JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2024JH000375

A Generative Super-Resolution Model for Enhancing Tropical Cyclone Wind Field Intensity and Resolution

Joseph W. Lockwood[1], **Avantika Gori**[2], **and Pierre Gentine**[3]

[1]The Data Science Institute at Columbia University, Columbia University, New York, NY, USA, [2]Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA, [3]Department of Earth and Environmental Engineering, Climate School, National Science Foundation Learning the Earth with Artificial intelligence and Physics, Columbia University, New York, NY, USA

Key Points:

- We demonstrate that a cascading deep learning generative model can effectively and efficiently model tropical cyclone wind maps
- The framework accurately models high-intensity TC wind speeds from low-resolution climate data
- The super-resolution model enhances resolution, producing realistic radial profiles and fine-scale spatial structures of wind fields

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. W. Lockwood,
jl6895@columbia.edu

Citation:

Lockwood, J. W., Gori, A., & Gentine, P. (2024). A generative super-resolution model for enhancing tropical cyclone wind field intensity and resolution. *Journal of Geophysical Research: Machine Learning and Computation*, *1*, e2024JH000375. https://doi.org/10.1029/2024JH000375

Received 30 JUL 2024
Accepted 30 OCT 2024

Abstract Extreme winds associated with tropical cyclones (TCs) can cause significant loss of life and economic damage globally, highlighting the need for accurate, high-resolution modeling and forecasting for wind. However, due to their coarse horizontal resolution, most global climate and weather models suffer from chronic underprediction of TC wind speeds, limiting their use for impact analysis and energy modeling. In this study, we introduce a cascading deep learning framework designed to downscale high-resolution TC wind fields given low-resolution data. Our approach maps 85 TC events from ERA5 data (0.25° resolution) to high-resolution (0.05° resolution) observations at 6-hr intervals. The initial component is a debiasing neural network designed to model accurate wind speed observations using ERA5 data. The second component employs a generative super-resolution strategy based on a conditional denoising diffusion probabilistic model (DDPM) to enhance the spatial resolution and to produce ensemble estimates. The model is able to accurately model intensity and produce realistic radial profiles and fine-scale spatial structures of wind fields, with a percentage mean bias of −3.74% compared to the high-resolution observations. Our downscaling framework enables the prediction of high-resolution wind fields using widely available low-resolution and intensity wind data, allowing for the modeling of past events and the assessment of future TC risks.

Plain Language Summary Extreme winds associated with tropical cyclones (TCs) can cause significant loss of life and economic damage globally, highlighting the need for accurate, high-resolution modeling and forecasting for wind. Traditional global climate models and weather forecasting models often lack the horizontal resolution needed to accurately model these extreme winds. We introduce a two-part deep learning method that enhances the quality of wind speed predictions from coarse data. The first model corrects inaccuracies in initial low-resolution data from 85 TC events. The second part uses an advanced generative deep learning model to refine these corrections, producing high-resolution wind field images. This process allows us to generate more detailed and accurate maps of TC wind speeds, capturing finer spatial details that are crucial for effective impact analysis and energy modeling. Our downscaling framework enables the prediction of high-resolution wind fields using widely available low-resolution and intensity wind data, allowing for the reconstruction of past events, the assessment of future TC risks, and improved understanding of historical TC impacts even in the absence of satellite data.

1. Introduction

Tropical cyclones (TCs) are among the most devastating natural phenomena, characterized by extreme winds speeds and significant loss of life and economic damage globally. Despite advances in weather prediction and climate modeling, accurately modeling TC intensity remains challenging due to TCs' significant and short-duration deviations from the mean atmospheric state and the complex dynamics within their inner and outer cores (Bloemendaal et al., 2019; Walsh et al., 2007). Many climate models, atmospheric reanalysis products, and long-range weather forecasting models use coarse resolutions ranging from 0.25 to 1°, which are insufficient for modeling the inner-core dynamics and physics of TCs. As a result, underestimation of TC intensity and lack of realistic structural features is ubiquitous in climate and weather data, particularly for the most intense events (Bloemendaal et al., 2019; Gori et al., 2023; Walsh et al., 2007).

Recent advances in deep learning-based weather models, such as Aurora (Bodnar et al., 2024), Pangu-Weather (Bi et al., 2023; Bouallègue et al., 2024), and GraphCast (Lam et al., 2023), offer significant benefits over traditional numerical weather prediction models, including faster computation, greater adaptability, scalability,

and often improved predictive accuracy (Mukkavilli et al., 2023). However, these models still face several challenges, particularly in forecasting extreme weather events. AI-based models tend to struggle with accurately predicting these events over short time scales due to their dependence on large datasets, limited scale awareness, coarse spatial resolution, and the lack of fine-tuning after general training. Additionally, they can suffer from blurring effects caused by averaging, forecast roll-out limitations (Bodnar et al., 2024), and the high computational costs and time required for training (Zhu et al., 2024).

Despite these challenges, AI-based models show great potential for improving the prediction of extreme events when deployed under specific climate or weather regimes, offering a more computationally efficient alternative to physics-based approaches. For instance, the Pangu-Weather model outperformed the European Center for Medium-Range Weather Forecasts (ECMWF) in tracking TC trajectories for 88 TC events in the year of 2018 (Bi et al., 2023). However, despite its superior track prediction, Pangu-Weather's 25-km horizontal resolution remains inadequate for modeling TCs physics, and the model systematically under-predicts TC maximum intensity and intensity changes (Bonavita, 2024; Bouallègue et al., 2024; Selz & Craig, 2023). Similar models such as Fourcastnet and NeuralGCM have also shown substantial underestimation of TC intensity and intensification rates (Kochkov et al., 2024; Pathak et al., 2022). Increasing the resolution of these models may address these issues, but imposes substantial computational burdens. Alternatively, the data used to train these models can be adjusted either prior to input (preprocessing) or following prediction (postprocessing) to more accurately reflect observed extreme events. Diffusion models, in particular, have shown great potential to predict a broad range of frequencies and extremes for turbulence (Gao et al., 2023; Oommen et al., 2024) and for precipitation downscaling (Asperti et al., 2023).

