https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

China Regional 3km Downscaling Based on Residual Corrective Diffusion Model

Honglu Sun[1,2], Hao Jing[1,2], Zhixiang Dai[3], Sa Xiao[1,2], Wei Xue[4], Jian Sun[1,2], and Qifeng Lu[1,2]

[1]State Key Laboratory of Severe Weather Meteorological Science and Technology (LaSW), Beijing, China
[2]CMA Earth System Modeling and Prediction Centre (CEMC), Beijing, China
[3]NVIDIA
[4]Tsinghua University, Beijing, China

Correspondence: Hao Jing (jingh@cma.gov.cn)

Abstract. A fundamental challenge in numerical weather prediction (NWP) is efficiently producing high-resolution forecasts. A widely adopted solution is to downscale coarser global model outputs. This study focuses on statistical downscaling, which uses historical data to learn mappings between low- and high-resolution meteorological fields. Deep learning has become a key tool for this, enabling powerful super-resolution models like diffusion models and Generative Adversarial Networks for downscaling applications. Herein, we leverage CorrDiff, a diffusion-based downscaling framework, with three key enhancements relative to its original implementation. First, the study domain is expanded to nearly 40 times the spatial extent of the original version. Second, the scope of target downscaling variables is extended beyond surface variables to incorporate upper-air variables across six pressure levels. Third, a global residual connection is integrated to further boost prediction accuracy. To produce 3km resolution forecasts for the China region, the trained CorrDiff model is applied to 25km global grid forecasts derived from two sources: a conventional global NWP model and a deep learning-based weather model developed using Spherical Fourier Neural Operators. The China Meteorological Administration Mesoscale Model (CMA-MESO) is selected as the baseline for comparative evaluation. Experimental results demonstrate that the downscaled forecasts generated by our framework outperform the direct forecasts of CMA-MESO for almost all target variables, as quantified by the Mean Absolute Error metric. Specifically, forecasts of radar composite reflectivity reveal that CorrDiff, as a generative model, is capable of capturing fine-grained meteorological details, yielding more physically realistic predictions than deterministic regression-based downscaling models.

1 Introduction

Gridded meteorological forecasts are important in various fields such as transportation, energy sector, agriculture, and scientific research. In particular, high spatial resolution forecasts are crucial for local studies and risk assessment. Traditionally, global gridded forecasts are obtained from numerical weather prediction models. Generating km-resolution forecasts is challenging for numerical weather prediction models due to the limitation of computation time. Currently, there are also data-driven deep learning models that generate global forecasts Bi et al. (2022); Lam et al. (2023); Pathak et al. (2022); Chen et al. (2023). However, the resolution of current data-driven models generally ranges from 10 to 25 kilometers. To generate high-resolution

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

forecasts, a practical approach is to derive regional high-resolution predictions by applying a downscaling technique to the low-resolution outputs of global models.

Downscaling methods can be categorized into three types: dynamical downscaling, statistical downscaling, and combined methods. Dynamical downscaling, similar to numerical weather prediction models, is based on a set of atmospheric dynamical equations. It derives high-resolution forecasts by solving these equations using initial and lateral boundary conditions provided by global models. Statistical downscaling establishes statistical relationships between low-resolution variables of global models and high-resolution variables using historical data. These relationships are then applied for future forecasting. Compared to dynamical downscaling, statistical downscaling offers advantages, including simpler implementation, lower computational cost, and potentially higher accuracy.

Many machine learning models have been applied for statistical downscaling, such as multiple linear regression Schoof and Pryor (2001), support vector machine Chen et al. (2010), random forest Davy et al. (2010), and artificial neural networks Laddimath and Patil (2019). Among these models, artificial neural networks, which have evolved into deep learning Sun et al. (2024), are likely the most promising approach.

The fast development in deep learning over the past decade has led to numerous active research areas, including super-resolution in computer vision. Various high-performance super-resolution models have been proposed, such as Generative Adversarial Networks (GANs) and diffusion models. Super-resolution is similarity to downscaling: the input for both is a low-resolution grid, and the output is a high-resolution grid. However, there is also a difference between super-resolution and downscaling: the goal of super-resolution is to generate visually realistic images, while meteorological downscaling must ensure accuracy and physical consistency. Given the similarity between super-resolution and downscaling, many researchers have investigated the application of such super-resolution models to downscaling Watt and Mansfield (2024); Addison et al. (2022). In parallel, there are also works that developed specific neural network structures for downscaling Wu et al. (2024).

This work investigates a diffusion-based downscaling model named Corrective Diffusion (CorrDiff) Mardani et al. (2025). CorrDiff is a two-step approach that includes the training of a regression model and the training of a diffusion model to improve the predictions of the regression model. In Mardani et al. (2025), CorrDiff is applied to the Taiwan region, the resolution of the inputs is 25km and the resolution of the outputs is 2km. The size of the 2km high-resolution grid is 448×448. In this study, we apply CorrDiff on the China region, the resolutions of the inputs and the outputs are 25km and 3km respectively. The size of our 3km high-resolution grid is 1600×2400, which is nearly 20 times the size in Mardani et al. (2025). Our models are trained on reanalysis data, including 25km ECMWF Reanalysis v5 (ERA5) Hersbach et al. (2020) (as low-resolution inputs of the downscaling models) and 3km reanalysis data (as high-resolution labels) that are produced by China Meteorological Administration Regional Reanalysis Atmospheric System (CMA-RRA). In contrast to Mardani et al. (2025), which focuses mainly on surface variables, in this work multiple variables are considered for downscaling, including surface variables and variables at six pressure levels. The prediction of radar composite reflectivity is also investigated. Different combinations of input and output variables are examined in order to understand the intervariable dependencies in the downscaling task. By connecting our downscaling models to global forecasts, 3km regional forecasts for the China region are obtained, which are evaluated through comparison with the China Meteorological Administration Mesoscale Model (CMA-MESO). CMA-MESO

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

is a high-resolution regional numerical weather prediction model that generates 3km and 1km resolution forecasts of the China region. Two global forecasts are considered: the China Meteorological Administration-Global Forecast System (CMA-GFS) and Sphere Fusion Forecast (SFF), which is a deep learning-based weather model. The experimental results demonstrate that, for the Mean Absolute Error (MAE), our forecasts generally outperform those of CMA-MESO for the target variables. Our assessment of the uncertainty estimated by CorrDiff shows that, for any downscaled variable, the uncertainty is correlated with the accuracy. For the prediction of radar reflectivity, our results show that deterministic models usually have a severe over-smoothing problem, while CorrDiff can generate realistic small-scale features that are similar to reanalysis data.

This paper is organized as follows. Section 2 introduces the high-resolution reanalysis data used for training and the CorrDiff framework. Section 3 presents the training setup, the evaluation on the validation data, covering prediction errors and the properties of the uncertainty estimated by our models, and assesses the 3km forecasts generated by applying our models to the forecasts of global models (CMA-GFS and SFF). Section 4 draws conclusions and discusses future directions.

2 Data and Method

2.1 3km Resolution Reanalysis Data

The 3km resolution reanalysis data used in this study are produced by CMA-RRA. CMA-RRA was developed by the CEMC and is based on the 3km rapid refresh cycling assimilation and forecast system of CMA-MESO. Its resolution is 3km/1hour and covers the same region as CMA-MESO: 10–60.1°N, 70–145°E.

CMA-RRA comprises multiple modules, including observational data preprocessing and quality control, three-dimensional variational data assimilation, cloud analysis, incremental analysis update (IAU), large-scale background field initialization, multi-scale hybrid assimilation, digital filtering, and mesoscale model forecasting. CMA-RRA operates through an inner analysis cycle and an outer cycle. The inner analysis cycle uses an hourly assimilation-forecast cycle and the hydrometeor variables are not updated during the analysis process. Using the results from the inner cycle analysis, the outer cycle conducts cloud analysis using networked Doppler weather radar 3D reflectivity products and Fengyun geostationary meteorological satellite cloud products. This process generates cloud analysis and hourly precipitation products.

2.2 Problem Formulation

$\{x_i \in \mathbb{R}^{c_{in} \times p \times q} \mid i \in \{1, ..., N\}\}$ and $\{y_i \in \mathbb{R}^{c_{out} \times m \times n} \mid i \in \{1, ..., N\}\}$ represent 25km and 3km resolution meteorological data over the China region, respectively. c_{in} (c_{out}) is the number of variables of the 25km (3km) data. p and q (m and n) represent the number of grid points in the meridional and zonal directions respectively of 25km (3km) data. N is the number of training samples. In this work, $p = 192$, $q = 288$, $m = 1600$, and $n = 2400$. The 25km data correspond to the region 12.25-60°N, 70-141.75°E and the 3km data correspond to the region 12.13-60.1°N, 70-141.97°E. Different combinations of c_{in} and c_{out} are investigated, which will be discussed in Section 3.1. The objective is to train a deep learning model to predict y_i from x_i.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

In the case of using a regression model (in this paper, regression models refer to deterministic deep learning models, such as UNet), first, the input x_i is interpolated onto the 3km resolution grid by bilinear interpolation, which is denoted as $bilinear(x_i)$, then $bilinear(x_i)$ is fed into a deep learning model that can be represented by a function $f : \mathbb{R}^{c_{in} \times p \times q} \mapsto \mathbb{R}^{c_{out} \times m \times n}$. $f(bilinear(x_i))$ is the prediction of the 3km grid that corresponds to x_i. The training of this deep learning model is to find parameters of f that minimize a loss function quantifying the distance between $f(bilinear(x_i))$ and y_i for $i \in \{1, ..., N\}$.

In this work, we also use diffusion models. Generally, a diffusion model involves two processes: noising and denoising. Noising is the process that progressively adds noise to a clean image until the image becomes random noise. Denoising is the process that transforms random noise into a structured image. Basically, the training of a diffusion model can be considered as the training of a denoising model. In the case of using diffusion models for downscaling, the low-resolution data are also fed into the diffusion models. Such models are also known as conditional diffusion models. The overall inference process of a conditional diffusion model for downscaling can be represented by a function $f : \mathbb{R}^{c_{in} \times p \times q}, E \mapsto \mathbb{R}^{c_{out} \times m \times n}$, which is a conditional sampling function of the probability of the 3km resolution data given a 25km resolution input. E represents the set of random noise that follows a certain distribution (which normally follows the normal distribution). Note that this function f is highly complex because the denoising process contains multiple steps that progressively decrease the noise level. For more details on diffusion models, see Ho et al. (2020).

