Working Paper

Hope or Hype: AI-driven Risk Modeling of Tropical Cyclones[1]

Project Period: September 2023 - August 2024

Initial Draft: August 2024

Revision: January 2025

Gan Zhang

Assistant Professor, Department of Climate, Meteorology & Atmospheric Sciences

Faculty Fellow, Office of Risk Management and Insurance Research, Gies Business School

Key Points

- AI technologies recently made breakthroughs in weather forecasting and showed promise in climate simulations.
- We develop two representative use cases to demonstrate that hurricane risk modeling can benefit from the recent AI advancements.
- Our baseline generative models trained with the atmospheric reanalysis dataset can emulate relatively realistic hurricane hazards (e.g., wind and precipitation).
- Our experimental system built on Google's AI-physics climate model can simulate realistic tropical cyclone activity and make skillful seasonal predictions.

[1] The original proposal focuses on modeling with Generative Adversarial Network (GAN). This exploration documented by this report exceeded the original scope substantially and covered more TC-related risk applications.

- The existing AI technologies still have limitations but will likely continue improving and unlock greater value for businesses seeking to enhance their risk modeling capability.

1 Introduction

The billion-dollar disasters caused by weather perils have been increasing in the US [1] and other parts of the world [2]. In 2024 alone, the US experienced five tropical cyclones (TCs), including Hurricanes Helene and Milton, making six Atlantic hurricane seasons in 2017–2024 ranking among the top ten costliest. With the ongoing climate change and societal shifts (e.g., migration and inflation), the financial risks associated with weather perils, such as hurricanes, severe thunderstorms, and wildfires, may experience profound changes [3]. These converging factors have led to a surge in demand for risk management services and instruments. For example, the insurance and reinsurance industry increasingly relies on instruments like catastrophe bonds, helping the market size double over 2014-2023 and hit a record high of more than $47 billion [4]. This growing market attracted new participants, including hedge funds, some of which devised highly profitable trading strategies by exploiting pricing inefficiencies [5]. Businesses unprepared for weather-climate catastrophe events face substantial losses (e.g., PG&E), while those who can accurately assess and price these risks stand to gain significant advantages.

The pricing of weather-climate catastrophe risks is challenging for conventional finance modeling, particularly with changes in physical and societal risk exposure. This complexity has led to the development of specialized catastrophe models, which attempt to combine the strengths of both physical and statistical approaches. While pure physical models offer the ability to simulate rare and extreme events, they are computationally expensive, require specialized expertise, and are prone to biases. Conversely, statistical models are more affordable and easier to implement, but they struggle to account for unprecedented events in a changing environment. Integrating the insights from these models is crucial for successful risk assessments and investments, but interested stakeholders often lack transparent, universally accepted solutions. The analytics by academic researchers and commercial vendors often are proprietary, tailored to specific tasks, and sometimes yield conflicting results [6]. To protect their investments and gain a competitive advantage, resourceful players like brokers and hedge funds acquire third-party products from multiple vendors and build internal modeling capabilities. Such a sophisticated approach is often out of reach for less resourceful market participants, leaving them in disadvantageous positions.

Recently, AI models made breakthroughs in image generation and weather simulations, raising interesting possibilities for weather-climate risk modeling. Similar to how AI can synthesize realistic-looking human and pet images, AI models can be trained to synthesize potential weather peril scenarios based on user prompts. The training can leverage historical weather data known as "reanalysis," a comprehensive dataset combining observational data with outputs from physical models [7]. When tailored to specific business needs, an AI model trained with reanalysis or other observational data can synthesize weather perils and contribute to flexible scenario analyses. In parallel to generative AI, another exciting breakthrough occurs in weather forecasting. Since 2022, the performance of AI models trained with reanalysis datasets improved quickly and has overtaken the best physical model in some forecasting tasks [8–11]. Compared with existing physical models, the simulation speed of these new AI models is also up to 10,000 times faster and thus more computationally affordable. The combination of accuracy and speed can unlock applications such as agile user-driven analytics and extensive sampling of rare, high-impact events (e.g., hurricane landfall). New AI models are also making strides in long-term climate simulations [11, 12]. If successful, they may offer an independent, cost-effective workflow to synthesize weather perils for risk analysis and complement existing tools in catastrophe modeling.

Building on the potential of AI outlined above, we now turn attention to two specific applications in hurricane risk modeling: 1) Generative modeling of hurricane hazards; and 2) AI-driven climate prediction. The first case seeks to complement existing catastrophe modeling tools by striking a balance of data granularity and computational affordability. The second task explores what may enable new AI models to model and predict risks across time scales ranging from days to seasons. While traditional catastrophe models often rely on complex, layered systems tailored to specific tasks, this paper explores how rapidly improving AI tools offer the opportunity to streamline these processes. Specifically, we investigate the feasibility of these AI-driven approaches and evaluate the potential for new players from the technology sector, such as tech giants and innovative start-ups, to enter the risk modeling space.

The rest of this work paper is organized as follows. Section 2 introduces the data, infrastructure, and modeling methods. Section 3 describes the preliminary results of generative modeling and AI-driven climate prediction. Section 4 provides a summary and discusses future directions.

2 Data and Methods

2.1 Observational Data

This study works with publicly accessible datasets to ensure the replicability and flexibility of future research. Following most existing weather-AI studies, we use the ERA5 reanalysis [7] for model training and validation. The reanalysis is generated by a numeric weather forecast model that follows physical laws and ingests multi-sourced observational data (e.g., weather station and satellite data). This approach allows the model to leverage observations as much as possible and infer the weather conditions where observations are scarce (e.g., oceanic regions that spawn hurricanes). Compared with raw observational data, the gridded reanalysis data provide a comprehensive, easy-to-access set of variables relevant to disaster damages. While the ERA5 reanalysis has limitations (e.g., sparse grid points and precipitation biases), the dataset is a useful starting point for experimenting with new concepts and tools.

The author also uses two additional datasets for hurricane research. The first is the International Best Track Archive for Climate Stewardship (IBTrACS) [13]. This dataset includes a collection of hurricane information (e.g., location and intensity) based on multi-sourced observations and quality-controlled by operational forecasters and experts. For hurricane research, the dataset is widely considered truth-like and more trustworthy than reanalysis datasets, which struggle with representing intense hurricanes [14]. The second dataset is Tropical Cyclone PRecipitation, Infrared, Microwave, and Environmental Dataset (TC PRIMED) [15]. This dataset combines information from the ERA5, the IBTrACS, and satellite imagery. Using the best track data as the reference, the TC PRIMED aligned individual storms from all the data sources via domain centering and cropping. The newly available TC PRIMED thus mitigates the data pre-processing burden and facilitates our generative modeling effort.

