WiREDiff: a Wind Resolution-Enhancing Diffusion Model

Rupa Kurinchi-Vendhan
California Institute of Technology
rkurinch@caltech.edu

Abstract

The transition to green energy grids depends, in part, on detailed wind forecasts to optimize the siting and scheduling of renewable energy generation. Operational forecasts from numerical weather prediction (NWP) models, however, only have a spatial resolution of 10 to 20-km, which leads to sub-optimal usage and development of wind turbine farms. Weather scientists have been developing super-resolution (SR) methods to increase the resolution of NWP outputs, but often rely on simple interpolation techniques or computationally expensive differential equation-based models. Recently, machine learning approaches, specifically diffusion models, have outperformed traditional and state-of-the-art downscaling methods. WiREDiff is a novel application of diffusion models to wind speed SR which makes use of quantile regression to computationally-efficiently produce and average an interval of predicted wind velocities. We provide a benchmark of traditional deep learning-based SR techniques on wind data fields, and demonstrate that diffusion models—and the WiREDiff in particular—have significant applications in climate scenarios.

1 Introduction

In the United States, the national Energy Information Administration (EIA) predicts that renewable energy, such as wind power, will contribute 42% of the country's electricity generation by 2050 (1). To achieve this goal, operational decision-makers must integrate forecasting models into local power systems to address the spatial variability of these clean energy forms. However, current climate simulations used to obtain high-resolution data are unable to resolve the spatial characteristics necessary for accurate future local energy assessments, as increasing their spatial resolution is computationally expensive and provides insufficient accuracy (2). NWPs provide short-term climatological forecasting data (i.e. wind speed fields) at a horizontal resolution of 10 to 20-km (3; 4), while energy planning requires this data at a smaller, more local scale, on the order of 2-km (5).

In the field of computer vision, researchers enhance the resolution of a data field through single-image super-resolution (SR) (6). However, this problem is inherently ill-posed—coarsened low-resolution (LR) input data can map to infinitely many high-resolution (HR) outputs. Machine learning-based approaches offer accurate and less expensive methods of generating the high-resolution data needed to predict the effect atmospheric dynamics have on power generation (7; 8; 9).

2 Related Work

In the context of super-resolving physics-related data, traditional convolutional neural networks (CNNs) have been shown to be successful at high fidelity SR (10; 11). Fukami et al. used the SRCNN network structure to super-resolve 2D laminar cylinder flow (12), and the MeshfreeFlowNet used a U-Net structure to reconstruct the Rayleigh-Bénard instability (13).

Preprint. Under review.

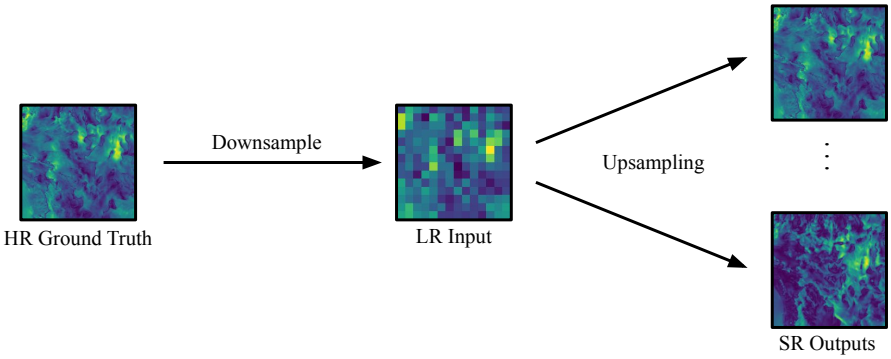

Figure 1: The problem of SR is inherently ill-posed, because infinitely many high-resolution data fields can look identical once you coarsen them to a lower-resolution.

Diffusion models, originally proposed in 2015, have seen an increase in interest due to their training stability and their promising sample quality results in generative tasks (14). Thus, they offer potentially favorable trade-offs compared to other types of deep generative models. Diffusion models work by corrupting the training data by progressively adding Gaussian noise, slowly wiping out details in the data until it becomes pure noise, and then training a neural network to reverse this corruption process. Running this reversed corruption process synthesizes data from pure noise by gradually denoising it until a clean sample is produced. This synthesis procedure can be interpreted as an optimization algorithm that follows the gradient of the data density to produce likely samples (15; 16). The notion of generating data using diffusion techniques originates from the concepts in physics, more specifically non-equilibrium thermodynamics, which deals with the compression and spread of fluids and gases based on energy (17; 18). This motivates the application of diffusion-based methods to wind speed prediction.

