BAMS
In Box

Closing the Gap—Hurricane Prediction Advances in the U.S. FV3-Based Models

Jan-Huey Chen,[a] Timothy Marchok,[a] Morris Bender,[a] Kun Gao,[b]
Sundararaman Gopalakrishnan,[c] Lucas Harris,[a] Andrew Hazelton,[d] Bin Liu,[e]
Avichal Mehra,[f] Matthew Morin,[a,g] Fanglin Yang,[f] Xuejin Zhang,[c]
Zhan Zhang,[f] and Linjiong Zhou[b]

KEYWORDS:
Hurricanes/typhoons;
Numerical weather prediction/forecasting;
Operational forecasting;
Dynamical system model;
Model evaluation/performance;
Numerical analysis/modeling

ABSTRACT: The Integrated Forecasting System (IFS) developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) has been regarded as the best guidance for hurricane track forecasts for years. However, the performance of U.S. models on hurricane forecasts has been catching up. Since 2019, various Finite-Volume Cubed-Sphere Dynamical Core (FV3)-based models, including the National Centers for Environmental Prediction (NCEP) operational Global Forecast System (GFS), newly operational Hurricane Analysis and Forecast System (HAFS), and research-oriented Geophysical Fluid Dynamics Laboratory (GFDL) System for High-resolution prediction on Earth-to-Local Domains (SHiELD), have consistently demonstrated improved hurricane forecasts in the North Atlantic basin, relative to the previous generation of National Oceanic and Atmospheric Administration (NOAA) operational and research models. This article presents the progress that has been made and identifies areas for improvement for U.S. model development on hurricane forecasts.

SIGNIFICANCE STATEMENT: Hurricane predictions in the North Atlantic basin performed by the numerical models developed in the United States have been improved in the past 5 years along with the major upgrade of the models' dynamical core. The gap in hurricane forecast skills between the U.S. models and the world-leading European model has been substantially decreasing. This article presents the 10-yr progression of the state-of-the-art operational and research-oriented U.S. models compared to the European models for North Atlantic tropical cyclone forecasts to demonstrate this achievement.

DOI: 10.1175/BAMS-D-24-0036.1

Corresponding author: Jan-Huey Chen, jan-huey.chen@noaa.gov

Manuscript received 1 February 2024, in final form 21 April 2025, accepted 14 May 2025

AFFILIATIONS: [a] NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey; [b] Cooperative Institute for Modeling the Earth System, Princeton University, Princeton, New Jersey; [c] NOAA/Atlantic Oceanographic and Meteorological Laboratory/Hurricane Research Division, Miami, Florida; [d] Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida; [e] Lynker at NOAA/NWS/NCEP/EMC, College Park, Maryland; [f] NOAA/Environmental Modeling Center, College Park, Maryland; [g] University Corporation for Atmospheric Research, Boulder, Colorado

1. Introduction

Tropical cyclones (TCs), also known as hurricanes and tropical storms in the North Atlantic basin, are one of the most destructive natural events on Earth. Improving the skill of TC forecasts, especially those generated in the North Atlantic basin, has always been an important objective for government weather forecast agencies, emergency managers, and the atmospheric science research community in the United States. The Integrated Forecasting System (IFS) developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) has been recognized as providing the most skillful deterministic model guidance for TC track forecasts for years (NOAA/NHC 2024), which sets an inspirational benchmark for model developers in the United States.

In 2016, the Finite-Volume Cubed-Sphere Dynamical Core (FV3; Harris et al. 2021; Lin 2004; Putman and Lin 2007) developed by National Oceanic and Atmospheric Administration (NOAA)/Geophysical Fluid Dynamics Laboratory (GFDL) was selected for the Next Generation Global Prediction System [NGGPS; NOAA/Office of Science and Technology Integration (OSTI) (NOAA/OSTI 2024)] for both global and regional configurations of the Unified Forecast System (UFS). After 3 years of development, in June 2019, the National Centers for Environmental Prediction (NCEP) launched the Global Forecast System (GFS), version 15, with the major upgrade being the replacement of the legacy spectral dynamical core with the FV3. The operational GFS was further upgraded to version 16 in March 2021, with numerous changes and upgrades made in model physics, infrastructure, and data assimilation, including raising the model top from the stratopause to the mesopause and the addition of wave model coupling (Yang 2020).

Meanwhile, the development of the System for High-Resolution Prediction on Earth-to-Local Domains (SHiELD) at GFDL continued the legacy of FV3-based global model development that was begun during the NGGPS selection period. SHiELD is a research-oriented UFS prototype atmosphere model with many subconfigurations to test and integrate cutting-edge dynamical and physical algorithms for a wide array of applications in atmospheric research (Harris et al. 2020). Both the flagship global SHiELD and the Atlantic basin–nested T-SHiELD have shown skillful TC forecasts which have demonstrated the capabilities of FV3-based modeling systems (Chen et al. 2019a,b; Gao et al. 2021; Hazelton et al. 2022; Chen et al. 2023; Gao et al. 2023).