2.2 Modeling

2.2.1 Computing Infrastructure and Preparation

The modeling and simulation of this study rely on graphics processing units (GPUs) and involve deploying new hardware and software infrastructure. While alternatives (e.g., cloud computing) exist, our initial development and concept proof use on-premise resources. This choice makes it straightforward to leverage existing local resources (e.g., IT support and researchers) and train future students. During the project implementation period, the PI secured funding from multiple sources (~$140,000 investments) to install high-performance computing nodes that consist of CPUs and GPUs. The ensuing software and data preparation were conducted by the IT support and the PI. The computing work of this project uses ~25% of the new resources (2× Nvidia L40S GPUs) over multiple weeks. We took a disciplined approach with the computational budget so that interested parties can easily procure resources to validate and improve our work.

2.2.2 Generative Modeling of Hurricane Hazards

Following the original proposal submitted to the Office of Risk Management and Insurance Research, the generative modeling uses the Generative Adversarial Network (GAN) [16] and its variants Conditional GAN [17] as the baseline models. The GANs build on the competition between a generative model and a discriminator model, which challenges both models to improve during the training. But for the data synthesis tasks, the GAN and its variants are usually hard to tune and improve. Leveraging the software infrastructure built for GAN-based experiments, we test additional generative model architectures, such as the diffusion probabilistic models [18] that serve as the foundation of many state-of-the-art image generators. The diffusion models take iterative steps to link labeled images to arbitrary noises. The diffusion models can attain realistic-looking synthesis more easily but tend to take more resources to train and deploy.

We train the GAN and the diffusion models using the northern-hemisphere hurricane samples in the TC-PRIMED from 1999 to 2020. For each hurricane case, we search the ERA5 subset of the TC-PRIMED and save three hazard variables (i.e., near-surface wind speed, precipitation, and sea level pressure). The wind speed and precipitation are key physical drivers of hurricane damage, while the sea-level pressure is a skillful predictor of hurricane damage [19]. The data is available on a 6-hourly basis with a grid spacing of 0.25 degrees in the latitude and longitude dimensions, which approximately corresponds to 25 km in the tropics. We group these 6-hourly snapshots (N=37754) based on the concurrent intensity information from the best track dataset. This results in six categories including Category 0-5 hurricanes. These intensity labels, along with the two-dimensional (latitude × longitude) of hazard information fields, serve as input for training the generative models. Each model is trained for approximately 8 hours on a single GPU. After the training, the models synthesize hazard data of potential hurricanes with user-specified random noise and intensity labels. This concept proof considers the intensity labels only and is possibly the simplest configuration. In principle, additional labeling (e.g., over land or not) can be embedded in the training to create models to accept relevant instructions.

2.2.3 AI-driven Climate Prediction

Compared with existing physical models, recent AI models perform well in forecasting weather for up to two weeks but still face notable challenges in climate prediction. Recent developments have shown models trained with the ERA5 reanalysis can support stable long-range simulations [11, 12]. A promising model is the NeuralGCM developed by Google, which uses the backbone of physical models but represents small-scale physical processes with machine learning methods [11]. Unlike the other data-driven models, the unique approach of the NeuralGCM makes it more straightforward to constrain the atmospheric simulation with climate forcings (e.g., sea surface temperature). Meanwhile, the speed-up of AI models is mostly retained, and the model can simulate realistic hurricane activity in a test using the climate forcings from the 2020 Atlantic hurricane season. After the NeuralGCM study was published by the Nature journal, the CEO of Alphabet and Google, Sundar Pichai, promoted the new hybrid model on social media:

> "NeuralGCM, a breakthrough in climate modeling. It combines physics-based modeling with AI, and is up to 100K times more efficient than other models for simulating the atmosphere, providing scientists with new tools for predicting climate change."

The PI coordinated with the Google research team and became one of the first external researchers to experiment with this new model. Noting that the NeuralGCM natively supports the use of time-varying climate forcing, we set up the NeuralGCM on the newly configured on-premise computer and run new seasonal prediction experiments. These experiments use the assumption of fixed anomalies of sea surface temperature [20], which helped approximate the evolution of the ocean state (e.g., the El Niño–Southern Oscillation) and establish the skill of high-resolution climate models in predicting hurricane activity. Specifically, we conduct prediction experiments initialized on July 1st for years between 1990 and 2023. Each year's prediction lasts for ~160 days and consists of twenty parallel simulations to sample the probability space. In terms of data record length, the 108,800 simulation days of the experiments are approximately 4.5 times that of the historical observation coverage and approximately 1/30 of the extensive physical experiments that the author used for earlier academic studies [21, 22].

This set of seasonal prediction experiments serves multiple purposes. First, this is an initial attempt at the community-driven development of the NeuralGCM and will provide helpful feedback to model developers. Second, this effort helps evaluate the potential of applying the NeuralGCM in seasonal prediction tasks, which may ultimately inform the pricing of hurricane risk. Third, a large amount of simulations can help assess whether the NeuralGCM can synthesize hurricanes and assist in catastrophe modeling. As will be shown in Section 3, our configuration with the current, public version of the NeuralGCM attained performance comparable to existing models, despite several hurdles in the model implementation. Considering license constraints, we briefly document initial results and refrain from discussing ongoing development activities[2].

3 Results

3.1 Generative Modeling of Hurricane Hazards

We evaluate the performance of generative models by comparing samples from the TC-PRIMED, the GAN, and the diffusion models (Figures 1-3). The GAN and diffusion models synthesize storms based on the input of hurricane categories and need random noise to generate diverse samples. We experiment with larger numbers of generated samples (~10^4) and present analyses with a fixed sample size to facilitate a fair comparison with the TC-PRIMED analysis. Specifically, we randomly draw samples from the TC-PRIMED from each intensity category and show the means of storm-centered data fields of near-surface wind speed, precipitation rate, and

[2] Google updated the user license in December 2024 and removed the previous "Non-Commercial" restriction.

sea level pressure. The size of each sample group is equal and set to 100. The sample size is relatively small since the TC PRIMED data have only ~330 samples in the Category 5 group. Accordingly, the means of TC-PRIMED fields are compared to the means of the outputs of the generative models of the same sample size.