2.3 Corrective Diffusion Model

This section briefly introduces the Corrective Diffusion (CorrDiff) model, for more details see Mardani et al. (2025). The code implementation is based on PhysicsNeMo (https://github.com/NVIDIA/physicsnemo). The training of CorrDiff has two steps. In the first step, we train a regression model (UNet), denoted as f, to minimize a loss function between its predictions $f(bilinear(x_i))$ and the targets y_i for all $i \in \{1, ..., N\}$. In the second step, we train an Elucidated Diffusion Model (EDM) Karras et al. (2022) to correct the predictions made by f. We denote the overall inference process of EDM by a function g. After training, the CorrDiff prediction $\hat{y}_i$ for the target y_i is given by $\hat{y}_i = g(bilinear(x_i), f(bilinear(x_i)), \epsilon) + f(bilinear(x_i))$, where ϵ is a random noise sampled from a standard normal distribution. By sampling multiple random noises, a CorrDiff model generates an ensemble of predictions, thereby producing a probabilistic prediction of y_i.

Following Mardani et al. (2025), we use a specific architecture of UNet Karras et al. (2022) for both the regression (first step) and diffusion (second step) networks. This architecture has 6 encoder and decoder layers, and it incorporates attention mechanisms and residual connections within its structure.

Fig. 1 shows the overall architecture of one of our trained CorrDiff models. Our regression networks differ slightly from the models in Mardani et al. (2025): for the downscaling of a variable, assuming that x_i^u and y_i^v correspond to this variable (where $x_i^u \in \mathbb{R}^{p \times q}$ is a channel of x_i, and $y_i^v \in \mathbb{R}^{m \times n}$ is a channel of y_i), instead of directly predicting y_i^v, we predict the residual $y_i^v - bilinear(x_i^u)$, because our experimental results show that this residual learning strategy could accelerate convergence speed and improve accuracy for the downscaling task. In contrast to Mardani et al. (2025) which only considers the downscaling of surface variables, this work also investigates the downscaling of pressure level variables. An additional input (3km resolution

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

orography) is included in our models as well. The implementation of our models is based on the code provided in Mardani et al. (2025).

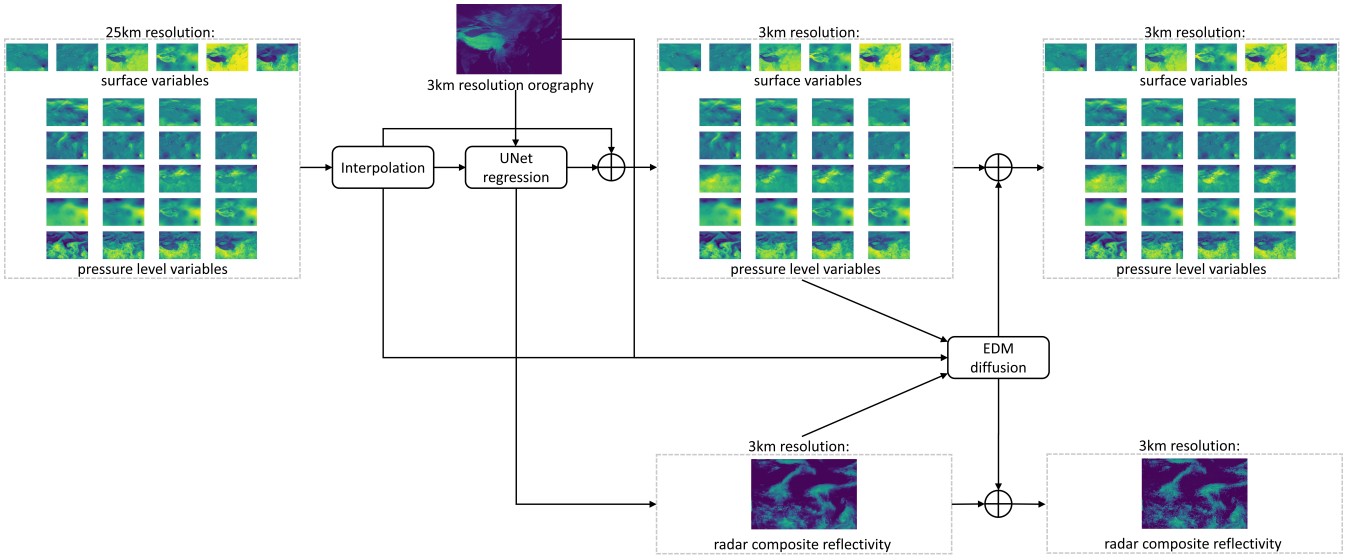

Figure 1. Illustration of the overall architecture of one of our trained CorrDiff models.

As stated in Mardani et al. (2025), the development of CorrDiff was motivated by the limitations of using conditional diffusion models for downscaling. It is assumed that there is a significant distribution shift between low-resolution and high-resolution data that hinders the learning process. To avoid this problem, the generation is decomposed into two steps. The first step aims to predict the conditional mean using regression (UNet), and the second step learns a correction using a diffusion model.

3 Experiments and Results

3.1 Training Setup

For the training of CorrDiff, we use 25km ERA5 as input data and 3km reanalysis data as target data. The size of the 25km input data is 192 × 288 covering the region 12.25-60°N, 70-141.75°E, and the size of the 3km target data is 1600 × 2400 covering the region 12.13-60.1°N, 70-141.97°E. The mismatch in spatial resolution (25km vs. 3km) (non-divisibility of 25 by 3) results in minor misalignment between input and target region, which could potentially reduce the accuracy on the boundary. We use the data from 2019 to 2022 as the training set, choosing eight time points each day: 00:00, 03:00, 06:00, 09:00, 12:00, 15:00, 18:00, and 21:00 UTC. In total, the training set contains 11688 samples.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

For deep learning models, we have the flexibility to select various combinations of variables as inputs and outputs. In this work, we investigate four distinct input/output configurations, detailed in Table 1. For example, for a model that corresponds to Combination 4, the input has 35 variables ($c_{in} = 35$) and the output has 24 variables ($c_{out} = 24$).

3.2 Training Results on the Validation Set

The data from January, April, July, and October 2023 are used as the validation set. We select eight time points each day as in the training set: 00:00, 03:00, 06:00, 09:00, 12:00, 15:00, 18:00, and 21:00 UTC. In total, there are 984 samples in the validation set.

3.2.1 Regression Models

Five regression models have been trained, denoted as Regression 1, Regression 2, Regression 3-1, Regression 3-2, and Regression 4. Regression 1, 2, and 4 correspond to Combination 1, 2, and 4 in Table 1, respectively. Regression 3-1 and 3-2 correspond to Combination 3.

- Regression 1 is the first trained model. It incorporates the most variables to test the feasibility of downscaling multiple variables with a single model. In order to ensure sufficient model complexity, we increase the size of Regression 1 such that only one sample can be computed at one time on a single NVIDIA H20 GPU with the use of checkpointing. The UNet embedding size of Regression 1 is [128, 256, 512, 512, 1024].

- Regression 2 builds upon Regression 1 by adding orography as an additional input variable in order to increase accuracy. Furthermore, radar composite reflectivity is added as an output for Regression 2, while high-altitude variables (100 and 200 hPa) are removed for both input and output, since high-altitude forecasting is not our main focus.

- Regression 3-1 modifies Regression 2 by excluding radar reflectivity as an output but reintroducing the 100 and 200 hPa variables as inputs, under the assumption that additional inputs would not harm performance. Regression 3-2 is a smaller version of Regression 3-1 with reduced model complexity, designed to test if a reduced-size network could maintain comparable accuracy. The UNet embedding size of Regression 3-2 is [32, 64, 128, 256, 256].

- Regression 4 is implemented based on Regression 3-1 but excludes total column integrated water vapour and surface pressure to ensure compatibility with SFF (see Section 3.3.1), which does not produce these two variables. For Regression 1, 2, 3-1 and 3-2, a batch size of 64 is used to ensure stable gradient estimations, while for Regression 4, we experimented with a smaller batch size (a batch size of 8) to explore its effects on training dynamics, acknowledging that this modification would decrease the time per step due to our use of gradient accumulation.

The training time per epoch on 8 NVIDIA H20 GPUs is approximately 3.5–4 hours for Regression 1, 2, and 3-1; 1.5 hours for Regression 3-2; and 3 hours for Regression 5.

The validation curves for the five regression models are shown in Fig. 2. Since only Regression 2 predicts radar reflectivity, the validation curve of radar reflectivity is excluded due to the absence of comparative benchmarks.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

Table 1. Different input/output variable combinations.

Combinations	Variables	Input		Output	
		Height Levels (m)	Pressure Levels (hPa)	Height Levels (m)	Pressure Levels (hPa)
Combination 1	Zonal Wind (u)	10	100 200 500 700 850 925	10	100 200 500 700 850 925
	Meridional Wind (v)	10	100 200 500 700 850 925	10	100 200 500 700 850 925
	Geopotential Height (z)	-	100 200 500 700 850 925	-	100 200 500 700 850 925
	Temperature (t)	2	100 200 500 700 850 925	2	100 200 500 700 850 925
	Specific Humidity (q)	-	100 200 500 700 850 925	-	100 200 500 700 850 925
	Total Column Integrated Water Vapour	-	Integrated	-	Integrated
	Mean Sea Level Pressure	-	Surface	-	Surface
	Surface Pressure	-	Surface	-	Surface
Combination 2	Zonal Wind (u)	10	500 700 850 925	10	500 700 850 925
	Meridional Wind (v)	10	500 700 850 925	10	500 700 850 925
	Geopotential Height (z)	-	500 700 850 925	-	500 700 850 925
	Temperature (t)	2	500 700 850 925	2	500 700 850 925
	Specific Humidity (q)	-	500 700 850 925	-	500 700 850 925
	Total Column Integrated Water Vapour	-	Integrated	-	Integrated
	Mean Sea Level Pressure		Surface	-	Surface
	Surface Pressure	-	Surface	-	Surface
	Orography	-	Surface	-	-
	Radar Composite Reflectivity	-	-	-	Surface
Combination 3	Zonal Wind (u)	10	100 200 500 700 850 925	10	500 700 850 925
	Meridional Wind (v)	10	100 200 500 700 850 925	10	500 700 850 925
	Geopotential Height (z)	-	100 200 500 700 850 925	-	500 700 850 925
	Temperature (t)	2	100 200 500 700 850 925	2	500 700 850 925
	Specific Humidity (q)	-	100 200 500 700 850 925	-	500 700 850 925
	Total Column Integrated Water Vapour	-	Integrated	-	Integrated
	Mean Sea Level Pressure		Surface	-	Surface
	Surface Pressure	-	Surface	-	Surface
	Orography	-	Surface	-	-
Combination 4	Zonal Wind (u)	10	100 200 500 700 850 925	10	500 700 850 925
	Meridional Wind (v)	10	100 200 500 700 850 925	10	500 700 850 925
	Geopotential Height (z)	-	100 200 500 700 850 925	-	500 700 850 925
	Temperature (t)	2	100 200 500 700 850 925	2	500 700 850 925
	Specific Humidity (q)	-	100 200 500 700 850 925	-	500 700 850 925
	Mean Sea Level Pressure		Surface	-	Surface
	Orography	-	Surface	-	-