The emergence of super-resolution techniques in machine learning, such as convolutional neural networks (Oktay et al., 2018), generative adversarial networks (GANs) (Goodfellow et al., 2014) and diffusion models (Ho et al., 2020) have demonstrated significant promise in enhancing spatial resolution beyond traditional interpolation methods. These methods hold great potential for applications beyond canonical computer vision applications, and their applications to downscaling extreme weather events have started to emerge (Lai et al., 2024; Rampal et al., 2024; Wang et al., 2021). For instance, Price and Rasp (2022) employed a conditional GAN to generate high-resolution, bias-corrected precipitation forecasts from coarse global forecasts, outperforming interpolation baselines methods, and nearing the performance of operational regional high-resolution models across various probabilistic metrics. Vosper et al. (2023) evaluated conditional super-resolution GANs and CNN models applied to TC precipitation fields, also revealing substantial improvements in capturing realistic spatial structures and power spectra over traditional interpolation methods. Additionally, a diffusion model's recent application to wind speed super-resolution over central Europe further validated its effectiveness in improving the spatial resolution of weather data (Merizzi et al., 2024).

In this paper, we propose a cascading deep learning framework designed to downscale high resolution TC wind fields from low-resolution climate and weather data. Our methodology leverages data from 85 TC events captured by both ERA5 reanalysis (low-resolution) and a high-resolution assimilated wind field product (HWIND). The framework comprises two primary components: a debiasing neural network (NN) to model TC wind fields and a super-resolution DDPM for spatial resolution enhancement and probabilistic ensemble generation (Figure 1).

2. Methodology

2.1. Data

The dataset comprises high-resolution observational target data (0.05° resolution) and low-resolution (0.25°) input data (Figure 1) of 85 TC events in the North Hemisphere that occurred from 2003 to 2014, see below. To ensure robust model evaluation, the test, training, and evaluation sets are divided such that there is no overlap between TC events, thereby limiting sampling bias resulting from correlated timesteps being split between the different sets (Lockwood, Loridan, et al., 2024). Finally, the dataset is split into 60% for training, 10% for validation and 30% test so that the intensity distribution of each set has the same range (Fig S1 in Supporting Information S1).

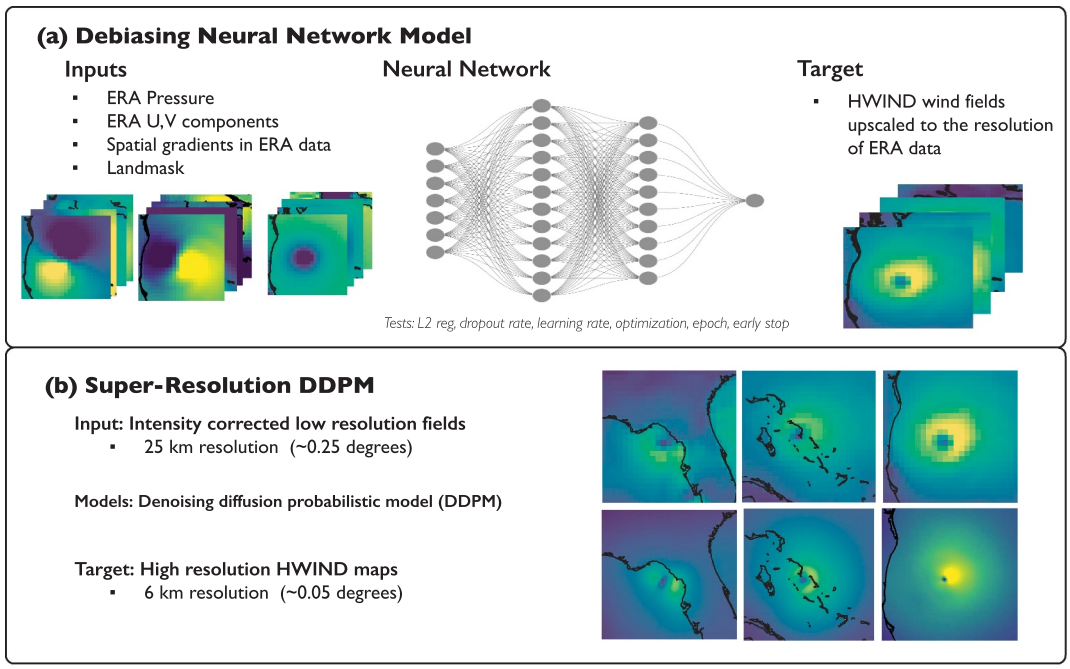

Figure 1. Overview of the (a) NN and (b) super-resolution model. The inputs to the NN corrector include the ERA surface pressure, U10 and V10 winds, spatial gradients in the former and a binary land mask.

2.1.1. Low Resolution Input Data

The input data for the ML model consists of ERA5 data on a regular latitude-longitude grid with a resolution of 0.25° at hourly intervals (Hersbach et al., 2023). The dataset is a state-of-the-art reconstruction of the past climate using an Earth System Model coupled with a data assimilation from a multitude of sources. From ERA5 we use the U and V components of the wind at 10 m, surface pressure, and the horizontal spatial gradients of all the aforementioned variables. Additionally, a binary land mask is included to distinguish between land and water surfaces.

2.1.2. High Resolution Target Data

Verification of TC surface winds is inherently difficult due to the poor temporal and spatial sampling of these events. The HWind-based dataset (Powell et al., 1998) provides valuable data at small and intermediate scales by assimilating aircraft measurements, surface measurements, and remotely sensed satellite data into a common model used in (Chavas et al., 2015, 2016). HWind data have approximate horizontal resolutions of 6 km at 6-hr time steps, and is known to be one of the most comprehensive independent surface wind comparison dataset available (Adams et al., 2005). There are some limitations to using the HWind dataset. HWind does not provide an instantaneous snapshot of a TC; rather, it assimilates data over a time window of three to 6 hr (Powell et al., 1998). This can introduce uncertainties in representing the exact structure of the TC at any given moment. Furthermore, our dataset is restricted to events in the North Atlantic basin, covering only the years from 2003 to 2013, which limits the temporal and spatial scope of the analysis.

To assess the accuracy of the HWIND data, the maximum wind speed at each timestep in the HWind dataset is compared to the International Best Track Archive for Climate Stewardship (IBTrACS) data for the 85 events used here. The maximum wind speed in HWind for these events matches IBTrACS data very well, with a correlation coefficient of 0.95, a mean squared error (MSE) of 20.99, and an error of only 5.92% (Figure S2 in Supporting Information S1). As shown in Figure S2 in Supporting Information S1, there are noticeable discrepancies between HWind and IBTrACS, particularly in the representation of both weak and strong TCs. These differences can introduce biases during model training and evaluation, especially in the extremes of the TC intensity spectrum.

2.2. Deep Learning Models

The proposed framework utilizes a two-step machine learning approach to map the low-resolution ERA5 data onto the high-resolution HWIND target data. The first component is a multi-layer perceptron debiasing neural network model that emulates high-intensity wind field observations using ERA5 data as the input. The second component employs a generative super-resolution denoising diffusion probabilistic model (DDPM) to enhance spatial resolution and produce ensemble estimates. By separating these steps, the DDPM can focus more effectively on the task of super-resolution, limiting the difference between the maps that can complicate the super-resolution process (Vosper et al., 2023). All models were trained on an 80 GB NVIDIA A100 Tensor Core GPU.