The generated samples are similar to those from the TC-PRIMED, even though some differences are apparent. In almost all the cases, the generated samples show distinct hurricane features, including a relatively calm eye and an eyewall with serious hazards. For each intensity category, the GAN produces storms with size and maximum wind speed like those in the TC-PRIMED. A close inspection suggests some samples generated by the GAN have artifacts of small scales. In comparison, the storm hazards from the diffusion model tend to be broader and more severe than those in the TC-PRIMED. The generated samples, however, show reasonable structures even at the fine scale. Finally, both models can generate many samples at a pace much faster than physical model simulations. Nonetheless, the GAN (~5,000 samples per minute) is notably faster than the diffusion model (~30 samples per minute). Overall, the two generative models have different strengths and weaknesses in replicating hurricane samples that resemble those in the training dataset (i.e., the TC-PRIMED).

The fidelity of the training dataset affects the performance of AI models and is also worth remarks. While some studies use the ERA5 as the observational truth in hurricane-related risk analysis [23], the dataset struggles with representing the most extreme wind [14]. This issue is also illustrated in Figure 1. For example, the maximum wind associated with Category-3 hurricanes is expected to be within 50–58 m s^{-1}, but the analyzed samples show much weaker wind speed on average. While higher extreme values appear in individual samples, they are all well below the maximum wind speed values from the IBTrACS. Therefore, all the AI models that use the ERA5 as the training dataset likely underestimate the severity of the extreme wind. Biases are likely present with the precipitation rate and the sea level pressure.

With additional bias correction or tuning, the performance of the generative models likely can improve. For example, a straightforward way for bias correction is to introduce a scaling factor to calibrate the training dataset or the model outputs against the IBTrACS or other datasets. The fine-scale issue of the GAN model may be overcome by moving beyond the baseline model and adopting more sophisticated models. The baseline diffusion model may also be improved via various strategies. For example, the sensitivity of the diffusion model to the intensity label input can be tuned with a parameter that controls the balance between fidelity to the original data and the intensity guidance. Figure 4 shows that higher values of this parameter can lead to more intense storms. We also speculate that the biases of the diffusion model can be mitigated by optimizing the training strategy (e.g., balancing the group sizes) and using more realistic training data.

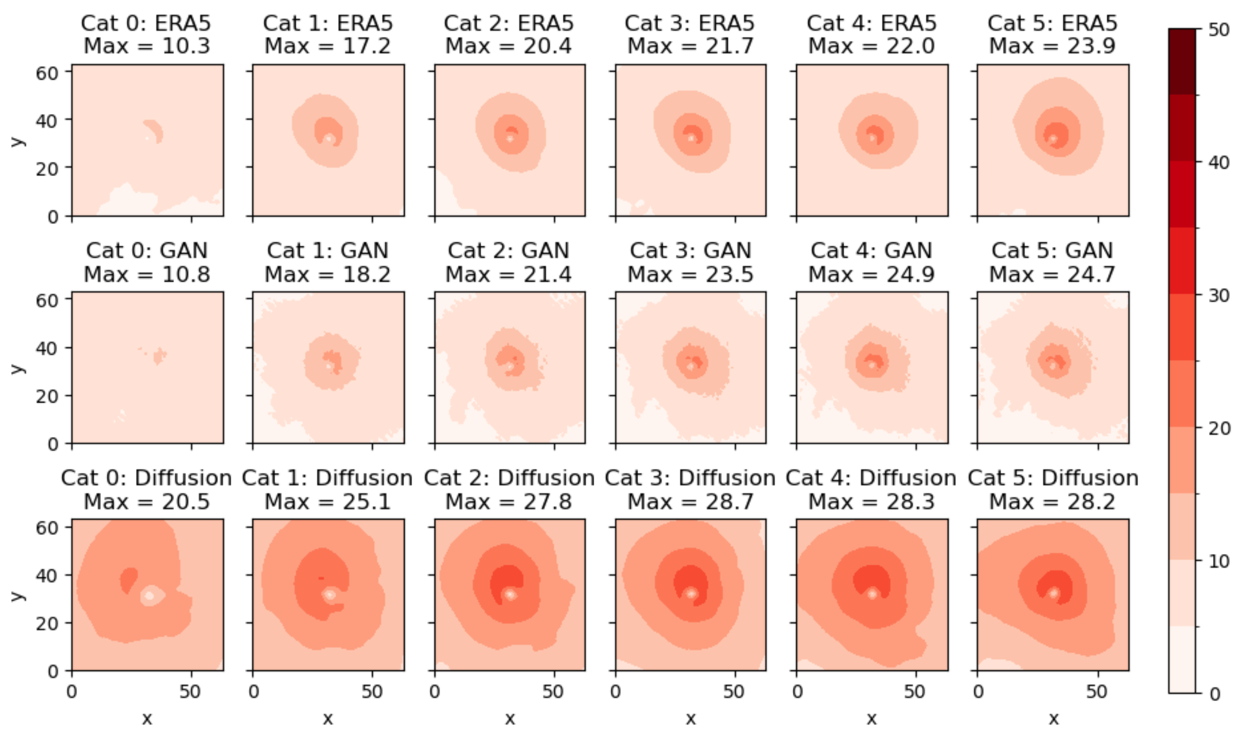

Figure 1 Near-surface wind data from the ERA5 (top), the GAN (middle), and the diffusion model (bottom). Each subplot is the mean of 100 samples. The columns are storms with the Saffir-Simpson intensity ranging from 0 to 5. The intensity label of the TC-PRIMED (ERA5) is based on the IBTrACS. The intensity label of the generative model outputs is user-specified. The extreme value on top of subplots is calculated based on the sample means. The unit of the wind speed is m s^{-1}.

