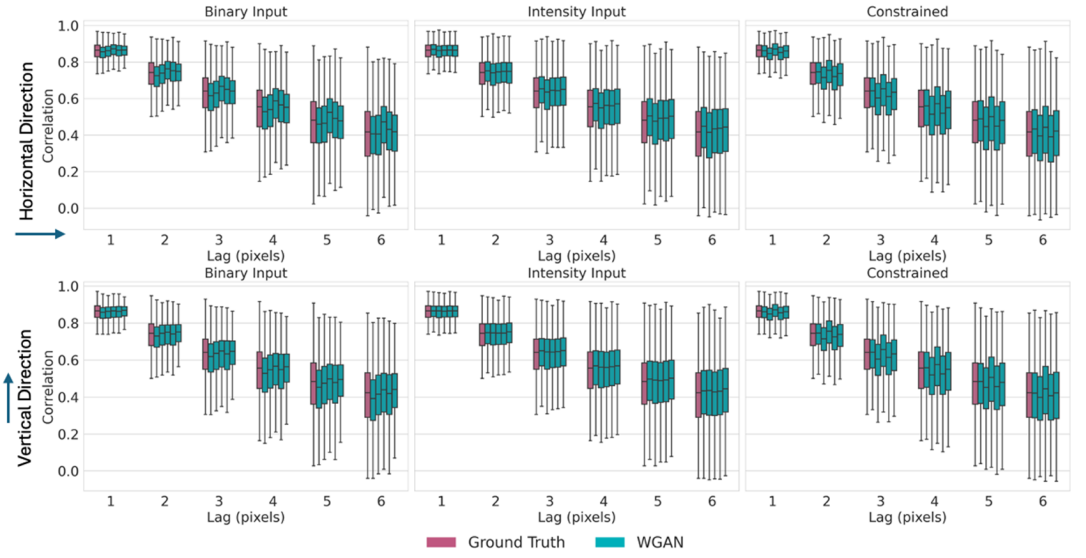


Figure 6. *Lagged spatial autocorrelation of binary dry/wet precipitation fields for increasing spatial lags (1–6 pixels) in the horizontal (top row) and vertical (bottom row) directions. Results are shown for the synthetic test dataset under the three conditioning strategies (binary input, intensity input, and constrained). Boxplots summarize the distribution of lagged spatial autocorrelation values across five independent training runs of the WGAN models with different random initializations (teal), compared against the ground truth fields (purple). Model outputs were converted to binary dry/wet maps using a probability threshold of 0.5 prior to analysis.*

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

5.2. Radar-based storm data

The training of both convolutional encoder–decoder and WGAN models on radar-based precipitation data reveals several important patterns, many of which align with what was observed during training on synthetic storm data (Supplementary Figure S3). For the convolutional encoder–decoder model, all three training settings (binary input, intensity input, and constrained input) demonstrate similar behavior to the training using synthetic data (Figure 2a) with steady reductions in training loss, while the validation loss reaches a plateau with the first 20 epochs (Supplementary Figure S3a). When using the dry constraint, the most favorable training is achieved, with both training and validation losses reaching consistently lower values than other settings and converging more smoothly, which reflects improved robustness and more physically consistent learning of dry/wet patterns. The WGAN also shows similar oscillatory training behavior to that observed on synthetic data (Supplementary Figure S3b). Generator loss exhibits large fluctuations under the Binary Input configuration, but these fluctuations become significantly narrower under the Intensity input and Constrained input settings. The difference in fluctuation range between synthetic and radar-based training may partly reflect differences in the scale of the conditioning inputs. In the synthetic experiments, the aggregated low-resolution intensity field can reach maximum values of approximately 1400 when the 60×60 synthetic precipitation fields are accumulated into 6×6 grids. In contrast, radar-based inputs correspond to hourly precipitation rates with maximum values on the order of several mm h^{-1}, resulting in a substantially smaller input magnitude. The difference in scale likely produces smoother gradients during optimization and contributes to differences in the observed fluctuation ranges (Figure S3b). Additionally, while the critic loss appears relatively flat across configurations, its magnitude alone does not provide a direct indication of training stability in WGAN-GP. The generator loss curves show different fluctuation patterns across conditioning settings, which may reflect differences in adversarial training dynamics under the different conditioning strategies.

Prediction results on radar-based precipitation data, shown in Figures 7 and 8, demonstrate the significant differences in dry/wet classification accuracy across conditioning strategies and model architectures. As seen previously with synthetic data, using binary input setting yields the poorest fidelity in both convolutional encoder-decoder and WGAN models. The predictions under this setting are overly coarse and tend to underestimate wet regions, largely replicating the blocky structure of the low-resolution binary input rather than the nuanced patterns observed in the high-resolution target. This issue is especially apparent for smaller storm cells and isolated wet areas, which are either entirely missed or exaggerated in extent. The convolutional encoder-decoder predictions based on binary input setting produce relatively large contiguous wet regions that do not align well with the ground truth, while WGAN predictions are more textured but still misrepresent the exact shape and spatial extent of wet zones (Figures 7, 8).

When using intensity as an input, both models exhibit some improvements. The added gradient and structure in the low-resolution intensity input may provide additional spatial information, which could help the networks to better localize precipitation regions. The convolutional encoder-decoder predictions become more confined, with the focus on the spatial extent of comparatively large contiguous wet zones matching the high-resolution target (Figure 7). The Intensity input setting improves spatial coherence, enabling more realistic storm patterns, but often overestimates wet area extent, especially in WGAN predictions. Notably, the constrained input setting consistently yields the most faithful reconstructions of high-resolution binary fields for both the convolutional encoder-decoder and the WGAN. It effectively suppresses false positives in dry regions and recovers sharper wet-dry boundaries, demonstrating that dry-region guidance significantly enhances physical realism in the generated outputs (Figure 8). Nevertheless, both models clearly benefit from the richer input, capturing small storm patches and distinguishing dry regions more accurately than in the binary input setting. The most accurate outputs are again achieved by the WGAN when it is constrained.