3. Neural Network Model

The neural network structure consists of an input layer, two hidden layers with ReLU activation functions, and a dropout layer. The final layer is a ReLU activation function. The input to the NN model are the ERA5 variables, whilst the target variable is HWIND wind speeds data interpolated to the same resolution of the ERA5 data using spline interpolation.

The model was optimized using various hyperparameters on the evaluation dataset. Specifically, dropout rates (5 values spanning homogeneously values between 0 and 0.1) were tested to prevent overfitting by randomly setting a fraction of input units to zero during training. L2 regularization (5 values spanning homogeneously 0 to 0.05) was applied to penalize large weights, helping to avoid overfitting. The number of epochs tested ranged from 400 to 4,000, defining how many times the learning algorithm works through the entire training dataset. Learning rates (5 values tested between 10^{-3} and 10^{-5}) controlled how much the model's weights were adjusted in response to the estimated error each time the model weights were updated. The ”best” model was chosen based on the minimum mean squared error on the evaluation dataset.

4. Super-Resolution Model

The second component of the framework involves super-resolution modeling to enhance the spatial resolution from a coarse 0.25° grid to a finer 0.05° grid (a five times increase). We utilizes the U-NET IMAGEN denoising diffusion probabilistic (DDPM) model of Saharia et al. (2022), modified for conditional super-resolution. We use the corrected low resolution ERA5 data as the input and the HWIND data as the target data of the DDPM model. Given that the model is probabilistic, we produce a 40-member ensemble of the DDPM model output.

The DDPM model, as described by Ho et al. (2020), is a diffusion probabilistic model that operates as a parameterized Markov chain. This chain is trained using variational inference to produce samples matching the data after a finite period (Ho et al., 2020). For practical purposes, small amounts of Gaussian noise are added during diffusion, and the sampling chain transitions are set to conditional Gaussians, allowing for a simple neural network parameterization.

The mathematical formulation of the diffusion model involves a forward process and a reverse process. The forward process gradually adds Gaussian noise to the data:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \alpha_t x_{t-1}, (1-\alpha_t) I) \tag{1}$$

where x_t is the noisy image data at time step t, and α_t is a parameter controlling the noise level. I represents the identity matrix, indicating that the Gaussian noise added during the diffusion process is isotropic.

The reverse process aims to denoise the data:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\big(x_{t-1}; \mu_\theta(x_t, t), \sigma_\theta(x_t, t)\big) \tag{2}$$

where μ_θ and σ_θ are neural network parameterizations that predict the mean and variance of the denoised data.

The DDPM model is trained to minimize the following objective:

$$E_{x,\epsilon\sim\mathcal{N}(0,I),t}\left[w_t \|x - \mu_\theta(\alpha_t x + \sigma_t \epsilon, t)\|^2\right] \tag{3}$$

Figure 2. (a) Scatter plot and (b) Q-Q plot for the test data at each gridpoint, comparing NN predictions to HWIND observations (blue points). Also shown are comparisons of ERA5 wind data to HWIND observations (red) - not the systematic under-prediction of high intensity windspeed in ERA5 data.

where w_t is a weighting function (which adjusts the importance of each timestep during training), x is the original data, ϵ is Gaussian noise, and α_t and σ_t are functions of t influencing the sample quality.

We use a classifier-free guidance, which enhances sample quality by balancing conditional and unconditional predictions during sampling. This technique adjusts the noise prediction as follows:

$$\tilde{\epsilon}_\theta(x_t, c) = w\epsilon_\theta(x_t, c) + (1 - w)\epsilon_\theta(x_t) \tag{4}$$

where $\tilde{\epsilon}_\theta$ is the guided noise prediction, c is the conditioning input, and w is the guidance weight. We modify the model using various configurations to identify the optimal setup that achieves the lowest Kullback-Leibler Divergence of the ensemble on the evaluation dataset. Our testing includes a comprehensive range of hyper-parameters and structural adjustments. Specifically, we experiment with different numbers of epochs, learning rates and diffusion timesteps. These extensive tests allow us to fine-tune the model for improved performance and accuracy.

5. Results

The results are organized into two subsections: the evaluation of the Neural Network (NN) debiasing model (Section 5.1) and the evaluation of the ensemble of super-resolution Denoising Diffusion Probabilistic Models (DDPM) model (Section 5.2).

5.1. Neural Network Model

To assess the performance of the NN corrector model, we compare the model predictions with HWIND observations on the test data (Figure 2). The NN model (blue points) significantly enhances the predicted wind intensities of the NN model compared to the raw ERA5 dataset (red points). The improvement is quantified by a reduction in Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and an increase in the correlation coefficient. Specifically, the NN corrector model achieved an MSE of 10.9 m^2s^{-2} and a correlation coefficient of 0.92 with respect to the target HWIND observations. These metrics indicate a substantial improvement compared to the raw ERA5 dataset, which has an MSE of 20.65 m^2s^{-2} and a correlation coefficient of 0.84 when compared to HWIND observations.

The distribution of the NN model predictions closely matches the distribution of the HWIND observations (Figure 2(b)) indicating good agreement between the two. To statistically validate this agreement, we performed a two-sample Kolmogorov–Smirnov (K–S) test. The K-S test results suggest that the predictions from the NN model and the HWIND observations are from the same distribution, reinforcing the reliability of the NN corrector model.