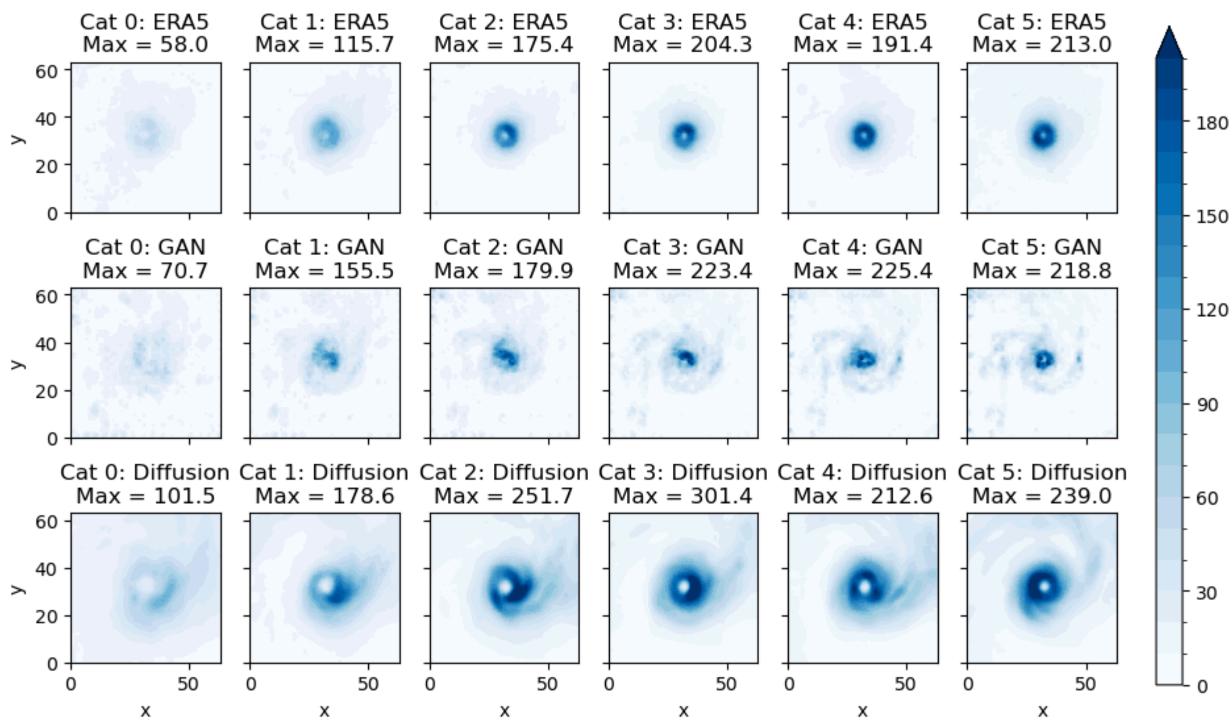

Figure 2 Same as Figure 1 but for precipitation (mm day^{-1}).

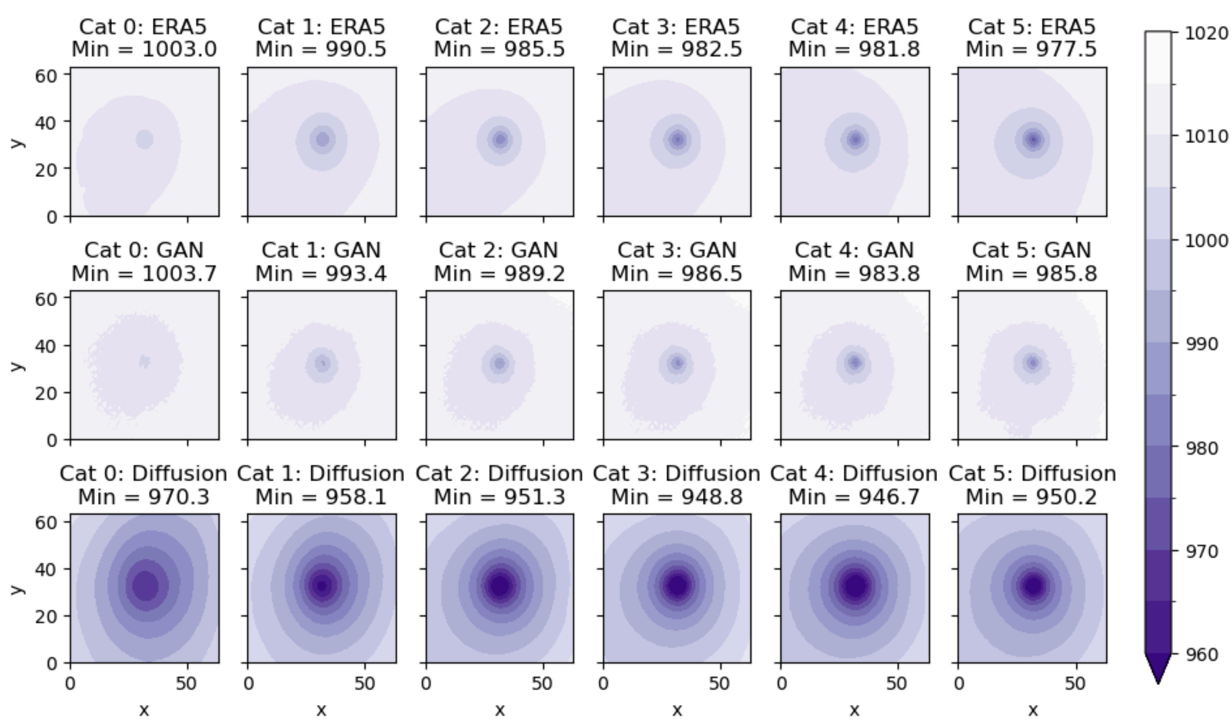

Figure 3 Same as Figure 1 but for sea level pressure (hPa).

