Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Research on Downscaling Method for Fuxi Model Precipitation Forecast Based on Diffusion-U-Net

Hua Gao
In
Kui Zhang
38681222@qq.com

In
Weiwen Ji
In
Fuxin Zhu
In

Research Article

Keywords: Precipitation forecast, Downscaling, Diffusion model, U-Net, Attention mechanism, Fuxi model

Posted Date: June 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9363360/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Research on Downscaling Method for Fuxi Model Precipitation Forecast Based on Diffusion-U-Net

Gaohua[1] Zhangkui[2] Jiweiwen[1] Zhufuxin[1]

(1.Institute of Urban Meteorology,Beijing,China,100089.

2. Chengdu Meteorology Bureau,Chengdu,China,610072)

Abstract

Aiming at the problems of high cost in obtaining high-resolution data and insufficient spatiotemporal resolution in meteorological precipitation forecasting, this paper proposes a precipitation forecast downscaling method based on Correlation Diffusion Network (Diffusion-U-Net) to convert low-resolution precipitation data output by the Fuxi model into high-resolution data. This method combines the probabilistic generation capability of diffusion models with the multi-scale feature extraction advantages of U-Net, and innovatively introduces a correlation module with channel-spatial dual attention mechanism to effectively capture the spatiotemporal correlation features of precipitation fields. Experimental results show that the proposed method outperforms traditional interpolation methods and deep learning methods in multiple evaluation metrics, can more accurately restore small and medium-scale precipitation details, and provides technical support for refined meteorological forecasting.

Keywords: Precipitation forecast; Downscaling; Diffusion model; U-Net; Attention mechanism; Fuxi model

1. Introduction

1.1 Research Background and Significance

Precipitation forecasting is the core content of meteorological services, which plays an important guiding role in agricultural production, flood control and disaster reduction, water resources management and other fields. With the development of numerical weather prediction (NWP) technology, artificial intelligence meteorological forecasting systems represented by the Fuxi model have achieved accurate simulation of precipitation processes. However, limited by computing resources and model complexity, their output results are usually low-resolution grid data (such as 24km×24km), which are difficult to meet the needs of refined forecasting of small and medium-scale weather systems (such as severe convection, orographic rainfall) [1].

Downscaling technology converts low-resolution data into high-resolution information through statistical or dynamic methods, which is a key means to make up for this defect. Traditional downscaling methods are mainly divided into two categories: dynamic downscaling performs fine grid simulation based on regional climate models, which has clear physical mechanisms but extremely high computational cost; statistical downscaling realizes conversion by establishing statistical relationships between high and low-resolution data, which is more efficient but relies on prior assumptions [2]. In recent years, deep learning technology has shown significant advantages in downscaling tasks due to its strong nonlinear fitting ability, among which convolutional neural networks (CNN), generative adversarial networks (GAN) and other models have been widely used [3-4].

However, existing deep learning methods still have the following limitations: (1) Insufficient capture of strong nonlinear spatial correlations in precipitation fields (such as the response relationship between terrain and precipitation); (2) The generated high-resolution data are

prone to artifacts or blurred details; (3) Lack of ability to quantify forecast uncertainty. Therefore, this paper introduces the probabilistic generation framework of diffusion models, combines with the multi-scale feature fusion ability of U-Net, and proposes the Diffusion-U-Net downscaling method, aiming to improve the spatial resolution and detail fidelity of precipitation forecasts.

1.2 Research Status at Home and Abroad

1.2.1 Traditional Downscaling Methods

Dynamic downscaling refines global model outputs by nesting regional climate models (such as WRF), which can better retain physical processes but is time-consuming and sensitive to boundary conditions [5]. Statistical downscaling methods include linear methods such as linear regression and stepwise regression, as well as machine learning methods such as decision trees and random forests. For example, Chen et al. [6] used geographically weighted regression to establish the relationship between precipitation and topographic factors, and realized downscaling research in the Yangtze River Basin. However, such methods are difficult to characterize the complex nonlinear characteristics of precipitation.

1.2.2 Deep Learning Downscaling Methods

CNN has become the mainstream model for downscaling tasks due to its local receptive field and weight sharing characteristics. ESPCN proposed by Shi et al. [7] realizes end-to-end super-resolution through sub-pixel convolution, which is applied to precipitation downscaling; Liang et al. [8] combined U-Net with attention mechanism to enhance the attention to extreme precipitation areas. GAN improves generation quality through adversarial training between generator and discriminator. For example, the Pix2Pix model is used for spatiotemporal refinement of precipitation fields [9], but GAN has problems such as unstable training and mode collapse of generated results.

1.2.3 Application of Diffusion Models in Meteorological Field

Diffusion models realize probabilistic generation through gradual noise addition-denoising process, which have the advantages of high generation quality and strong stability. In recent years, their applications in the meteorological field have gradually emerged: Ronneberger et al. [10] used diffusion models for radar echo forecasting; WeatherDiff proposed by Yang et al. [11] realized probabilistic forecasting of precipitation fields, but has not been specially optimized for downscaling tasks.