As a stochastic process, diffusion-based methods can produce a wide distribution of potential SR outputs. Therefore, these models do not provide a way of statistically guaranteeing the degree of confidence they have in their generated results. In the context of renewable energy integration, it is important for infrastructural planners and policy-makers to be able to trust that the high resolution wind speed predictions generated by the latent diffusion model are true to reality. To this end, we can make use of quantile regression construct an pixel-wise interval around each generated pixel such that the true wind speed value corresponding to that pixel lies within the interval with a probability set by the user (19; 20; 21). This approach is far more computationally-efficient than the naive sampling method, in which each test image needs multiple sampled variations and each diffusion model inference step requires iterative denoising steps.

Thus, in this work we contribute: (1) a novel application of state-of-the-art diffusion-based SR techniques to a specific task from the physical sciences and (2) an implementation of quantile regression methodologies to produce SR outputs within a tighter interval of predicted wind speeds.

3 Methods

In the following sections, we describe the WiREDiff approach, the wind speed dataset we evaluate it on, and the models we compare it to.

3.1 Approach

Towards our first objective, the WiREDiff produces SR outputs using the approach of the known diffusion model SR3 (22). As depicted in Figure 2, we use a T-step diffusion model which contains a forward process and a reverse process. Residual predictions guide the model by inferring the difference between ground truth HR data field and upsampled outputs, denoted by the input residual data field x_0. In the forward process, the posterior $q(x_1, \cdots, x_T|x_0)$ converts the data distribution $q(x_0)$ to the latent variable distribution $q(x_T)$ through a Markov chain which gradually adds isotropic Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$. The reverse process, determined by the conditional noise predictor ϵ_θ, transforms the latent variable distribution $p_\theta(x_T)$ to the data distribution $p_\theta(x_0)$. Taking the latent variable x_T as input, we iteratively denoise in finite step T using ϵ_θ to produce a residual data x_r. We

then add these residual values in x_r to the upsampled LR data field to generate the final SR output. As with SR3, the architectures of the conditional noise predictor and LR encoder and the training and inference procedures in WiREDiff are the same as those proposed in SRDiff (14).

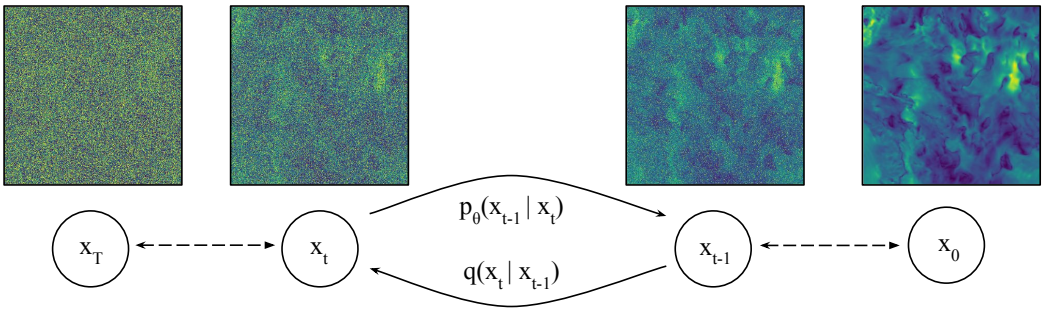

Figure 2: Overview of two processes in WiREDiff. The diffusion process is from right to left and the reverse process is left from left to right. θ in p_θ denotes the learnable components including conditional noise predictor and low-resolution encoder in the WiREDiff.