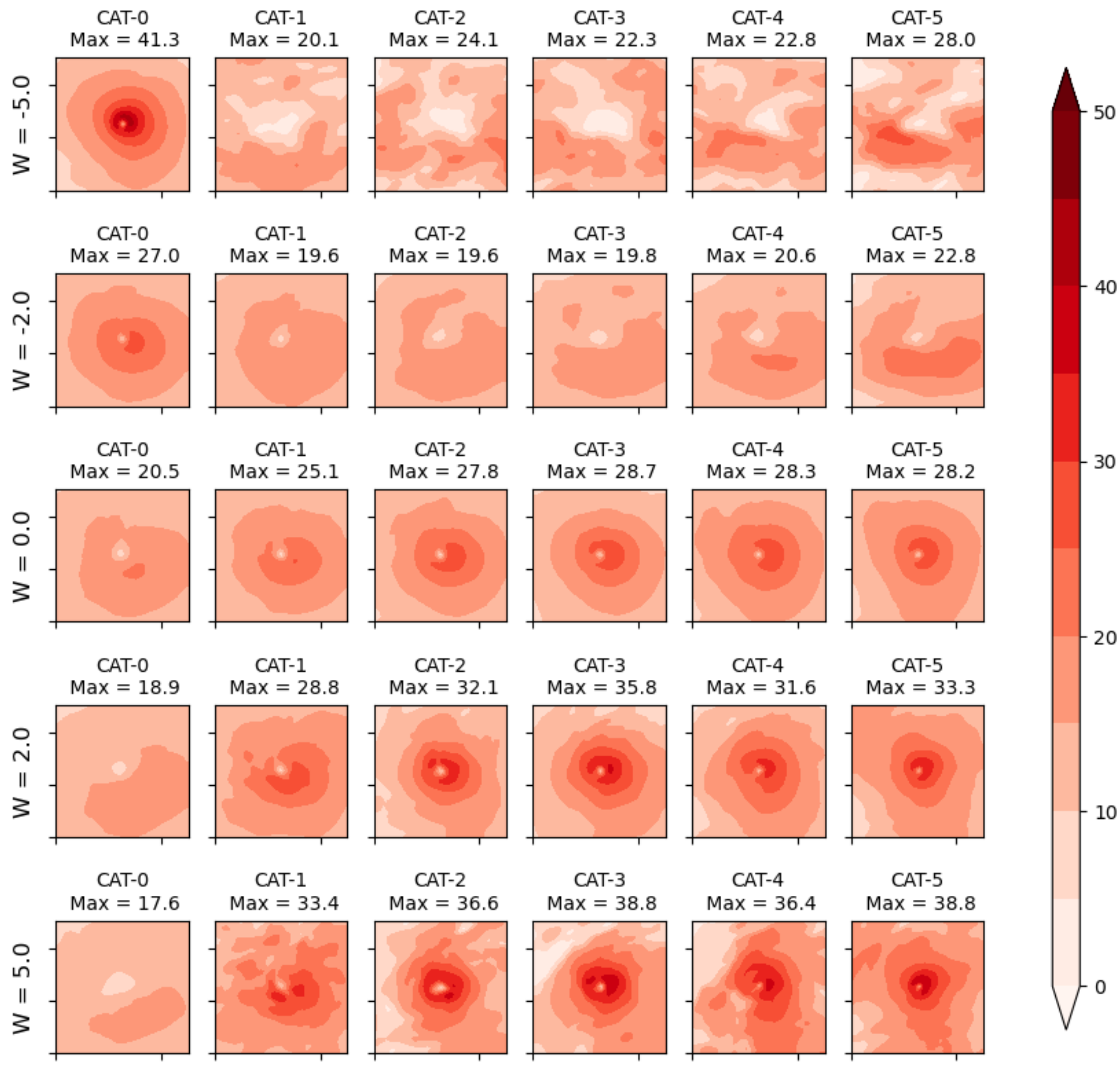

Figure 4 Sensitivity of diffusion model to the choice of the guidance weight (W). The columns show the results for each category of hurricane intensity. The rows show the results corresponding to different W values. Higher W values mean stronger guidance, and lower W values mean weaker guidance. The other settings are the same as in Figure 1.

3.2 AI-driven Climate Prediction

Here we seek to extend the capability of the NeuralGCM to experimental seasonal prediction. Our strategy resembles the development of physical climate models and starts with an experiment with prescribed sea surface temperature (SST) forcing. The slowly-varying SST forcing includes key information such as the El Niño/Southern Oscillation (ENSO) and is crucial for the seasonal climate prediction of the climate statistics [24], such as the activity of hurricanes and other extremes (e.g., atmospheric rivers). Because of the thermal inertia of the ocean water, the simplest representation of the SST forcing can be generated via the persistence or autoregressive models of anomalies. The SST forcing can also be generated via more sophisticated statistical or physical models. The results presented here are generated with the persistence of anomalies and serve as a demonstration of feasibility.

We initialize the NeuralGCM on July 1 of the years between 1990 and 2023. The initial conditions are acquired from the ERA5, which is quasi-realtime with a lag of several days for public users. We generate the SST and sea ice forcings using the assumption of persistence anomalies. Therefore, all the information needed for operational predictions would be available near the start date, unlike the original NeuralGCM experiments that rely on knowledge of future SST and sea ice forcings. We then run twenty parallel simulations (i.e., ensemble size) using the NeuralGCM with slightly different model parameters. To better sample the probability space, more parallel simulations can be introduced by perturbing the model parameters or using slightly different initial conditions. The size of the 20-ensemble simulation is larger than most operational physical models and can be extended affordably. With our initial implementation, a 5-month simulation with a 1.4-degree spatial grid can finish in approximately 14 minutes on a single L40S GPU. The execution would be ~50% faster on Google's cloud infrastructure.

Figure 5 shows the skill of the simple configuration in forecasting the atmosphere state. The target variable is the geopotential height at 500 hPa, which is approximately 5-6 km above sea level. This variable describes the atmosphere circulation state (e.g., the North Atlantic Oscillation) and is frequently used to evaluate the quality of weather forecasts. We evaluate the consistency of the July to September means between the NeuralGCM prediction and the ERA5 reference, which corresponds to the peak of the Northern Hemisphere hurricane activity. This result and analyses of other variables (not shown) suggest the simple configuration can skillfully forecast the atmosphere state months ahead. Such skills are essential for the seasonal prediction of TC activity.

Despite the relatively coarse resolution of the NeuralGCM configuration, the seasonal simulation produced vortices that resemble hurricanes (Figure 6). In the forecast snapshot of August 28, 2020, a simulation generated two vortices with tropical storm winds (>17 m s^{-1}) in the northern Atlantic. The first one is east of New York state of the US, and the second one is weaker and travels in the tropical Atlantic. A closer examination of these vortices suggests that they follow typical trajectories of Atlantic hurricanes. The first vortex resembles a recurving hurricane that moves along the East Coast, and the second vortex resembles a westward-moving hurricane that develops from the African coast and moves toward the West Indies. Such vortices are common in the other basins and simulations initialized in other years. The skillful simulation of a large-scale environment (Figure 5) and hurricane-like vortices suggest that our configuration simulates relatively realistic hurricanes[3].

