

# TAUDiff: Highly efficient kilometer-scale downscaling using generative diffusion models

---

**Rahul Sundar**  
Verisk Analytics  
Hyderabad, IND 500081  
rahul.sundar@verisk.com

**Yucong Hu**  
Verisk Analytics  
Ottawa, Ontario  
Canada K1J 9J8  
yucong.hu@verisk.com

**Nishant Parashar**  
Verisk Analytics  
Hyderabad, IND 500081  
nparashar@verisk.com

**Antoine Blanchard**  
Verisk Analytics  
Boston, MA 02115  
ablanchard@verisk.com

**Boyko Dodov**  
Verisk Analytics  
Boston, MA 02115  
bdodov@verisk.com

## Abstract

Deterministic regression-based downscaling models for climate variables often suffer from spectral bias, which can be mitigated by generative models like diffusion models. To enable efficient and reliable simulation of extreme weather events, it is crucial to achieve rapid turnaround, dynamical consistency, and accurate spatio-temporal spectral recovery. We propose an efficient correction diffusion model, TAUDiff, that combines a deterministic spatio-temporal model for mean field downscaling with a smaller generative diffusion model for recovering the fine-scale stochastic features. We demonstrate the efficacy of this approach on downscaling atmospheric wind velocity fields obtained from coarse GCM simulations. We then extend TAUDiff for computationally efficient kilometer-scale downscaling of atmospheric wind velocity fields. Owing to low inference times, our approach can ensure quicker simulation of extreme events necessary for estimating associated risks and economic losses.

## 1 Introduction

Weather extremes are on the rise due to accelerated climate change [1]. Given their potential to severely damage life and property, it is becoming increasingly important to estimate their frequency, associated risks and economic losses beforehand [2–4]. By insuring for such losses, we can become more resilient towards extreme events [5]. Climate risk modeling often relies on historical Earth system observations [6] or physics-based general circulation models (GCMs) [7] to generate climate projections. Typically, GCMs operate at a coarse resolution ( $O(10) - O(10^2)$ km) due to compute limitations. This leads to incorrect characterization of weather extremes. In recent years, machine-learning based statistical downscaling approaches have been explored to obtain realistic well-resolved climate data over specific regions [8–10]. These methods leverage historical Earth system observation data to create a non-linear mapping from bias-corrected coarse GCM simulations to the desired higher-resolution outputs.

While deterministic regression models effectively capture large-scale features, they struggle with fine-scale stochastic atmospheric processes due to low-frequency spectral bias [11]. This limitation has recently led to the adoption of generative models like GANs [10, 12, 13], and denoising diffusion models for downscaling tasks [14, 15]. Denoising diffusion models [16–18] are particularly promising due to their stability in training, reliable convergence, and high output quality. However, sampling is often time consuming. Addressing this, Karras *et al.* [19] explored the design space of such diffusion

models, and proposed the elucidated diffusion model (EDM) which successfully reduced the number of model evaluations (from  $O(10^3)$  to  $O(10)$ ) required to generate a single sample. Motivated by this, a correction diffusion model (CorrDiff) [14] was proposed for kilometer-scale downscaling. CorrDiff combined a UNet-based deterministic model to map the mean field and an EDM correction to capture fine-scale stochastic content. More recently, Merizzi *et al.* [20] proposed an ensemble-diffusion model for kilometre-scale downscaling 10-metre wind speeds. Moreover, with their ensemble-diffusion model, they reported an inference time of approximately two hours for one year worth of data.

In the context of extreme-event simulation, it is vital that both short- and long-term event statistics of downscaled data be consistent with historical observations. So, the lack of temporal modeling in downscaling models may affect dynamical consistency of downscaled data (e.g. distorted propagation of storm fronts). One could address this issue by borrowing techniques from video generation/prediction [21], as explored by Yoon *et al.* [22] for regional weather forecasting. However, such techniques have not yet been explored for downscaling. Moreover, large models as in [14] are computationally intensive to train and infer. This prohibits the generation of even relatively small ( $O(10^3)$ ) extreme-event datasets, which are crucial for accurately quantifying climate tail risk. Given a good mean-field model, it is possible that a smaller and computationally efficient diffusion model would suffice. This would reduce overall computational demands, inference times, and improve efficiency for real-time use.