Besides the development of global models, there is a long history of using specialized hurricane forecasting modeling systems to predict TC track and intensity in the North Atlantic basin in the United States. Building on the successful GFDL operational hurricane model (Bender et al. 2019), and through the support of the Hurricane Forecast Improvement Project (HFIP; Gall et al. 2013; Gopalakrishnan et al. 2021), the Hurricane Weather Research and Forecasting Model (HWRF) had served as an operational TC forecast model since 2007. HWRF is considered one of the most skillful ocean-coupled, regional numerical weather prediction models for hurricane intensity and structure forecasts (Gopalakrishnan et al. 2011, 2012, 2013; Alaka et al. 2024). Since 2019, the HFIP has pivoted to develop an FV3-based new-generation hurricane forecasting system, the Hurricane Analysis and Forecast System (HAFS; Dong et al. 2020; Hazelton et al. 2021, 2022). The two developmental configurations of HAFS (HAFS-A and HAFS-B) became operational in June 2023.

This article illustrates and summarizes the improvement of North Atlantic hurricane forecasts achieved by the abovementioned new-generation U.S. models. Forecast skill of TC track and intensity in the FV3-based GFS, SHiELD, T-SHiELD, HAFS-A, and HAFS-B models is compared to the performance of previous NOAA operational models and to the ECMWF IFS during the past decade.

2. Models and TC track datasets

The NCEP operational GFS was upgraded from version 14, which consisted of a spectral dynamic core, to version 15 with the FV3 dynamic core in June 2019. We therefore focus on the North Atlantic hurricane seasons before and after this important transition to investigate the U.S. model performance for TC track and intensity forecasts from 2014 to 2023. TC forecasts from the operational GFS, HWRF, HAFS-A, and HAFS-B, and from research-oriented SHiELD and T-SHiELD, are compared with the performance of the operational ECMWF IFS in these 10 hurricane seasons in the North Atlantic basin.

From 2014 to 2023, the operational IFS has been upgraded through nine cycles from CY40R1 to CY47R3 (ECMWF 2024). For the last five hurricane seasons since 2019, the IFS forecasts were from the versions of CY46r1, CY47r1, CY47r2, CY47r3, and CY48r1 implemented in June 2019, June 2020, May 2021, October 2021, and June 2023, respectively (ECMWF 2023). In this article, the IFS TC forecasts are used as the benchmark to verify the performance of the new U.S. models.

For the operational GFS, the legacy spectral dynamical core (version 14 and older) was used in 2014–18. The FV3-based GFS versions 15 and version 16 were launched on 12 June 2019 and 22 March 2021, respectively. Detailed GFS upgrade information can be found on its official website [NOAA/Environmental Modeling Center (EMC) 2024].

HWRF served as a NOAA operational TC forecast model beginning in 2007. The detailed history of HWRF upgrades can be found in Alaka et al. (2024). NOAA has developed and transitioned to the next-generation FV3-based HAFS, which became operational in 2023. HAFS is the first ocean–wave coupled UFS implementation with multiple moving nest capability and a sophisticated vortex initialization and data assimilation package. The detailed model description of HAFS and its two operational configurations (HAFS-A and HAFS-B) can be found in Dong et al. (2020), in Hazelton et al. (2021, 2022), and in Tables 2 and 3 in Zhang et al. (2023).

The GFDL 13-km global SHiELD near-real-time forecast system has been advanced through four model versions, v2019, v2020, v2021, and v2022 since 2019. The detailed configuration of the SHiELD v2019 can be found in Harris et al. (2020). Based on v2019, v2020 included the improved version 2 of the GFDL microphysics scheme and the turbulent kinetic energy (TKE)-based hybrid eddy-diffusivity mass-flux (EDMF) planetary boundary layer (PBL) scheme (TKE-EDMF; Han and Bretherton 2019). Compared to v2019, the lower-troposphere cold bias was significantly reduced while the 500-hPa geopotential height forecast skill was also improved in v2020. The initial conditions used in the SHiELD v2019 and v2020 were from the analyses generated by the GFS version 15. The SHiELD v2021 kept the same model configuration as v2020 but used the GFS version 16 analyses as the model's initial condition. The most recent version, v2022, further updated the GFDL microphysics scheme from version 2 (Harris et al. 2020) to version 3 (Zhou et al. 2022), adopted a revised mixing length formula and cap for the mixing length in the TKE-EDMF, and updated land surface data. These updates significantly reduced the cloud fraction errors and biases, improved the liquid water forecasts, and generally enhanced the large-scale medium-range weather prediction.

The GFDL global-nested T-SHiELD configuration has been run near–real time in the past four consecutive hurricane seasons since 2020. Version v2021, an upgrade of v2020 (Gao et al. 2021), replaced the Yonsei University PBL scheme (Hong 2010) with the TKE-EDMF scheme and adjusted the vertical levels and shallow convection scheme in the nested domain. These changes eliminated the mean eastward track bias and significantly reduced the mean track error in TC track forecasts compared to v2020. Building upon v2021, the latest version, v2022, employs a revised shear-dependent mixing length formula and an increased mixing length cap within the TKE-EDMF scheme. The mean intensity error beyond 72-h lead time was significantly reduced in this version.