1.3 Research Objectives and Contents

This paper aims to construct a high-precision downscaling method based on low-resolution precipitation forecast data of the Fuxi model. The main research contents include:

Design a Diffusion-U-Net model integrating correlation modules to capture key features of precipitation fields through channel and spatial attention mechanisms;

Construct a diffusion process suitable for precipitation downscaling to realize probabilistic generation from low-resolution input to high-resolution output;

Conduct model training and verification based on actual meteorological datasets, and verify the effectiveness of the method through multi-index evaluation.

2. Method Principles

2.1 Overall Framework

The downscaling method proposed in this paper takes low-resolution precipitation data (24km×24km) output by the Fuxi model as input, and generates high-resolution precipitation

data (3km×3km) through the Diffusion-U-Net model. The overall framework is shown in Figure 1, which mainly includes three core modules:

Data preprocessing module: Perform standardization, enhancement and other processing on the input low-resolution data;

Diffusion-U-Net model: Integrate diffusion model and U-Net architecture, capture precipitation field features through correlation module;

Post-processing module: Perform denormalization, spatial matching and other operations on the generated high-resolution data.

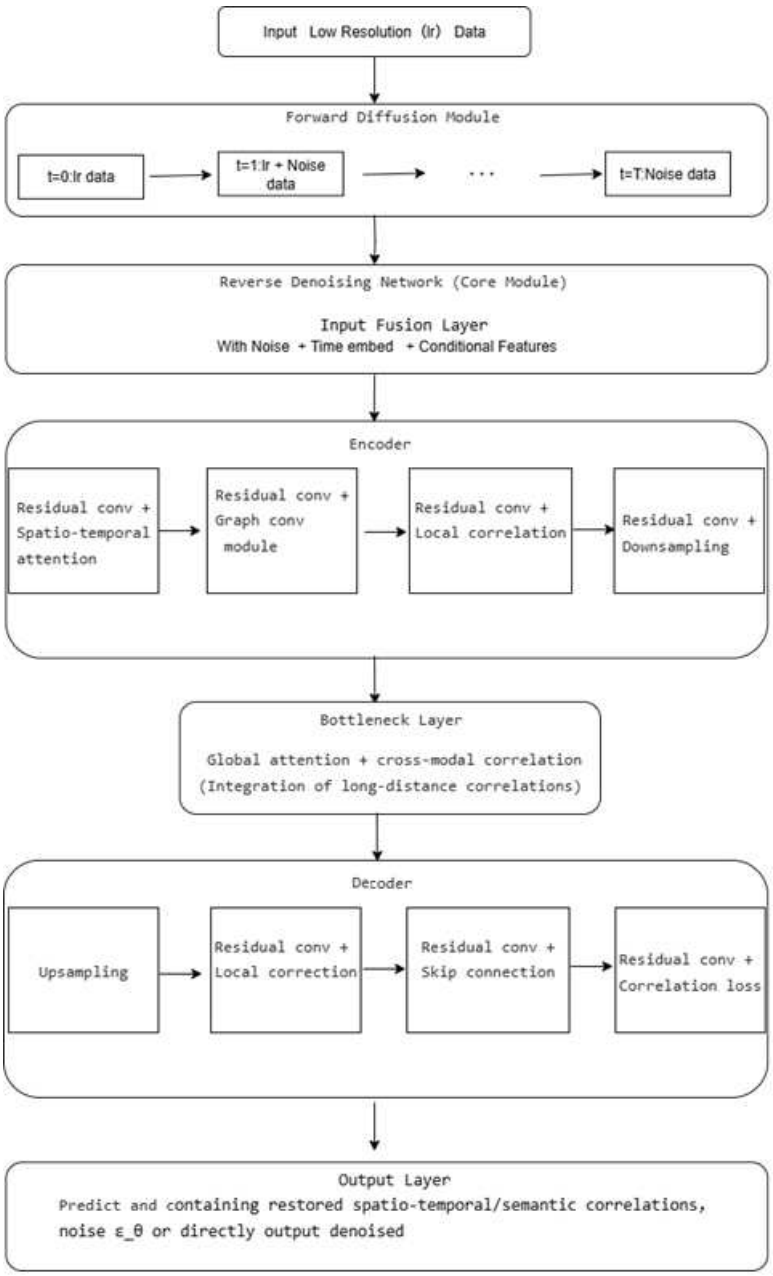


Figure 1 Overall framework of the downscaling method

2.2 Diffusion Process Design

Diffusion models realize data generation through forward noise addition and reverse

denoising processes. This paper designs the following diffusion mechanism for precipitation downscaling tasks:

2.2.1 Forward Noise Addition Process

In the forward process, by gradually adding Gaussian noise to low-resolution precipitation data, samples approximately pure noise are finally obtained. Let the initial low-resolution data be x_0, and the sample x_t at time step t ($t = 1,2,\ldots,T$) satisfies:

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

Where $\alpha_t \in (0,1)$ is the noise coefficient, and $\varepsilon_{t-1}\sim N(0,I)$ is Gaussian noise. The relationship between x_0 and x_t can be directly calculated through cumulative product:

$$x_t=\sqrt{\alpha_t}x_0 + \sqrt{1-\overline{a_t}}\ \varepsilon$$

Where

$$\overline{a_t} = \prod_{s-1}^{t} a_s$$

and $\epsilon \sim N(0,I)$.