After pre-training the WiREDiff, we obtain a SR3 model which is capable of producing SR outputs given an LR input. The lower and upper bounds of the SR3 outputs are approximated using quantile regression (QR). We can adjust the model from the SR outputs that the SR3 produces by minimizing the quantile loss for a quantile α which is given in Equation 1.

$$\mathcal{L} = \mathcal{L}_{\alpha/2}(\tilde{l}(x), y) + \mathcal{L}_{1-\alpha/2}(\tilde{u}(x), y) + \mathcal{L}_{mse}(x, y) \tag{1}$$

where the quantile loss for a particular quantile estimator $\hat{q}_\alpha(x)$ is shown below (21)

$$\mathcal{L}_\alpha(\hat{q}_\alpha(x), y) = (y - \hat{q}_\alpha(x))\alpha \mathbb{1}\{y > \hat{q}_\alpha(x)\} + (\hat{q}_\alpha(x) - y)(1-\alpha)\{y \leq \hat{q}_\alpha(x)\} \tag{2}$$

Since we require $\tilde{u}$ and $\tilde{l}$ to estimate different quantiles, the quantile loss becomes (21)

$$\mathcal{L}_{QR}(x, y) = \mathcal{L}_{\alpha/2}(\tilde{l}(x), y) + \mathcal{L}_{1-\alpha/2}(\tilde{u}(x), y), \tag{3}$$

and the final objective combines the quantile loss for bound estimation with $\mathcal{L}_1$ for pointwise prediction:

$$\mathcal{L}(x, y) = \mathcal{L}_{QR}(x, y) + \mathcal{L}_1. \tag{4}$$

The quantile regression provides us with the $\tilde{u}$ and $\tilde{l}$ that approximate the $1 - \alpha/2$ and α quantiles, respectively. We then calibrate these values to produce $\hat{u}$ and $\hat{l}$.

For some calibration set $\{x_i, y_i\}_{i=1}^N$ where x_i are the matrices of size $M \times N$ of the corrupted images and y_i are the matrices of size $M \times N$ of the target images. During calibration, we are interested in value λ such that the number of pixels falling outside the interval $[\lambda\tilde{l}, \lambda\tilde{u}]$ is below α. In other words, we want to solve Equation 5 to determine the calibration constant $\hat{\lambda}$ (21).

$$\hat{\lambda} = \arg\min_\lambda 1 - \frac{\left|\{(m,n)\} : y_{(m,n)} \in [\hat{l}(x_{i_{mn}}), \hat{u}(x_{i_{mn}})]_{(m,n)}\right|}{MN} \tag{5}$$

This process of finetuning the SR3 model is called the Narrow Con*ffusion* (N-Con*ffusion*) model. This is a powerful tool as we can utilize the pre-trained SR3 model outputs for quantile regression, and it separates the model from the computationally expensive process of diffusion (21). After acquiring the lower and upper bounds from N-Con*ffusion*, we then average the bounds to produce the outputs in the test set for the WiREDiff model. Thus, the WiREDiff performs the SR task by applying conformal prediction intermediately, taking an LR data field as input and generating an HR data field that is the mean of the calibrated upper and lower bounds of the output.

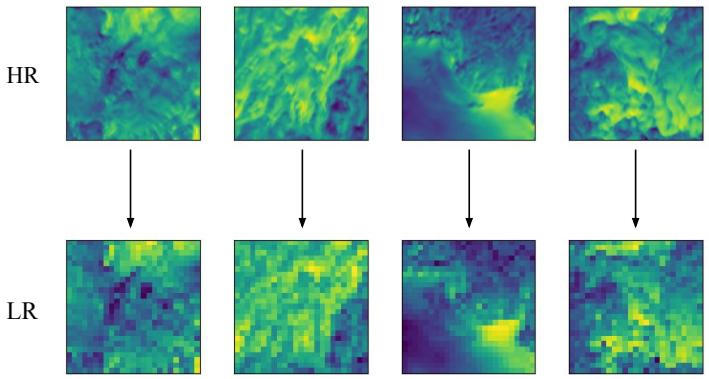

Figure 3: The LR and HR pairs from the wind dataset, with the coarsened data being downsampled by a factor of 5.

3.2 Data

Wind fields will be obtained from the National Renewable Energy Laboratory's (NREL's) Wind Integration National Database (WIND) Toolkit, with a focus on the continental United States (23; 24). The wind velocity data is comprised of westward (ua) and southward (va) components, calculated from wind speeds and directions 100-m from Earth's surface (25; 26). The WIND Toolkit has a spatial resolution of 2km $\times$ 1hr spatiotemporal resolution (25; 26). The training dataset will contain wind fields sampled at a 4-hourly temporal resolution (every fourth data point) between 2007 and 2013. Wind test data will be sampled at a 4-hourly temporal resolution for 2014. Information about the dataset is summarized in Table 1. All 20 TB of relevant data is publicly available for download (23; 24). Data files are available as .NPY files, where each file contains two 100$\times$100 data arrays of ua and va wind speed components. Low resolution data will be generated by applying bilinear interpolation on the 100×100 HR patches to create 20$\times$20 patches. Examples of LR and HR data fields for the wind dataset are visualized in the Figure 3.