The SHiELD and T-SHiELD model versions employed for near-real-time TC forecasts typically align with the corresponding calendar year. For example, SHiELD v2019 was used for the TC forecasts in the 2019 hurricane season. The only exception is that both SHiELD v2022 and T-SHiELD v2022 were used for the hurricane seasons of 2022 and 2023.

The observed TC tracks used to compute the errors of model forecasted TC location and intensity are from the best track data (b-deck) in the Automated Tropical Cyclone Forecasting System (ATCF) dataset (Miller et al. 1990; Sampson and Schrader 2000). The official TC track forecasts in the operational GFS, IFS, HWRF, and HAFS are retrieved from the ATCF a-deck files. All TC track and intensity forecasts from the dynamical models evaluated in this study are diagnosed using the GFDL vortex tracker (Marchok 2021). The TC track and intensity forecasts from the climatology and persistence models, CLIPER5 and SHIFOR5, respectively, retrieved from the ATCF a-deck files, are used as the no-skill benchmarks for model forecast skill to account for interannual variation of TC behaviors (Neumann 1972; Jarvinen and Neumann 1979; Aberson 1998; Knaff et al. 2003). All TC track and intensity forecast error comparisons in this article are homogeneous, i.e., a forecast case for a given lead time is only included in the statistics if all models have a valid forecast position and intensity for the TC at that lead time for this case.

3. 10-yr evolution of TC forecast skill

The errors of the North Atlantic TC track forecasts in GFS and IFS are homogeneously compared before and after the GFS major upgrade in 2019 (Figs. 1a,b). Compared to the seasons of 2014 to 2018 (lower panel in Fig. 1a), the gap between the GFS and IFS TC track forecast errors after the 72-h lead time largely decreased during the 2019–23 seasons (lower panel in Fig. 1b). Although the differences between the two models in the latter five seasons are larger at early lead times, the general improvement of GFS is very encouraging. Moreover, the research-oriented SHiELD performed even better than GFS consistently during the 7-day forecast from 2019 to 2023, which demonstrates the potential of future NOAA operational global models.

The improvement of TC track forecasts in the U.S. regional models is even more significant. Using the same IFS forecasts as the benchmark, we first see that the gap between TC track forecast errors in HWRF and IFS largely decreased after 2019 (Figs. 1c,d), especially after the 48-h lead time, similar to the performance of GFS. Since GFS provides the boundary condition to the HWRF forecasts, the major GFS upgrade should also benefit the HWRF track forecasts. Very encouraging results are shown in the new-generation regional models, HAFS-A, HAFS-B, and T-SHiELD, that their TC track forecast errors are very close to the IFS. HAFS-A achieved the lowest track forecast error of the three, and its track errors are smaller than those in the IFS at 60-, 72-, and 96-h lead times. This result represents a considerable success of HFIP.