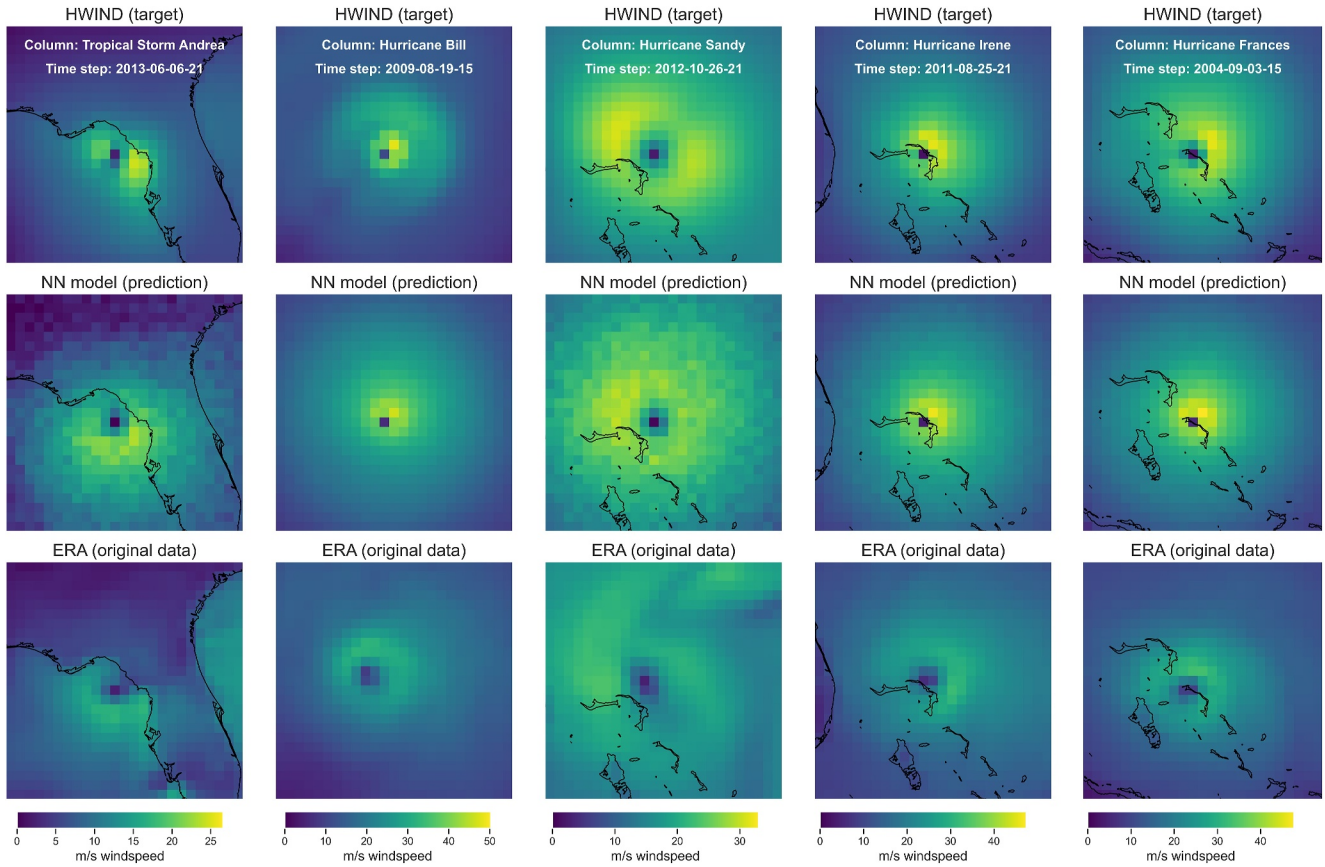

Figure 3. Comparison of wind speeds for five tropical cyclones in the North Atlantic for five randomly selected time steps and storms from the test data. The top row shows observed wind speeds; the middle row shows corrected wind speeds predicted by the NN model; and the bottom row shows raw ERA5 data.

Additionally, we analyze the model predictions visually (Figure 3). The NN model produces wind intensities that are more realistic and closer to the HWIND observations compared to the raw ERA5 data. The NN corrector model effectively captures the spatial patterns and intensities of the wind fields, demonstrating a clear improvement over the ERA5 dataset. In summary, the NN model significantly improves the accuracy of wind speed predictions compared to the raw ERA5 data. The model achieves lower MSE values and a higher correlation coefficient, indicating a closer match to the HWIND observations. The Q-Q plot and K-S test further confirm the model's accuracy, and the visual analysis shows that the NN model produces realistic and spatially accurate wind fields.

To better understand how the model predicts TC intensity despite the coarse resolution of ERA5, we conducted an importance analysis using the Integrated Gradients method. The results, presented in Fig. S3 in Supporting Information S1, highlight the most significant features driving the model's intensity predictions. Notably, surface pressure emerged as the most important feature, followed by U Wind and V Wind Gradient Y. These findings suggest that surface pressure fields, even at coarse resolution, provide valuable information about the storm's intensity. Additionally, wind components and their gradients contribute substantially to the model's predictions, indicating that the broader wind structure around the cyclone helps capture critical dynamics related to intensity. This analysis demonstrates that while ERA5 may not fully resolve the inner core of TCs, key features in its fields are still indicative of intensity and are leveraged effectively by the model.

5.2. Super-Resolution DDPM

The performance of the super-resolution model DDPM in capturing the intricate details of TCs is illustrated in Figure 4. This figure shows randomly selected timesteps for test events, highlighting the model's ability to reveal details in the TC structures, including very fine asymmetries in the structure of TCs. Compared to a bilinear