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

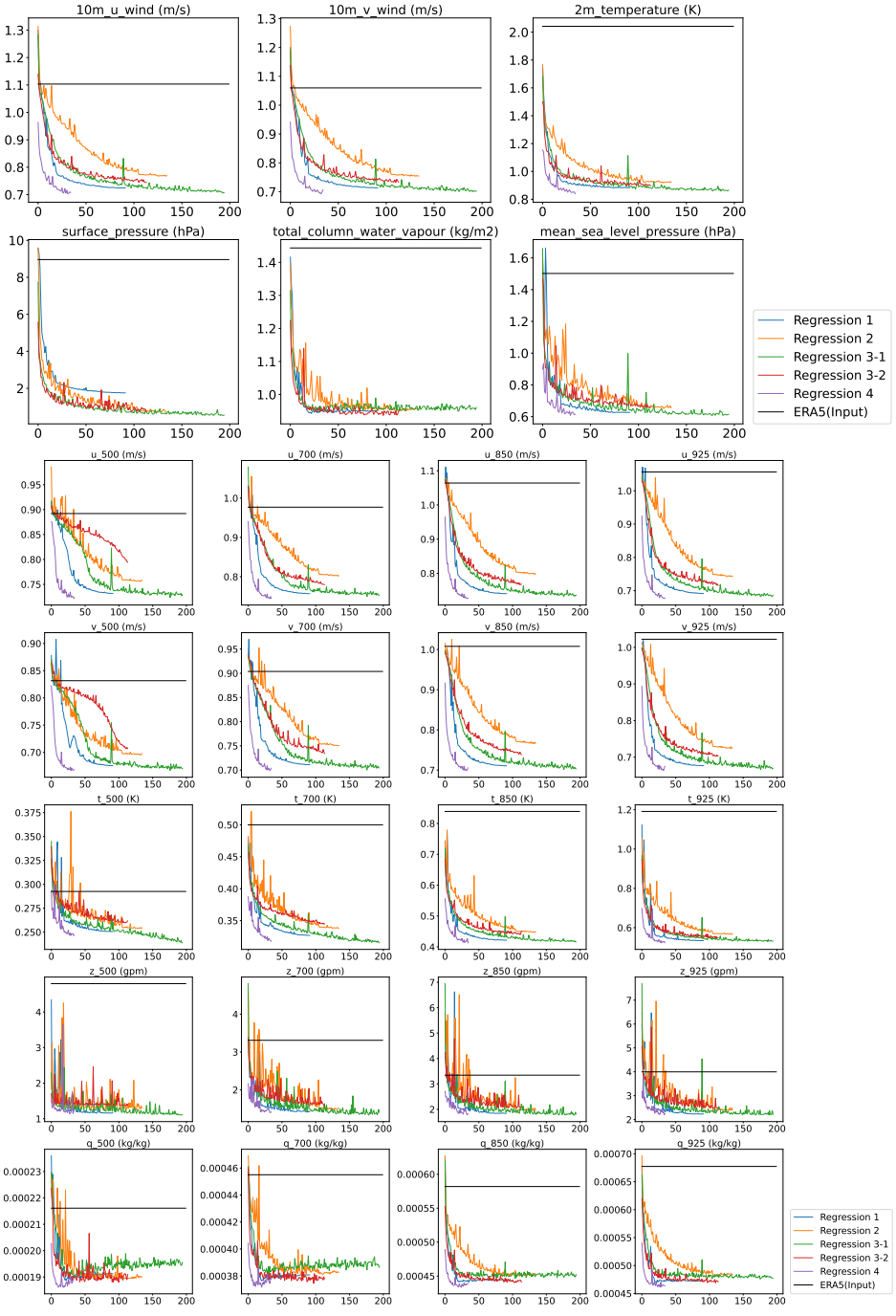

Figure 2. Validation curves of regression models for surface variables (top) and pressure level variables (bottom). Each subplot corresponds to a target variable. The curves plot the Mean Absolute Error (y-axis) against the number of training epochs (x-axis). ERA5 is used as the ground truth.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

In each subplot, the black horizontal line denotes the MAE of the ERA5 input for a variable (MAE between $bilinear(x_i^u)$ and y_i^v for all (x_i, y_i) in the validation set, where x_i^u and y_i^v correspond to this variable). After the first few epochs, all validation curves remain below the black line, indicating that all regression models consistently outperform simple interpolation. The validation curves exhibit distinctive patterns between variables. For example, the validation curves of the 10m wind decrease with a progressively decreasing rate of decline, while, for the 500 hPa wind, the rate of decline decreases at first and increases after several epochs. In addition, all validation curves display a globally decreasing trend, except for specific humidity (q), suggesting the existence of overfitting for this variable. These inter-variable differences imply that employing variable-specific embedding methods could potentially enhance both the convergence speed and model accuracy.

The comparative analysis between the validation curves of Regression 1 and Regression 2 shows that Regression 1 significantly outperforms Regression 2 on all variables except surface pressure. It is important to note that we did not exhaustively explore all combinations, for example, between Combination 1 and Combination 2, we can also examine another combination that does not include radar composite reflectivity as well as the 100 and 200 hPa variables. We assume that the impact of the 100 and 200 hPa variables on the downscaling of other variables exists but is minor and that including orography as an input should not harm the performance. Therefore, the results suggest that integrating radar reflectivity inference with downscaling in a single model might compromise downscaling accuracy, and adding orography to the inputs increases the accuracy of surface pressure.

As expected, Regression 3-1, which excludes radar composite reflectivity but includes orography as an additional input, outperforms both Regression 1 and 2. The reduction of model complexity in Regression 3-2 causes a performance degradation for most variables, indicating that this reduced version does not have sufficient model complexity.

The curves of Regression 4 are significantly lower than those of the previous models. This improvement can be attributed to two potential reasons: first, the smaller batch size may provide better gradient estimations for this downscaling task; second, excluding surface pressure and total column integrated water vapour likely reduces task complexity, freeing up model capacity for the remaining variables. Consequently, further adjustment to the batch size and a more sophisticated selection of input/output variables could lead to additional performance gains.

In the rest of this paper, we mainly focus on analyzing Regression 2 and Regression 4, as only Regression 2 can output radar composite reflectivity, and Regression 4 is compatible with SFF and has the highest accuracy on the validation set.

3.2.2 CorrDiff Models

The CorrDiff models that correspond to Regression 1, 2, and 4 are denoted as CorrDiff 1, 2, and 4, respectively. Their MAE scores on the validation set are presented in Table 2. All reported MAE scores for the CorrDiff models are computed using a single sample. Our results indicate that the MAE scores of the regression models are consistently lower than those of the CorrDiff models. In fact, the degradation in MAE of the CorrDiff models compared to that of the regression models is also reported in Mardani et al. (2025). As stated in Mardani et al. (2025), the degradation is expected as the diffusion models optimize the Kullback-Leibler divergence as opposed to the regression models that minimize the MSE loss. Following Mardani et al. (2025), we also compute the Continuous Ranked Probability Score (CRPS) Hersbach (2000), see Table 3. CRPS can be

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

considered as a generalization of MAE for probability prediction. Since the computation of CRPS is slow with our current implementation (primarily due to the large grid size), we select a more restricted set of time points for its calculation: 00:00 and 12:00 UTC on the first day of each month in 2023 and we only focus on CorrDiff 4. As expected, the CRPS values are generally the lowest.

Table 2. MAE on the validation set.

Variable (Unit)	ERA5 (Input)	Regression 1	Regression 2	Regression 4	CorrDiff 1	CorrDiff 2	CorrDiff 4
10m Zonal Wind (m/s)	1.10	0.72	0.77	**0.70**	1.05	1.25	0.98
10m Meridional Wind (m/s)	1.05	0.71	0.75	**0.70**	1.04	1.16	0.97
2m Temperature (K)	2.04	0.88	0.92	**0.83**	1.10	1.21	1.05
Mean Sea Level Pressure (hPa)	1.50	0.62	0.66	**0.61**	0.88	1.07	0.81
Surface Pressure (hPa)	8.96	1.75	**0.77**	–	1.88	2.01	–
Total Column Integrated Water Vapour (kg/m2)	1.44	**0.95**	0.95	–	1.29	1.47	–
u100 (m/s)	0.79	**0.71**	–	–	1.10	–	–
u200 (m/s)	0.99	**0.94**	–	–	1.43	–	–
u500 (m/s)	0.89	0.73	0.75	**0.72**	1.11	1.21	1.06
u700 (m/s)	0.97	0.75	0.80	**0.74**	1.13	1.31	1.08
u850 (m/s)	1.06	0.74	0.80	**0.72**	1.09	1.28	1.03
u925 (m/s)	1.05	0.69	0.74	**0.67**	1.01	1.13	0.98
v100 (m/s)	0.78	**0.63**	–	–	0.95	–	–
v200 (m/s)	1.00	**0.89**	–	–	1.35	–	–
v500 (m/s)	0.83	0.67	0.69	**0.66**	1.02	1.08	0.96
v700 (m/s)	0.90	0.71	0.75	**0.70**	1.08	1.22	1.02
v850 (m/s)	1.00	0.71	0.76	**0.70**	1.07	1.25	1.01
v925 (m/s)	1.02	0.67	0.72	**0.67**	0.99	1.14	0.96
z100 (gpm)	12.91	**2.08**	–	–	4.91	–	–
z200 (gpm)	9.49	**1.75**	–	–	4.42	–	–
z500 (gpm)	4.79	**1.16**	1.53	1.27	2.67	3.44	2.55
z700 (gpm)	3.30	1.43	1.50	**1.41**	2.74	3.23	2.43
z850 (gpm)	3.34	1.84	1.96	**1.82**	3.45	3.97	3.28
z925 (gpm)	3.99	**2.23**	2.39	2.24	3.97	6.49	4.33
t100 (K)	0.58	**0.45**	–	–	0.64	–	–
t200 (K)	0.36	**0.29**	–	–	0.44	–	–
t500 (K)	0.29	0.25	0.25	**0.24**	0.40	0.46	0.38
t700 (K)	0.50	0.32	0.33	**0.31**	0.49	0.56	0.45
t850 (K)	0.83	0.42	0.45	**0.41**	0.61	0.68	0.57
t925 (K)	1.19	0.53	0.56	**0.52**	0.75	0.92	0.69
q100 (kg/kg)	8.55×10^{-7}	$\mathbf{6.90 \times 10^{-7}}$	–	–	1.32×10^{-6}	–	–
q200 (kg/kg)	7.52×10^{-6}	$\mathbf{6.86 \times 10^{-6}}$	–	–	1.06×10^{-5}	–	–
q500 (kg/kg)	0.00021	0.00019	0.00018	**0.00018**	0.00026	0.00028	0.00025
q700 (kg/kg)	0.00045	0.00038	0.00038	**0.00037**	0.00051	0.00058	0.00052
q850 (kg/kg)	0.00058	0.00044	0.00045	**0.00043**	0.00060	0.00068	0.00058
q925 (kg/kg)	0.00067	0.00047	0.00048	**0.00046**	0.00062	0.00071	0.00061
Radar Composite Reflectivity (dBz)	–	–	6.64	–	–	8.17	–

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

Table 3. MAE and CRPS for the data at 00:00 and 12:00 UTC on the first day of each month in 2023 of Regression 4 and CorrDiff 4.