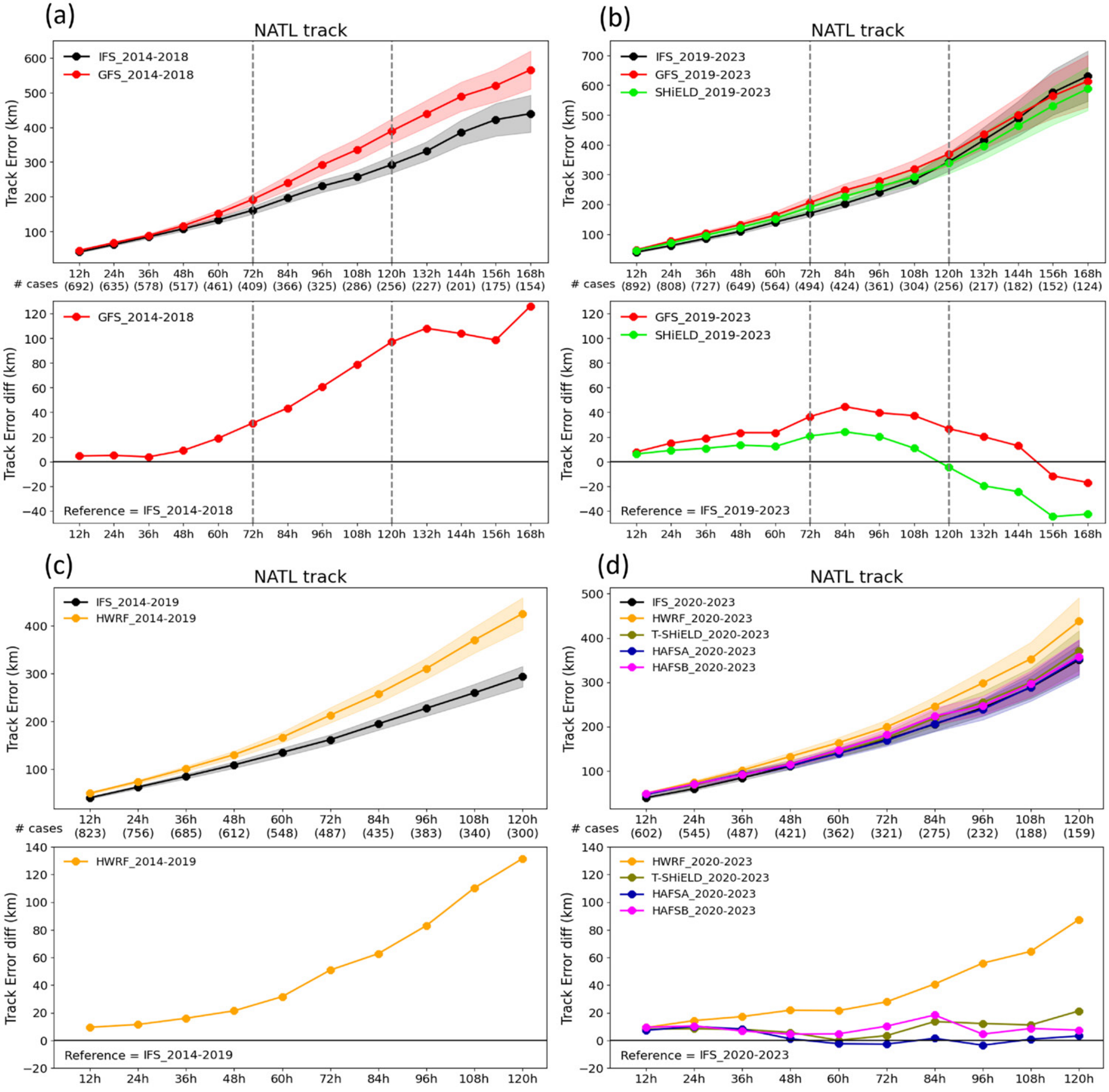

Fig. 1. **North Atlantic TC track errors. (a) Upper panel shows the mean TC track errors (km) at 12-h forecast lead time intervals for IFS (black) and GFS (red) during the years of 2014 to 2018. The 95% confidence levels for each model are indicated by the same transparent color shading. Numbers of homogeneous comparison cases for individual lead times are listed in the brackets at the bottom of the abscissa. The vertical gray dotted lines indicate 72- and 120-h forecast lead times. The error differences in GFS compared to IFS are shown in the lower panel. (b) As in (a), but for the years of 2019 to 2023, with the SHiELD forecast in green. (c) As in (a), but for IFS (black) and HWRF (yellow) during the years of 2014 to 2019. (d) As in (c), but for the years of 2020 to 2023, with the forecasts in T-SHiELD (dark green), HAFS-A (blue), and HAFS-B (magenta).**

The results for track forecast skill relative to the CLIPER5 benchmark forecast shown in Fig. 2 indicate that the 2023 Atlantic season was the most challenging for all the dynamical models during the past five seasons, while the models as a group achieved higher skill during the 2021 and 2022 seasons. The skill differences between GFS and IFS were relatively large in 2019 and 2023. However, during the 2020 and 2021 seasons, GFS showed slightly higher skill than IFS which is encouraging. Furthermore, SHiELD performed slightly better than or