Regarding the overall predicted probability of zero (P_0; see Figure 9), for the convolutional encoder–decoder, the binary input setting shows the largest spread from the 1:1 line, with a near-zero bias (0.02%) and an RMSE of 7.67% (Figure 9a). The spread is particularly pronounced at mid-to-high observed P_0

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

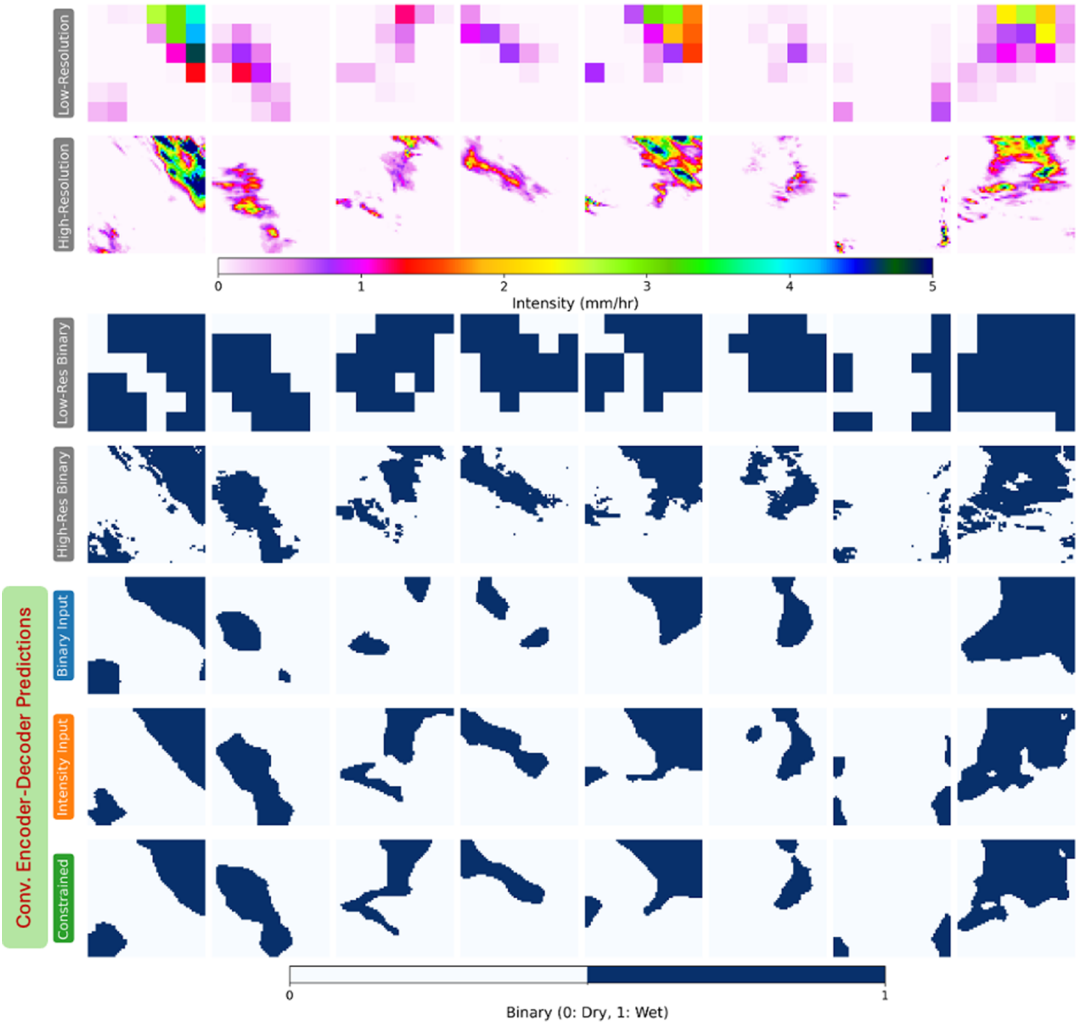


Figure 7. *Spatial representation of storm structures from the Radar dataset test sample and predictions from convolutional encoder–decoder models trained under the three conditioning settings. The first and second rows display the low-resolution and high-resolution precipitation fields, whereas the third and fourth rows show their corresponding binary fields. The fifth through seventh rows illustrate the predicted high-resolution fields generated under the binary, intensity, and constrained Input settings. The results presented here are obtained after converting predicted probabilities into dry (0) and wet (1) using a threshold of 0.5.*

values, indicating greater variability in predicted dry fractions under this input configuration. Introducing Intensity input improves the agreement with observations: RMSE drops to 2.80% and predictions align more closely with the 1:1 line, though a slight positive bias of 1.07% is observed. Under the constrained input setting, predictions remain closely clustered around the diagonal with reduced outliers, yielding an RMSE of 3.06% and a bias of 1.48%.

Across all input settings, the convolutional encoder–decoder tends to slightly overestimate dry region in cases with higher observed dryness ($P_0 > 50\%$), indicating a tendency to underpredict wet conditions in relatively dry fields. The WGAN exhibits similar overall behavior but with smaller errors (Figure 9b). Under binary input setting, the WGAN predictions exhibit a negative bias of –1.36% and an RMSE of