The experiments and analyses also revealed limitations of the experimental system. One issue is the model instability that results in unrealistic data fields (not shown). The instability in these failed simulations is similar to that documented by the original NeuralGCM paper. It often originates from the tropics and appears to arise from gravity waves associated with convection. The instability happens more frequently in some predictions than in others. The instability issues are also accompanied by mean state drifts that are common in physical model simulations. The current study discards simulations with those problems, and fixes are being tested by the Google team. Another notable issue is that the public version of the NeuralGCM does not output surface variables[4] such as wind speed. The lack of surface wind speed makes it hard to directly compare the model output with the best track observations.

Despite those challenges, we managed to apply an objective tracking algorithm [25] to identify hurricanes and showed that our configuration can make skillful seasonal predictions of Atlantic hurricane activity. Figure 7 shows the seasonal hurricane counts in the observation and the model

[3] Later analyses in December 2024 suggest model deficiencies in simulating some aspects of hurricane structure. This issue will be discussed by an upcoming publication.
[4] An update in late 2024 allows users to access the surface pressure variable and partly alleviates the issue.

simulations. The correlation between the observation and the ensemble mean is 0.67, with high statistical significance ($p<0.01$). High skills are also found for intensity metrics (e.g., accumulated cyclone energy) and in the Northeastern Pacific basin (not shown). The author notes that such performance is similar to top-performing physical climate models[26] and might be near the limit of potential prediction skills[27]. More details about the model performance and limitations will be documented in an upcoming peer-reviewed publication.

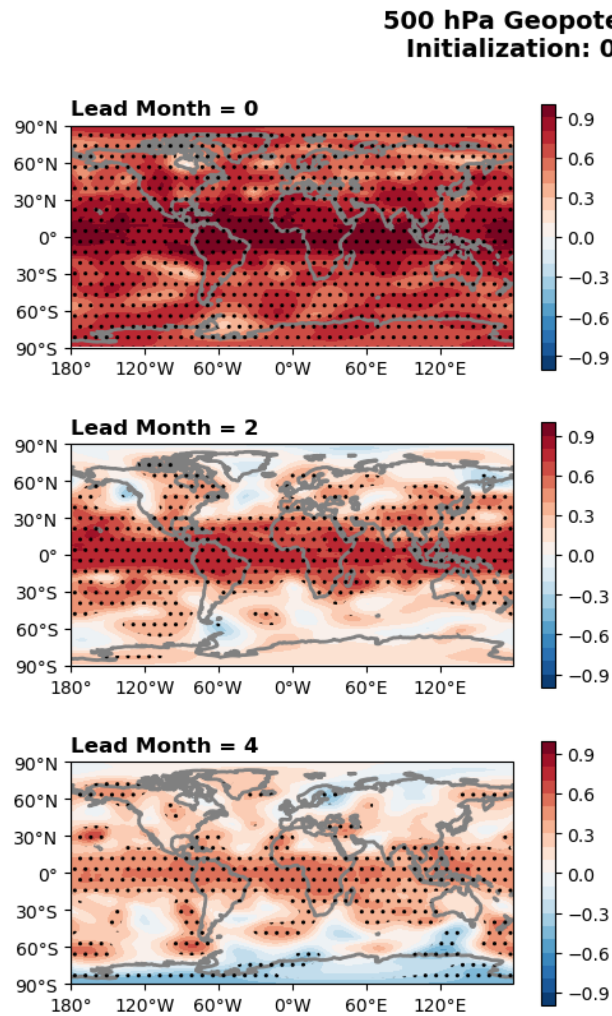

Figure 5 Skills of the NeuralGCM configuration in predicting the 500-hPa geopotential height. The predictions are initialized on July 1 of each year between 1990 and 2023. For each grid point, the predictions are validated against the July-September means of the ERA5 data. The skill metric is the anomaly correlation coefficient, with the 95% confidence level signals highlighted with hatching.

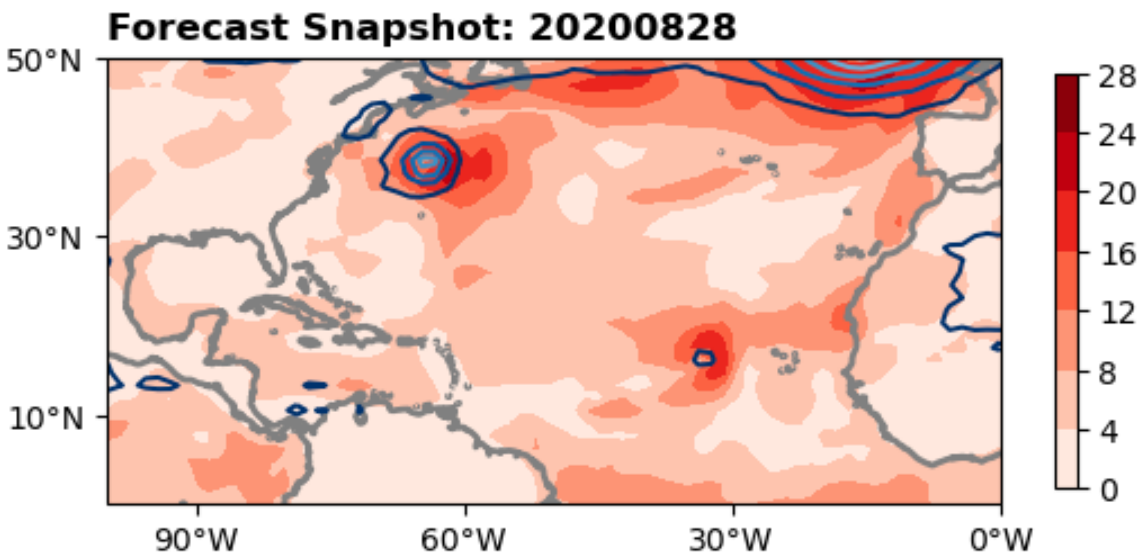

Figure 6 Snapshot of the atmospheric state in one prediction simulation. The color shading shows the near-surface wind speed (m s^{-1}). The contours show sea-level pressure. Hurricane features include strong wind and low values of sea-level pressure. Other parallel simulations yield active hurricanes during the hurricane season.