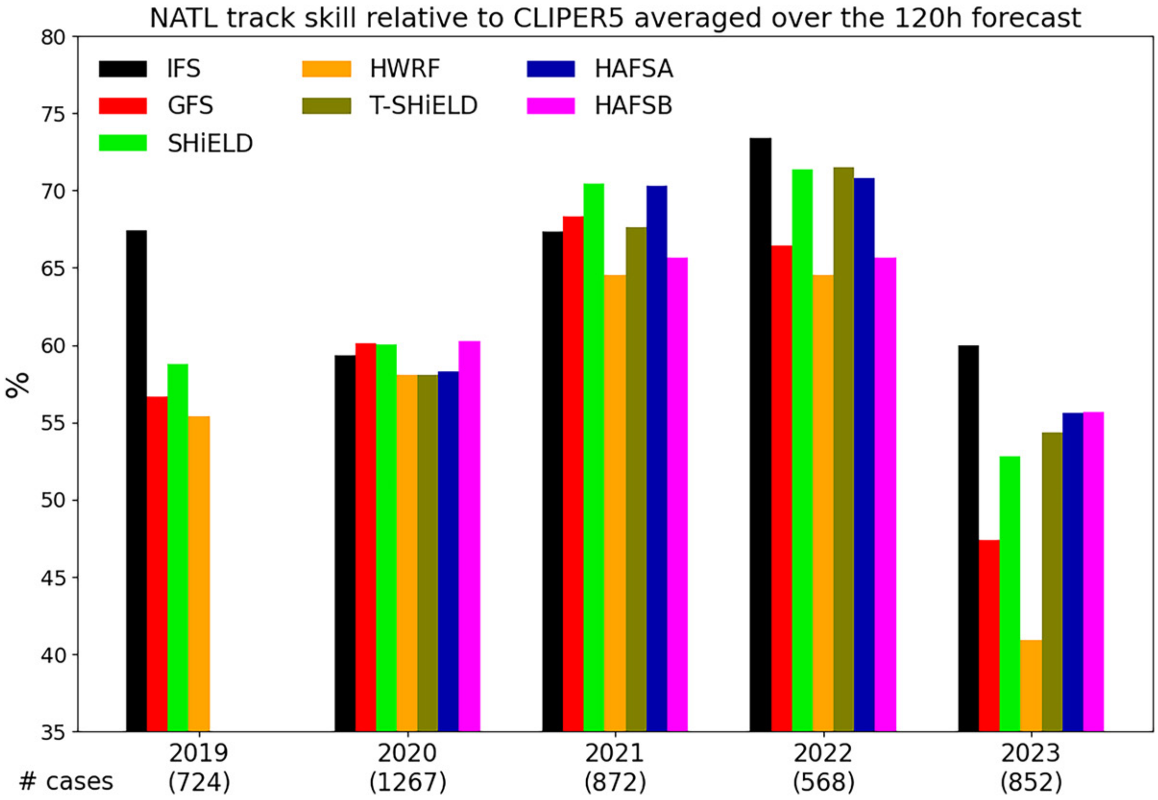

Fig. 2. **Mean TC track forecast skill relative to the CLIPER5 track forecast error, averaged over the 120-h forecast for the seven models during the 2019–23 Atlantic hurricane seasons. The mean track forecast errors are computed at 6-h intervals, ranging from 6 to 120 h, and then, the values shown are computed by averaging skill values over those lead times. The colors used for the models are the same as in Fig. 1. Total numbers of homogeneous cases are listed in the brackets at the bottom of the abscissa for each hurricane season.**

comparable to IFS during the 2020–22 seasons, and the difference between SHiELD and IFS was much smaller than that between GFS and IFS during 2023.

The new-generation regional hurricane models have performed very well since the 2020 season. HAFS-B showed higher skill than IFS in 2020; HAFS-A and T-SHiELD showed either better or comparable skills to IFS in 2021 and 2022. In 2023, compared to HWRF, both HAFS models and T-SHiELD showed much closer skills to the IFS, and their skills are better than the global GFS or SHiELD.

Figure 3 shows the intensity forecast skill based on the errors of maximum 10-m wind speed in each model relative to the SHIFOR5 forecasts averaged over the 120-h forecast period. Compared to the negative skill IFS, GFS and SHiELD showed positive skill in the years 2019, 2020, and 2022, which is a significant achievement for global model development. Also, it is very encouraging to see that the new HAFS models performed as well as the leading HWRF model for TC intensity forecasts from 2020 onward, which demonstrates a successful transition to NOAA's new-generation operational hurricane models.

4. Summary and prospects

Along with the major upgrades of the dynamical core, physics, and data assimilation in the operational GFS, U.S. model development for weather forecasts has been steadily advancing since 2019. The improved hurricane track and intensity forecasts described here can be taken as an indicator to demonstrate the progress made by the entire UFS community. The gap in hurricane track forecast skill between the GFS and IFS has been substantially reduced. The research-oriented SHiELD performed even better than the GFS for TC track forecasting in the North Atlantic basin, which demonstrates the potential of future NOAA operational

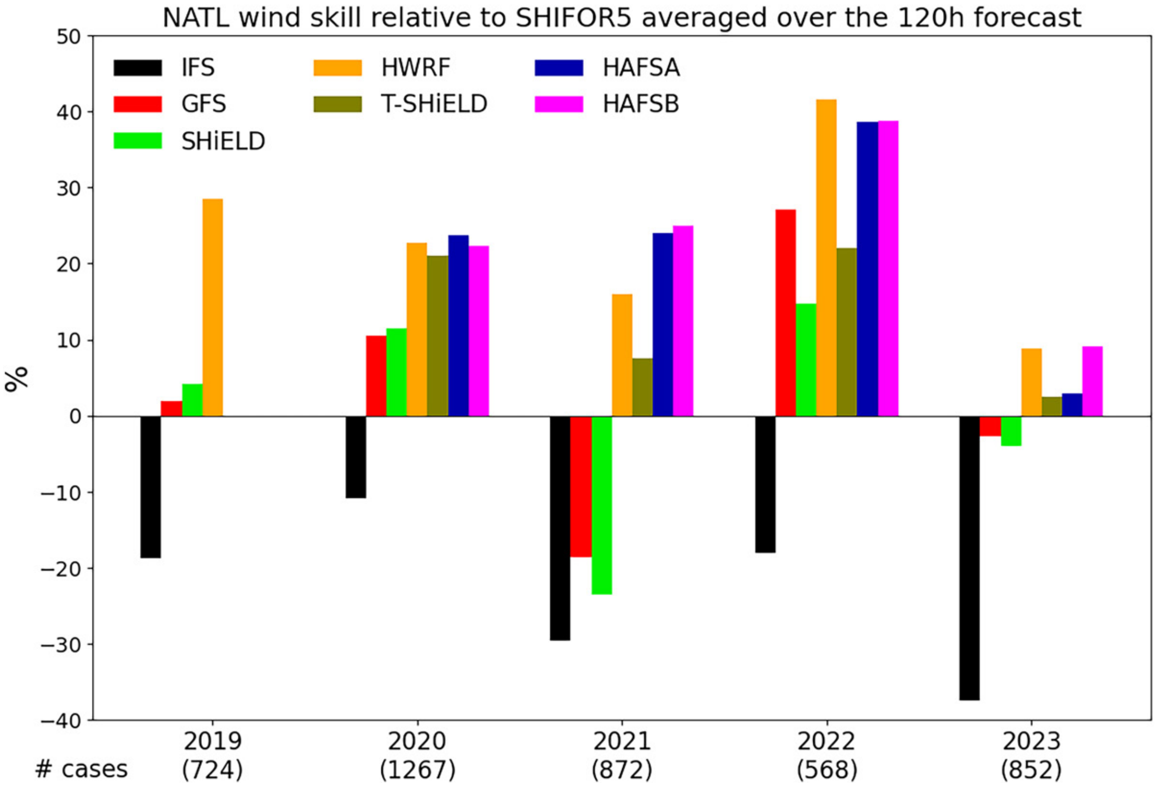

Fig. 3. As in Fig. 2, but for the mean TC intensity forecast skill relative to the SHIFOR5 10-m wind speed forecast error.