Variable (Unit)	Regression 4 (MAE)	CorrDiff 4 (MAE)	CorrDiff 4 (CRPS)
10m Zonal Wind (m/s)	0.73	1.02	**0.55**
10m Meridional Wind (m/s)	0.73	1.01	**0.55**
2m Temperature (K)	0.88	1.10	**0.66**
Mean Sea Level Pressure (hPa)	0.58	0.78	**0.45**
u500 (m/s)	0.73	1.09	**0.56**
u700 (m/s)	0.76	1.10	**0.59**
u850 (m/s)	0.75	1.08	**0.57**
u925 (m/s)	0.70	1.01	**0.53**
v500 (m/s)	0.68	0.99	**0.52**
v700 (m/s)	0.72	1.06	**0.55**
v850 (m/s)	0.73	1.05	**0.55**
v925 (m/s)	0.70	1.02	**0.54**
t500 (K)	0.25	0.38	**0.20**
t700 (K)	0.32	0.46	**0.24**
t850 (K)	0.41	0.57	**0.31**
t925 (K)	0.52	0.70	**0.40**
z500 (gpm)	1.26	2.53	**1.14**
z700 (gpm)	1.47	2.44	**1.22**
z850 (gpm)	1.84	3.33	**1.83**
z925 (gpm)	**2.23**	4.26	2.57
q500 (kg/kg)	0.00019	0.00026	**0.00015**
q700 (kg/kg)	0.00039	0.00054	**0.00031**
q850 (kg/kg)	0.00045	0.00060	**0.00035**
q925 (kg/kg)	0.00048	0.00064	**0.00038**

Some examples of the predictions of Regression 4 and CorrDiff 4 are given in Fig. 3. The local patterns that do not occur in the low-resolution data (top-right) are captured by both the regression and CorrDiff models. An example of typhoon downscaling is given in Fig. 4. The typhoon structures generated by both Regression 4 and CorrDiff 4 are more similar to the 3km reanalysis data compared to ERA5. The comparison between the outputs of Regression 4 and CorrDiff 4 reveals that the CorrDiff models generate more high-frequency details.

3.2.3 Uncertainty Estimation based on CorrDiff

A key advantage of diffusion models is that they can provide probabilistic predictions. The variance of the N-member ensemble predictions from CorrDiff can be used to quantify the predictive uncertainty. Fig. 5 illustrates the correlation between the absolute errors of Regression 4 and the variance of the 20-member CorrDiff 4 predictions for the data from 2023-10-01-00 UTC. The shape of the point sets reveals that grid points that exhibit both high absolute error and very low variance are rare, suggesting that low variance is a potential indicator of high accuracy. Due to the high density of points, it is difficult to clearly see the distribution of the points. So we plot three horizontal black lines that mark the 75th, 50th, and 25th percentile of the variance. These percentiles partition the data into four groups, each representing a different level of uncertainty (for example,

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

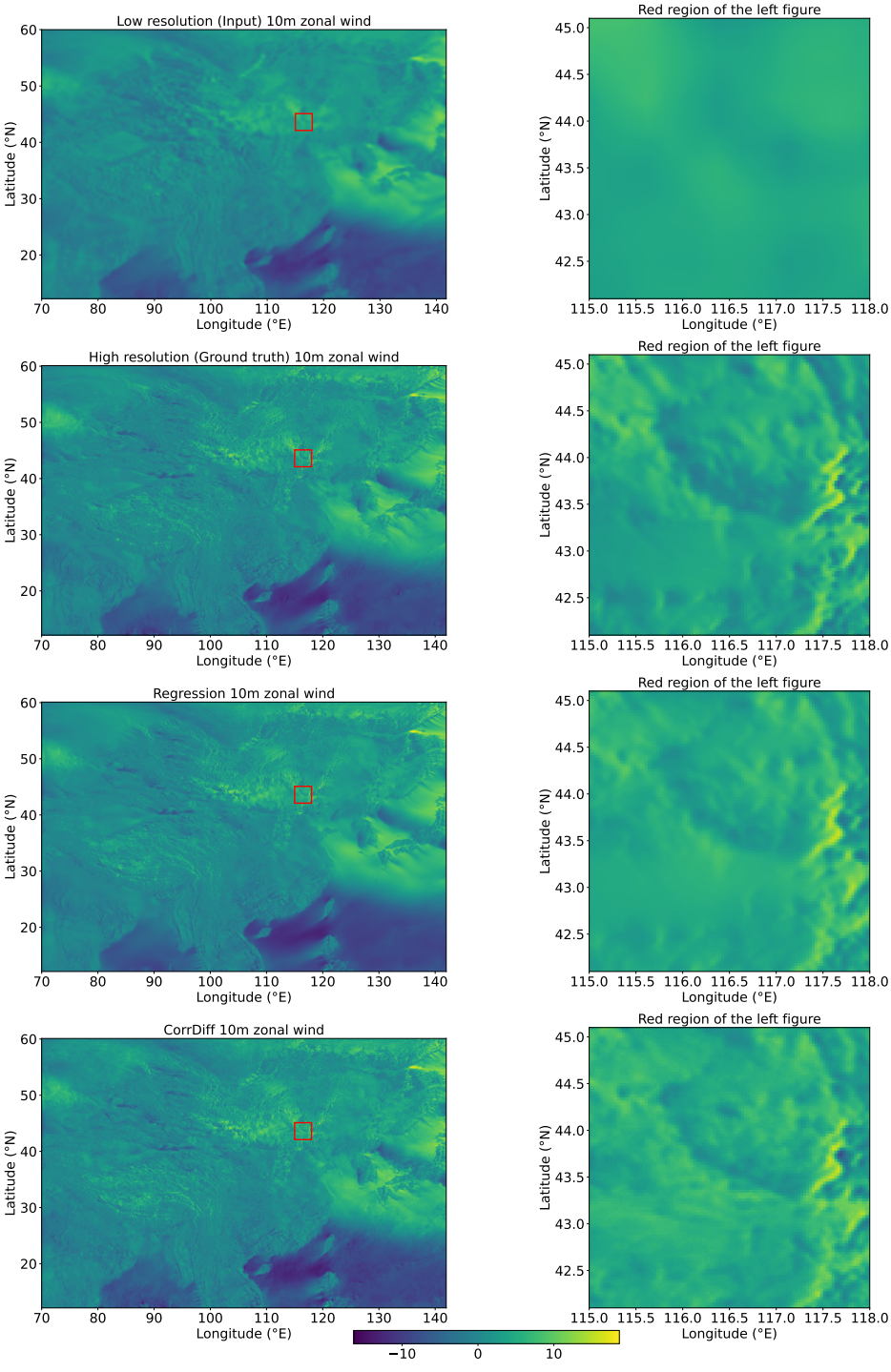

Figure 3. Illustration of the 10m zonal wind inference of Regression 4 and CorrDiff 4 for data from 2023-12-01-00 UTC. Each figure on the right is a zoomed view of the red box area in the left figure.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

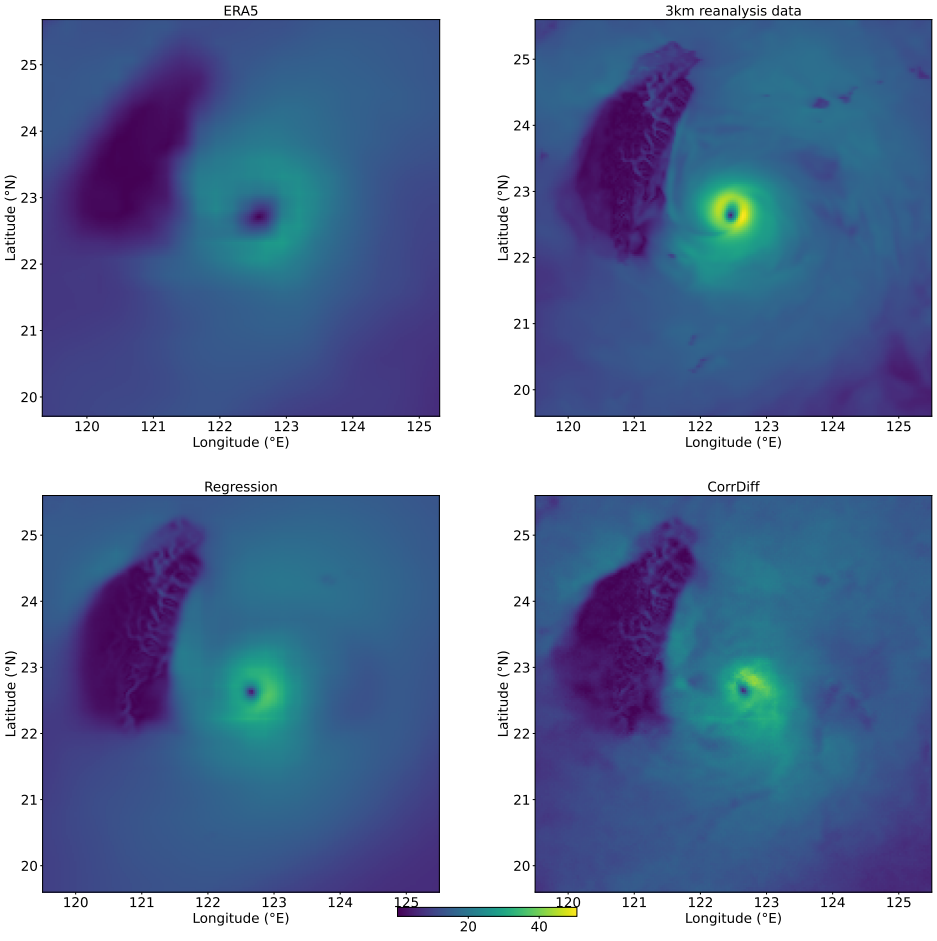

Figure 4. Illustration of the downscaling of Typhoon Haikui on 2023-09-03-00 UTC by Regression 4 and CorrDiff 4. Figures show the 10m wind speed.