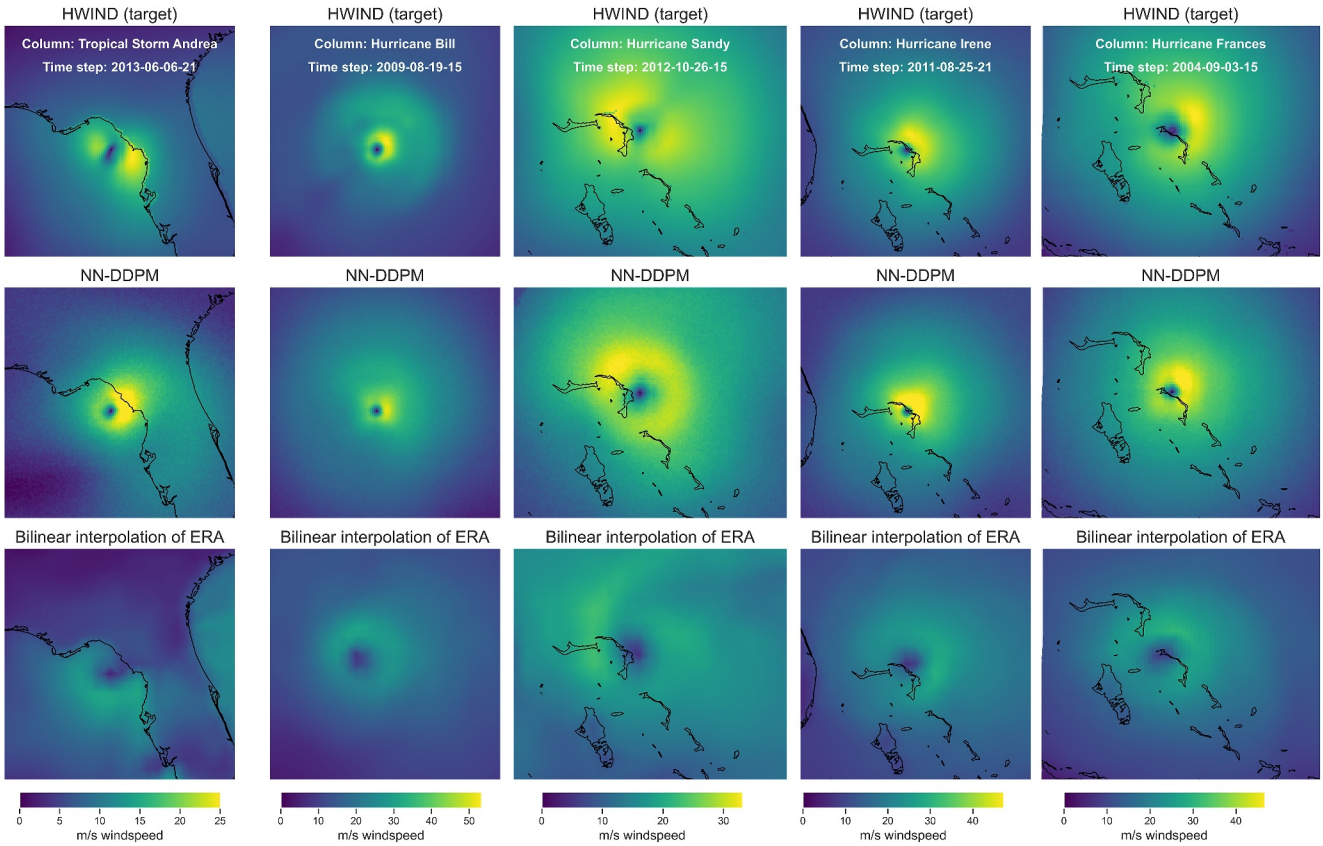

Figure 4. Wind speeds of five tropical cyclones in the North Atlantic for five randomly chosen time steps and storms from the test data. The top row shows observed wind speeds; the middle row shows a randomly chosen NN-DDPM model ensemble; and the bottom row shows bilinear interpolated ERA data.

interpolation of low-resolution ERA data, Figure 4, the DDPM more accurately captures the inner, most intense structures of the TCs. However, the model ensemble is shown to sometimes overestimate (e.g., Hurricane Frances) or underestimate (e.g., Hurricanes Irene and Bill) the wind speed magnitudes, as shown in Figure 4. Our DDPM framework may overestimate or underestimate wind speeds due to the stochastic nature of the denoising process. While DDPMs are designed to capture complex patterns in the data, their probabilistic structure can introduce variability in the predictions. This randomness, coupled with the inherent uncertainty in TC dynamics and limited input data for certain events, may contribute to these discrepancies.

To quantitatively evaluate the NN-DDPM model's performance, we analyzed the overall wind speed metrics and compared them with HWIND data. The maximum wind speed observed in the HWIND dataset is 72.54 m/s, whereas the model predictions peak at 69.32 m/s (Table 1). Similarly, the overall mean wind speed for HWIND is 14.96 m/s, while the model predictions slightly underestimate this with a mean value of 14.40 m/s. The mean bias, calculated as the difference between predicted and observed mean wind speeds, is −0.56 m/s, indicating a consistent underestimation in the predictions by the NN-DDPM model. The percentage mean bias is −3.74%, reflecting the relative difference in mean wind speeds as a percentage of the observed HWIND mean. These metrics suggest that while the predictions are relatively accurate, there is a systematic slight underestimation of wind speeds in our model.

Table 1
Comparison of Overall Wind Speed Metrics and Bias for HWIND Observations and NN-DDPM Model Predictions

Metric	Value
Overall Max Wind Speed (HWIND)	72.54 m/s
Overall Mean Wind Speed (HWIND)	14.96 m/s
Overall Max Wind Speed (Predictions)	69.32 m/s
Overall Mean Wind Speed (Predictions)	14.40 m/s
Overall Bias	−0.56 m/s
Percentage Mean Bias	−3.74%

We also analyze the radially averaged wind speeds, comparing HWIND observations with the DDPM model, as shown in Figure 5, along with a bilinear interpolation of the ERA5 data as a common downscaling benchmark (Vosper et al., 2023). For major TC category 3–5 events, the ensemble DDPM wind speeds at the radius of maximum wind are slightly higher than the

10.1029/2024JH000375

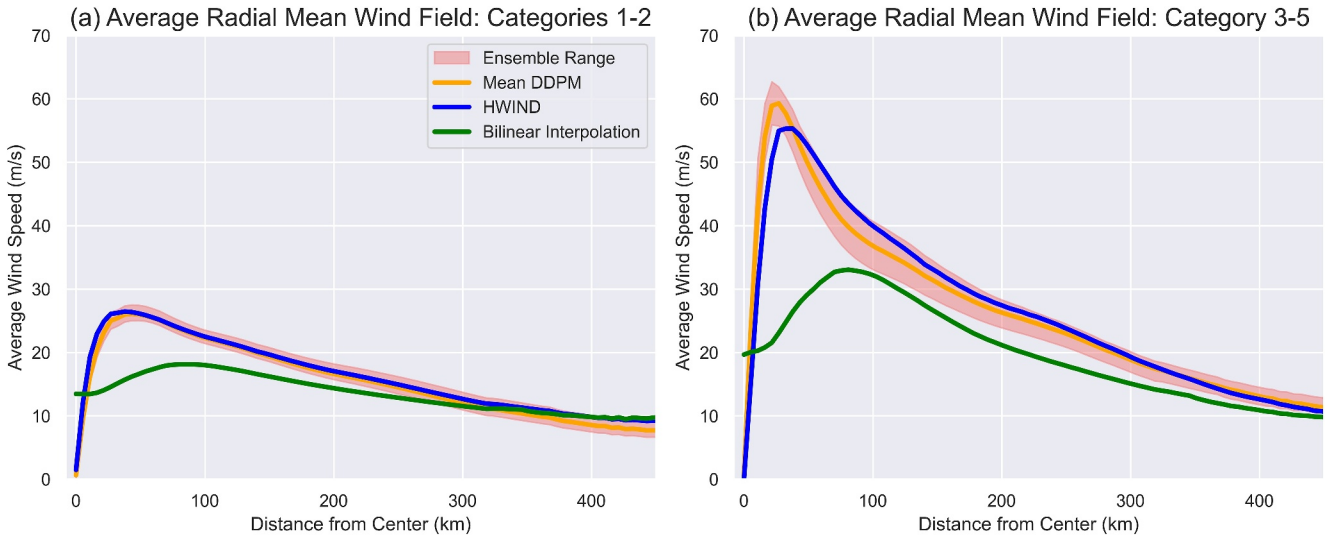

Figure 5. Radially averaged wind speeds as a function of distance for TCs (a) category 1–2 and (b) between Category 3 or above for the NN-DDPM ensemble and range. Green lines show a bilinear downscaled interpolation of ERA5 data.