Data	Wind
Institute	NREL
Model	WIND Toolkit
Spatial Resolution	2 km
Temporal Resolution	4-hr
Years	2007-2014
HR Dimensions	100$\times$100
LR Dimensions	20$\times$20

Table 1: Wind dataset specifications

3.3 Experiments

Our goal is to examine the ability of standard SR diffusion models to perform 5$\times$ spatial super-resolution, e.g., 10-km to 2-km spatial resolution, on wind data fields. We will compare the performance of the WiREDiff against the SR3 and SRCNN (27), with bicubic interpolation included as a baseline for comparison. The SR3 (which is also in the WiREDiff) and SRCNN have previously only been applied to image data, so we modified the data ingestion pipeline of existing implementations to handle raw wind speed arrays of shape (2, H, W) as opposed to RGB images of shape (3, H, W). Note that the difference in the number of channels in the raw data is due to the concatenation of 2D wind fields for both the ua and va directions as opposed to the RGB channels in image data. To generate an output that's most representative of the data distribution for the SR3, we run the diffusion model for an arbitrarily large number of repetitions (chosen to be 500), and report the pixel-wise average across all samples as the final SR data field. Note that we do not run inference on the WiREDiff multiple times, since we approximate the data distribution by averaging the upper and lower bounds produced by the N-Con*ffusion* approach.

4 Results

We first apply Con*ffusion* to generate lower and upper bounds for the generated SR outputs, and then demonstrate how averaging across this interval to produce WiREDiff SR outputs compares against the benchmarking models outlined in Section 3.3.

4.1 Interval Prediction

First, we applied the naive method of interval construction: we passed the LR data from the test set as input to the SR3 to generate a distribution of potential SR outputs, and found the pixel-wise lower and upper bounds of these data fields (where $\alpha = 0.1$). Then, we ran a single forward pass of the fine-tuned N-Con*ffusion* model to generate a calibrated upper and lower bound for each output. Figure 4 shows an example of how the interval size decreases when applying the Con*ffusion* approach to wind speed data.

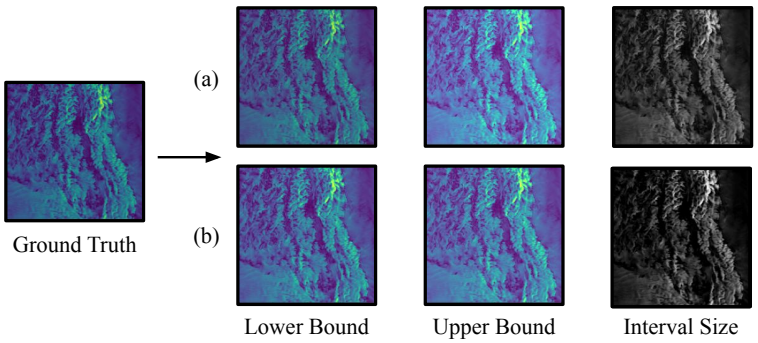

Figure 4: A comparison of two interval construction approaches: (a) the naive method and (b) Con*ffusion*. Con*ffusion* results in a tighter interval.

This validates the results demonstrated by Horwitz and Hoshen, as the Con*ffusion* approach yields tighter bounds than the naive approach (21). As in the Con*ffusion* paper, we set $\alpha = 0.1$. With the sample above, we achieve a coverage rate of 0.891 with N-Con*ffusion*, which reasonably approximates $1 - \alpha$. This as an improvement from the coverage rate with the naive approach, 0.882. This appears to be consistent with our knowledge on Con*ffusion* as the process of minimizing the upper and lower intervals such that only $\alpha = 0.1$ of the pixels fall outside of the interval should increase our coverage closer to $1 - \alpha = 0.9$.