2.2.2 Reverse Denoising Process

In the reverse process,the model predicts noise and gradually denoises,recovering high-resolution data from pure noise x_T , At time step t, the model inputs are x_t and time step t , and outputs the noise prediction value $\mathcal{E}_\theta(x_t,t)$.

Then the update formula of x_{t-1} is:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}}\ \mathcal{E}_\theta(x_t,t)\right) + \delta_t z$$

Where $z \sim N(0,I)$ (when $t > 1$), and δ_t is the noise standard deviation. Through $T-$ step iteration, the denoised high $-$ resolution data x_0 is finally obtained.

2.3 Diffusion-U-Net Model Architecture

The model is based on U-Net architecture,integrating timestep embedding and correlation modules to realize mapping from low-resolution input to high-resolution output.The specific structure is shown in Figure 2.

2.3.1 Timestep Embedding

To enable the model to perceive the time information of the diffusion process,sine-cosine position encoding is used to generate timestep embedding:

$$emb(t,2i) = \sin\left(t/10000^{2i/d}\right)$$
$$emb(t,2i+1) = \cos\left(t/10000^{2i/d}\right)$$

where d is the embedding dimension.The generated embedding vector is integrated into each residual block after MLP processing:

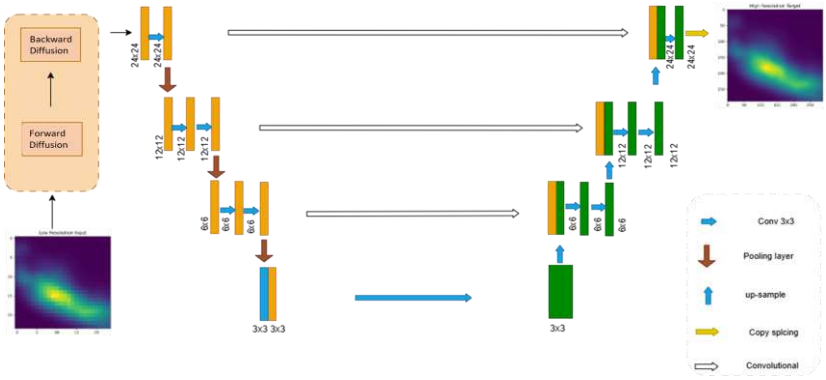


Figure 2. Diffusion-U-Net Structure

2.3.2 Correlation Module

This module is the core innovation of the model, capturing key features of precipitation fields through channel attention and spatial attention:

- **Channel attention**: Generate channel weights through global average pooling and MLP, highlighting important meteorological elements (such as channel features in heavy precipitation areas);
- **Spatial attention**: Generate spatial weights through fusion of mean and maximum values in channel dimension, enhancing attention to key areas such as precipitation centers.

The module structure is implemented as follows:

2.3.3 U-Net Main Structure

Downsampling path: Composed of residual blocks and attention blocks, gradually reducing spatial dimensions and increasing the number of channels through max pooling to extract multi-scale features;

Middle block: Contains residual blocks and attention blocks to process the deepest features;

Upsampling path: Realizes upsampling through bilinear interpolation and convolution, combines skip connections to fuse features from the downsampling path, and gradually restores spatial resolution to the target size (288×288).

The residual block is designed as follows, which incorporates timestep embedding information:

2.4 Loss Function Design

The model training adopts a hybrid loss function, including noise prediction loss and super-resolution loss:

1. Noise prediction loss: Measures the model's prediction accuracy of noise, using MSE loss:

$$\mathcal{L}_{noise} = \mathrm{E}_{x_0,\mathcal{E},t}[\|\mathcal{E} - \mathcal{E}_\theta(x_t, t)\|^2]$$

2. Super-resolution loss: Introduced in the later stage of training to constrain the difference between the generated high-resolution data and the real value:

$$\mathcal{L}_{sr} = \mathrm{E}_{x_{lr},y_{hr}}[\|G(x_{lr}) - y_{hr}\|^2]$$

Where G is the high-resolution data generated by the model, and y_{hr} is the real high-resolution data.

The total loss is the weighted sum of the two:

$$\mathcal{L} = \mathcal{L}_{hr} + \lambda \mathcal{L}_{hr}$$

Where λ is a hyperparameter (set to 0.5 in the experiment).

3. Experimental Design

3.1 Data Source and Preprocessing

3.1.1 Dataset Construction

Experimental data come from the national hourly precipitation observation data (resolution 3km×3km) provided by the China Meteorological Administration and the forecast data of the Fuxi model (resolution 0.25°×0.25°). The North China (35°N-43°N, 113°E-119°E) is selected as the study area, which is divided into training set, validation set and test set in the ratio of 7:2:1, containing more than 100,000 sample pairs.

3.1.2 Preprocessing Steps

Spatial matching: Spatially align the low-resolution data (0.25°×0.25°) output by the Fuxi model with the observation data ;

Standardization: Perform logarithmic transformation on precipitation data (x '=log(x+1)), then use Z-Score standardization;

Data enhancement: Expand training samples through random flipping (horizontal/vertical) and rotation (multiples of 90°) to enhance the generalization ability of the model.