HWIND observations on average (Figure 5b). Outside these regions, the DDPM ensemble accurately captures the profiles of HWIND observations. The DDPM model effectively captures significant changes in wind speed over short distances, particularly the substantial increase in wind speed from the center of the storm to the radius of maximum wind. These results suggest that the super-resolution model produces realistic radial profiles and fine-scale spatial structure of wind fields.

We next explore the azimuthal asymmetry of the predictions to complement the azimuthal-mean performance, providing a more detailed understanding of the model's ability to capture asymmetric structures in hurricane intensity (Figure 6). Category 3+ hurricanes exhibit higher RMSE values and lower correlation compared to their lower intensity counterparts in Category 1–2.

Finally, we assess the capability of our framework to model the evolution of TC maximum intensity over time and rate of change, which are important to hazard modeling (Lockwood, Lin, et al., 2024). Figure 7 illustrates the timeseries of TC intensity change for six randomly selected events. The framework not only captures the overall temporal evolution of TC intensity and its changes but also accurately represents the large fluctuations in intensification and de-intensification rates over time.

6. Discussion

In this study, we developed a cascading Neural Network (NN) and Denoising Diffusion Probabilistic Model (DDPM) super-resolution generative framework to unbias and model high intensity TC wind speeds at fine resolutions, using low resolution data as inputs. Our findings contribute to the growing body of evidence that deep learning methods hold significant promise in weather and climate science, particularly in downscaling applications. The models developed here produce high-resolution TC wind predictions could be invaluable for forecasting and hazard assessments.

The NN-DDPM super-resolution model demonstrates a strong ability to enhance the resolution and accuracy of wind speed predictions for TCs. While the model shows a slight underestimation in wind speed predictions, its ability to model high intensity wind fields and fine-scale structures is noteworthy. Overall, these developments have the potential to significantly improve weather and climate model outputs, and for global TC risk quantification. Accurate modeling of these variations is essential for disaster preparedness, energy modeling and impact modeling.

Despite these advances, several areas for improvement and further research remain. One such area is the incorporation of additional inputs to enhance predictions, such as the influence of vertical wind shear, upper atmospheric dynamics, and wind field asymmetries on TC wind fields, yet these variables are less readily

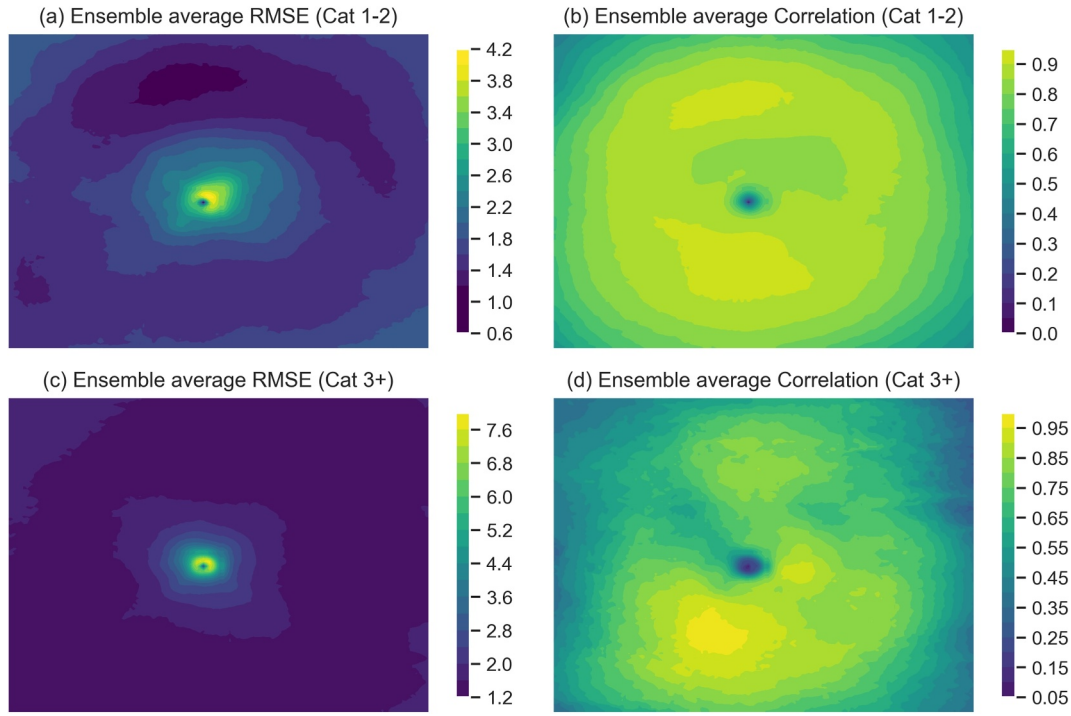

Figure 6. Ensemble-averaged comparison of NN-DDPM predictions and HWIND data for hurricane intensity. Subplots (a) and (b) show RMSE and correlation for Category 1–2 hurricanes, while (c) and (d) display the same for Category 3+.

available in observations. A key limitation of machine learning (ML) models in this context is their non-stationarity (Lai et al., 2024): the tendency for ML models to produce significant errors when applied to data distributions that differ from those they were trained on. Climate change, characterized by changing mean states and variability, may limit the application of these ML models in this context. In particular, previous studies have shown that the accuracy of machine learning models applied to TC hazards significantly diminishes when these models are used outside their training data (Lockwood, Loridan, et al., 2024).