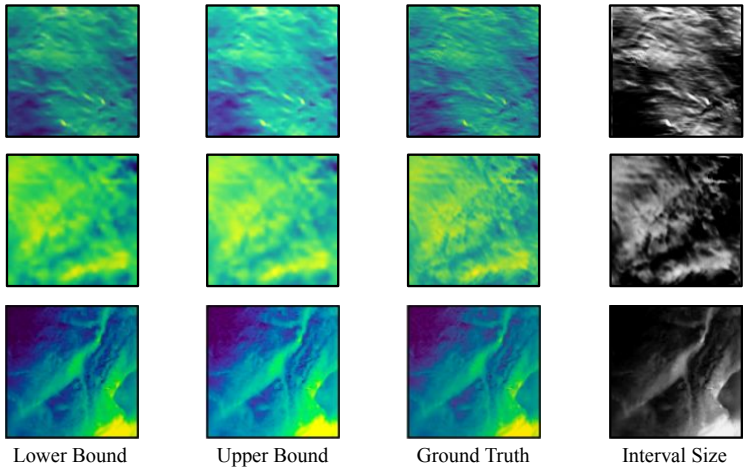

Figure 5: The bounds for SR extracted via N-Con*ffusion*. The left and right columns correspond with the lower and upper bounds of the interval, respectively, and the rightmost column contains the interval size (the difference between the lower and upper bounds).

In Figure 5, we apply N-Con*ffusion* to generate lower and upper bounds for several images in the test set. We can see that this approach provides meaningful bounds, covering a wide range when needed. Examining the interval sizes, we observe that regions with higher frequency data correspond

to wider bounds and higher model uncertainty. Thus, N-Con*ffusion* constructs the tightest intervals while maintaining realistic-looking bounds.

4.2 Benchmark Analysis

Figure 6 compares sample outputs from each model. Qualitatively, the results show that deep learning models—most noticeably the SR3 and WiREDiff—produce smoother outputs with small-scale details. Diffusion models have higher fidelity than traditional CNN methods because by iteratively denoising samples, some noise may remain to emulate the high frequency features of the ground truth data. While the SRCNN produces outputs that have sharper structures, the range of wind speed values is inaccurate relative to those generated by the diffusion models. We observe that the wind speeds in the diffusion outputs are dynamically compressed (i.e. the predictions are between a more limited range of values). The output from the SR3 smooths out some high frequency features as compared to the WiREDiff because this result is obtained by averaging over a large number of diffusion outputs.

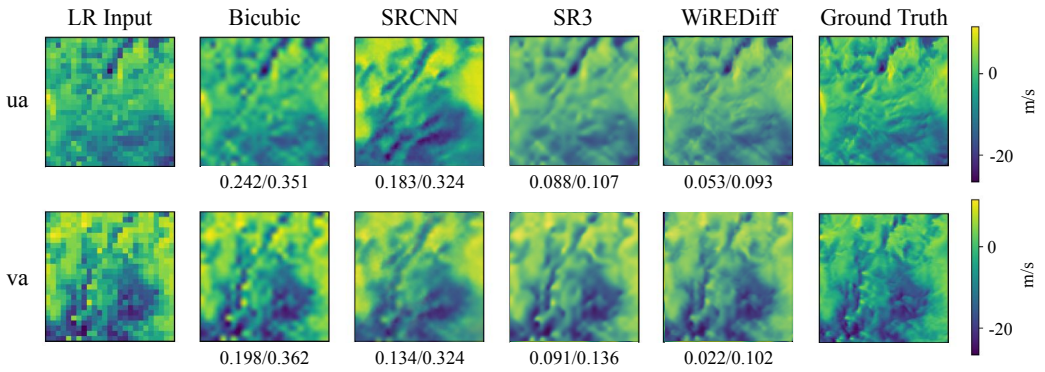

Figure 6: A comparison of wind (top) and solar (bottom) outputs from each model, with reported MSE/MAE values. While the outputs of current models are pixelated (bicubic) or high-contrast (SR CNN) our contributed diffusion model is the most accurate. The WiREDiff output has slightly more higher frequency details than the SR3.

To assess the ability of each model to accurately recreate the ground truth data, we applied data similarity metrics such as normalized mean-squared error (MSE) and mean absolute error (MAE). Table 2 below shows the average values of each metric across all test wind data fields for each model. All deep learning methods assessed outperform bicubic interpolation. Both diffusion models sample noise in the forward process by preserving some of the original input signal, which may drive outputs to more closely represent the ground truth data. The WiREDiff outperforms the SR3 in terms of both metric values.