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

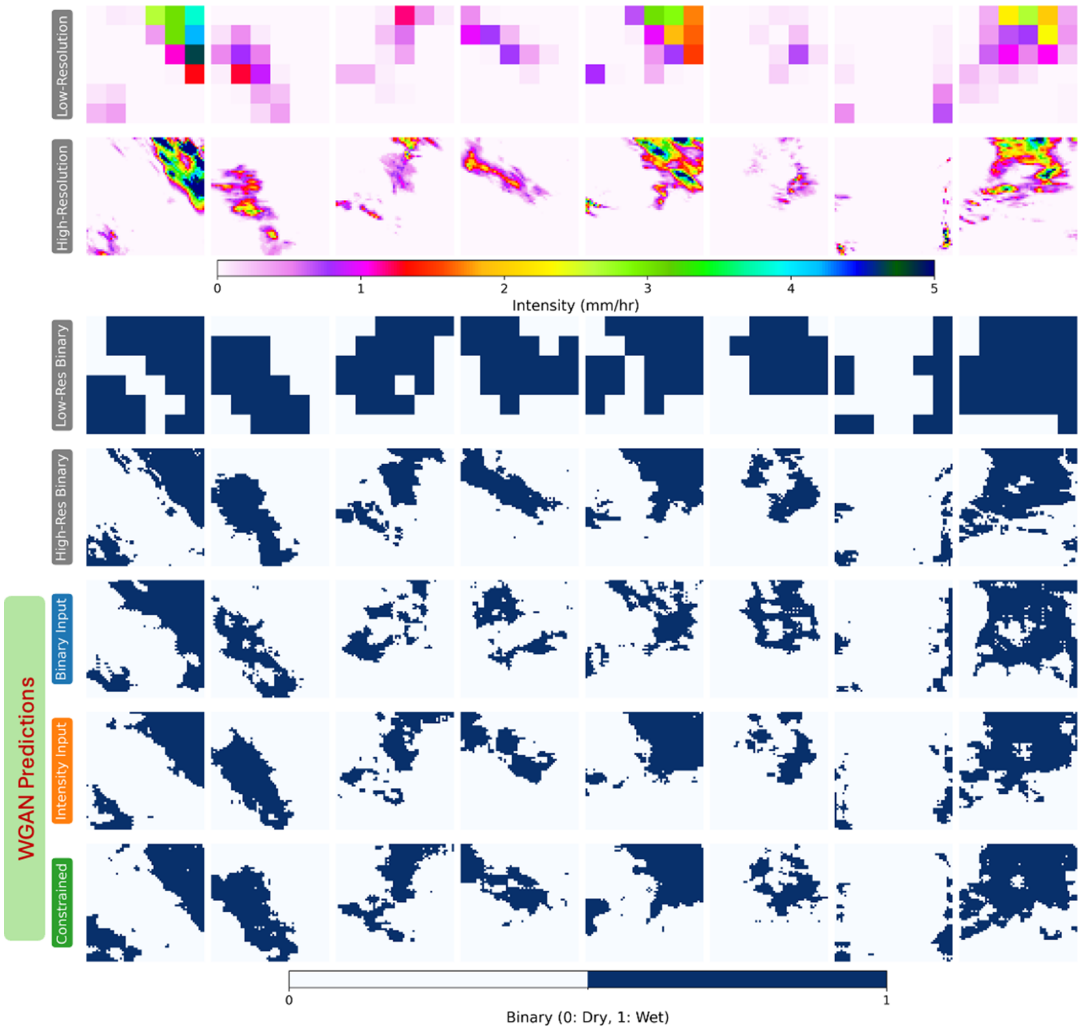


Figure 8. *Spatial representation of storm structures from the Radar dataset test sample and predictions from the WGAN trained under the three conditioning settings. The first and second rows display the low-resolution and high-resolution precipitation fields, whereas the third and fourth rows show their corresponding binary fields. The fifth to seventh rows illustrate the predicted high-resolution fields generated under the binary, intensity, and constrained input settings. The results presented here are obtained after converting predicted probabilities into dry (0) and wet (1) using a threshold of 0.5.*

6.13%, with a greater spread relative to the intensity-based setting. The performance improves significantly under intensity input setting, where RMSE decreases to 1.80% and the bias is reduced to –0.16%, with predictions closely aligned along the 1:1 line. The Constrained input setting achieves similar strong performance with an RMSE of 1.98% and a bias of –0.49%. While a slight underestimation of P_0 is observed for very dry cases ($P_0 > 50\%$), but overall the results indicate improved consistency in predicted dry-region fractions. In general, both models benefit from intensity-based inputs, while the addition of dry-region constraints provides further improvements in prediction consistency, particularly for the WGAN framework.

For the convolutional encoder–decoder, predicted fields consistently exhibit higher median correlations than the ground truth across all input settings and lag distances (Supplementary Figure S4). This is

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

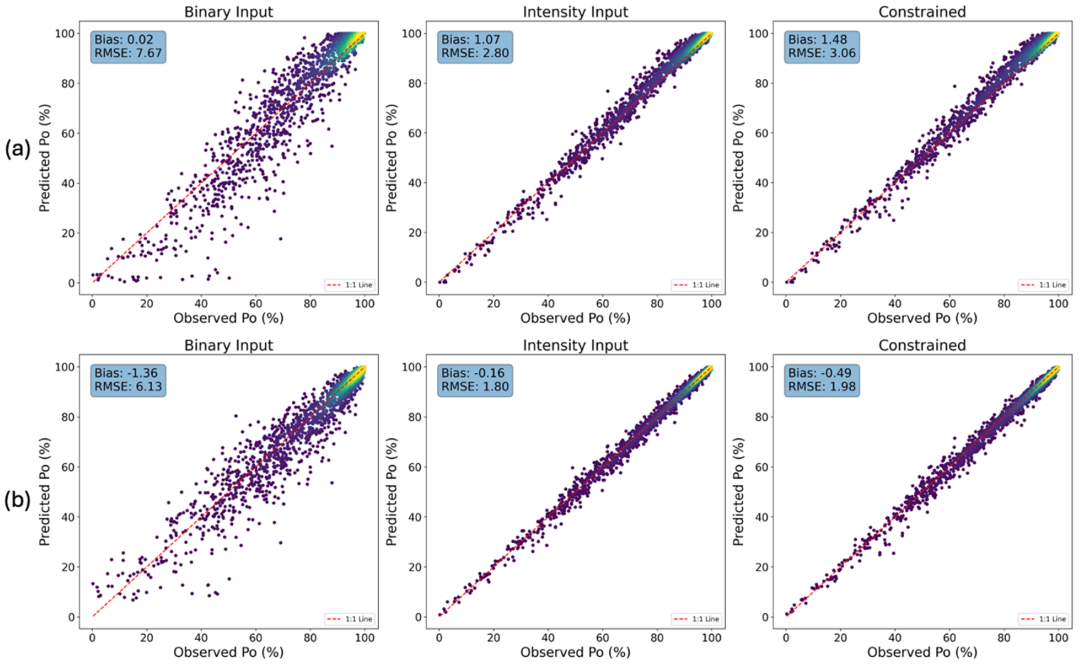


Figure 9. *Comprehensive evaluation of model performance in predicting the Probability of Zero (P_0) from a test sample of radar data. The evaluation was conducted across five independently trained models for each of the three conditioning settings: Binary, Intensity, and Dry Constraint. (a) Convolutional encoder–decoder model. (b) WGAN model.*

especially evident at mid-to-high lags, where the ground truth correlations naturally decay due to fine-scale variability, whereas convolutional encoder-decoder predictions maintain artificially elevated correlation. This behavior reflects the model's tendency to produce overly smooth outputs, which suppress local noise and enhance spatial continuity beyond what is present in the actual data. Among input types, the Intensity input setting and the Constrained input setting yield slightly more controlled correlation trends compared to binary input setting, but the overall over-smoothing pattern persists. In all three settings, the convolutional encoder–decoder shows elevated interquartile variability, suggesting greater uncertainty and instability in capturing long-range spatial structures. The WGAN exhibit reduced interquartile variability compared to the convolutional encoder–decoder, in both directions, suggesting more stable spatial pattern learning (Figure 10). Notably, the intensity input and constrained input setting yield autocorrelation whose medians closely track the ground truth, across different lags. The dry constraint regularizes spatial features effectively but introduces slightly higher seed-to-seed variability. These observations highlight that while the convolutional encoder–decoder tends to over-persist spatial patterns, WGAN better capture the true spatial extent and variability of dry/wet zones.

Similarly to the conclusions from the synthetic dataset, the constrained training leads to enhanced accuracy, physical realism, and structural alignment of dry/wet separations, especially for WGAN models. While predictions on radar data exhibit overall slightly lower performance due to real-world complexity, the relative gains introduced by using the intensity input setting and constrained input setting remain evident (see Supplementary Figures S7 and S8). Models trained with intensity input setting and constrained input setting not only achieve lower RMSE and reduced bias in terms of predicting the overall probability of dry conditions but also produce spatial patterns that better match the observed spatial autocorrelation structure. These results highlight the value of integrating a consistency constraint into data-driven downscaling frameworks to improve predictive performance, even in challenging real-world applications.