global models from the integration of current cutting-edge dynamical and physical algorithms. Moreover, different from IFS and the previous generation of U.S. global models, the FV3-based GFS and SHiELD showed much improved TC intensity forecast skill, marking a significant achievement of U.S. global model development. This article focuses on the model performance in forecasting North Atlantic TCs due to their large impact on the United States. It will be worthwhile to explore the improvement of U.S. models in TC prediction for other global basins in a future study.

The next-generation FV3-based regional hurricane models showed outstanding performance in both TC track and intensity forecasts. The newly operational HAFS-A and HAFS-B successfully reproduced the intensity forecast skill achieved by the HWRF which has been one of the world's leading regional hurricane intensity and structure forecast models for many years. Furthermore, both HAFS models and T-SHiELD showed comparable or even better TC track forecasts than global models from 2020 to 2023. The achievement of HAFS represents a considerable success for HFIP in upgrading the NOAA regional hurricane models.

The abovementioned progress cannot be attributed to a single change in the model development. All model components, including dynamical core, grid configuration, physics, and data assimilation, play important roles. Also, the setup and adjustments among the model components can also largely contribute to the model performance. The progress shown in this article is the bit-by-bit accumulated model development effort from the entire community. Accordingly, a sufficiently large sample of cases from multiple hurricane seasons is necessary when assessing the performance of the TC forecasts among models. In other words, a single case or a single hurricane season cannot fully represent the models' performance. In addition, natural yearly variation is an important factor that affects the models' performance. Figure 2 shows that even the world-leading IFS does not show consistent improvement on TC track forecasts over the years, even though it had a new model version for each hurricane season during that period. This indicates that a consistent year-by-year skill improvement cannot and should not be expected, but a long-term improvement from one 5-yr period to the next demonstrating the efforts devoted by the model development community is what we aimed to present in this article.

To further improve the U.S. models' hurricane prediction, many aspects are worth exploring and approaching. A holistic modeling system with well-integrated physics, dynamics, and data assimilation techniques has been a goal which still requires strong community engagement. Skillful dynamical models that can accurately depict these processes are also the key foundation for probabilistic forecasts using single- or multiple-model ensembles. From a scientific perspective, the year-to-year variation of TC track forecast skill, i.e., "more difficult" or "easier" years for TC prediction, is definitely an interesting research topic to explore. As we push the global model resolutions to 10 km or higher, e.g., current IFS (CY49r1), GFS (version 16), and SHiELD (v2024) are at 9, 13, and 6.5 km, respectively, it will be necessary to address the impact from the resolved grayscale or kilometer-scale phenomena to TC track or intensity forecasts. It is expected that the global nested modeling approach will better represent the connection between the large-scale circulation and individual storms.

Recently, with the enhancement of computer power and the appearance of new technologies, e.g., graphical processing unit (GPU) and machine learning techniques, it is time for us to consider adjusting the current model development strategy. In recent years, data-driven machine learning (ML) weather models have emerged, including Fourier ForeCasting Neural Network (FourCastNet; Pathak et al. 2022) in the United States and Artificial Intelligence/Integrated Forecasting System (AIFS; Lang et al. 2024) at ECMWF. Some of these ML models already demonstrate TC track forecast skills on par with both the IFS and GFS while having very low computational cost (DeMaria et al. 2024). While ML models are still in need of more extensive evaluation, and current-generation ML models struggle with intensity, mesoscale TC structure (Bonavita 2024), and out-of-sample events (Sun et al. 2024), these provide a promising complement to dynamical models. We believe that properly adopting the new technologies to best complement the physical models could maximize the benefits from both and possibly be the next great advance in hurricane prediction.

Acknowledgments. The authors would like to thank Dr. Shian-Jiann Lin who retired from GFDL in 2019 for his years of development of the FV3 dynamical core and numerous contributions to NOAA modeling systems. The authors would also thank Dr. Frank Marks for his leadership in the Hurricane Forecast Improvement Program and Dr. Vijay Tallapragada for his support and leadership of operational hurricane modeling efforts in NCEP. The authors would like to acknowledge support from the Hurricane supplemental and the NOAA Research Global-Nest Initiative. Finally, the authors would like to thank Jie Chen and Mingjing Tong for GFDL internal reviews and thank Dr. Chris Landsea and three anonymous reviewers for their insightful comments that led to a much-improved manuscript.