220 points above the 75th percentile have the highest uncertainty). The MAE for each group is computed and annotated within the subplots. For all variables, the MAE decreases markedly as the variance decreases across these four groups. The results demonstrate that variance thresholds can be used to identify regions with higher or lower MAE.

The spatial distribution of these four uncertainty groups is illustrated in Fig. 6, where brighter areas represent higher variance. These results provide a more intuitive understanding of the predictive uncertainty. For example, the variance of the 2m
225 temperature over the ocean is significantly lower than that over the land; the variance of the geopotential height over the Qinghai-Tibet Plateau is generally higher than that in other regions; in the bottom row of Fig. 6, the variance in the lower part (which corresponds to areas with more water vapour) of each subplot is higher.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

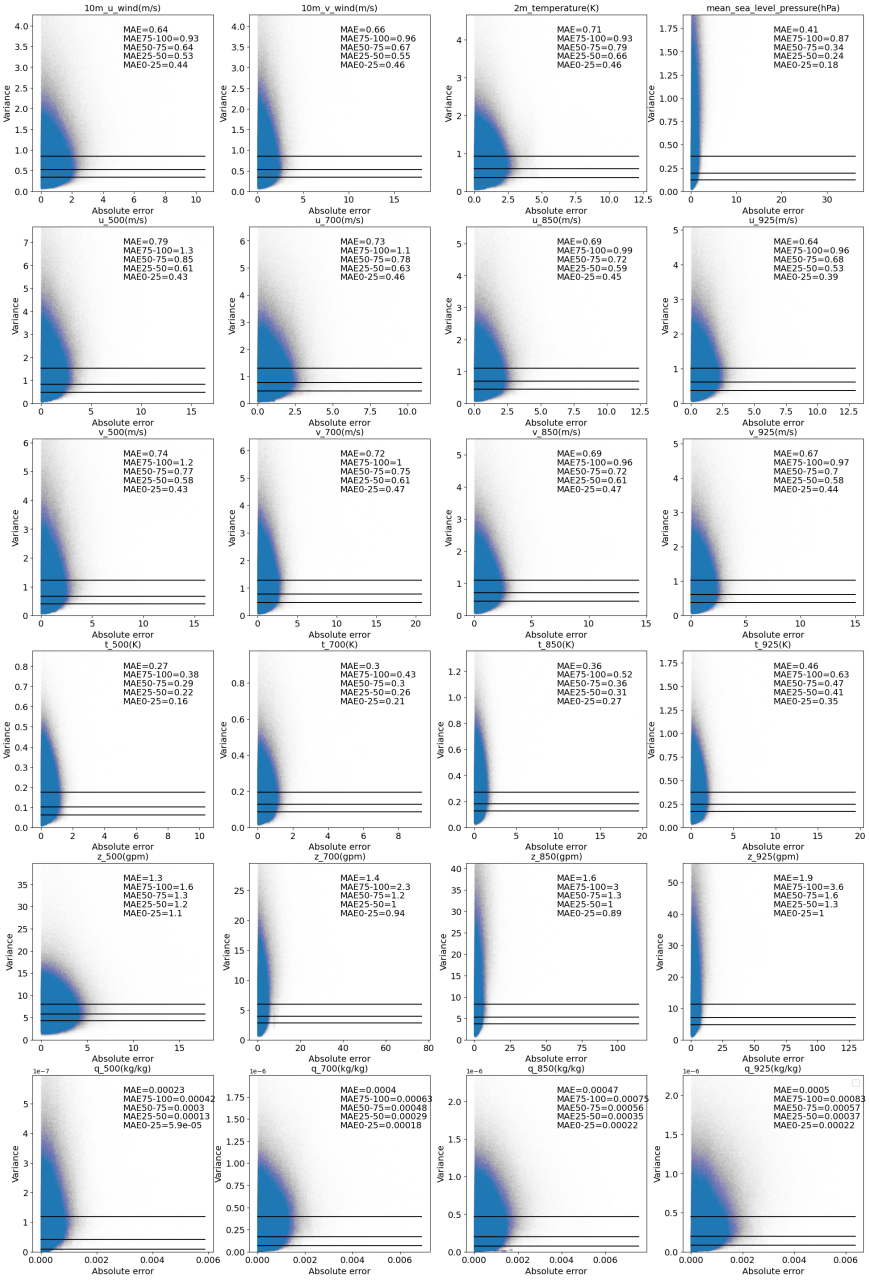

Figure 5. Correlation between the absolute error of Regression 4 and the variance of 20-member CorrDiff 4 predictions for the data from 2023-10-01-00 UTC. Each point in the figure corresponds to a grid point. The three black lines in each subplot represent the 75th, 50th and 25th percentile of the variance. These three lines separate the grid points into four groups. MAE75-100 is the MAE of the grid points above the 75th line, MAE50-75 is the MAE of the grid points between the 50th and 75th line, etc.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

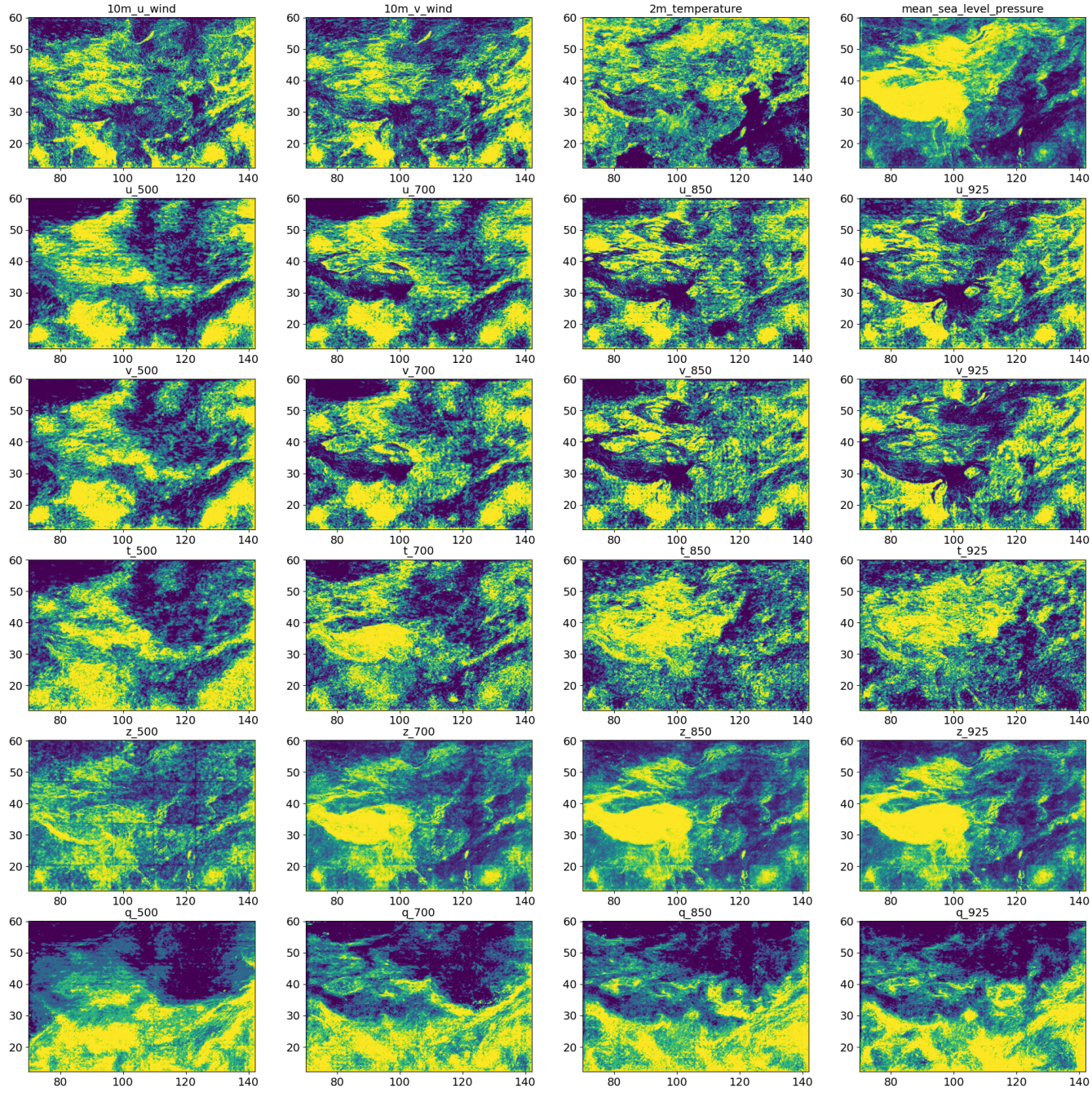

Figure 6. Illustration of the four uncertainty groups of grid points in Fig. 5. Brighter areas correspond to higher variance.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

3.3 Downscaling the SFF and CMA-GFS Forecasts

This section applies the trained downscaling models to the low-resolution outputs of the global forecast systems to generate 3km forecasts. Two global forecast systems are considered: SFF and CMA-GFS. The resulting 3km forecasts are evaluated against the CMA-MESO model, which serves as our baseline model. The 3km reanalysis data are used as the ground truth.

3.3.1 SFF

Sphere Fusion Forecast (SFF) is a data-driven deep learning-based weather model developed from Spherical Fourier Neural Operators (SFNO) Bonev et al. (2023). Compared to SFNO, two major improvements are made in SFF: the up-sampling and down-sampling operators between the SFNO blocks are added, allowing the initial and final stages of the SFNO block chain to handle broader frequency spectra, while the middle layers focus on relatively low-frequency information; a Vision Transformer–like architecture between the encoder and decoder is introduced as the skip connection, which improves the model's capacity for local feature learning, producing more robust and accurate forecasts.

3.3.2 CMA-GFS

China Meteorological Administration-Global Forecast System (CMA-GFS) is a global model developed by the CEMC. It comprises a semi-implicit semi-Lagrangian (SISL) non-hydrostatic dynamical core, a physical parameterization package, and a four-dimensional variational data assimilation system Chen et al. (2008). Currently, the horizontal resolution of CMA-GFS is $0.125° \times 0.125° (\approx 12.5\text{km})$.

3.3.3 3km Forecast Evaluation

Fig. 7 shows the MAE scores for the 3km forecasts that are based on the 25km forecasts of different global models. The current operational CMA-MESO provides forecasts up to 72 hours, so we also consider 72 hour forecasts for comparision. The downscaling results for ERA5 are also given in each subplot, which can be considered as an upper-bound for the downscaling task. For SFF, forecasts are generated from two initial fields: ERA5 and CRA1.5. In fact, the current version of SFF is trained on CRA1.5, which is a reanalysis dataset developed by the CMA and is utilized in the AIM-FDP project. The daily updates of CRA1.5 are available 2 hours behind real time. So, the SFF forecasts using CRA1.5 as initial fields can be considered as real-time forecasts.