To address these challenges, we propose a computationally efficient Temporal Attention Unit enhanced Diffusion model (TAUDiff) that integrates (a) a video prediction model for dynamically consistent mean-field downscaling, and (b) a smaller guided denoising diffusion model for stochastically generating the fine-scale features. We train the models on atmospheric wind fields obtained from reanalysis dataset. The performance of TAUDiff is first compared against separate mean-field regression and end-to-end diffusion models under a fixed training budget. We evaluate the downscaling performance of TAUDiff on an ensemble of bias-corrected coarse GCM outputs using various spectral statistics. As a proof of concept, we also evaluate if TAUDiff is capable of mapping a reanalysis dataset of  $0.25^\circ$  (about 25km) grid resolution to another dataset at  $0.0625^\circ$  (about 5km) grid resolution. We finally discuss its potential in producing accurate and computationally efficient extreme-event datasets, and the implications of model inference times and carbon footprint offsetting.

## 2 Methods

**Overview** We demonstrate the efficacy of TAUDiff in downscaling atmospheric wind fields over the European region through two case studies: (i) model performance validation and GCM output downscaling to  $0.25^\circ$  resolution, and (ii) extension of TAUDiff to kilometer-scale downscaling ( $0.25^\circ$  to  $0.0625^\circ$ ). Instead of a single time instance input like in [14], our approach uses a deterministic mean-field regression component that takes a temporal sequence of coarse wind velocity snapshots with orography data as input. Here, the high-resolution wind fields from the final time step of the sequence serves as the target. We then train a generative diffusion model for correcting the outputs of the mean-model. We discuss the mean-field and diffusion model components of TAUDiff below.

**TAUDiff Component 1: Mean model** We adopt the Simple yet better Video Prediction (SimVP) architecture [23] consisting of a spatial backbone, and a translator for temporal modelling, ensuring temporal coherence and simplicity as compared to the more complex transformer based architectures [21]. We specifically consider a UNet for the spatial backbone, and the temporal attention unit (TAU) [24] for the translator. The TAU first independently models spatial dependency via static, and both cross-channel and temporal dependencies using dynamical attention units, respectively, and then combines them. We train the mean model using a weighted combination of mean absolute error (MAE), mean squared error (MSE), and also to maintain dynamical consistency, physics-based losses on advection ( $\mathbf{u} \cdot \nabla \mathbf{u}$ ), vorticity ( $\nabla \times \mathbf{u}$ ) and divergence ( $\nabla \cdot \mathbf{u}$ ) of wind fields ( $\mathbf{u}$ ) are considered. Although dynamically consistent predictions are possible with this mean model, the downscaled fields still lack the stochastic fine scale features. This is where Component 2 comes into play.

**TAUDiff Component 2: Correction diffusion model** To capture the residual stochastic fine scale features (which cannot be captured by the mean model), we build a relatively small correction diffusion model ( $\sim O(1)$  million (M) parameters) trained using a score-matching loss [25]. To maintain consistency of our approach, we use a SimVP architecture as in the mean model but with

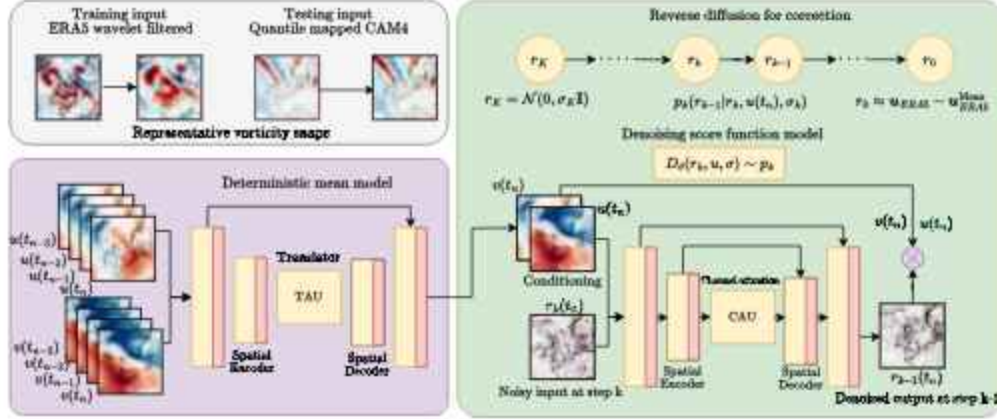


Figure 1: Schematic of the TAUDiff model

a residual dense UNet as the spatial backbone. Once the model is trained, a data sample can be generated by solving a stochastic differential equation modelling a reverse diffusion process [17, 19]. Since the conditional input to the diffusion model is the mean model output (for a single time instance), the TAU morphs into a Channel Attention Unit (CAU) where the dynamical attention unit now models cross-channel dependencies and their relative importance (see figure 1 for a detailed schematic of TAUDiff).

