Evaluation of the Outstanding Track Performance of the GFDL SHiELD Global Model During the 2021 Atlantic Hurricane Season

Linjiong Zhou[1], Morris Bender[1], Matt Morin[2], Lucas Harris[3], Alex Kaltenbaugh[2], and Jie Chen[1]

[1]Princeton University
[2]UCAR/GFDL
[3]NOAA/Geophysical Fluid Dynamics Laboratory

November 24, 2022

Abstract

The 13 km SHiELD (System for High-resolution prediction on Earth-to-Local Domains) global model that is under development at the Geophysical Fluid Dynamics Laboratory (GFDL) and runs in near real time, produced outstanding tropical cyclone track forecasts during the 2021 Atlantic hurricane season, compared to both the upgraded National Weather Service Global Forecast System (GFSv16), the Hurricane Weather Research and Forecasting (HWRF) model and the European Centre for Medium-Range Weather Forecast Integrated Forecasting System (IFS). SHiELD's average track forecast errors were 10% and 15% less than the GFSv16 and HWRF, respectively, for the 3-5 day forecast lead times. SHiELD's track forecast skill was comparable to the National Hurricane Center's official forecast at several forecast lead times, and approached 70% skill relative to the Climatology and Persistence Model (CLIPER) at 3 and 4 days. Similar improvements were found in the western Pacific basin in 2021, with improvements also seen in the eastern Pacific at days 4 and 5. Improved performance was also found in the 2019 Atlantic hurricane season, with neutral performance in 2020, when SHiELD was run retrospectively from the GFSv16 initial conditions. Distribution of the spatial errors and biases showed that in both the 2021 Atlantic hurricane season and the previous two years, the largest track forecast errors from both SHiELD and GFSv16 occurred in the subtropical eastern Atlantic, associated with a distinct northeast bias. Analysis indicated that some of the excessive north bias in the GFSv16 is associated with lower geopotential height fields compared to those in SHiELD.

Evaluation of the Outstanding Track Performance of the GFDL SHiELD Global Model During the 2021 Atlantic Hurricane Season

Linjiong Zhou[a] , Morris Bender[a], Matthew Morin[b], Lucas Harris[c], Alex Kaltenbaugh[b], and Jie Chen[a]

[a] *Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey*

[b] *University Corporation for Atmospheric Research, Boulder, Colorado*

[c] *National Oceanic and Atmospheric Administration / Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey*

Corresponding author: Linjiong Zhou, linjiong.zhou@noaa.gov

April, 2022

This Work has been submitted to Monthly Weather Review.

ABSTRACT

The 13 km SHiELD (System for High-resolution prediction on Earth-to-Local Domains) global model that is under development at the Geophysical Fluid Dynamics Laboratory (GFDL) and runs in near real time, produced outstanding tropical cyclone track forecasts during the 2021 Atlantic hurricane season, compared to both the upgraded National Weather Service Global Forecast System (GFSv16), the Hurricane Weather Research and Forecasting (HWRF) model and the European Centre for Medium-Range Weather Forecast Integrated Forecasting System (IFS). SHiELD's average track forecast errors were 10% and 15% less than the GFSv16 and HWRF, respectively, for the 3-5 day forecast lead times. SHiELD's track forecast skill was comparable to the National Hurricane Center's official forecast at several forecast lead times, and approached 70% skill relative to the Climatology and Persistence Model (CLIPER) at 3 and 4 days. Similar improvements were found in the western Pacific basin in 2021, with improvements also seen in the eastern Pacific at days 4 and 5. Improved performance was also found in the 2019 Atlantic hurricane season, with neutral performance in 2020, when SHiELD was run retrospectively from the GFSv16 initial conditions. Distribution of the spatial errors and biases showed that in both the 2021 Atlantic hurricane season and the previous two years, the largest track forecast errors from both SHiELD and GFSv16 occurred in the subtropical eastern Atlantic, associated with a distinct northeast bias. Analysis indicated that some of the excessive north bias in the GFSv16 is associated with lower geopotential height fields compared to those in SHiELD.

Keywords: tropical cyclones, prediction/forecasting, model evaluation/performance

1. Introduction

Tropical Cyclone (TC) track prediction has shown dramatic improvements in the past 30 years, with average 24-72h track forecast errors in the Atlantic and eastern Pacific basins decreasing nearly 70% during this time (Cangialosi 2021). It is well known that most of these improvements can be attributed to the improvements in the accuracy of the track forecasting performance of numerical models used for TC prediction. Initially, limited-area models with higher resolution that could more adequately resolve the hurricane inner-structure provided the most accurate hurricane track prediction (Bender et al. 2019; Bender et al. 2007; Bender and Ginis 2000). For example, as shown by Bender et al. (2019), when the National Oceanic and Atmospheric Administration (NOAA)'s Geophysical Fluid Dynamics Laboratory (GFDL) regional hurricane model became operational for the National Weather Service (NWS) in 1995, its 72h track forecast error that year was 210 nautical miles (n mi for short hereafter) or 389 km compared to 360 n mi (667 km) for the NWS's operational global model (then called the Aviation model, or "AVN"). However, within 5 years, the average track forecast skill of the global and regional models became comparable. By the time the NWS's new limited area Hurricane Weather Research and Forecasting (HWRF; Tallapragada et al. 2016) model became fully operational in 2007, the NWS operational global model (then renamed the Global Forecast System, or "GFS") was consistently exhibiting superior track forecasting performance compared to either the GFDL or HWRF model. As the NWS continued to upgrade the GFS model, it has remained one of the most skillful track prediction models in the world (Cangialosi 2021). On 12 June 2019, a new version of the GFS model was implemented into operations at the NWS which replaced the model's spectral dynamical core with the non-hydrostatic Finite-Volume Cubed-Sphere Dynamical Core (FV3; Putman and Lin 2007). The dynamical core was transitioned from GFDL where it serves as the backbone of the GFDL's global seamless weather-to-climate modeling system. In this initial implementation of the FV3-based GFS model (referred to as GFSv15 in the NWS implementation), most of the physics suite running in the previous spectral version of the GFS (referred to as the GFSv14) was transitioned to GFSv15, except for the microphysics scheme, which was replaced with the single-moment five-category cloud microphysics scheme developed at GFDL (Zhou et al. 2019).