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

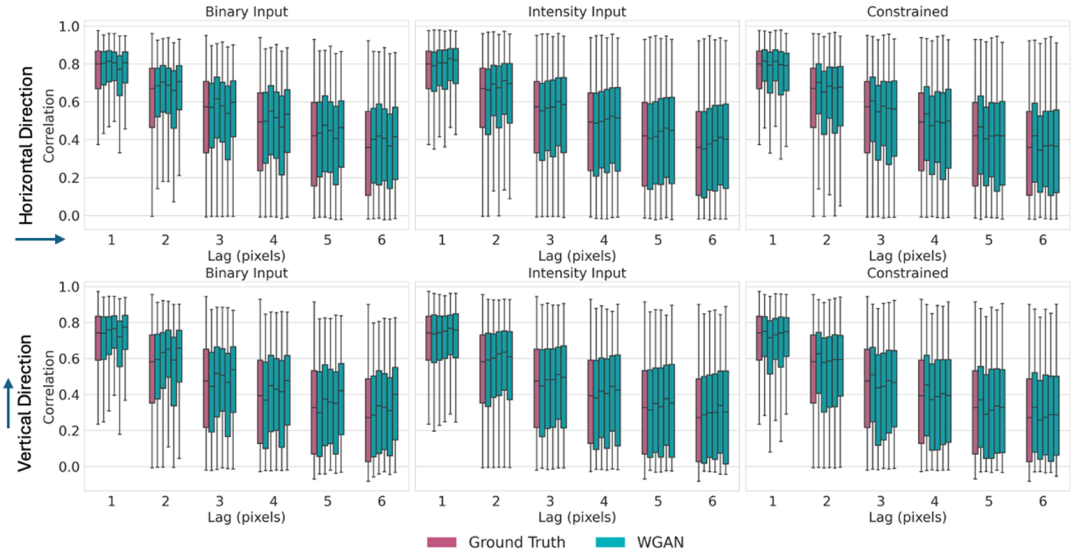


Figure 10. *Lagged spatial autocorrelation of binary dry/wet precipitation fields for increasing spatial lags (1–6 pixels) in the horizontal (top row) and vertical (bottom row) directions. Results are shown for the radar test dataset under the three conditioning strategies (binary input, intensity input, and constrained). Boxplots summarize the distribution of lagged spatial autocorrelation values across five independently trained WGAN with different random initializations (teal), compared against the ground truth fields (purple). Model outputs were converted to binary dry/wet maps using a probability threshold of 0.5 prior to analysis.*

6. Discussion

To demonstrate the value of explicitly predicting dry/wet classification fields, we applied the dry/wet mask generated by the WGAN trained under the constrained input setting to post-process high-resolution precipitation intensity fields produced by separately trained intensity prediction models. These intensity models use the same architectures and training datasets described earlier for the dry/wet experiments, but are trained to predict continuous precipitation intensity rather than binary occurrence. The convolutional encoder–decoder is optimized using mean squared error (MSE) loss, while the WGAN retains the adversarial training framework with intensity fields as targets. Results for synthetic data are presented in Figures 11, 12, while corresponding results for radar data are provided in Supplementary Figures S10–S11. Both the convolutional encoder-decoder (trained with mean squared error loss) and the WGAN (trained using the Wasserstein loss with gradient penalty) were used for storm intensity prediction; however, these models commonly exhibit drizzle bias, producing low, non-zero precipitation values in regions that should be entirely dry (Lazoglou et al., 2024). This behavior has been reported in previous studies of machine learning–based precipitation prediction and downscaling, where models trained with MSE or adversarial losses tend to underestimate the occurrence of truly dry events. To address this, we applied the binary dry/wet mask generated by the WGAN trained in the constrained input setting to the predicted intensity fields, effectively eliminating spurious low-intensity values and enforcing the spatial dryness constraint. This correction significantly improved the realism and accuracy of the downscaled precipitation fields. In the synthetic data use-case, the uncorrected WGAN outputs exhibited a pronounced drizzle bias, which is also reflected in the flat-line behavior in the probability of zero (P_0) plots, where predicted P_0 values remained systematically lower than the observed values. Applying the dry/wet mask corrected this behavior. Specifically, the raw downscaled precipitation fields showed a substantial negative bias *in* P_0 (−50.35%) and a high RMSE (58.45%), indicating a consistent overestimation of wet conditions (Figure 11a). After correction, the predicted P_0 values aligned closely with the 1:1 reference line, with the bias reduced to −0.54% and the

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

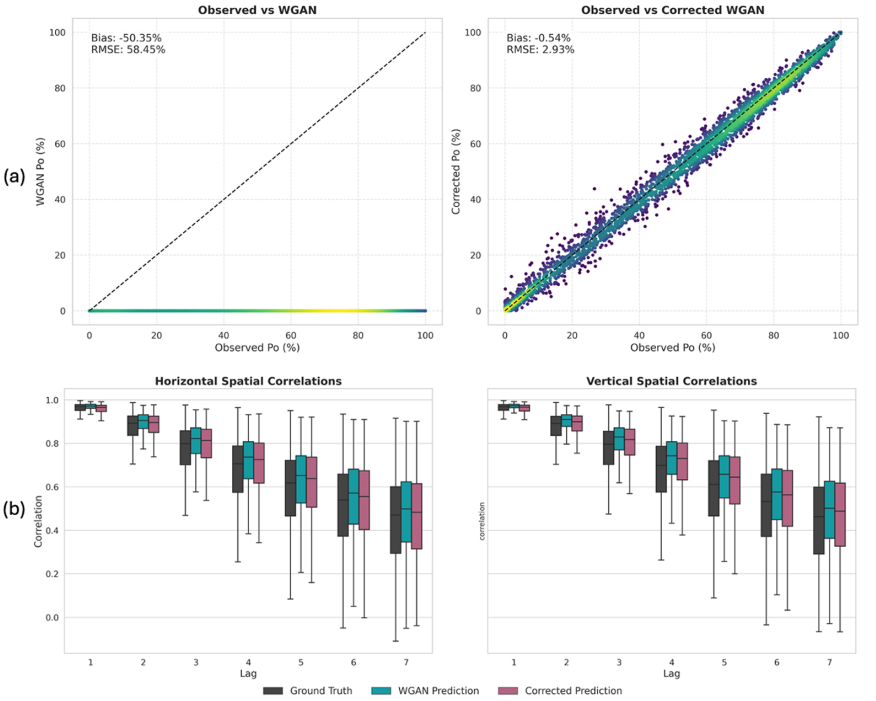


Figure 11. *Improved dry/wet prediction and spatial structure following correction. (a) Applying a predicted dry/wet mask to WGAN-generated precipitation fields substantially reduces bias and RMSE in dry-pixel prediction (P_0). (b) Corrected outputs exhibit lagged spatial autocorrelations that more closely track the ground truth in both horizontal and vertical directions, indicating improved spatial coherence.*

RMSE to 2.93%. This substantial improvement demonstrates that integrating explicit dry-region awareness through classification masks meaningfully enhances the model's ability to reproduce realistic dry/wet distributions in high-resolution downscaled precipitation fields.