Model	MSE	MAE
WiREDiff	**0.043**	**0.091**
SR3	*0.082*	*0.11*
SRCNN	0.15	0.31
Bicubic	0.21	0.36

Table 2: Summary of Average Metric Values. The diffusion models we introduce to this field are most accurate, with the WiREDiff achieving the best performance.

However, data similarity metrics are limited in scope and do not offer a comprehensive method of evaluating wind data fields as they cannot capture the ability to replicate high-frequency features. Thus, we validated the super-resolved wind speed outputs by generating kinetic energy spectra for each model which measure the distribution of energy across the various wavenumbers, k (28; 29).

In Figure 7, energy is conserved in the inertial range of each energy spectrum and cascades at higher wavenumbers. Bicubic interpolation and the SRCNN perform visibly worse than the WiREDiff in capturing high-frequency data consistent with turbulence theory. The WiREDiff deviates the least from the energy spectrum of the ground truth data, followed closely by the SR3, which suggests that diffusion-based wind downscaling approaches most successfully learn physical relationships across various frequencies that classical techniques are less capable of recreating.

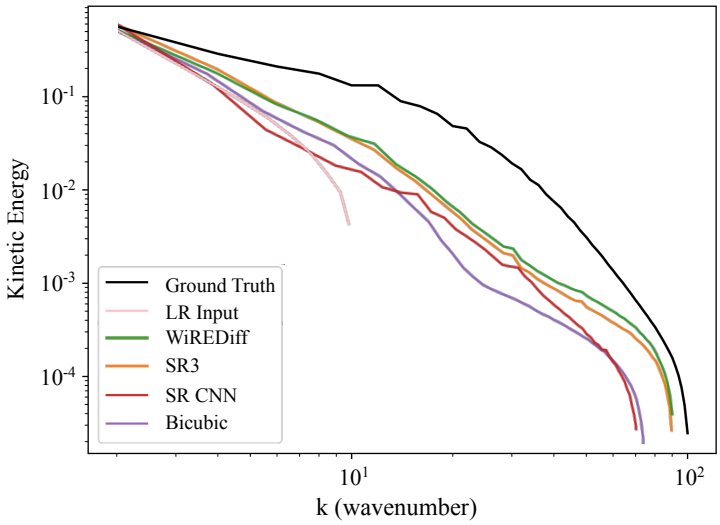

Figure 7: Kinetic energy spectra for the data fields corresponding to each HR output, averaged over all wind test data. The WiREDiff most closely matches the turbulent physics of the ground truth data.

5 Discussion and Future Works

These results indicate that diffusion-based models have significant applications in climate scenarios, as they more accurately and reliably generate results that match the spectral dynamics of the ground truth than traditional CNNs or interpolation techniques. In particular, our proposed model, the WiREDiff, demonstrates promising improvements upon standard SR techniques. When compared with averaging the distribution of SR3 outputs, the WiREDiff approach is less computationally expensive—inference can be accomplished with a single forward diffusion pass as opposed to running inference multiple times and averaging the generated outputs. Additionally, our benchmarking results show that the WiREDiff generates the most accurate and physically realistic SR data fields. We suspect that the WiREDiff model may outperform the SR3 model because N-Con*ffusion* produces well-defined lower and upper bounds for the confidence interval of each pixel. Averaging out the lower and upper bounds provides less room for error, so on average across all the pixels, we are producing consistently smaller errors. Without this calibration, the SR3 model has higher tendencies for outliers that stray further away from the true model.

To better determine the success of diffusion models in super-resolving wind speeds, it is important that in future works, we extend the benchmark to other classes of SR models. In particular, we aim to examine the potential of generative adversarial networks (GANs) (30) and different stochastic SR techniques such as variational auto-encoders (VAEs) (31) and normalizing flows (32). Validating these models on data outside of NREL's WIND Toolkit and NSRDB will also widen the scope of this study and evaluate their generalization to various geographic locations and timespans. In the context of climate change, augmenting more recent data to our training and validation dataset may improve generalizability in predicting fluctuations in wind speeds across longer periods of time.