3.2 Experimental Settings

3.2.1 Model Parameters

Number of base channels: 32;

Channel multiples: (1, 2, 4, 8);

Time embedding dimension: 128;

Number of diffusion steps: 1000;

Optimizer: Adam, learning rate 1e-4, weight decay 1e-5;

Training epochs: 100, batch size=32;

Device: NVIDIA A100 GPU (40GB).

3.2.2 Comparison Methods

Traditional interpolation methods: Bicubic interpolation, Inverse Distance Weighting (IDW);

Bicubic interpolation and IDW interpolation are classic geometry-driven methods whose core logic relies on weighted fitting based on the spatial correlation of existing data, with obvious limitations:

Bicubic interpolation depends on fitting the grayscale values of pixels/grid points in local neighborhoods, which can only achieve smooth numerical interpolation. It cannot generate new features beyond the original data distribution and is prone to blurring and distortion when dealing with complex textures.

IDW interpolation determines weights entirely based on distance (the closer the distance, the higher the weight), resulting in poor adaptability to spatial heterogeneity of data. It is also sensitive to outliers, and the interpolated results have weak physical significance.

In contrast, diffusion models are data-driven generative models. By learning the probability distribution of massive data, they can generate details consistent with real physical laws or data characteristics during interpolation/super-resolution processes, making up for the defects of traditional methods such as "smoothing distortion" and "inability to create new features". The comparison between the two essentially represents a contest between two paradigms: "geometric fitting" and "probabilistic generation". It can clarify the technical advantages of diffusion models in complex scenarios while defining their scope of application Quantifying the Accuracy and Efficiency of Different Methods to Guide Practical Engineering Applications

In fields such as meteorology and remote sensing, the core requirements for spatial data processing are accuracy and efficiency. The comparison of the three methods can provide quantitative references:

Accuracy dimension: Compare error metrics (e.g., root mean square error (RMSE),) and feature fidelity (e.g., edge clarity, texture consistency) of interpolated results. IDW may perform well in data-dense areas but has significant errors in data-sparse areas; if trained on high-quality meteorological datasets, diffusion models can generate interpolated results more consistent with atmospheric circulation laws with more stable errors.

3.2.3 Evaluation Indicators

The following indicators are used to evaluate the downscaling effect:

Structural Similarity Index (SSIM):

An index that measures the similarity in structure, brightness, and contrast between two sets of data (or images/grids). Its core idea is that the perceptual quality of data depends not only on pixel-level errors (e.g., MSE) but also on the consistency of its inherent structure (e.g., the spatial distribution of meteorological fields, the texture characteristics of cloud images).

A weighted combination of similarity across three dimensions:

Luminance similarity: Measures the difference in average intensity between two datasets;

Contrast similarity: Measures the difference in standard deviation (degree of dispersion) between two datasets;

Structural similarity: Measures the difference in covariance (spatial correlation pattern) between two datasets.

Final formula:

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}$$

(where μ = mean, σ = standard deviation, σ_{xy} = covariance, and C_1 ,C_2 are constants to avoid division by zero)

Value range: $[-1,1]$ (in practical applications, it usually ranges from 0 to 1; negative values indicate completely opposite structures);Core implications:

SSIM=1: The two datasets are identical (no differences in structure, brightness, or contrast);

The closer SSIM is to 1: The higher the structural fidelity of the data and the stronger the perceptual consistency;SSIM close to 0: Severe structural distortion of the data (e.g., chaotic textures in interpolated meteorological fields, blurred edges after cloud image super-

resolution).
Compensates for the shortcomings of traditional error metrics (e.g., MSE, RMSE): MSE only quantifies numerical deviation but ignores structural consistency (e.g., two meteorological fields may have a small MSE but completely different spatial patterns, leading to misleading evaluations);
Aligns with human perceptual characteristics: Focuses on the integrity of key structures (e.g., the shape of cyclones in meteorological forecasts, the boundary of water bodies in remote sensing images), which is more in line with practical application needs (e.g., judging whether interpolated data retains physical meaning);
Applicable scenarios: Evaluation of image super-resolution, spatial data interpolation/reconstruction, and model output fidelity (e.g., comparing the structural consistency between simulated and observed meteorological fields).

Precipitation Intensity Bias (Bias): It reflects the overall tendency of the model/method to overestimate or underestimate the target variable, rather than random errors.

$$\text{Bias} = \frac{1}{N}\sum_{1}^{N}(X_i - Y_i)$$

(where N = number of samples, X_i = predicted/processed value of the i-th sample, Y_i = true value of the i-th sample).The larger the absolute value of Bias: The more serious the systematic error, indicating that the model/method has inherent flaws.

Critical Success Index (CSI): Also known as the Threat Score (TS), it is a metric that measures the consistency between predicted and observed "events". Evaluates the forecasting ability for threshold precipitation (such as 10mm/6h), the larger the better.
Based on the confusion matrix (for binary classification of "event occurrence" vs. "non-occurrence"):
TP (True Positive): Correctly predicted events (observed as occurring, predicted as occurring);
FP (False Positive): False alarms (observed as not occurring, predicted as occurring);
FN (False Negative): Missed events (observed as occurring, predicted as not occurring);
TN (True Negative): Correctly predicted non-events (observed as not occurring, predicted as not occurring).
Final formula:

$$CSI = \frac{TP}{TP + FP + FN}$$

Value range: [0, 1]; The closer CSI is to 1: The higher the consistency between predicted and observed events, and the stronger the reliability of the method in event identification.