The benefits of the dry/wet mask correction extend beyond pixel-wise accuracy to improvements in spatial structure, as assessed through spatial autocorrelation in both horizontal and vertical directions. The corrected fields exhibit more stable and realistic autocorrelation patterns across all spatial lags, closely tracking the ground truth (Figure 11b, Supplementary Figure S10b). This indicates that the correction not only enhances dry/wet classification but also helps preserve spatial coherence and the structural integrity of storm systems. These quantitative improvements are further supported by visual inspection of sample storm fields (Figure 12, Supplementary Figure S11). Uncorrected WGAN outputs tend to diffusely extend light rainfall into surrounding dry areas, a clear manifestation of drizzle bias. In contrast, applying the dry/wet mask effectively delineates dry regions and storm boundaries. The corrected fields exhibit sharper edges, cleaner separation between wet and dry zones, and overall stronger visual agreement with the ground truth. Notably, fine-scale dry regions, often completely absent in the uncorrected predictions, are successfully recovered, highlighting the mask's value in restoring both large- and small-scale storm features.

Overall, these results demonstrate the value of combining high-resolution intensity prediction with an explicit binary dry/wet mask. The predicted mask acts as a consistency constraint that guides the precipitation field toward more realistic dry/wet patterns. This two-stage modeling approach substantially reduces overestimation of wet conditions, improves the spatial organization of rainfall features, and ensures better alignment with both statistical and structural aspects of the observed data. Such physically constrained post-processing offers a practical and effective strategy for improving the realism and interpretability of deep learning-based downscaling models.

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

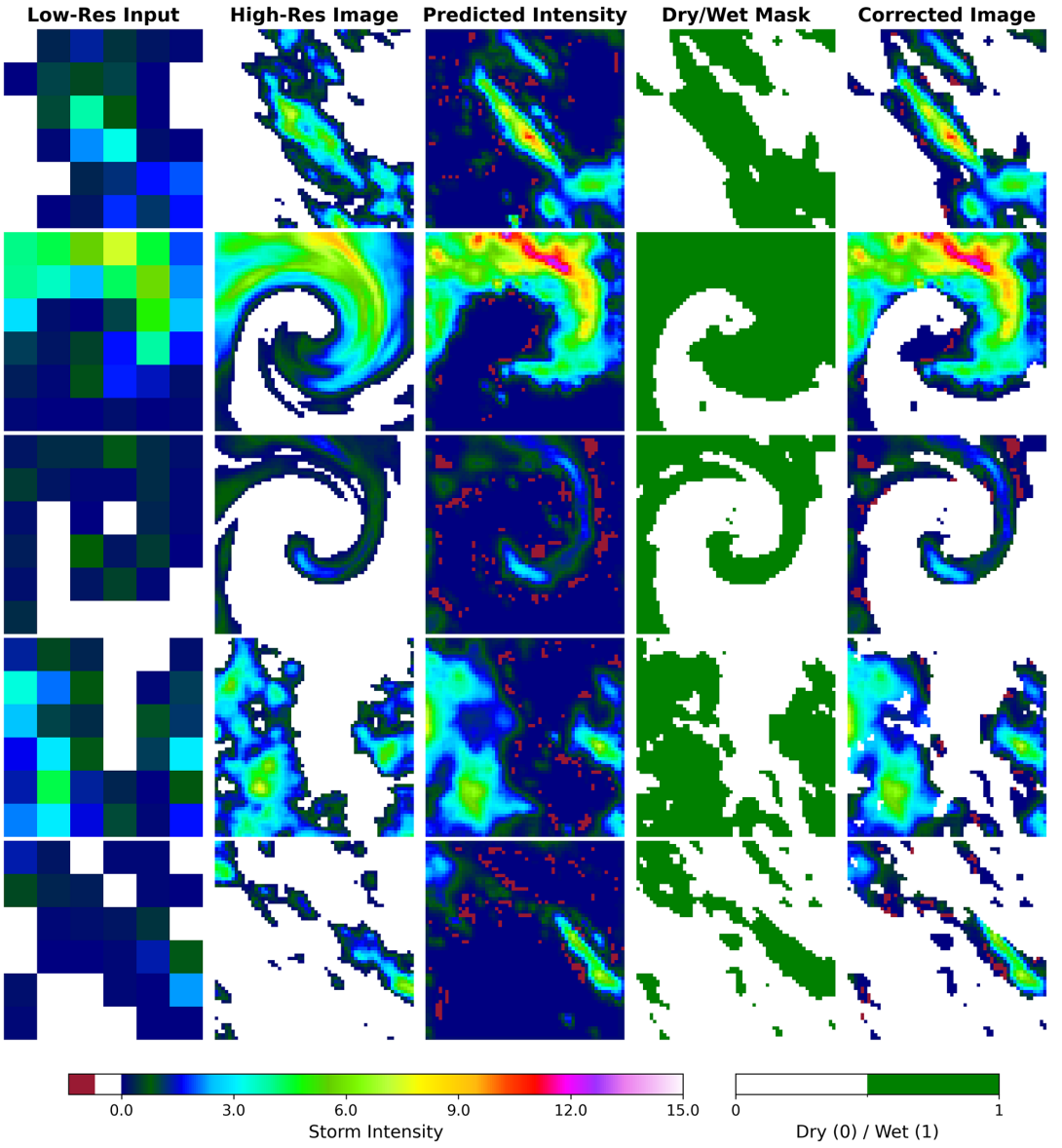


Figure 12. *Enhanced realism in corrected precipitation fields. Corrected images, generated by masking WGAN-predicted intensities using the dry/wet classification, show improved alignment with the ground truth, eliminating false wet areas and preserving sharper storm boundaries. (White space is represented as dry regions, whereas negative predictions from WGAN are represented with brown color).*

7. Conclusion

The study provides a comprehensive evaluation of deep learning-based downscaling approaches for high-resolution dry/wet classification using both synthetic and radar-based precipitation datasets. The convolutional encoder-decoder and WGAN were rigorously tested under three distinct training settings—using binary input, intensity input, and a constrained input, in order to assess their ability to reproduce spatially accurate and physically consistent dry/wet patterns. Two primary evaluation metrics were employed: the probability of zero (P_0), which measures the accuracy of dry pixel prediction, and the lagged Pearson spatial autocorrelation, which quantifies spatial coherence and structural fidelity across multiple spatial