In each subplot, regression models exhibit lower MAE scores than the corresponding CorrDiff models, which is consistent with the results on the validation set. In all figures, the curves of the downscaled ERA5 (black curves) exhibit the lowest MAE scores, followed by the downscaled SFF forecasts that take ERA5 as initial fields (red curves). For most variables except specific humidity, generally, the downscaled SFF forecasts that uses CRA1.5 as initial fields (purple curves) are better than the downscaled CMA-GFS forecasts (green curves), which in turn are better than the CMA-MESO forecasts (blue curves). In general, these results in Fig. 7 indicate that, in terms of the MAE scores, our data-driven models can potentially outperform CMA-MESO for most variables.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

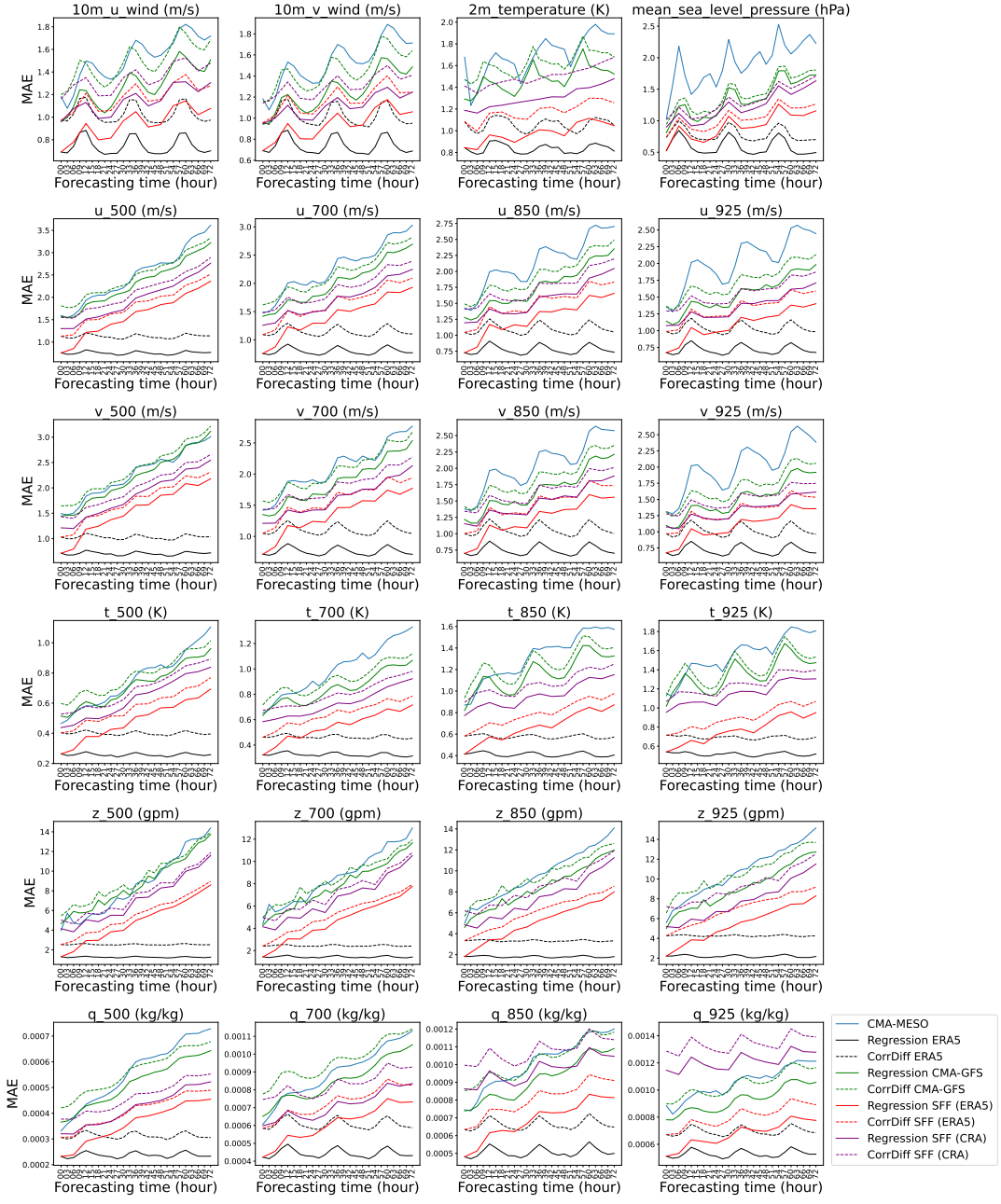

Figure 7. The MAE scores of the 3km forecasts that include the downscaled SFF forecasts (taking CRA1.5 or ERA5 as initial fields), the downscaled CMA-GFS forecasts and the CMA-MESO forecasts, using the 3km reanalysis data as the ground truth. The downscaled ERA5 are also present for comparison. The downscaling models are Regression 4 and CorrDiff 4. Seven initialization times are considered: 2023-03-01-00, 2023-06-01-00, 2023-09-01-00, 2023-12-01-00, 2023-05-24-00, 2023-07-30-00, and 2023-09-04-00 UTC. The last three times correspond to the periods of Typhoon Mawar, Khanun and Haikui, respectively.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

Fig. 8 show the area average of the 3km reanalysis data and 3km forecasts for the 2m temperature and 10m wind. The forecasts of the 2m temperature are fairly consistent with the 3km reanalysis data, but there are significant inter-model differences for the 10m wind. The 10m wind trend in the 3km reanalysis data clearly differs from that in the CMA-MESO forecasts. All forecasts based on the downscaling models exhibit trends more closely aligned with the 3km reanalysis data. Among them, the downscaled CMA-GFS forecasts produce higher wind speeds than the downscaled SFF forecasts. This difference is likely due to the over-smoothing of the outputs of data-driven models like SFF, which reduces intensity.

3.3.4 Case Study: Tropical Cyclone

Fig. 9 shows the 10m wind speed of the downscaling of Typhoon Khanun by Regression 4 and CorrDiff 4. The downscaling models are applied to the forecasts of CMA-GFS and SFF that use CRA1.5 as initial fields. Generally, the outputs of the regression models are smoother and less sharp than those of CMA-MESO, which should be due to the over-smoothing of data-driven models. Consequently, the Regression SFF forecasts, which involves two data-driven models (SFF and Regression 4), exhibit the most smoothness. The use of diffusion model (CorrDiff 4) notably reduces this effect. Because of the blurriness, the fine-scale structure of the typhoon cannot be clearly captured by data-driven models.

For the last row in Fig. 9, using the 3km reanalysis as the ground truth, the MAE scores of CMA-MESO, Regression CMA-GFS, and Regression SFF are 5.24 m/s, 4.82 m/s, and 2.18 m/s, respectively. This indicates that a lower MAE does not automatically yield a more realistic visual representation of typhoon winds. Note that the Khanun storm center location of SFF forecasts is closer to that of the 3km reanalysis data than that of CMA-GFS and CMA-MESO, which should be a reason for the lower MAE scores. In fact, typically data-driven models outperform numerical weather prediction models in the inference of the path of a typhoon, as data-driven models are usually better at predicting large-scale patterns. In order to improve the visual quality of the 10m wind speed forecasts for typhoons, designing a more sophisticated loss function could be a promising direction. Although, for these results, data-driven downscaling models do not generate visually realistic results of the 10m wind, their inference of the typhoon eye radius might be better than CMA-MESO, indicating that such data-driven models have the potential to refine the typhoon structure in coarse-resolution forecasts.

Fig. 10 shows the 3km radar composite reflectivity forecasts for Typhoon Khanun, obtained by applying Regression 2 and CorrDiff 2 to the CMA-GFS forecasts. The regression model exhibits overly smooth results, failing to predict high radar reflectivity, and it cannot reproduce small reflectivity cores, indicating the limitation of regression models to predict the fine-scale extreme convective precipitation. By contract, the CorrDiff model exhibits a more realistic spatial distribution. The high-frequency features of the CorrDiff model are closer to those of the 3km reanalysis data, compared to CMA-MESO.

Fig. 11 (top) shows the Probability Density Functions (PDF) for the 3km radar composite reflectivity forecasts of Typhoon Khanun. Overall, the PDF of the data-driven models aligns more closely with that of the 3km reanalysis than with that of CMA-MESO. This is particularly evident during the first 24 hours for reflectivity exceeding 50 dBZ, where the CorrDiff model (blue curve) more accurately replicates the reanalysis distribution (black curve) compared to CMA-MESO (green curve). For any forecast time, the regression model (red lines) dramatically underestimates the distribution beyond 50 dBz, which is a