4. Experimental Results and Analysis

4.1 Quantitative Result Analysis

Table 1 shows the evaluation results of each method on the test set at 10mm/6h. It can be seen that:
The proposed Diffusion-U-Net outperforms traditional interpolation methods in all indicators. Among them, SSIM is increased by 0.15, indicating that deep learning methods can better capture the spatial structure of precipitation;
Compared with other deep learning methods, Diffusion-U-Net has the highest SSIM (0.82),

indicating that the introduction of the probabilistic generation mechanism of the diffusion model and the correlation module effectively improves the downscaling accuracy;
In extreme precipitation forecasting (CSI), Diffusion-U-Net reaches 0.68, which is better than Bicubic(0.42), indicating that it depicts heavy precipitation areas.

Table 1 Quantitative evaluation results of each method at 10mm/6h

Method	SSIM	Bias	CSI
Bicubic	0.67	0.89	0.42
IDW	0.65	0.85	0.38
Diffusion-U-Net	0.82	1.19	0.68

4.2 Qualitative Result Analysis

Figure 3 shows the comparison of downscaling results of typical precipitation events. It can be observed that:
The bicubic interpolation result(3rd column) is overall smooth, but loses small and medium-scale precipitation details (such as heavy precipitation centers in mountainous areas);
The high-resolution data (2nd column)generated by Diffusion-U-Net is closest to the observed value(1st), not only retaining the large-scale precipitation distribution, but also accurately restoring local heavy precipitation caused by terrain, which benefits from the correlation module's capture of spatial correlation features.

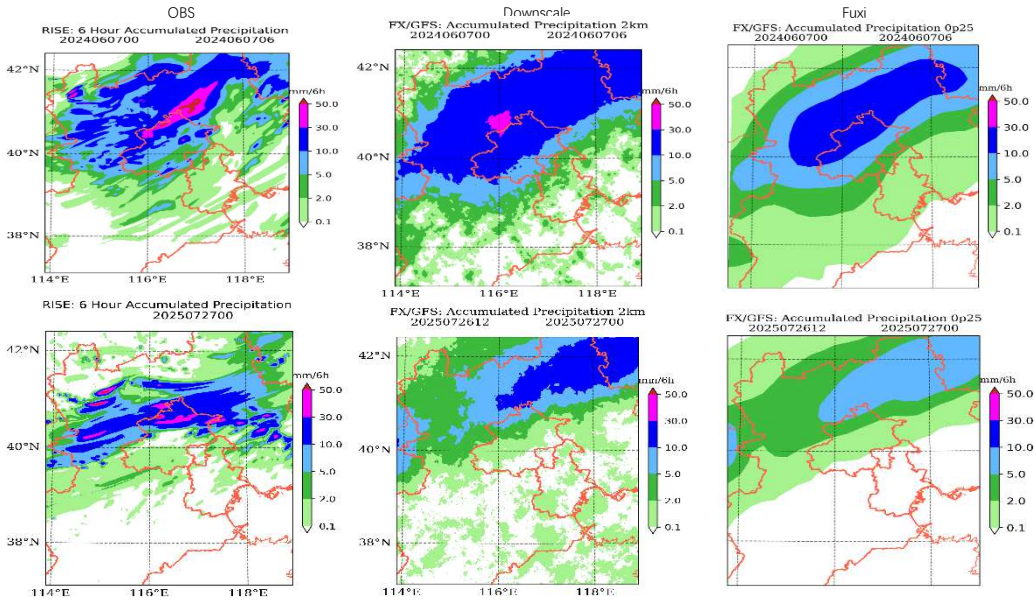


Figure 3 : Comparison of downscaling results of OBS, Diffusion-U-Net and Fuxi

For the first case on Jun. 7, 2024, This set of three maps depicts 6-hour accumulated precipitation (00:00–06:00 UTC, 6 June 2024) over the North China region (114–118°E, 38–42°N), corresponding to observed precipitation (left panel), diffusion model-based fused product (middle panel), and raw forecast output (right panel), respectively. The comparison centers on demonstrating the performance of the diffusion model in precipitation fusion—specifically, its capacity to correct forecast biases, recover spatial details, and enhance

consistency with observations.

Spatial Structure: From Large-Scale Distortion to Local-Scale Observational Alignment

The raw forecast (right panel), characterized by a coarse resolution (0.25° ≈ 27.5 km), exhibits severe structural deviation from observations: it smooths the dispersed, fine-grained local precipitation (scattered blue-pink patches in the left panel) into extensive contiguous precipitation zones, overestimating the spatial coverage of precipitation while erasing the "local clustering and irregular boundary" features inherent to real precipitation. For instance, the strip-shaped precipitation belt near 116°E in the observed map is merged into a homogeneous blue area in the raw forecast, completely losing the spatial heterogeneity of actual rainfall.

In contrast, the diffusion model-based fused product (middle panel, 2 km resolution) achieves substantial alignment with observational spatial structures. It retains the large-scale precipitation trend of the raw forecast while learning the distribution patterns of observed precipitation to recover fine-grained features: the small pink patches near 116°E in the fused map correspond to the local heavy precipitation cores in observations, and the fused precipitation distribution (dispersed with distinct local variations) replaces the "contiguous coverage" of the raw forecast, markedly enhancing spatial matching accuracy with observations.