### 3 Results and discussion

#### 3.1 Model validation and GCM downscaling

**Experimental protocol** We demonstrated the potential of our framework by comparing three models: a deterministic mean-field regression, an end-to-end diffusion, and our TAUDiff model, each with  $O(10)$ M trainable parameters overall. For our training, we use the atmospheric reanalysis dataset (ERA5) at  $0.25^\circ$  lat-lon resolution produced by the European Center for Medium-range Weather Forecasts (ECMWF) [6]. Instead of using coarse interpolation, we use lowpass spherical wavelet filtering [26, 27] to create band-limited low-resolution ERA5 fields to ensure proper scale separation. This approach closely mirrors real-world scenarios where bias-corrected GCM data lacks fine-scale spatio-temporal features. The models were trained over 40 years of ERA5 atmospheric wind data over Europe (1980-2020) (figure 2(a)), and validated over 2021-23. All the models were trained on a single T4 GPU over 50 epochs, with training times of 24, 48, and 60 hours for the mean, end-to-end diffusion, and the TAUDiff models, respectively.

**Validation** Diffusion models can create noisy artifacts in the samples and they only get magnified when plotting the gradients of the outputs. Hence, for a rigorous qualitative evaluation, we consider vorticity snapshots instead of using wind speeds like in other works [20]. Qualitatively, the vorticity contour predictions of mean and TAUDiff models demonstrate dynamical consistency of storm fronts, whereas the end-to-end diffusion model distorts them due to noise injection (circled zone in figure 2(b)). Quantitatively, pointwise statistics computed over validation years 2021-23 show good recovery of spatial and temporal spectrum for TAUDiff, while mean model underrepresents, and end-to-end diffusion overrepresents higher temporal frequencies, respectively (figures 2(c) and 2(d)).

**Testing** The performance of our TAUDiff model was then evaluated on downscaling bias-corrected coarse GCM obtained wind fields over 40 years. We use the Community Atmosphere Model 4.0 (CAM4) [28] (at  $1^\circ$  resolution) as the coarse GCM in this study. Bias correction is done by quantile-mapping [29] the 40-year distribution of each grid cell to that of ERA5, wavelet-filtered to GCM resolution. As earlier, we obtain physically consistent output, and remarkable spectral recovery (figures 3(a-c)). Although only a simple quantile mapping [29] is adopted for bias correction, we see good agreement with ERA5 ground truth in the local storm counts (see figure 3(d)). This cements the need for a stochastic correction using a diffusion model for accurate extreme-event risk estimation.



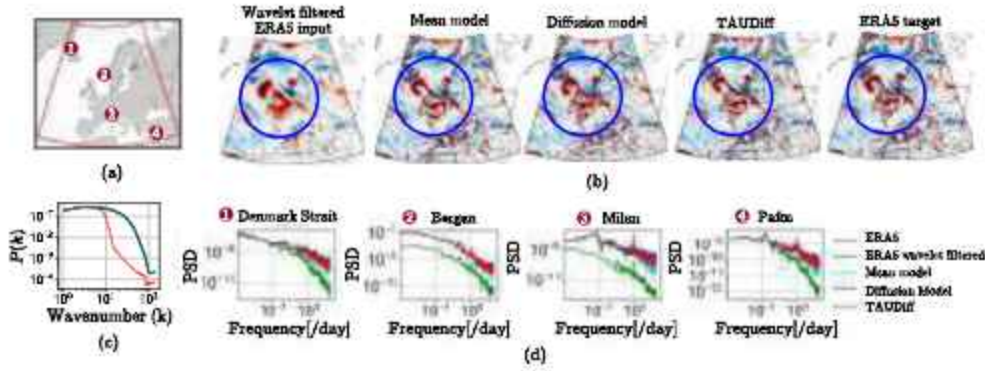


Figure 2: (a) European region used for training the downscaling models, with select locations used for evaluating performance. Comparison of model predictions: (b) vorticity snapshot at UTC: 2023-12-31 21:00, (c) spatial spectrum, and (d) temporal spectra at select locations shown in (a).