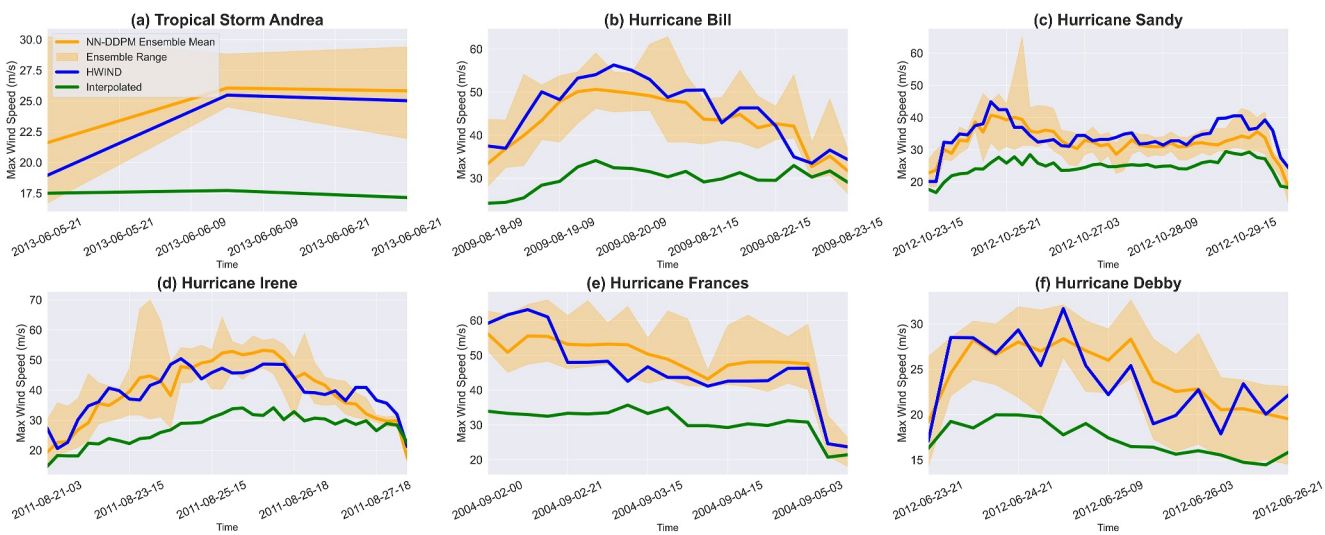

Figure 7. Time series plot of maximum wind speed for six randomly chosen events as above comparing the NN-DDPM ensemble mean and range (orange), HWIND observations (blue) and bilinear downscaled interpolation of ERA5 data (green).

The emergence of physics-informed neural networks (PINNs) offers a potential solution (Beucler et al., 2021; Lai et al., 2024; Rampal et al., 2024) or the inclusion of climate-invariant variables offer potential solutions to this extrapolation and generalization issue (Beucler et al., 2024). Those physics-informed strategies integrate first-principle equations, such as momentum or continuity equations, or non-dimensionalization into the training process, and have shown promise in modeling planetary-scale geophysical flows, including TCs (Eusebi et al., 2024). Future research should explore the extent to which incorporating physical constraints impacts model predictions in changing climates.

We selected the DDPM framework because it offers significant advantages in capturing complex, probabilistic patterns in atmospheric data, which are essential for accurate downscaling. Unlike traditional neural networks, which may struggle with the diversity of weather phenomena, or GANs, which can suffer from mode collapse, DDPMs generate detailed, realistic high-resolution wind fields by iteratively refining low-resolution inputs. This makes them particularly well-suited for weather downscaling, where small-scale variability and extremes plays a critical role in accurate forecasting, despite occasional over/underestimations.

Other generative models, such as Wasserstein Generative Adversarial Networks (WGANs), have shown promise in modeling weather and TC hazards (Vosper et al., 2023). This study focused on DDPMs due to their probabilistic ensemble capabilities and their stability during training, which is often a challenge with GAN-based approaches (Mardani et al., 2024). Additionally, DDPMs are particularly well-suited for capturing small-scale variability in weather data, making them ideal for high-resolution wind field reconstruction. Future research should evaluate the performance of other super-resolution models, such as GANs and conditional CNNs, in this context. Future work will also explore the application of this method for estimating high-resolution TC windfields using Global Climate Model (GCM) data as inputs. This would require high-resolution dynamical model simulations to develop and validate the model, which may pose limitations.

Our framework has demonstrated the ability to reconstruct high-resolution wind fields from past events and to forecast future wind fields, particularly during periods without satellite observations. This capability is essential for improving the accuracy of TC modeling and risk management, especially in correcting wind intensity underestimations from low-resolution models such as CMIP6, Pangu-Weather, and GraphCast or subseasonal to seasonal forecasts (S2S in Supporting Information S1). By enhancing the detail and accuracy of TC wind field predictions, our approach provides a valuable tool for better understanding and mitigating TC-related risks.

Data Availability Statement

The code used in this study can be accessed via the following Zenodo repository (Lockwood, Gori, & Gentine, 2024). HWIND data can be found open-access from RMS (RMS, 2024), whilst ERA5 can be found at Hersbach et al. (2023).

Acknowledgments
Gentine and Lockwood acknowledge funding from the Data Science Institute at Columbia university and from National Science Foundation Learning the Earth with Artificial intelligence and Physics (LEAP), Award 2019625 - STC.

References

Adams, I., Hennon, C., Jones, W., & Ahmad, K. (2005). Hurricane wind vector estimates from windsat polarimetric radiometer. *Geoscience and remote sensing symposium, 2005. Igarss '05. Proceedings. 2005 IEEE international*, *6*, 4149–4152. https://doi.org/10.1109/IGARSS.2005.1525794

Asperti, A., Merizzi, F., Paparella, A., Pedrazzi, G., Angelinelli, M., & Colamonaco, S. (2023). Precipitation nowcasting with generative diffusion models. *arxiv*. Retrieved from https://arxiv.org/abs/2308.06733

Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., et al. (2024). Climate-invariant machine learning. *Science Advances*, *10*(6), eadj7250. https://doi.org/10.1126/sciadv.adj7250

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, *126*(9), 098302. https://doi.org/10.1103/PhysRevLett.126.098302

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). 07). Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, *619*, 1–538. https://doi.org/10.1038/s41586-023-06185-3

Bloemendaal, N., Muis, S., Haarsma, R., Verlaan, M., Irazoqui, M., Moel, H., et al. (2019). 04). Global modeling of tropical cyclone storm surges using high-resolution forecasts. *Climate Dynamics*, *52*(7–8), 5031–5044. https://doi.org/10.1007/s00382-018-4430-x

Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Brandstetter, J., Garvan, P., et al. (2024). Aurora: A foundation model of the atmosphere. *arxiv*. Retrieved from https://arxiv.org/abs/2405.13063

Bonavita, M. (2024). On some limitations of current machine learning weather prediction models. *Geophysical Research Letters*, *51*(12), e2023GL107377. https://doi.org/10.1029/2023GL107377

Bouallègue, Z. B., Clare, M. C. A., Magnusson, L., Gascón, E., Maier-Gerber, M., Janoušek, M., et al. (2024). The rise of data-driven weather forecasting: A first statistical assessment of machine learning–based weather forecasts in an operational-like context. *Bulletin of the American Meteorological Society*, *105*(6), E864–E883. https://doi.org/10.1175/BAMS-D-23-0162.1