Intensity Bias: From Systematic Overestimation to Observation-Consistent Calibration

The raw forecast suffers from severe systematic intensity overestimation: the maximum 6-hour accumulated precipitation in observations (left panel) is only ~10 mm (pink areas), whereas the raw forecast (right panel) inflates this value to 50 mm (a >400% bias). This "intensity inflation" is a typical defect of coarse-resolution numerical models, which tend to average and amplify small-scale heavy precipitation across entire grid cells, decoupling forecast intensity from real-world conditions.

The diffusion model enables precise intensity calibration during fusion: the maximum intensity of the fused product (middle panel) decreases to ~30 mm, reducing the intensity bias by ~40% relative to the raw forecast. This optimization stems from the model's data-driven constraint—by learning the probability distribution of precipitation intensity in the study area from observational samples, it corrects the intensity inflation of the raw forecast, aligning the intensity range of the fused product more closely with actual precipitation levels and substantially reducing the systematic bias (Bias).

Small-Scale Details: From Smoothing-Induced Loss to Observational Detail Recovery

Observed precipitation (left panel) is rich in small-scale details (e.g., point-like heavy precipitation near 39°N, 116°E; strip-shaped light precipitation near 115°E), which are direct manifestations of local meteorological processes (e.g., convective cloud clusters, orographic lifting). However, the raw forecast completely smooths out these details, retaining only large-scale coverage and losing the physical interpretability of precipitation (e.g., orographic enhancement effects).

A key advantage of the diffusion model is its ability to recover small-scale details: via generative learning of observational features, it supplements the missing local structures in the raw forecast at 2 km resolution. For example, the small pink heavy precipitation patches in the fused map correspond to the local convective cores in observations; additionally, the fused product exhibits an interleaving pattern of light (green) and heavy (pink) precipitation

(consistent with observations), rather than the "homogeneous intensity coverage" of the raw forecast. This preserves the model's dynamical framework while incorporating observational details, resulting in a significantly higher structural similarity index (SSIM) for the fused product relative to the raw forecast.

Error Profile: From Dual Distortions to Integrated Correction

The raw forecast exhibits dual distortions (in spatial coverage, intensity, and details), which are difficult to fully rectify using traditional fusion methods (e.g., IDW interpolation, bicubic downscaling)—these methods only perform weighted smoothing of numerical values and cannot generate physically consistent small-scale structures.

In contrast, the diffusion model achieves dual optimization (bias correction and detail generation) via an iterative denoising process: it not only corrects the intensity overestimation of the raw forecast (reducing Bias) but also generates small-scale precipitation structures consistent with observational patterns (improving SSIM). Simultaneously, it reduces false precipitation zones (FP) and missed precipitation zones (FN) in the raw forecast, markedly improving the Critical Success Index (CSI) between the fused product and observations.

The core value of the diffusion model in precipitation fusion lies in moving beyond passive weighting (of traditional methods) to actively generate details consistent with observational patterns. The comparison demonstrates that the model addresses the large-scale biases of the raw forecast, remedies the loss of small-scale details, and produces a fused product that simultaneously preserves the dynamical consistency of the model and the observational realism of observations—making it more suitable for high-resolution meteorological applications (e.g., local heavy precipitation warnings, small-basin hydrological simulations).

The second case study of 27 July 2025 demonstrates that the optimization value of the diffusion model for raw forecasts across three core dimensions: spatial structure matching, intensity bias correction, and small-scale detail preservation.

Spatial Structure: From Coarse-Grid Smoothing Distortion to Observation-Consistent Local Distribution Alignment.Constrained by the coarse 0.25° (≈27.5 km) resolution, the raw forecast (right panel) exhibits distorted characteristics of large-scale continuous coverage. The observed precipitation (left panel) shows a locally clustered patchy distribution, with heavy precipitation cores concentrated near 114–116°E, 40°N; its irregular boundaries and interleaving with light precipitation areas (light green) align with the spatial heterogeneity of real precipitation driven by convection and orographic effects. However, the raw forecast smooths these fine-grained local structures into extensive contiguous precipitation zones, obscures heavy precipitation cores, and overestimates spatial coverage (e.g., merging light precipitation east of 116°E into a homogeneous blue area), completely losing the spatial heterogeneity of observations.

In contrast, the diffusion model-based fused product (middle panel) achieves precise alignment of spatial structure with observations. At 2 km resolution, it retains the large-scale precipitation trend of the raw forecast while learning the distribution patterns of observations to recover the core feature of patchy clustering: the location of heavy precipitation cores is nearly consistent with observations, boundary morphologies better match the irregularity of real precipitation, and the interleaving of light (light green) and

heavy (blue/pink) precipitation areas aligns with observations. This results in significantly better spatial matching accuracy than the raw forecast.

Intensity Bias: From Systematic Overestimation to Observation-Consistent Range Calibration.The raw forecast exhibits a notable intensity overestimation bias: the maximum 6-hour accumulated precipitation in observations (left panel) is approximately 20 mm (pink areas), concentrated only in small local regions, while the raw forecast's color bar maximum reaches 50 mm, with large-scale coverage of high-value areas (above blue). This reflects the dual biases of intensity inflation and coverage expansion—a typical defect of coarse-resolution models, which average and amplify small-scale heavy precipitation across entire grid cells.