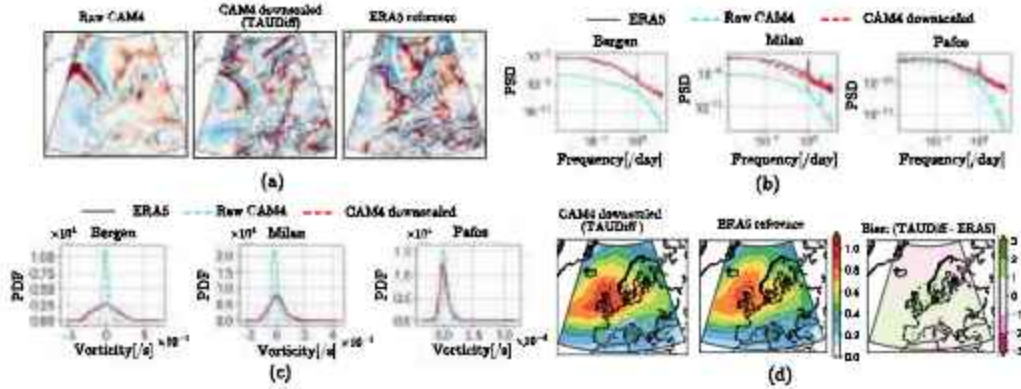


Figure 3: Assessment of downscaling performance on bias corrected CAM4 data: (a) Vorticity contours at a representative time instance, (b) temporal spectrum, (c) vorticity distributions, and (d) local storm counts.

### 3.2 Extension to km-scale downscaling

**Methodology extension** Diffusion models require multiple function evaluations while sampling [18, 16]. As a result, in case of km-scale regional downscaling, ensemble methods [20], a large model size [19], and high image resolutions can vastly increase the inference times. Hence, for applications necessitating km-scale downscaling, it would be beneficial to have TAUDiff operate at a coarser resolution to reduce inference times. Since the models can be trained on reanalysis data, a single ensemble member of the diffusion model should be representative of the field-statistics. The generated samples at coarser resolution can then be downsampled using deterministic UNet based regression model to recover the fine-resolution data as depicted in the schematic figure 4.

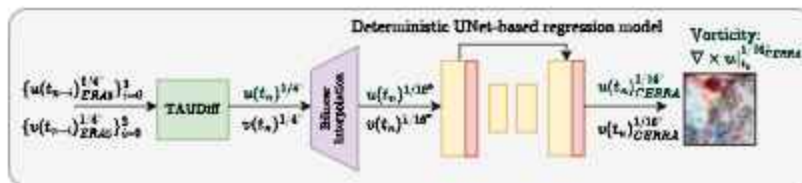


Figure 4: Schematic of the km-scale downscaling pipeline.

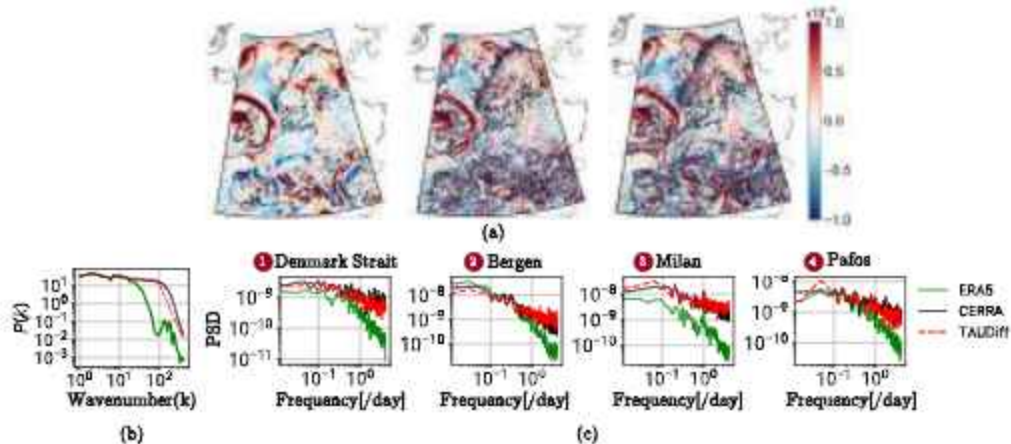


Figure 5: Assessment of ERA5 to CERRA downscaling performance: (a) Vorticity contours at UTC: 2010-11-10 21:00, (b) spatial spectrum, and (c) temporal spectra at select locations as shown in 2(a).