Chavas, D., Lin, N., Dong, W., & Lin, Y. (2016). Observed tropical cyclone size revisited. *Journal of Climate*, *29*(8), 2939. https://doi.org/10.1175/JCLI-D-15-0731.1

Chavas, D., Lin, N., & Emanuel, K. (2015). A model for the complete radial structure of the tropical cyclone wind field. Part I: Comparison with observed structure. *Journal of the Atmospheric Sciences*, *72*(9), 3662. https://doi.org/10.1175/JAS-D-15-0014.1

Eusebi, R., Vecchi, G., Lai, C.-Y., & Tong, M. (2024). Realistic tropical cyclone wind and pressure fields can be reconstructed from sparse data using deep learning. *Communications Earth Environment*, *5*(1), 8. https://doi.org/10.1038/s43247-023-01144-2

Gao, Z., Shi, X., Han, B., Wang, H., Jin, X., Robinson, D. M., et al. (2023). Prediff: Precipitation nowcasting with latent diffusion models. In *Neurips 2023*. Retrieved from https://www.amazon.science/publications/prediff-precipitation-nowcasting-with-latent-diffusion-models

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. Retrieved from https://arxiv.org/abs/1406.2661

Gori, A., Lin, N., Schenkel, B., & Chavas, D. (2023). North Atlantic tropical cyclone size and storm surge reconstructions from 1950-present. *Journal of Geophysical Research: Atmospheres*, *128*(5), e2022JD037312. https://doi.org/10.1029/2022JD037312

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023). Era5 hourly data on single levels from 1940 to present [dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.adbb2d47

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *arxiv*. Retrieved from https://arxiv.org/abs/2006.11239

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., et al. (2024). Neural general circulation models for weather and climate. *Nature*, *632*(8027), 1060–1066. https://doi.org/10.1038/s41586-024-07744-y

Lai, C.-Y., Hassanzadeh, P., Sheshadri, A., Sonnewald, M., Ferrari, R., & Balaji, V. (2024). Machine learning for climate physics and simulations. *arxiv*. Retrieved from https://arxiv.org/abs/2404.13227

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., et al. (2023). Graphcast: Learning skillful medium-range global weather forecasting. Retrieved from https://arxiv.org/abs/2212.12794

Lockwood, J. W., Gori, A., & Gentine, P. (2024). A generative super-resolution model for enhancing tropical cyclone wind field intensity and resolution. *Zenodo*. https://doi.org/10.5281/zenodo.14029551

Lockwood, J. W., Lin, N., Gori, A., & Oppenheimer, M. (2024). Increasing flood hazard posed by tropical cyclone rapid intensification in a changing climate. *Geophysical Research Letters*, *51*(5), e2023GL105624. https://doi.org/10.1029/2023GL105624

Lockwood, J. W., Loridan, T., Lin, N., Oppenheimer, M., & Hannah, N. (2024). A machine learning approach to model over-ocean tropical cyclone precipitation. *Journal of Hydrometeorology*, *25*(1), 207–221. https://doi.org/10.1175/JHM-D-23-0065.1

Mardani, M., Brenowitz, N., Cohen, Y., Pathak, J., Chen, C.-Y., Liu, C.-C., et al. (2024). Residual corrective diffusion modeling for km-scale atmospheric downscaling. *arxiv*. Retrieved from https://arxiv.org/abs/2309.15214

Merizzi, F., Asperti, A., & Colamonaco, S. (2024). Wind speed super-resolution and validation: From era5 to cerra via diffusion models. Retrieved from https://arxiv.org/abs/2401.15469

Mukkavilli, S. K., Civitarese, D. S., Schmude, J., Jakubik, J., Jones, A., Nguyen, N., et al. (2023). Ai foundation models for weather and climate: Applications, design, and implementation. *arxiv*. Retrieved from https://arxiv.org/abs/2309.10808

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arxiv*. Retrieved from https://arxiv.org/abs/1804.03999

Oommen, V., Bora, A., Zhang, Z., & Karniadakis, G. E. (2024). Integrating neural operators with diffusion models improves spectral representation in turbulence modeling. *arxiv*. Retrieved from https://arxiv.org/abs/2409.08477

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., et al. (2022). Fourcastnet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arxiv*. Retrieved from https://arxiv.org/abs/2202.11214

Powell, M., Houston, S., Amat, L., & Morisseau-Leroy, N. (1998). The hrd real-time hurricane wind analysis system. *Journal of Wind Engineering and Industrial Aerodynamics*, *77–78*, 53–64. https://doi.org/10.1016/S0167-6105(98)00131-7

Price, I., & Rasp, S. (2022). Increasing the accuracy and resolution of precipitation forecasts using deep generative models. Retrieved from https://arxiv.org/abs/2203.12297

Rampal, N., Hobeichi, S., Gibson, P. B., Baño-Medina, J., Abramowitz, G., Beucler, T., et al. (2024). Enhancing regional climate downscaling through advances in machine learning. *Artificial Intelligence for the Earth Systems*, *3*(2), 230066. https://doi.org/10.1175/AIES-D-23-0066.1

RMS. (2024). HWind Legacy Archive [dataset]. *RMS*. Retrieved from https://www.rms.com/event-response/hwind/legacy-archive

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., et al. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *arxiv*. Retrieved from https://arxiv.org/abs/2205.11487

Selz, T., & Craig, G. C. (2023). Can artificial intelligence-based weather prediction models simulate the butterfly effect? *Geophysical Research Letters*, *50*(20), e2023GL105747. https://doi.org/10.1029/2023GL105747

Vosper, E., Watson, P., Harris, L., McRae, A., Santos-Rodriguez, R., Aitchison, L., & Mitchell, D. (2023). Deep learning for downscaling tropical cyclone rainfall to hazard-relevant spatial scales. *Journal of Geophysical Research: Atmospheres*, *128*(10), e2022JD038163. https://doi.org/10.1029/2022JD038163

Walsh, K. J. E., Fiorino, M., Landsea, C. W., & McInnes, K. L. (2007). Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *Journal of Climate*, *20*(10), 2307–2314. https://doi.org/10.1175/JCLI4074.1

Wang, C., Tang, G., & Gentine, P. (2021). Precipgan: Merging microwave and infrared data for satellite precipitation estimation using generative adversarial network. *Geophysical Research Letters*, *48*(5), e2020GL092032. https://doi.org/10.1029/2020GL092032

Zhu, X. X., Xiong, Z., Wang, Y., Stewart, A. J., Heidler, K., Wang, Y., et al. (2024). On the foundations of earth and climate foundation models. *arxiv*. Retrieved from https://arxiv.org/abs/2405.04285