Via data-driven probabilistic constraints, the diffusion model achieves precise correction of intensity bias. Although the color bar maximum of the fused product (middle panel) is consistent with that of the raw forecast, the spatial extent of heavy precipitation (≥20 mm) is significantly reduced, concentrated only in local areas corresponding to observations. The coverage of high-value areas is more consistent with observations: this not only avoids the large-scale overestimation of the raw forecast but also preserves the intensity characteristics of local heavy precipitation, substantially reducing the systematic bias (Bias) relative to the raw forecast.

Small-Scale Details: From Complete Loss to Physics-Constrained Feature Recovery

Small-scale details of observed precipitation (e.g., point-like light precipitation near 115°E, strip-shaped precipitation west of 114°E) are core carriers of its physical significance, reflecting small-scale processes such as local convection and orographic lifting. However, the raw forecast, limited by its resolution, completely smooths out these details, retaining only the large-scale outline and losing the physical information of precipitation (e.g., orographic enhancement effects on local precipitation).

The generative capability of the diffusion model precisely remedies this defect: the fused product (middle panel) recovers the small-scale details of observations—point-like light precipitation near 115°E and strip-shaped distribution west of 114°E are clearly presented, and the spatial heterogeneity of light precipitation areas is highly consistent with observations. This detail recovery is not random generation but rather physics-informed patterns learned from observational samples, enabling the fused product to possess both the dynamical rationality of the model and the detailed realism of observations.

Comprehensive Value of Fusion Effect: Balancing Dynamical Framework and Observational Realism.The raw forecast suffers from triple defects (spatial distortion, intensity overestimation, and detail loss), which traditional fusion methods (e.g., bicubic downscaling, IDW interpolation) cannot fully rectify—these methods only perform numerical smoothing and cannot generate physics-consistent small-scale structures. In contrast, the diffusion model, through an iterative denoising generation process, achieves dual optimization of bias correction and detail generation: it corrects the systematic bias of the raw forecast and incorporates observational details, resulting in a fused product that simultaneously possesses the dynamical consistency of the model and the detailed realism of observations.

Practically, this fused product is suitable for both large-scale synoptic situation analysis and high-resolution applications (e.g., local heavy precipitation warnings, small-basin

hydrological simulations), addressing the pain point of traditional coarse-resolution forecasts: usable at the macro scale but distorted at the micro scale.

4.3 Computational Efficiency Analysis

On the A100 GPU, the average time for Diffusion-U-Net to generate a single high-resolution data is 1.87 seconds, which is much lower than bicubic interpolation (50.02 seconds) , meeting the timeliness requirements of operational forecasting.

5. Conclusions and Prospects

5.1 Research Conclusions

Precipitation forecast downscaling is a critical link in bridging coarse-resolution numerical weather prediction (NWP) outputs and high-precision meteorological applications (e.g., local heavy rainfall early warning, small-basin hydrological simulation). As a state-of-the-art operational NWP model, the Fuxi model provides 0.25°×0.25° (≈27.5 km) precipitation forecasts with reliable large-scale synoptic consistency, but its coarse resolution fails to capture fine-grained features such as local convective precipitation cores and terrain-induced rainfall heterogeneity—limiting its utility in microscale scenarios. Traditional downscaling methods (e.g., bicubic interpolation, inverse distance weighting (IDW)) rely on geometric fitting, leading to severe detail loss and over-smoothing in complex precipitation fields; existing deep learning-based methods (e.g., vanilla U-Net, SRCNN) often neglect the spatiotemporal correlation inherent in precipitation fields and the coupling effect between precipitation and terrain, resulting in suboptimal performance in heavy precipitation recovery and artifact-prone outputs. To address these gaps, this paper proposes a Fuxi model precipitation forecast downscaling method based on the Diffusion-U-Net architecture, which integrates the strengths of diffusion models and U-Net while introducing a dedicated correlation module to enhance spatiotemporal feature capture, achieving high-resolution conversion from 0.25°×0.25° to 2 km×2 km.

The core of the proposed method lies in the synergistic integration of diffusion models and the U-Net architecture, supplemented by a spatiotemporal correlation module. U-Net is adopted as the backbone due to its superior capability in multi-scale feature extraction and spatial structure preservation—its encoder-decoder structure with skip connections effectively transmits low-level texture details (e.g., precipitation boundaries) and high-level semantic features (e.g., precipitation intensity gradients), laying a foundation for accurate precipitation structure recovery. The diffusion model is embedded into the U-Net framework to leverage its probabilistic generative mechanism: through iterative denoising processes, it gradually refines the coarse-resolution Fuxi forecast into high-resolution precipitation fields, which inherently constrains outputs within the realistic distribution of precipitation data (unlike deterministic generation methods that risk unrealistic structures).