**Experimental protocol** As a proof of concept, we consider the example of downscaling ERA5 atmospheric wind velocity fields at  $0.25^\circ$  resolution to the Copernicus European Regional Reanalysis (CERRA) [30] dataset resolution of  $0.0625^\circ$ . The CERRA dataset is natively obtained on a cartesian grid. However, in our study, we project the CERRA data onto a lat-lon grid of similar resolution. Unlike Merrizi *et al.* [20], where only a sub-region encompassing Italy and Alps was considered, we consider the entire European region for training as shown in figure 2(a). While the size of mean model component of TAUDiff remains the same as in the earlier experiments, the correction-diffusion, and the deterministic UNet based regression models now consist  $O(10)$ M, and  $O(10^2)$  thousand trainable parameters, respectively. This is to ensure that the finer scales are well captured by the models. We train this TAUDiff model over 10 years (2011-2020) of input-target pairs of ERA5, and  $0.25^\circ$  interpolated CERRA over the European region (see figure 2(a)), and tested over the year 2010. The deterministic UNet based regression model was independently trained over the same domain using high resolution CERRA wind velocity fields as targets, and interpolated CERRA wind velocity fields as the model inputs. At inference, we chain TAUDiff, and the regression model together to generate a sample.

**Testing** We obtain physically consistent fields with good qualitative, and quantitative agreement with CERRA data (see figure 5). If one were to pass inputs at  $0.0625^\circ$  resolution to the diffusion correction model, it can take approximately 76 minutes for downscaling one year on a single NVIDIA H100 GPU. However, since TAUDiff is now operating at  $0.25^\circ$  resolution, we obtain a reasonable inference time of approximately 4 minutes per one year of data. In both the cases, 20 reverse diffusion steps were considered. With frameworks like NVIDIA TensorRT, our preliminary investigations indicate that it is also possible to further reduce the inference times by up to three times the original.

## 4 Conclusion

Overall, our proposed video-prediction-based TAUDiff model and its km-scale downscaling extension demonstrates dynamically consistent downscaling, remarkable reconstruction of spatio-temporal fine scale features, and viable inference times with the use of a small correction diffusion model. Since coarse and fine scale content of the atmospheric fields are resolved well, accurate estimation of storm statistics was possible and excellent performance on spectrum and storm statistics were obtained. It was also demonstrated as a proof of concept that even when TAUDiff operated on coarser resolutions, the ERA5-CERRA downscaling performance was remarkable. Thus, we show that TAUDiff has immense potential in accurate, and computationally efficient estimation of extreme weather events. Our future work would involve staging of TAUDiff models to obtain multi-resolution outputs for extreme weather event simulations while maintaining reasonable inference times.