While traditional super resolution (SR) techniques can boost model granularity by minimizing pixel-level accuracy, their outputs may not capture the underlying physical processes at work. In Figure 7, while the WiREDiff outperformed the simpler approaches included in the benchmark, there is much room for improvement in recreating the energy spectra of the ground truth data. The physics-informed neural network for SR (PINNSR) demonstrated the potential for incorporating a physics-based loss in diffusion models to produce realistic outputs that preserve the laws of turbulent flow and kinetic energy (33; 34). Extending this methodology to the domain of wind speed predictions may result in improved spatial SR performance with SR3 and WiREDiff. One potential implementation for a kinetics-based loss function could compute the Kullback-Leibler (KL) divergence between the true posterior kinetic energy distribution and that of an SR model (35), averaged over all wavenumbers k. Incorporating this term in the training objective of SR3 or other SR models may drive outputs to be more physically consistent with the ground truth data in the kinetic energy spectra generated in Figure 7.

In line with using conformal prediction in diffusion models, future exploration may experiment with applying quantile regression on partially diffused samples (i.e. in an earlier intermediate step of the denoising process as opposed to the final step) to drive final outputs to fall within a tighter confidence interval. Additionally, designing a model to super-resolve spatiotemporally using conformal timeseries forecasting may more comprehensively improve the physical fidelity of outputs across timesteps and be more applicable to power grid planning than single-frame SR (36).

6 Conclusion

In this paper, we introduce and evaluate diffusion methods on national wind data. We introduce conformal quantile regression methodologies to produce SR outputs by averaging the upper and lower bounds of the interval of potential wind speed values. Comprehensive experimental results demonstrate how this approach outperforms traditional approaches to SR with respect to accuracy and physical fidelity. The benchmarking assessments show the qualitative and quantitative performance and limitations of the model proposed. Our GitHub repository[1] provides information on how to access the machine learning-ready dataset of wind speed data fields used in our experiments as well as all base code for training the WiREDiff and running evaluations on the model outputs.

References

[1] *Annual Energy Outlook 2021*. U.S. Energy Information Administration, 2021.

[2] W. J. Gutowski, P. A. Ullrich, A. Hall, L. R. Leung, T. A. O'Brien, C. M. Patricola, R. Arritt, M. Bukovsky, K. V. Calvin, Z. Feng *et al.*, "The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information," *Bulletin of the American Meteorological Society*, vol. 101, no. 5, pp. E664–E683, 2020.

[3] C. Sweeney and P. Lynch, "Adaptive post-processing of short-term wind forecasts for energy applications," *Wind Energy*, vol. 14, no. 3, pp. 317–325, 2011.

[4] A. Clifton, B.-M. Hodge, C. Draxl, J. Badger, and A. Habte, "Wind and solar resource data sets," *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 7, no. 2, p. e276, 2018.

[5] S. L. Cox, A. J. Lopez, A. C. Watson, N. W. Grue, and J. E. Leisch, "Renewable energy data, analysis, and decisions: A guide for practitioners," 3 2018. [Online]. Available: https://www.osti.gov/biblio/1427970

[6] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *European conference on computer vision*. Springer, 2014, pp. 372–386.

[7] J. Baño-Medina, R. Manzanas, and J. M. Gutiérrez, "Configuration and intercomparison of deep learning neural models for statistical downscaling," *Geoscientific Model Development*, vol. 13, no. 4, pp. 2109–2124, 2020.

[8] C. D. Watson, C. Wang, T. Lynar, and K. Weldemariam, "Investigating two super-resolution methods for downscaling precipitation: Esrgan and car," 2020.

[9] J. Leinonen, D. Nerini, and A. Berne, "Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, 2021.

[10] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "Deepsd: Generating high resolution climate change projections through single image super-resolution," in *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, 2017, pp. 1663–1672.

[11] R. Kaltenboeck, G. Croonen, H. Ganster, M. Gruber, K. Hennermann, M. Kerschbaum, C. Nowak, H. Mayer, S. Mayer, and M. Uray, "Image processing for weather radar data correction for aeronautical meteorology," 01 2012.