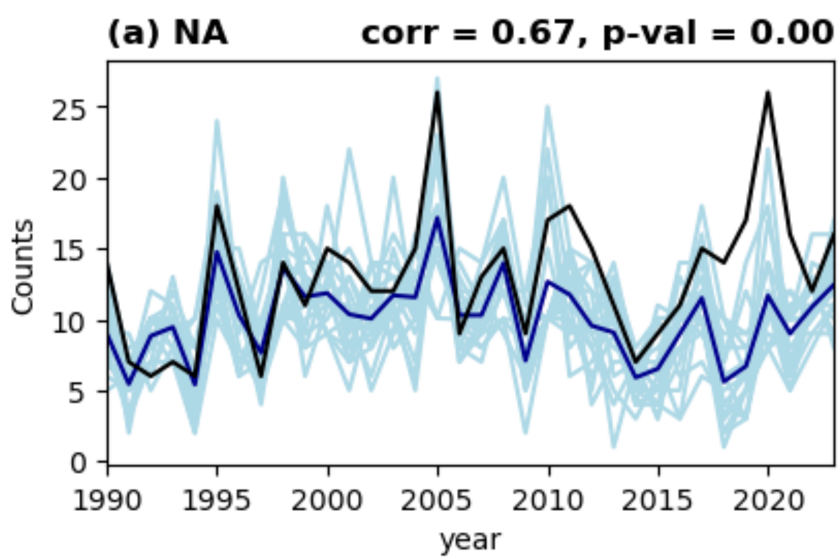

Figure 7 Seasonal Prediction Skill of the Experimental Configuration. The black line shows the annual North Atlantic hurricane counts in the observation. The blue lines show the same variable in the seasonal prediction experiments. The dark blue line shows the ensemble mean, and the light blue line shows the individual ensemble members.

4 Summary and Discussion

This report documents the progress of the PI in leveraging AI techniques to model hurricanes. The one-year project sets up the research infrastructure and investigates the generative modeling of hurricane hazards and AI-driven climate prediction. Despite challenges related to the new technologies, the findings about their applications are encouraging. Specifically, the baseline implementation of generative models can reasonably emulate hazards similar to those in the training dataset. The simple implementation of the Neural GCM prediction can skillfully predict the large-scale environment during the Northern Hemisphere hurricane season and simulate hurricane-like extreme features. The results suggest that these AI technologies can complement—if not disrupt—the existing technologies in hurricane risk modeling soon.

Future development should conduct more thorough evaluations, particularly against existing technologies and user needs, to prioritize the most valuable development efforts. For example, the outputs of the GAN respect the user-specified intensity label and show good consistency with the training data, but the fine-scale features have noisy artifacts. While physical scientists may consider such fine-scale issues unacceptable, practitioners may find the baseline GAN model is adequate to complement the widely used parametric models. Another example is the evaluation of the NeuralGCM prediction. A thorough evaluation with more initialization months and more target variables is necessary to clarify the strengths and weaknesses of the NeuralGCM relative to physical models. Depending on evaluation and user needs, future development may prioritize analyzing other weather perils (e.g., flood risks), improving prediction skills (e.g., refining the SST forcing), or expanding model capabilities (e.g., increasing spatial resolutions).

Besides the follow-up development suggested earlier, some bottlenecks need to be addressed to fully realize the potential of these AI technologies. With additional tuning and optimization, the GAN and the diffusion models likely can better emulate the hurricane features in training datasets. However, the training dataset ERA5, which is used by most AI models, has notable biases in representing hurricane-related extreme conditions. Without improving or replacing the training dataset, the performance of AI models trained with the ERA5 will be universally bottlenecked[5].

[5] Nvidia released a diffusion model in June 2024. The model has a more advanced architecture and was trained using better observational data (https://blogs.nvidia.com/blog/weather-forecast-corrdiff/).

The data interface is also an area that needs more resource investments. The development of the NeuralGCM configuration encountered various technical obstacles. While the expertise to address those issues exists in academic institutions and technology companies, such expertise remains relatively scarce among the end users in the finance sector. Addressing this issue can help ambitious players to make the best use of the rapidly evolving technologies.

We also recognize that alternative approaches to leverage AI technologies exist. For example, the risk modeling of TC activity or the seasonal prediction tasks can be addressed with task-specific solutions [25, 26]. In comparison, our efforts with the NeuralGCM envisioned a general-purpose solution that can potentially model all hazards and make hours-to-year predictions in a single, unified framework. The compelling benefits of such an approach include the simplicity of future infrastructure maintenance and the capability of natively addressing multi-hazard compounding risk. We believe such features, together with the high accuracy and fast speed of AI-driven simulations, can provide a unique business edge to risk-sensitive users in the finance sector.

The first numeric weather prediction was implemented in the 1950s and took over half a century to perfect. The current wave of AI weather model innovation started around 2022 [28, 29] and has achieved remarkable skills over two years. Given the results of this project and other peer studies, the author believes that AI models will be ready to play important roles in risk modeling within this decade and help stakeholders address weather-climate challenges.

5 Acknowledgment

G.Z. thanks Ethan Cai for assistance with the GAN modeling and Megha Rao for assistance with the storm tracking. Helpful discussions with Randy Chase, Renzhi Jing, and Stephen Hoyer helped inspire this study. G.Z. is grateful for the financial support of the ORMIR faculty fellowship of the Gies Business School.