Data availability statement. The Automated Tropical Cyclone Forecasting System (ATCF) a-deck and b-deck files and the GFDL vortex tracker outputs can be downloaded from https://doi.org/10.5281/zenodo.13314471 (Chen 2024).

References

Aberson, S. D., 1998: Five-day tropical cyclone track forecasts in the North Atlantic basin. *Wea. Forecasting*, **13**, 1005–1015, https://doi.org/10.1175/1520-0434(1998)013<1005:FDTCTF>2.0.CO;2.

Alaka, G. J., and Coauthors, 2024: Lifetime performance of the operational Hurricane Weather Research and Forecasting Model (HWRF) for North Atlantic tropical cyclones. *Bull. Amer. Meteor. Soc.*, **105**, E932–E961, https://doi.org/10.1175/BAMS-D-23-0139.1.

Bender, M. A., T. Marchok, R. E. Tuleya, I. Ginis, V. Tallapragada, and S. J. Lord, 2019: Hurricane Model Development at GFDL: A collaborative success story from a historical perspective. *Bull. Amer. Meteor. Soc.*, **100**, 1725–1736, https://doi.org/10.1175/BAMS-D-18-0197.1.

Bonavita, M., 2024: On some limitations of current machine learning weather prediction models. *Geophys. Res. Lett.*, **51**, e2023GL107377, https://doi.org/10.1029/2023GL107377.

Chen, J.-H., 2024: TC track data for BAMS In-Box article. Zenodo, accessed 13 August 2024, https://doi.org/10.5281/zenodo.13314471.

——, S.-J. Lin, L. Magnusson, M. A. Bender, X. Chen, and L. Zhou, 2019a: Advancements in hurricane prediction with NOAA's next generation forecast system. *Geophys. Res. Lett.*, **46**, 4495–4501, https://doi.org/10.1029/2019GL082410.

——, ——, L. Zhou, X. Chen, S. Rees, M. A. Bender, and M. Morin, 2019b: Evaluation of tropical cyclone forecasts in the next generation global prediction system. *Mon. Wea. Rev.*, **147**, 3409–3428, https://doi.org/10.1175/MWR-D-18-0227.1.

——, L. Zhou, L. Magnusson, R. McTaggart-Cowan, and M. Köhler, 2023: Tropical cyclone forecasts in the DIMOSIC project—Medium-range forecast models with common initial conditions. *Earth Space Sci.*, **10**, e2023EA002821, https://doi.org/10.1029/2023EA002821.

DeMaria, M., and Coauthors, 2024: Evaluation of tropical cyclone track and intensity forecasts from Artificial Intelligence Weather Prediction (AIWP) models. arXiv, 2409.06735v1, https://doi.org/10.48550/arXiv.2409.06735.

Dong, J., and Coauthors, 2020: The evaluation of real-time Hurricane Analysis and Forecast System (HAFS) Stand-Alone Regional (SAR) model performance for the 2019 Atlantic hurricane season. *Atmosphere*, **11**, 617, https://doi.org/10.3390/atmos11060617.

ECMWF, 2023: Changes to the forecasting system. Accessed 5 March 2025, https://confluence.ecmwf.int/display/FCST/Changes+to+the+forecasting+system.

——, 2024: IFS documentation. Accessed 12 November 2024, https://www.ecmwf.int/en/publications/ifs-documentation.

Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.*, **94**, 329–343, https://doi.org/10.1175/BAMS-D-12-00071.1.

Gao, K., L. Harris, L. Zhou, M. Bender, and M. Morin, 2021: On the sensitivity of hurricane intensity and structure to horizontal tracer advection schemes in FV3. *J. Atmos. Sci.*, **78**, 3007–3021, https://doi.org/10.1175/JAS-D-20-0331.1.

——, ——, M. Bender, J.-H. Chen, L. Zhou, and T. Knutson, 2023: Regulating fine-scale resolved convection in high-resolution models for better hurricane track prediction. *Geophys. Res. Lett.*, **50**, e2023GL103329, https://doi.org/10.1029/2023GL103329.

Gopalakrishnan, S., and Coauthors, 2021: 2020 HFIP R&D activities summary: Recent results and operational implementation. NOAA HFIP Tech. Rep. HFIP2021-1, 49 pp., https://hfip.org/sites/default/files/documents/hfip-annual-report-2020-final_0.pdf.

Gopalakrishnan, S. G., F. Marks, X. Zhang, J.-W. Bao, K.-S. Yeh, and R. Atlas, 2011: The experimental HWRF system: A study on the influence of horizontal resolution on the structure and intensity changes in tropical cyclones using an idealized framework. *Mon. Wea. Rev.*, **139**, 1762–1784, https://doi.org/10.1175/2010MWR3535.1.

——, S. Goldenberg, T. Quirino, F. Marks, X. Zhang, K.-S. Yeh, R. Atlas, and V. Tallapragada, 2012: Toward improving high-resolution numerical hurricane forecasting: Influence of model horizontal grid resolution, initialization, and physics. *Wea. Forecasting*, **27**, 647–666, https://doi.org/10.1175/WAF-D-11-00055.1.