[1] https://github.com/RupaKurinchiVendhan/WiREDiff

[12] K. Fukami, K. Fukagata, and K. Taira, "Super-resolution reconstruction of turbulent flows with machine learning," *Journal of Fluid Mechanics*, vol. 870, pp. 106–120, may 2019. [Online]. Available: https://doi.org/10.1017%2Fjfm.2019.238

[13] C. M. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, K. Kashinath, M. Mustafa, H. A. Tchelepi, P. Marcus, Prabhat, and A. Anandkumar, "Meshfreeflownet: A physics-constrained deep continuous space-time super-resolution framework," *CoRR*, vol. abs/2005.01463, 2020. [Online]. Available: https://arxiv.org/abs/2005.01463

[14] H. Li, Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Srdiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222000522

[15] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *CoRR*, vol. abs/1907.05600, 2019. [Online]. Available: http://arxiv.org/abs/1907.05600

[16] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *CoRR*, vol. abs/2006.11239, 2020. [Online]. Available: https://arxiv.org/abs/2006.11239

[17] T. Hoar and D. Nychka, "Statistical downscaling of the community climate system model (ccsm) monthly temperature and precipitation projections," April 2008. [Online]. Available: https://gisclimatechange.ucar.edu/sites/default/files/users/Downscaling.pdf

[18] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," *CoRR*, vol. abs/1503.03585, 2015. [Online]. Available: http://arxiv.org/abs/1503.03585

[19] A. N. Angelopoulos, A. P. Kohli, S. Bates, M. I. Jordan, J. Malik, T. Alshaabi, S. Upadhyayula, and Y. Romano, "Image-to-image regression with distribution-free uncertainty quantification and applications in imaging," 2022.

[20] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978. [Online]. Available: http://www.jstor.org/stable/1913643

[21] E. Horwitz and Y. Hoshen, "Conffusion: Confidence intervals for diffusion models," 2022.

[22] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *arXiv:2104.07636*, 2021.

[23] C. Draxl, B. Hodge, A. Clifton, and J. McCaa, "Overview and meteorological validation of the wind integration national dataset toolkit," National Renewable Energy Laboratory, Tech. Rep., 2015.

[24] C. Draxl, A. Clifton, B.-M. Hodge, and J. McCaa, "The wind integration national dataset (wind) toolkit," *Applied Energy*, vol. 151, p. 355–366, 2015.

[25] W. Lieberman-Cribbin, C. Draxl, and A. Clifton, "Guide to using the wind toolkit validation code," National Renewable Energy Laboratory, Tech. Rep., 2015.

[26] J. King, A. Clifton, and B. Hodge, "Validation of power output for the wind toolkit," National Renewable Energy Laboratory, Tech. Rep., 2014.

[27] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," 2015.

[28] A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers," *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 434, no. 1890, pp. 9–13, 1991.

[29] A. N. Kolmogorov, V. Levin, J. C. R. Hunt, O. M. Phillips, and D. Williams, "Dissipation of energy in the locally isotropic turbulence," *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 434, no. 1890, pp. 15–17, 1991.

[30] K. Stengel, A. Glaws, D. Hettinger, and R. N. King, "Adversarial super-resolution of climatological wind and solar data," *Proceedings of the National Academy of Sciences*, vol. 117, no. 29, pp. 16 805–16 815, 2020.

[31] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *2nd International Conference on Learning Representations, ICLR*, 2014.

[32] A. Lugmayr, M. Danelljan, L. Van Gool, and R. Timofte, "Srflow: Learning the super-resolution space with normalizing flow," in *ECCV*, 2020.

[33] M. Bode, M. Gauding, Z. Lian, D. Denker, M. Davidovic, K. Kleinheinz, J. Jitsev, and H. Pitsch, "Using physics-informed super-resolution generative adversarial networks for subgrid modeling in turbulent reactive flows," *CoRR*, vol. abs/1911.11380, 2019. [Online]. Available: http://arxiv.org/abs/1911.11380

[34] C. Wang, E. Bentivegna, W. Zhou, L. J. Klein, and B. Elmegreen, "Physics-informed neural network super resolution for advection-diffusion models," *CoRR*, vol. abs/2011.02519, 2020. [Online]. Available: https://arxiv.org/abs/2011.02519

[35] J. M. Joyce, *Kullback-Leibler Divergence*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 720–722. [Online]. Available: https://doi.org/10.1007/978-3-642-04898-2_327

[36] K. Stankeviciute, A. M. Alaa, and M. van der Schaar, "Conformal time-series forecasting," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 6216–6228. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/312f1ba2a72318edaaa995a67835fad5-Paper.pdf