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

lags. Across all experiments, using intensity input setting consistently improved model performance over the binary input setting, reducing drizzle bias and enhancing the spatial localization of wet regions for both the convolutional encoder–decoder and WGAN. The constrained input setting yielded the most balanced and generalizable results, especially for the WGAN, by reinforcing the extent and continuity of dry regions and reducing the occurrence of false wet predictions. While the convolutional encoder-decoder model tended to produce smoother outputs with somewhat overestimated spatial autocorrelation, it performed stably across different random initializations, with particularly strong results when trained under the intensity input setting. In contrast, the WGAN demonstrated sharper storm boundaries and greater inter-seed consistency, particularly when conditioned on both intensity and dry-region constraint, and their spatial autocorrelations were more closely aligned with the ground truth, especially in the horizontal direction for both types of storm data, synthetic as well as radar-based. Moreover, our results clearly demonstrated that incorporating a dry/wet mask significantly improves the downscaling outputs, effectively correcting the overestimation of wet conditions and allowing for the recovery of fine-scale dry features that were often missed in raw high-resolution predictions. Ultimately, a two-stage approach of separately predicting intensity and dry/wet masks proved highly effective, demonstrating that physically informed conditioning can substantially enhance the spatial accuracy and realism of deep learning-based precipitation downscaling.

These findings underscore the potential of deep learning models, particularly WGAN-based architectures with physically informed constraints for downscaling coarse-resolution precipitation fields into spatially detailed and physically consistent dry/wet patterns. The demonstrated improvements in both statistical accuracy and spatial structure, especially through the use of a predicted dry/wet mask, offer a practical pathway toward more reliable and interpretable downscaling outputs. Such advances are especially valuable in applications requiring high-fidelity dry/wet delineation, including drought monitoring, hydrological modeling, and storm impact assessments. Future work should explore further architectural enhancements and extend evaluations to diverse climatic regimes and storm types to better understand the models' generalizability and robustness for real-world applications.

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/eds.2026.10039.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/eds.2026.10039.

Acknowledgements. We gratefully acknowledge the financial support provided by the Environmental Institute, University of Virginia, through the Strategic Investment Fund for carrying out this research work.

Author contribution. Conceptualization: S.S., S.M.P., H.M.A., A.M.; Data curation: S.S.; Formal analysis: S.S.; Investigation: S.S., S.M.P., H.M.A.; Methodology: S.S., S.M.P., A.M.; Validation: S.S., A.M.; Visualization: S.S., H.M.A.; Writing - original draft: S.S.; Writing - review & editing: S.S., S.M.P., H.M.A., T.H., A.M.; Funding acquisition: S.M.P., T.H., A.M.; Resources: S.M.P., A.M.; Supervision: S.M.P., T.H., A.M.; Project administration: T.H., A.M.

Competing interests. The authors declare none.

Data availability statement. The CoSMoS R package (Papalexiou et al., 2021) used for generating storms, is available at CRAN (R Core Team, 2021; https://cran.r-project.org/web/packages/CoSMoS/vignettes/vignette.html).

The hourly precipitation data used in this study were obtained from the NCEP Multi-Radar/Multi-Sensor (MRMS) system, accessed via the Iowa Environmental Mesonet (IEM) platform (https://mesonet.agron.iastate.edu/GIS/rasters.php?rid=4). The MRMS dataset provides high-resolution precipitation estimates at a 1 km × 1 km spatial scale and is publicly available for research purposes. These data were used to evaluate model performance on real-world storm structures and dry/wet classification accuracy. The code used for model training, inference, and evaluation is publicly available on GitHub at https://github.com/shivamsinghhada/Binary_precipitation_downscaling.

References

Abdelmoaty HM, **Papalexiou SM**, **Mamalakis A**, **Singh S**, **Coia V**, **Hairabedian M**, **Szeftel P and Grover P** (2025) Generative adversarial networks for downscaling hourly precipitation in the Canadian Prairies. *Journal of Geophysical Research: Machine Learning and Computation 2*(4). https://doi.org/10.1029/2025JH000678.

AghaKouchak A, **Mehran A**, **Norouzi H and Behrangi A** (2012) Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters 39*(9). https://doi.org/10.1029/2012GL051592.

Akinsanola AA, **Chen Z**, **Kooperman GJ and Bobde V** (2024) Robust future intensification of winter precipitation over the United States. *npj Climate and Atmospheric Science 7*(1), 212. https://doi.org/10.1038/s41612-024-00761-8.

Arjovsky M, **Chintala S and Bottou L** (2017) Wasserstein GAN. Preprint. http://arxiv.org/abs/1701.07875.

Badrinarayanan V, **Kendall A and Cipolla R** (2017) SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615.

Baghanam AH, **Nourani V**, **Bejani M**, **Pourali H**, **Kantoush SA and Zhang Y** (2024) A systematic review of predictor screening methods for downscaling of numerical climate models. *Earth-Science Reviews 253*, 104773. https://doi.org/10.1016/j.earscirev.2024.104773.

Banõ-Medina J, **Manzanas R**, **Cimadevilla E**, **Fernandez J**, **Gonzalez-Abad J**, **Cofinõ AS and Gutierrez JM** (2022) Downscaling multi-model climate projection ensembles with deep learning (DeepESD): Contribution to CORDEX EUR-44. *Geoscientific Model Development 15*(17), 6747–6758. https://doi.org/10.5194/gmd-15-6747-2022.

Baño-Medina J, **Manzanas R and Gutiérrez JM** (2020) Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development 13*(4), 2109–2124. https://doi.org/10.5194/gmd-13-2109-2020.

Berrang-Ford L, **Ford JD and Paterson J** (2011) Are we adapting to climate change? *Global Environmental Change 21*. https://doi.org/10.1016/j.gloenvcha.2010.09.012.

Chen J, **Janke T**, **Steinke F and Lerch S** (2024) Generative machine learning methods for multivariate ensemble postprocessing. *The Annals of Applied Statistics 18*(1). https://doi.org/10.1214/23-AOAS1784.

Chen X, **Leung LR**, **Gao Y**, **Liu Y and Wigmosta M** (2023) Sharpening of cold-season storms over the western United States. *Nature Climate Change 13*(2), 167–173. https://doi.org/10.1038/s41558-022-01578-0.

Choubey D, **Patil V and Anand Kumar M** (2024) Cognitive chromatic image synthesis using UNET and GAN. In *2024 2nd International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS)*, Manipal, India. Piscataway, NJ, USA: IEEE, pp. 178–183. https://doi.org/10.1109/ICRAIS62903.2024.10811702.

Creswell A, **White T**, **Dumoulin V**, **Arulkumaran K**, **Sengupta B and Bharath AA** (2018) Generative adversarial networks: An overview. *IEEE Signal Processing Magazine 35*(1), 53–65. https://doi.org/10.1109/MSP.2017.2765202.