——, F. Marks, J. A. Zhang, X. Zhang, J.-W. Bao, and V. Tallapragada, 2013: A study of the impacts of vertical diffusion on the structure and intensity of the tropical cyclones using the high-resolution HWRF system. *J. Atmos. Sci.*, **70**, 524–541, https://doi.org/10.1175/JAS-D-11-0340.1.

Han, J., and C. S. Bretherton, 2019: TKE-based moist Eddy-Diffusivity Mass-Flux (EDMF) parameterization for vertical turbulent mixing. *Wea. Forecasting*, **34**, 869–886, https://doi.org/10.1175/WAF-D-18-0146.1.

Harris, L., and Coauthors, 2020: GFDL SHiELD: A unified system for weather to seasonal prediction. *J. Adv. Model. Earth Sci.*, **12**, e2020MS002223, https://doi.org/10.1029/2020MS002223.

——, X. Chen, W. Putman, L. Zhou, and J.-H. Chen, 2021: A scientific description of the GFDL finite-volume cubed-sphere dynamical core. NOAA Tech. Memo. OAR GFDL 2021-001, 109 pp., https://doi.org/10.25923/6nhs-5897.

Hazelton, A., and Coauthors, 2021: 2019 Atlantic hurricane forecasts from the global-nested hurricane analysis and forecast system: Composite statistics and key events. *Wea. Forecasting*, **36**, 519–538, https://doi.org/10.1175/WAF-D-20-0044.1.

——, and Coauthors, 2022: Performance of 2020 real-time Atlantic hurricane forecasts from high-resolution global-nested hurricane models: HAFS-globalnest and GFDL T-SHiELD. *Wea. Forecasting*, **37**, 143–161, https://doi.org/10.1175/WAF-D-21-0102.1.

Hong, S.-Y., 2010: A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon. *Quart. J. Roy. Meteor. Soc.*, **136**, 1481–1496, https://doi.org/10.1002/qj.665.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity for the North Atlantic basin. NOAA Tech. Memo. NWS NHC-10, 22 pp., https://repository.library.noaa.gov/view/noaa/6555.

Knaff, J. A., M. DeMaria, B. Sampson, and J. M. Gross, 2003: Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92, https://doi.org/10.1175/1520-0434(2003)018<0080:SDTCIF>2.0.CO;2.

Lang, S., and Coauthors, 2024: AIFS—ECMWF's data-driven forecasting system. arXiv, 2406.01465v2, https://doi.org/10.48550/arXiv.2406.01465.

Lin, S.-J., 2004: A "Vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2.

Marchok, T., 2021: Important factors in the tracking of tropical cyclones in operational models. *J. Appl. Meteor. Climatol.*, **60**, 1265–1284, https://doi.org/10.1175/JAMC-D-20-0175.1.

Miller, R. J., A. J. Scrader, C. R. Sampson, and T. L. Tsui, 1990: The Automated Tropical Cyclone Forecasting system (ATCF). *Wea. Forecasting*, **5**, 653–660, https://doi.org/10.1175/1520-0434(1990)005<0653:TATCFS>2.0.CO;2.

Neumann, C. B., 1972: An alternate to the HURRAN (hurricane analog) tropical cyclone forecast system. NOAA Tech. Memo. NWS SR-62, 28 pp., https://repository.library.noaa.gov/view/noaa/3605.

NOAA/EMC, 2024: GFS home page. Accessed 31 July 2024, https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs.php.

NOAA/NHC, 2024: National Hurricane Center forecast verification. Accessed 31 July 2024, https://www.nhc.noaa.gov/verification/verify6.shtml.

NOAA/OSTI, 2024: NGGPS dynamic core evaluation and selection. Accessed 31 July 2024, https://vlab.noaa.gov/web/osti-modeling/dynamiccoreselection.

Pathak, J., and Coauthors, 2022: FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv, 2202.11214v1, https://doi.org/10.48550/arXiv.2202.11214.

Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, https://doi.org/10.1016/j.jcp.2007.07.022.

Sampson, C. R., and A. J. Schrader, 2000: The automated tropical cyclone forecasting system (version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240, https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2.

Sun, Y. Q., P. Hassanzadeh, M. Zand, A. Chattopadhyay, J. Weare, and D. S. Abbot, 2024: Can AI weather models predict out-of-distribution gray swan tropical cyclones? arXiv, 2410.14932v3, https://doi.org/10.48550/arXiv.2410.14932.

Yang, F., 2020: Development and evaluation of NCEP's Global Forecast System GFSv16. UFS Webinar, accessed 7 August 2024, https://ufs.epic.noaa.gov/2020/10/development-and-evaluation-of-nceps-global-forecast-system-gfsv16/.

Zhang, Z., and Coauthors, 2023: A review of recent advances (2018–2021) on tropical cyclone intensity change from operational perspectives, Part 1: Dynamical model guidance. *Trop. Cyclone Res. Rev.*, **12**, 30–49, https://doi.org/10.1016/j.tcrr.2023.05.004.

Zhou, L., and Coauthors, 2022: Improving global weather prediction in GFDL SHiELD through an upgraded GFDL cloud microphysics scheme. *J. Adv. Model. Earth Sci.*, **14**, e2021MS002971, https://doi.org/10.1029/2021MS002971.