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

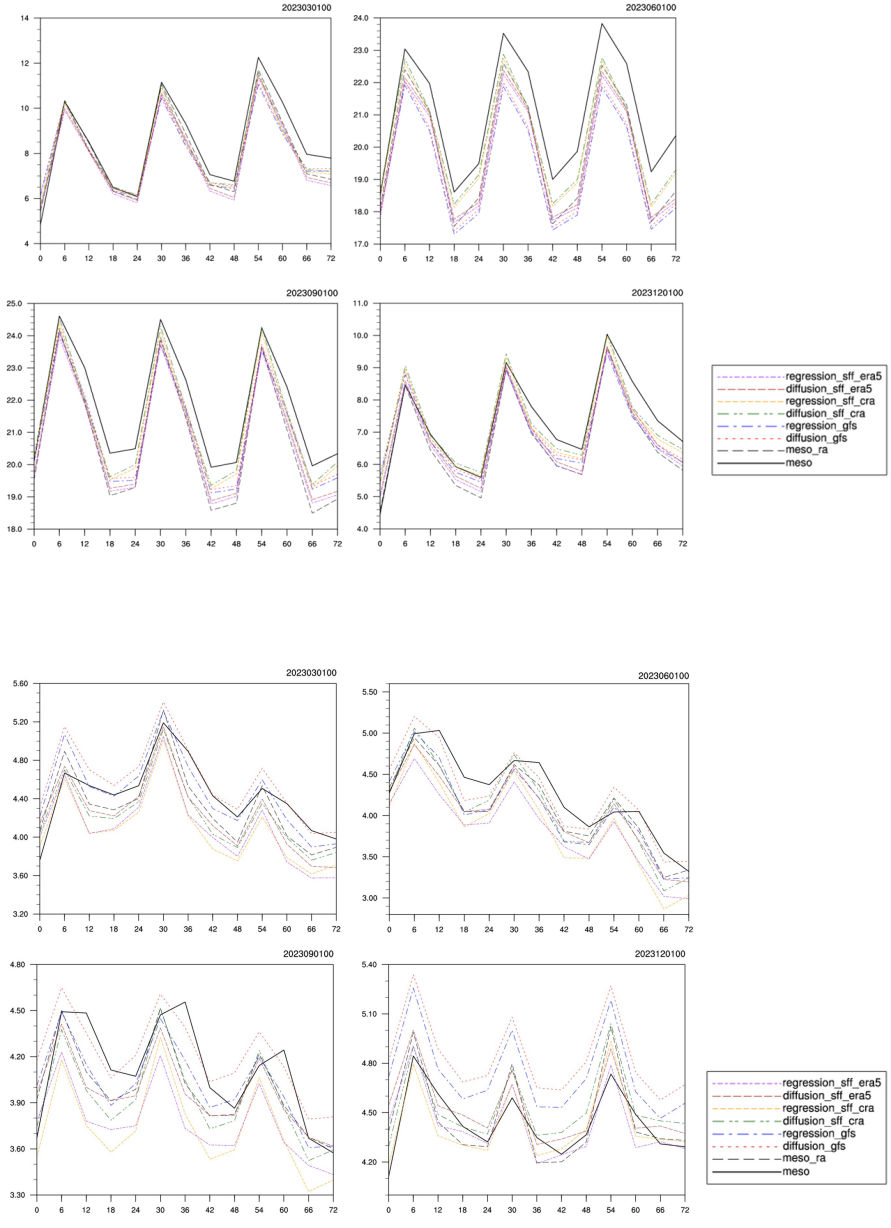

Figure 8. Area average of the 2m temperature (top) and the 10m wind (bottom) of the 3km reanalysis data and the 3km forecasts of CMA-MESO, downscaled SFF (taking CRA1.5 or ERA5 as initial fields), and downscaled CMA-GFS. Four initialization times are considered: 2023-03-01-00, 2023-06-01-00, 2023-09-01-00, and 2023-12-01-00 UTC. "meso_ra" denotes the 3km reanalysis; "regression" and "diffusion" refer to the Regression 4 and CorrDiff 4, respectively.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

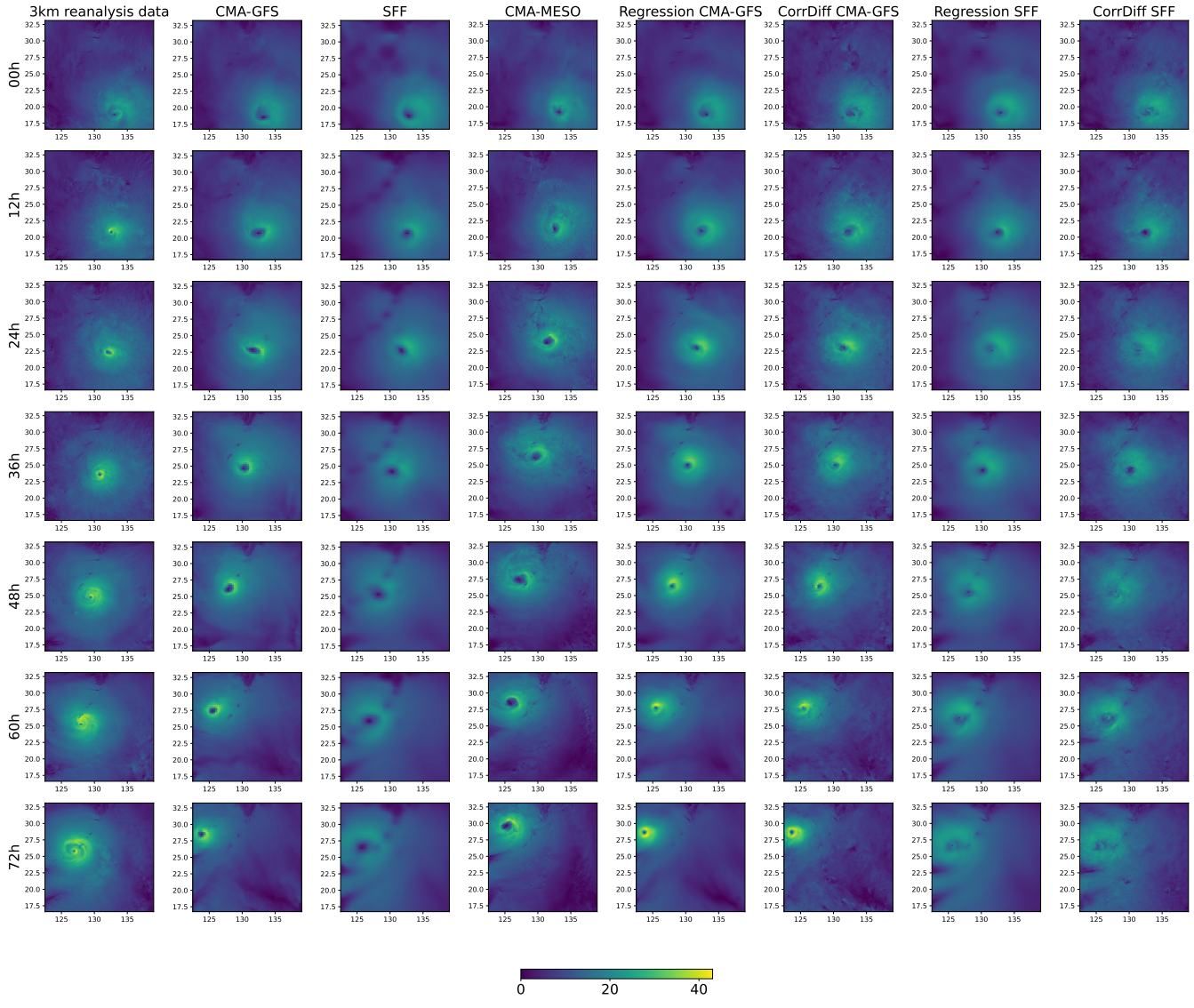

Figure 9. Illustration of the downscaling of CMA-GFS and SFF forecasts of Typhoon Khanun by Regression 4 and CorrDiff 4, compared to the 3km reanalysis data, CMA-GFS, SFF and CMA-MESO. Figures represent 10m wind speed. The initialization time is 2023-07-30-00 UTC. The initial fields of SFF are CRA1.5.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

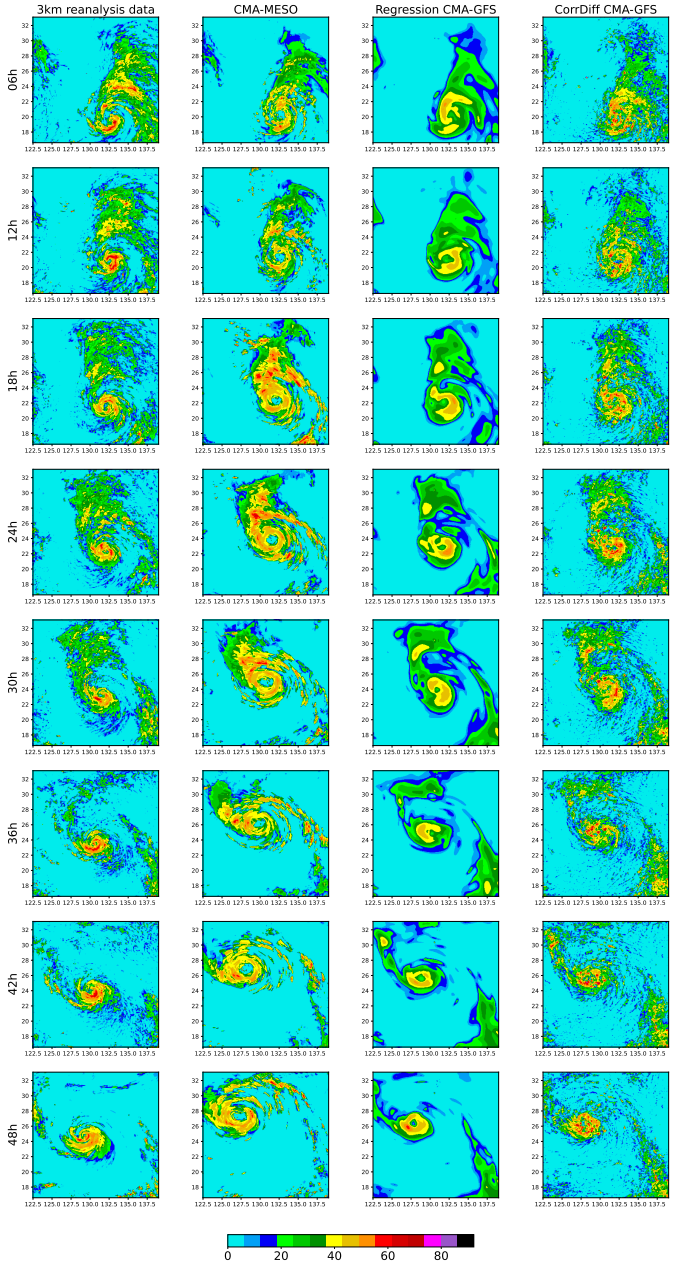

Figure 10. Illustration of the 3km radar composite reflectivity forecasts of Typhoon Khanun by applying Regression 2 and CorrDiff 2 on the CMA-GFS forecasts. The initialization time is 2023-07-30-00 UTC.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

consequence of over-smoothing, proving that diffusion-based models outperform deterministic models in the prediction of high radar reflectivity or precipitation.