Dueben PD and Bauer P (2018) Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development 11*(10), 3999–4009. https://doi.org/10.5194/gmd-11-3999-2018.

Duncan J, **Subramanian S and Harrington P** (2022) Generative modeling of high-resolution global precipitation forecasts. Preprint. http://arxiv.org/abs/2210.12504.

Gao XJ, **Shi Y**, **Zhang D**, **Wu J**, **Giorgi F**, **Ji Z and Wang Y** (2012) Uncertainties in monsoon precipitation projections over China: Results from two high-resolution RCM simulations. *Climate Research 52*. https://doi.org/10.3354/cr0108.

Giorgi F and Gutowski WJ (2015) Regional dynamical downscaling and the CORDEX initiative. *Annual Review of Environment and Resources 40*. https://doi.org/10.1146/annurev-environ-102014-021217.

Glawion L, **Polz J**, **Kunstmann H**, **Fersch B and Chwala C** (2023) spateGAN: Spatio-temporal downscaling of rainfall fields using a cGAN approach. *Earth and Space Science 10*(10). https://doi.org/10.1029/2023EA002906.

González-Abad J, **Baño-Medina J and Cachá IH** (2023) On the use of deep generative models for perfect prognosis climate downscaling. Preprint. http://arxiv.org/abs/2305.00974.

Goodfellow I (2016) NIPS 2016 tutorial: Generative adversarial networks. Preprint. http://arxiv.org/abs/1701.00160.

Gulrajani I, **Ahmed F**, **Arjovsky M**, **Dumoulin V and Courville AC** (2017) Improved training of Wasserstein GANs. In *Advances in Neural Information Processing Systems (NeurIPS 2017)*, vol. *30*.

Gutowski WJ, **Ullrich PA**, **Hall A**, **Leung LR**, **O'Brien TA**, **Patricola CM**, **Arritt RW**, **Bukovsky MS**, **Calvin KV**, **Feng Z**, **Jones AD**, **Kooperman GJ**, **Monier E**, **Pritchard MS**, **Pryor SC**, **Qian Y**, **Rhoades AM**, **Roberts AF**, **Sakaguchi K**, et al. (2020) The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information. *Bulletin of the American Meteorological Society 101*(5), E664–E683. https://doi.org/10.1175/BAMS-D-19-0113.1.

Han Y and Guan L (2023) GAN-based vertical federated learning for label protection in binary classification. arXiv preprint arXiv: 2302.02245. Available at https://arxiv.org/abs/2302.02245.

Harder P, **Hernandez-Garcia A**, **Ramesh V**, **Yang Q**, **Sattegeri P**, **Szwarcman D**, **Watson CD**, **Rolnick D and Watson C** (2023) Hard-constrained deep learning for climate downscaling. *Journal of Machine Learning Research 24*, 1–40.

Harris L, **McRae ATT**, **Chantry M**, **Dueben PD and Palmer TN** (2022) A generative deep learning approach to stochastic downscaling of precipitation forecasts. *Journal of Advances in Modeling Earth Systems 14*(10). https://doi.org/10.1029/2022MS003120.

Hobeichi S, **Nishant N**, **Shao Y**, **Abramowitz G**, **Pitman A**, **Sherwood S**, **Bishop C and Green S** (2023) Using machine learning to cut the cost of dynamical downscaling. *Earth's Future 11*(3). https://doi.org/10.1029/2022EF003291.

Kalantar R, **Messiou C**, **Winfield JM**, **Renn A**, **Latifoltojar A**, **Downey K**, **Sohaib A**, **Lalondrelle S**, **Koh D-M and Blackledge MD** (2021) CT-based pelvic T1-weighted MR image synthesis using UNet, UNet++ and cycle-consistent generative adversarial network (cycle-GAN). *Frontiers in Oncology 11*. https://doi.org/10.3389/fonc.2021.665807.

Kingma DP and Ba J (2014) Adam: A method for stochastic optimization. Preprint. http://arxiv.org/abs/1412.6980

Kumar B, **Atey K**, **Singh BB**, **Chattopadhyay R**, **Acharya N**, **Singh M**, **Nanjundiah RS and Rao SA** (2023) On the modern deep learning approaches for precipitation downscaling. *Earth Science Informatics 16*(2), 1459–1472.

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

Lazoglou G, **Economou T**, **Anagnostopoulou C**, **Zittis G**, **Tzyrkalli A**, **Georgiades P and Lelieveld J** (2024) Multivariate adjustment of drizzle bias using machine learning in European climate projections. *Geoscientific Model Development 17*(11), 4689–4703. https://doi.org/10.5194/gmd-17-4689-2024.

Leinonen J, **Nerini D and Berne A** (2020) Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing 59*, 7211–7223. https://doi.org/10.1109/TGRS.2020.3032790.

Lesnikowski AC, **Ford JD**, **Berrang-Ford L**, **Barrera M and Heymann J** (2015) How are we adapting to climate change? A global assessment. *Mitigation and Adaptation Strategies for Global Change*, *20*. https://doi.org/10.1007/s11027-013-9491-x.

Liu Y, **Ganguly AR and Dy J** (2020) Climate downscaling using YNet. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: ACM, pp. 3145–3153. https://doi.org/10.1145/3394486.3403366.

Lopez-Gomez I, **Wan ZY**, **Zepeda-Núñez L**, **Schneider T**, **Anderson J and Sha F** (2024) Dynamical-generative downscaling of climate model ensembles. Preprint. http://arxiv.org/abs/2410.01776.

Luo Y, **Zhang K**, **Wang W**, **Chen X**, **Feng J**, **Wang H**, **Liu W**, **Guo C**, **Chen C and Wang X** (2025) An improved statistical bias correction method for global climate model (GCM) precipitation projection: A case study on the CMCC-CM2-SR5 model projection in China's Huaihe River basin. *Journal of Hydrology: Regional Studies 57*, 102146. https://doi.org/10.1016/j.ejrh.2024.102146.

Magnan AK (2016) Climate change: Metrics needed to track adaptation. *Nature 530*. https://doi.org/10.1038/530160d.

Mamalakis A, **Langousis A**, **Deidda R and Marrocu M** (2017) A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall. *Water Resources Research 53*(3), 2149–2170. https://doi.org/10.1002/2016WR019578.

Maraun D (2016) Bias correcting climate change simulations—a critical review. *Current Climate Change Reports 2*. https://doi.org/10.1007/s40641-016-0050-x.

Maraun D, **Wetterhall F**, **Ireson AM**, **Chandler RE**, **Kendon EJ**, **Widmann M**, **Brienen S**, **Rust HW**, **Sauter T**, **Themeßl M**, **Venema VKC**, **Chun KP**, **Goodess CM**, **Jones RG**, **Onof C**, **Vrac M and Thiele-Eich I** (2010) Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics 48*(3), RG3003. https://doi.org/10.1029/2009RG000314.