6 References

1. Smith AB (2020) U.S. Billion-dollar Weather and Climate Disasters, 1980 - present (NCEI Accession 0209268). https://doi.org/10.25921/STKW-7W73

2. Lörinc M, Hotový O, Hotový O, Čejka T (2024) AON Climate and Catastrophe Insight. https://assets.aon.com/-/media/files/aon/reports/2024/climate-and-catastrophe-insights-report.pdf

3. Wing OEJ, Lehman W, Bates PD, Sampson CC, Quinn N, Smith AM, Neal JC, Porter JR, Kousky C (2022) Inequitable patterns of US flood risk in the Anthropocene. *Nature Climate Change*, 12(2):156–162. https://doi.org/10.1038/s41558-021-01265-6

4. Evans S (2024) Cat bond & related ILS market hits record size at over $47bn, up 5% so far this year - Artemis.bm. *Artemis.bm - The Catastrophe Bond, Insurance Linked Securities & Investment, Reinsurance Capital, Alternative Risk Transfer and Weather Risk Management site*, https://www.artemis.bm/news/cat-bond-related-ils-market-hits-record-size-at-over-47bn-up-5-so-far-this-year/

5. Naik G, Lee STT (2024) Risk Models Behind World's Best Hedge Fund Strategy Are Getting a Lot Harder to Crack. *Bloomberg.com*, https://www.bloomberg.com/news/articles/2024-02-25/catastrophe-bonds-behind-record-hedge-fund-returns-face-new-era-of-risk

6. Roston E, Karra K, Kaufman L, Rangarajan S (2024) The Risky Business of Predicting Where Climate Disaster Will Hit. *Bloomberg.com*, https://www.bloomberg.com/graphics/2024-flood-fire-climate-risk-analytics/

7. Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049. https://doi.org/10.1002/qj.3803

8. Bi K, Xie L, Zhang H, Chen X, Gu X, Tian Q (2023) Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970):533–538. https://doi.org/10.1038/s41586-023-06185-3

9. Lam R, Sanchez-Gonzalez A, Willson M, Wirnsberger P, Fortunato M, Alet F, Ravuri S, Ewalds T, Eaton-Rosen Z, Hu W, Merose A, Hoyer S, Holland G, Vinyals O, Stott J, Pritzel A, Mohamed S, Battaglia P (2023) Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421. https://doi.org/10.1126/science.adi2336

10. Price I, Sanchez-Gonzalez A, Alet F, Ewalds T, El-Kadi A, Stott J, Mohamed S, Battaglia P, Lam R, Willson M (2023) GenCast: Diffusion-based ensemble forecasting for medium-range weather. https://doi.org/10.48550/ARXIV.2312.15796

11. Kochkov D, Yuval J, Langmore I, Norgaard P, Smith J, Mooers G, Klöwer M, Lottes J, Rasp S, Düben P, Hatfield S, Battaglia P, Sanchez-Gonzalez A, Willson M, Brenner MP, Hoyer S (2024) Neural general circulation models for weather and climate. *Nature*, https://doi.org/10.1038/s41586-024-07744-y

12. Bonev B, Kurth T, Hundt C, Pathak J, Baust M, Kashinath K, Anandkumar A (2023) Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere. https://doi.org/10.48550/ARXIV.2306.03838

13. Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ (2010) The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data. *Bulletin of the American Meteorological Society*, 91(3):363–376. https://doi.org/10.1175/2009BAMS2755.1

14. Dulac W, Cattiaux J, Chauvin F, Bourdin S, Fromang S (2024) Assessing the representation of tropical cyclones in ERA5 with the CNRM tracker. *Climate Dynamics*, 62(1):223–238. https://doi.org/10.1007/s00382-023-06902-8

15. Naufal Razin M, Slocum CJ, Knaff JA, Brown PJ, Bell MM (2022) Tropical Cyclone Precipitation, Infrared, Microwave, and Environmental Dataset (TC PRIMED). *Bulletin of the American Meteorological Society*, https://doi.org/10.1175/BAMS-D-21-0052.1

16. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative Adversarial Networks. https://doi.org/10.48550/ARXIV.1406.2661

17. Mirza M, Osindero S (2014) Conditional Generative Adversarial Nets. https://doi.org/10.48550/ARXIV.1411.1784

18. Ho J, Jain A, Abbeel P (2020) Denoising Diffusion Probabilistic Models. https://doi.org/10.48550/ARXIV.2006.11239

19. Klotzbach PJ, Bell MM, Bowen SG, Gibney EJ, Knapp KR, Schreck CJ (2020) Surface Pressure a More Skillful Predictor of Normalized Hurricane Damage than Maximum Sustained Wind. *Bulletin of the American Meteorological Society*, 101(6):E830–E846. https://doi.org/10.1175/BAMS-D-19-0062.1

20. Chen J-H, Lin S-J (2013) Seasonal Predictions of Tropical Cyclones Using a 25-km-Resolution General Circulation Model. *Journal of Climate*, 26(2):380–398. https://doi.org/10.1175/JCLI-D-12-00061.1

21. Zhang G, Murakami H, Knutson TR, Mizuta R, Yoshida K (2020) Tropical cyclone motion in a changing climate. *Science Advances*, 6(17):eaaz7610. https://doi.org/10.1126/sciadv.aaz7610

22. Zhang G (2023) Warming-induced contraction of tropical convection delays and reduces tropical cyclone formation. *Nature Communications*, 14(1):6274. https://doi.org/10.1038/s41467-023-41911-5

23. Pryor SC, Barthelmie RJ (2021) A global assessment of extreme wind speeds for wind energy applications. *Nature Energy*, 6(3):268–276. https://doi.org/10.1038/s41560-020-00773-7

24. Shukla J (1998) Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting. *Science*, 282(5389):728–731. https://doi.org/10.1126/science.282.5389.728

25. Ullrich PA, Zarzycki CM, McClenny EE, Pinheiro MC, Stansfield AM, Reed KA (2021) TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in

large datasets. *Geoscientific Model Development*, 14(8):5023–5048. https://doi.org/10.5194/gmd-14-5023-2021

26. Zhang G, Murakami H, Gudgel R, Yang X (2019) Dynamical Seasonal Prediction of Tropical Cyclone Activity: Robust Assessment of Prediction Skill and Predictability. *Geophysical Research Letters*, 46(10):5506–5515. https://doi.org/10.1029/2019GL082529

27. Zhang G, Murakami H, Yang X, Findell KL, Wittenberg AT, Jia L (2021) Dynamical Seasonal Predictions of Tropical Cyclone Activity: Roles of Sea Surface Temperature Errors and Atmosphere–Land Initialization. *Journal of Climate*, 34(5):1743–1766. https://doi.org/10.1175/JCLI-D-20-0215.1

28. Keisler R (2022) Forecasting Global Weather with Graph Neural Networks. https://doi.org/10.48550/ARXIV.2202.07575

29. Pathak J, Subramanian S, Harrington P, Raja S, Chattopadhyay A, Mardani M, Kurth T, Hall D, Li Z, Azizzadenesheli K, Hassanzadeh P, Kashinath K, Anandkumar A (2022) FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators. https://doi.org/10.48550/ARXIV.2202.11214