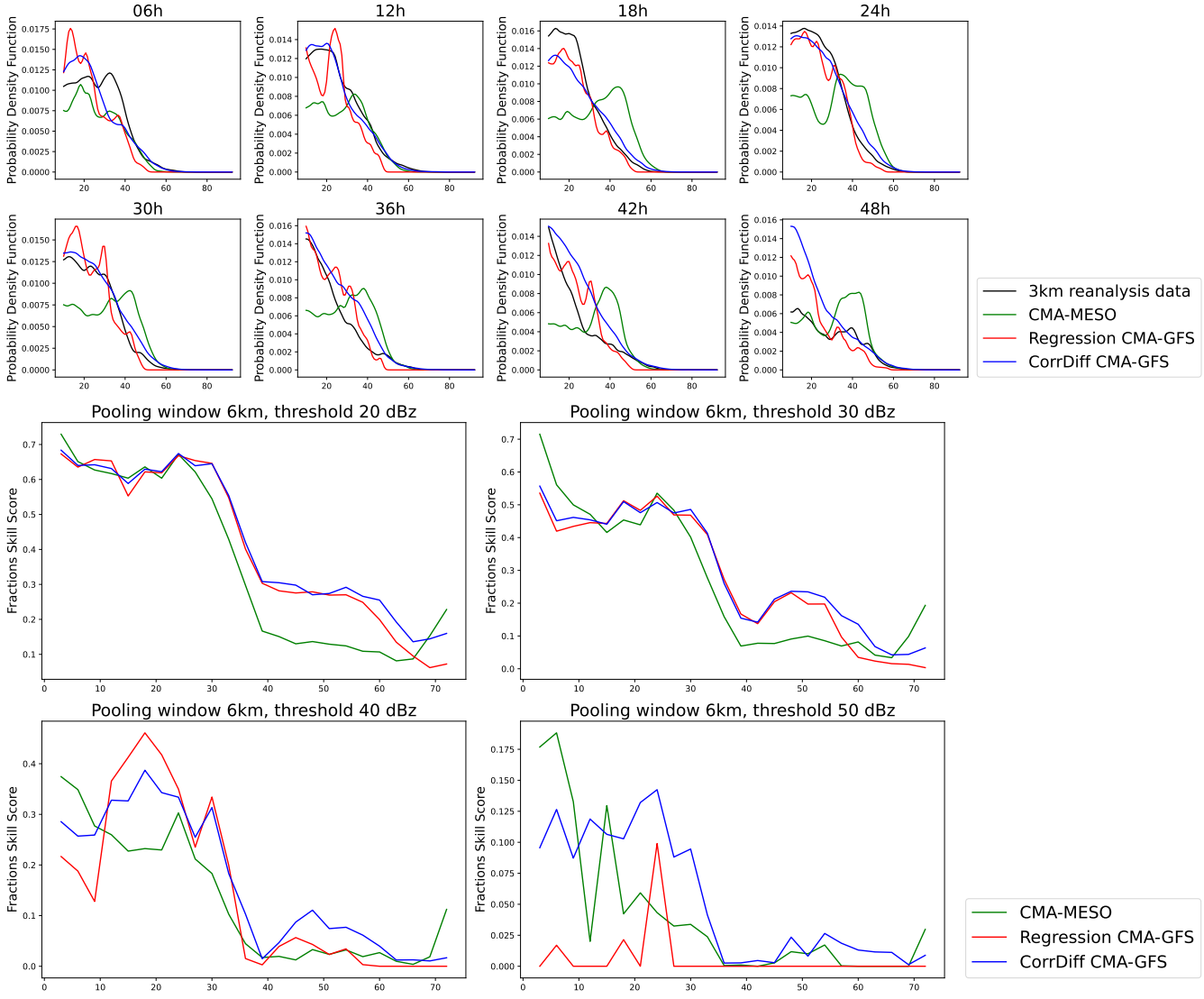

Figure 11. Top: Probability Density Functions of the 3km radar composite reflectivity forecasts of Typhoon Khanun by applying Regression 2 and CorrDiff 2 on CMA-GFS forecasts, compared with 3km reanlysis data and CMA-MESO. Only the probability density of the reflectivity that is higher than 10 dBz is shown for clarity. Bottom: Fractions Skill Scores of the 3km radar composite reflectivity forecasts of Typhoon Khanun by applying Regression 2 and CorrDiff 2 on CMA-GFS forecasts, compared with CMA-MESO, using 3km reanalysis data as ground truth. The initialization time is 2023-07-30-00 UTC.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

The Fractions Skill Scores (FSS) for the 3km radar composite reflectivity forecasts of Typhoon Khanun are given in Fig. 11 (bottom). For all models, there is a dramatic decrease in performance around 36 hours. In the first 9 hours, CMA-MESO achieves the best FSS. Between 12 and 36 hours, the CorrDiff model can outperform CMA-MESO. CorrDiff and regression models have similar FSS for lower reflectivity thresholds, while CorrDiff substantially outperforms the regression model for high thresholds, since the regression model fails to reproduce high reflectivity.

4 Conclusion

In this work, we train multiple CorrDiff models with different input/output combinations for 3km downscaling. Evaluation against the CMA-MESO baseline confirms the success of our approach, which generally outperforms CMA-MESO in terms of MAE for the target variables. However, it is important to note that a lower MAE does not necessarily correspond to better forecasts, particularly for extreme weather events such as typhoons. Experimental results also show that, for radar composite reflectivity inference, the CorrDiff model can produce more realistic high-frequency details than the regression model.

One major drawback of our CorrDiff models is that they have higher MAE scores compared to the regression models. This phenomenon was also observed in the original work of CorrDiff. However, in our experiments, the difference in the MAE scores between CorrDiff and regression models is more obvious, which might be due to the fact that the size of our high-resolution grid is much larger than that in the original work. Another limitation is the inference speed of CorrDiff models. Owing to the iterative denoising process, the inference time for a diffusion model is multiple times longer than that of a regression model. This computational overhead increases considerably when generating an ensemble of downscaling results.

Future work could explore several promising directions:

- In current work, only reanalysis data are used to train the models, in order to improve model accuracy, pre-trained models can be further finetuned on operational data. Moreover, they can be finetuned together with SFF.
- Our demonstration of a correlation between the predictive uncertainty quantified by CorrDiff and the MAE enables the future development of methods that use uncertainty estimates to mitigate forecast errors.
- A challenge remains in the interpretability of deep learning models Zhang et al. (2021). Various methods (for example, gradient-based approaches Simonyan et al. (2013); Sundararajan et al. (2017)) have been developed for computer vision tasks such as image classification. These techniques could be adapted to elucidate the predictions of CorrDiff by incorporating physical principles.

Code and data availability. The model code is available from https://doi.org/10.5281/zenodo.18604660 Sun (2026). The ERA5 data are publicly available from the Climate Data Store (CDS). The 3km resolution China regional reanalysis data are provided by the Chinese Meteorological Administration Earth System Modeling and Prediction Centre and can be obtained upon request.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

Author contributions. Honglu Sun developed the methodology and wrote the original draft. Hao Jing contributed to the methodology and reviewed and edited the manuscript. Zhixiang Dai conceived the study, contributed to the methodology, and participated in manuscript review. Sa Xiao contributed to conceptualization, resources, and data curation. Wei Xue contributed to conceptual design, supervised the entire research project, secured funding, and reviewed the manuscript. Jian Sun and Qifeng Lu administered the project and provided essential resources. All authors reviewed and approved the final manuscript.

Competing interests. The authors have no relevant competing interests to disclose.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Grant No. U2242210 and No. U2342220). We would like to thank Zhifang Xu and Jilin Wang for offering the 3km reanalysis data. We thank Li Zhang and Siyuan Sun for the evaluation of our 3km forecasts. We thank Simeng Qian and Chenyu Wang for providing the forecasts of SFF. We also thank Zhiyan Jin, Huadong Xiao, Qilong Jia and Tongda Xu for the fruitful discussions.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

References

Addison, H., Kendon, E., Ravuri, S., Aitchison, L., and Watson, P. A.: Machine learning emulation of a local-scale UK climate model, arXiv preprint arXiv:2211.16116, 2022.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Pangu-weather: A 3d high-resolution model for fast and accurate global weather forecast, arXiv preprint arXiv:2211.02556, 2022.

Bonev, B., Kurth, T., Hundt, C., Pathak, J., Baust, M., Kashinath, K., and Anandkumar, A.: Spherical fourier neural operators: Learning stable dynamics on the sphere, in: International conference on machine learning, pp. 2806–2823, PMLR, 2023.

Chen, D., Xue, J., Yang, X., Zhang, H., Shen, X., Hu, J., Wang, Y., Ji, L., and Chen, J.: New generation of multi-scale NWP system (GRAPES): General scientific design, Chinese Science Bulletin, 53, 3433–3445, 2008.

Chen, H., Guo, J., Xiong, W., Guo, S., and Xu, C.-Y.: Downscaling GCMs using the Smooth Support Vector Machine method to predict daily precipitation in the Hanjiang Basin, Advances in Atmospheric Sciences, 27, 274–284, 2010.

Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., and Li, H.: FuXi: a cascade machine learning forecasting system for 15-day global weather forecast, npj climate and atmospheric science, 6, 190, 2023.

Davy, R. J., Woods, M. J., Russell, C. J., and Coppin, P. A.: Statistical downscaling of wind variability from meteorological fields, Boundary-layer meteorology, 135, 161–175, 2010.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather and Forecasting, 15, 559–570, 2000.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly journal of the royal meteorological society, 146, 1999–2049, 2020.

Ho, J., Jain, A., and Abbeel, P.: Denoising diffusion probabilistic models, Advances in neural information processing systems, 33, 6840–6851, 2020.

Karras, T., Aittala, M., Aila, T., and Laine, S.: Elucidating the design space of diffusion-based generative models, Advances in neural information processing systems, 35, 26 565–26 577, 2022.

Laddimath, R. S. and Patil, N. S.: Artificial neural network technique for statistical downscaling of global climate model, Mapan, 34, 121–127, 2019.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al.: Learning skillful medium-range global weather forecasting, Science, 382, 1416–1421, 2023.

Mardani, M., Brenowitz, N., Cohen, Y., Pathak, J., Chen, C.-Y., Liu, C.-C., Vahdat, A., Nabian, M. A., Ge, T., Subramaniam, A., et al.: Residual corrective diffusion modeling for km-scale atmospheric downscaling, Communications Earth & Environment, 6, 124, 2025.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., et al.: Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators, arXiv preprint arXiv:2202.11214, 2022.

Schoof, J. T. and Pryor, S.: Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks, International Journal of Climatology: A Journal of the Royal Meteorological Society, 21, 773–790, 2001.

Simonyan, K., Vedaldi, A., and Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps, arXiv preprint arXiv:1312.6034, 2013.

Sun, H.: Downscaling model based on CorrDiff, https://doi.org/10.5281/zenodo.18604660, 2026.

https://doi.org/10.5194/egusphere-2026-822
Preprint. Discussion started: 25 February 2026

Sun, Y., Deng, K., Ren, K., Liu, J., Deng, C., and Jin, Y.: Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review, ISPRS Journal of Photogrammetry and Remote Sensing, 208, 14–38, 2024.

Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic attribution for deep networks, in: International conference on machine learning, pp. 3319–3328, PMLR, 2017.

Watt, R. A. and Mansfield, L. A.: Generative diffusion-based downscaling for climate, arXiv preprint arXiv:2404.17752, 2024.

Wu, X., Zhao, R., Chen, H., Wang, Z., Yu, C., Jiang, X., Liu, W., and Song, Z.: GSDNet: A deep learning model for downscaling the significant wave height based on NAFNet, Journal of Sea Research, 198, 102 482, 2024.

Zhang, Y., Tiňo, P., Leonardis, A., and Tang, K.: A survey on neural network interpretability, IEEE transactions on emerging topics in computational intelligence, 5, 726–742, 2021.