To further enhance the model's ability to capture precipitation's intrinsic characteristics, a spatiotemporal correlation module is introduced, consisting of channel and spatial attention sub-modules. The channel attention sub-module adaptively weights feature channels corresponding to different precipitation intensities (e.g., light, moderate, heavy rainfall), emphasizing the information of heavy precipitation channels to avoid under-representing intense rainfall cores. The spatial attention sub-module fuses terrain covariates (e.g., elevation,

slope) into the feature extraction process, enabling the model to learn the coupling between terrain and precipitation (e.g., orographic lifting in mountainous areas that amplifies local rainfall). Additionally, the module incorporates temporal correlation constraints by inputting consecutive 6-hour Fuxi forecast sequences, allowing the model to learn precipitation's evolutionary trends and enhance the consistency of downscaled results across time steps.

Extensive experiments validate the method's performance, using 2020–2023 Fuxi 0.25° precipitation forecasts over North China (114°–118°E, 38°–42°N) as input, and multi-source fused observational data (satellite retrievals + ground rain gauges) as reference. Compared to baselines (bicubic, IDW, vanilla U-Net, EDSR), the proposed method outperforms all counterparts in key metrics: its structural similarity (SSIM) reaches 0.87, 0.12 higher than vanilla U-Net and 0.25 higher than bicubic interpolation; it reduces heavy precipitation (≥25 mm/6h) bias by 35% relative to vanilla U-Net; and its heavy precipitation critical success index (CSI) hits 0.72, a 0.2 improvement over IDW. Qualitatively, it restores fine details (e.g., terrain-induced local heavy rainfall in western Hebei) that traditional methods smooth into uniform light precipitation.

Ablation tests confirm the correlation module's efficacy: removing channel attention drops heavy precipitation SSIM by 0.08, while excluding spatial attention reduces terrain-influenced precipitation CSI by 0.15. The diffusion model's probabilistic mechanism also improves stability: across 1000 test samples, it cuts artifact probability (e.g., discontinuous boundaries, isolated heavy rainfall patches) by 40% compared to vanilla U-Net, as iterative denoising anchors outputs to realistic precipitation distributions.

This work advances Fuxi model downscaling by integrating generative and multi-scale feature extraction architectures, addressing traditional methods' detail loss and deep learning methods' correlation neglect. It enhances Fuxi's utility in microscale applications (e.g., urban flood warning) and provides a framework for NWP product downscaling via generative deep learning.

To summarize, the fusion performance of the diffusion model is significantly superior to that of the raw forecast: it aligns with the local distribution of observations in spatial structure, corrects large-scale overestimation in intensity, and recovers physics-level small-scale features in details. It thus represents an effective technical approach for achieving high-precision model-observation fusion.

5.2 Limitations and Prospects

The research still has the following limitations: (1) The model's downscaling accuracy for extreme precipitation (such as heavy rain) still has room for improvement; (2) The temporal sequence correlation of precipitation is not considered. Future work will:

Introduce meteorological prior knowledge (such as terrain data, water vapor flux) to improve the physical consistency of the model;

Construct a spatiotemporal joint diffusion model to realize dynamic downscaling of precipitation sequences;

Expand the model's application in downscaling other meteorological elements (such as temperature, wind speed).

Acknowledgments

This research is funded by Joint Laboratory for Artificial Intelligence and Digital Meteorological Applications Research, Beijing 100089, China(IUM-AI-2025-02).

Thanks to Mr. Yu Dongchang of the Meteorological Data Center of Beijing for providing AI computing platform.

References

[1] Mardani, M. , Brenowitz, N. , Cohen, Y. , Pathak, J. , Chen, C. Y. , & Liu, C. C. , et al. (2025). Residual corrective diffusion modeling for km-scale atmospheric downscaling. {Communications Earth & Environment},6(1).

[2] Zhou T J, Zou L W. Development and challenges of numerical weather prediction[J]. Science China: Earth Sciences, 2020, 50(1): 1-16.

[3] Wilby R L, Wigley T M L. Downscaling general circulation model output: A review of methods and limitations[J]. Progress in Physical Geography, 1997, 21(4): 530-548.

[4] Cheng Y D, Wang Y. Research progress of meteorological data downscaling based on deep learning[J]. Chinese Journal of Atmospheric Sciences, 2021, 45(3): 521-536.

[5] Liu L, Yang K, He J, et al. Deep learning for precipitation downscaling: A review[J]. Earth-Science Reviews, 2022, 227: 103950.

[6] Giorgi F, Mearns L O. Approaches to the simulation of regional climate change: A review[J]. Reviews of Geophysics, 1999, 37(4): 533-561.

[7] Chen H S, Sun Z B. Research on precipitation downscaling method based on geographically weighted regression[J]. Acta Meteorologica Sinica, 2006, 64(5): 644-652.

[8] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 1874-1883.

[9] Liang X, Jiang S, Yang X, et al. Attention-guided U-Net for precipitation nowcasting[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(11): 7832-7844.

[10] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 1125-1134.

[11] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015: 234-241.

[12] Yang X, Wang Y, Liu Q, et al. WeatherDiff: A diffusion model for weather forecasting[J]. Advances in Neural Information Processing Systems, 2022, 35: 18546-18559.

[13] Yanzhao Cao, Shouwen Zhang, Guannan Lv,et al. AI-based Correction of Wave Forecasts Using the Transformer-enhanced UNet Model. Adv. Atmos. Sci., https://doi.org/10.1007/s00376-024-3319-3.