## References

- [1] P. Hoeppe. Trends in weather related disasters—consequences for insurers and society. *Weather and Climate Extremes*, 11:70–79, 2016.
- [2] T. Houser, S. Hsiang, R. Kopp, K. Larsen, M. Delgado, A. Jina, M. Mastrandrea, S. Mohan, R. Muir-Wood, D. J. Rasmussen, et al. *Economic risks of climate change: an American prospectus*. Columbia University Press, 2015.
- [3] C. B. Field, V. Barros, T. F. Stocker, and Q. Dahe. *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2012.
- [4] The Securities and Exchange Commission. SEC proposes rules to enhance and standardize climate-related disclosures for investors. <https://www.sec.gov/news/press-release/2022-46>, March 21, 2022.
- [5] P. J. Robinson, W. J. W. Botzen, S. Duijndam, and A. Molenaar. Risk communication nudges and flood insurance demand. *Climate Risk Management*, 34:100366, 2021.
- [6] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.
- [7] C. Wang, L. Zhang, S.-K. Lee, L. Wu, and C. R. Mechoso. A global perspective on CMIP5 climate model biases. *Nature Climate Change*, 4(3):201–205, 2014.
- [8] S. Park, K. Singh, A. Nellikkattil, E. Zeller, and M. Mai, T. D. and Cha. Downscaling earth system models with deep learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3733–3742, 2022.
- [9] A. Blanchard, N. Parashar, B. Dodov, C. Lessig, and T. Sapsis. A multi-scale deep learning framework for projecting weather extremes. *arXiv preprint arXiv:2210.12137*, 2022.
- [10] K. Daust and A. Monahan. Capturing climatic variability: Using deep learning for stochastic downscaling. *arXiv preprint arXiv:2406.02587*, 2024.
- [11] Z. J. Xu, Y. Zhang, and T. Luo. Overview frequency principle/spectral bias in deep learning. *arXiv preprint arXiv:2201.07395*, 2022.
- [12] N. Rampal, P. B. Gibson, S. Sherwood, G. Abramowitz, and S. Hobeichi. A robust generative adversarial network approach for climate downscaling and weather generation. *Auhorea Preprints*, 2024.
- [13] G. Li and G. Cao. Generative Adversarial Models for Extreme Downscaling of Climate Datasets. *arXiv preprint arXiv:2402.14049*, 2024.
- [14] M. Mardani, N. Brenowitz, Y. Cohen, J. Pathak, C.Y. Chen, C.C. Liu, A. Vahdat, K. Kashinath, J. Kautz, and M. Pritchard. Generative residual diffusion modeling for km-scale atmospheric downscaling. *arXiv preprint arXiv:2309.15214*, 2023.
- [15] R. A. Watt and L. A. Mansfield. Generative diffusion-based downscaling for climate. *arXiv preprint arXiv:2404.17752*, 2024.
- [16] J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.
- [17] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.
- [18] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.
- [19] T. Karras, M. Aittala, T. Aila, and S. Laine. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35:26565–26577, 2022.
- [20] F. Merizzi, A. Asperti, and S. Colamonaco. Wind speed super-resolution and validation: from era5 to cerra via diffusion models. *arXiv preprint arXiv:2401.15469*, 2024.
- [21] C. Tan, S. Li, Z. Gao, W. Guan, Z. Wang, Z. Liu, L. Wu, and S. Z. Li. Openstl: A comprehensive benchmark of spatio-temporal predictive learning. *Advances in Neural Information Processing Systems*, 36:69819–69831, 2023.

- [22] D. Yoon, M. Seo, D. Kim, Y. Choi, and D. Cho. Deterministic guidance diffusion model for probabilistic weather forecasting. *arXiv preprint arXiv:2312.02819*, 2023.
- [23] Z. Gao, C. Tan, L. Wu, and S. Z. Li. Simvp: Simpler yet better video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3170–3180, 2022.
- [24] C. Tan, Z. Gao, L. Wu, Y. Xu, J. Xia, S. Li, and S. Z. Li. Temporal attention unit: Towards efficient spatiotemporal predictive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18770–18782, 2023.
- [25] Y. Song and S. Ermon. Improved techniques for training score-based generative models. *Advances in Neural Information Processing Systems*, 33:12438–12448, 2020.
- [26] P. Schröder and W. Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 161–172, 1995.
- [27] C. C. da Silva, C. Lessig, B. Dodov, H. Dijkstra, and T. Sapsis. A local spectral exterior calculus for the sphere and application to the shallow water equations. *arXiv preprint arXiv:2005.03598*, 2020.
- [28] R. B. Neale, J. H. Richter, A. J. Conley, S. Park, P. H. Lauritzen, A. Gettelman, D. L. Williamson, P. J. Rasch, S. J. Vavrus, M. A. Taylor, et al. Description of the NCAR Community Atmosphere Model (CAM 4.0), NCAR Tech. Note, TN-485, 212, 2010.
- [29] D. Maraun. Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, 26(6):2137–2143, 2013.
- [30] S. Schimanke, M. Ridal, P. Le Moigne, L. Berggren, P. Undén, R. Randriamampianina, U. Andrea, E. Bazile, A. Bertelsen, P. Brousseau, et al. Cerra sub-daily regional reanalysis data for europe on single levels from 1984 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, 2021.