Meghani S, **Singh S**, **Kumar N and Goyal MK** (2023) Predicting the spatiotemporal characteristics of atmospheric rivers: A novel data-driven approach. *Global and Planetary Change 231*, 104295. https://doi.org/10.1016/j.gloplacha.2023.104295.

Mendoza Paz S and Willems P (2023) The skill of statistical downscaling in future climate with high-resolution climate models as pseudo-reality. *Journal of Hydrology: Regional Studies 48*, 101477. https://doi.org/10.1016/j.ejrh.2023.101477.

Miralles O, **Steinfeld D**, **Martius O and Davison AC** (2022) Downscaling of historical wind fields over Switzerland using generative adversarial networks. *Artificial Intelligence for the Earth Systems 1*(4). https://doi.org/10.1175/aies-d-22-0018.1.

Murukesh M, **Golla S and Kumar P** (2024) Downscaling and reconstruction of high-resolution gridded rainfall data over India using deep learning-based generative adversarial network. *Modeling Earth Systems and Environment 10*(2), 2221–2237. https://doi.org/10.1007/s40808-023-01899-9.

Nishant N, **Hobeichi S**, **Sherwood S**, **Abramowitz G**, **Shao Y**, **Bishop C and Pitman A** (2023) Comparison of a novel machine learning approach with dynamical downscaling for Australian precipitation. *Environmental Research Letters 18*(9), 094006. https://doi.org/10.1088/1748-9326/ace463.

Pan B, **Anderson GJ**, **Goncalves A**, **Lucas DD**, **Bonfils CJW**, **Lee J**, **Tian Y and Ma H** (2021) Learning to correct climate projection biases. *Journal of Advances in Modeling Earth Systems 13*(10). https://doi.org/10.1029/2021MS002509.

Papalexiou SM (2022) Rainfall generation revisited: Introducing CoSMoS-2s and advancing copula-based intermittent time series Modeling. *Water Resources Research 58*(6). https://doi.org/10.1029/2021WR031641.

Papalexiou SM and Mamalakis A (2025) Machine unlearning: Bias correction in neural network downscaled storms. *Journal of Hydrology*, 134689. https://doi.org/10.1016/j.jhydrol.2025.134689.

Papalexiou SM, **Serinaldi F and Porcu E** (2021) Advancing space-time simulation of random fields: From storms to cyclones and beyond. *Water Resources Research 57*(8). https://doi.org/10.1029/2020WR029466.

Potter NJ, **Chiew FHS**, **Charles SP**, **Fu G**, **Zheng H and Zhang L** (2020) Bias in dynamically downscaled rainfall characteristics for hydroclimatic projections. *Hydrology and Earth System Sciences 24*(6), 2963–2979. https://doi.org/10.5194/hess-24-2963-2020.

Rahimi S, **Huang L**, **Norris J**, **Hall A**, **Goldenson N**, **Risser M**, **Feldman DR**, **Lebo ZJ**, **Dennis E and Thackeray C** (2024) Understanding the Cascade: Removing GCM biases improves dynamically downscaled climate projections. *Geophysical Research Letters 51*(9). https://doi.org/10.1029/2023GL106264.

Rampal N, **Hobeichi S**, **Gibson PB**, **Baño-Medina J**, **Abramowitz G**, **Beucler T**, **González-Abad J**, **Chapman W**, **Harder P and Gutiérrez JM** (2024) Enhancing regional climate downscaling through advances in machine learning. *Artificial Intelligence for the Earth Systems 3*(2). https://doi.org/10.1175/aies-d-23-0066.1.

Scher S and Peßenteiner S (2021) Technical note: Temporal disaggregation of spatial rainfall fields with generative adversarial networks. *Hydrology and Earth System Sciences 25*(6), 3207–3225. https://doi.org/10.5194/hess-25-3207-2021.

Singh S and Goyal MK (2023) An innovative approach to predict atmospheric rivers: Exploring convolutional autoencoder. *Atmospheric Research 289*, 106754.

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press

Stocker T (2014) *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press.

Suliman AHA, **Awchi TA**, **Al-Mola M and Shahid S** (2020) Evaluation of remotely sensed precipitation sources for drought assessment in semi-arid Iraq. *Atmospheric Research 242*, 105007. https://doi.org/10.1016/j.atmosres.2020.105007.

Tabari H, **Paz SM**, **Buekenhout D and Willems P** (2021) Comparison of statistical downscaling methods for climate change impact analysis on precipitation-driven drought. *Hydrology and Earth System Sciences 25*(6), 3493–3517. https://doi.org/10.5194/hess-25-3493-2021.

Vogel E, **Johnson F**, **Marshall L**, **Bende-Michl U**, **Wilson L**, **Peter JR**, **Wasko C**, **Srikanthan S**, **Sharples W**, **Dowdy A**, **Hope P**, **Khan Z**, **Mehrotra R**, **Sharma A**, **Matic V**, **Oke A**, **Turner M**, **Thomas S**, **Donnelly C and Duong VC** (2023) An evaluation framework for downscaling and bias correction in climate change impact studies. *Journal of Hydrology 622*, 129693. https://doi.org/10.1016/j.jhydrol.2023.129693.

Vrac M, **Stein ML**, **Hayhoe K and Liang XZ** (2007) A general method for validating statistical downscaling methods under future climate change. *Geophysical Research Letters*, *34*(18). https://doi.org/10.1029/2007GL030295.

Wang F, **Tian D**, **Lowe L**, **Kalin L and Lehrter J** (2021) Deep learning for daily precipitation and temperature downscaling. *Water Resources Research 57*(4). https://doi.org/10.1029/2020WR029308.

Xu W, **Jang-Jaccard J**, **Liu T**, **Sabrina F and Kwak J** (2022) Improved bidirectional GAN-based approach for network intrusion detection using one-class classifier. *Computers 11*(6), 85. https://doi.org/10.3390/computers11060085.

Yazdandoost F, **Moradian S**, **Izadi A and Aghakouchak A** (2021) Evaluation of CMIP6 precipitation simulations across different climatic zones: Uncertainty and model intercomparison. *Atmospheric Research 250*, 105369. https://doi.org/10.1016/j.atmosres.2020.105369.

Cite this article: Singh S, Papalexiou SM, Abdelmoaty HM, Hartvigsen T and Mamalakis A (2026). Correcting dry/wet classification bias in precipitation downscaling via generative adversarial networks. *Environmental Data Science*, 5: e12. doi:10.1017/eds.2026.10039

https://doi.org/10.1017/eds.2026.10039 Published online by Cambridge University Press