Meanwhile, advancements in global model development at GFDL have continued since the transition of the FV3 dynamical core to the NWS and the implementation of GFSv15. Most of these model advancements at GFDL have focused on improved weather prediction through the development of the System for High-resolution prediction on Earth-to-Local Domains (SHiELD), an atmospheric global prediction system that initially coupled the non-hydrostatic FV3 with the physics suite in GFSv14. Since then, the Weather and Climate Dynamics Division at GFDL has continued to advance SHiELD with improved physics and dynamics (Harris et al. 2020). In order to investigate advanced model capabilities and improve the skill of FV3-based models in the forecast and prediction of weather phenomena on various time and spatial scales, a hierarchy of models has also been developed at GFDL over the past few years from centennial-scale earth-system simulations (Dunne et al. 2020) to very high-resolution weather prediction. For example, a global-nested configuration of the SHiELD system, with a high resolution 3 km nest spanning the tropical Atlantic (called T-SHiELD), has been developed and used to predict TC track and intensity over the past five years (Hazelton et al. 2022), in support of NOAA's Hurricane Advanced Forecast System (HAFS) and the Hurricane Forecast Improvement Project (HFIP) programs.

On 22 March 2021, the NWS operational GFS was upgraded to a new version (hereafter referred to as GFSv16) that included a doubling of the vertical resolution from 63 to 127 levels, improved model physics and major advancements in the data assimilation. These upgrades are summarized in detail in Han et al. (2021; 2022). The 2020 version of the SHiELD model continued to be run at GFDL in near real time throughout 2021 with the GFSv16 fields providing the initial conditions for these forecasts. As we will show in this paper, SHiELD provided outstanding hurricane track guidance in all northern hemisphere TC basins, particularly the Atlantic where it produced lower track forecast errors compared to all operational models.

The purpose of this study is to quantify the outstanding track forecasting performance of the SHiELD model particularly in the Atlantic basin, making extensive analysis of its performance compared to the NWS's GFSv16 and HWRF operational forecast systems as well as the Integrated Forecast System (IFS) of the European Centre for Medium-Range Weather Forecasting (ECMWF), which has been the top performing TC track prediction model in all northern hemisphere basins over the past several years (e.g., Cangialosi 2018; 2019; 2020). It is hoped that this analysis, which pinpoints some of SHiELD's strengths and weaknesses, may facilitate transition of improvements to the GFS and advancement in global modeling, with the goal of assisting in improved numerical weather prediction (NWP) on all weather time scales. In section 2 the developmental efforts of SHiELD will be briefly outlined, focusing on the improved physics and advancements in the dynamical core that may have led to the improved prediction in hurricane track. These results will be presented in detail in section 3, starting with the improved anomaly correlation coefficient (ACC) which is often used to evaluate NWP model skill (e.g., Harris et al. 2020). Finally, we end with a summary and discussion in section 4.

2. Summary of the SHiELD Model and Experimental Design

As discussed previously, advancements in global model development have continued within the Weather and Climate Dynamics Division at GFDL since the 12 June 2019 implementation of the GFSv15 at the NWS's National Centers for Environmental Prediction (NCEP) through development of the SHiELD atmospheric model. The second column of Table 1 summarizes some of the upgrades that have been implemented in the 2020 version of SHiELD that was used in this study. First, continuous efforts have been put into the non-hydrostatic FV3 dynamical core development in order to improve its numerical and computational performance and enhance its capability of seamless predictions from convective-scale to seasonal-scales (Harris et al. 2020). Through code sharing, the 2020 version of SHiELD and GFSv16 use the same version of the FV3 dynamical core. However, the GFSv16 uses higher vertical resolution, with 127 vertical levels topped at 80 km, while SHiELD continues to use 91 vertical levels topped at 55 km. Other significant upgrades for SHiELD include the application of a 1-D mixed-layer ocean (MLO) model (Pollard et al. 1973) together with an ocean surface roughness modification from HWRF to improve the prediction of TC intensity (Biswas et al. 2018). Finally, updating to the inline GFDL cloud microphysics scheme (Harris et al. 2020) was done to improve the simulation of moist processes as well as cloud and weather predictions. These developments were added to SHiELD and have not been transitioned to the GFS yet. GFS had significant upgrades to its convection scheme and the boundary layer turbulence scheme in version 16 (Han et al. 2021; 2022), which were not implemented in the version of SHiELD used in this study.

Model	**SHiELD**	**GFSv16**	**IFS**	**HWRF**
Dynamical Core	Non-hydrostatic FV3[1]	Non-Hydrostatic FV3	Hydrostatic Spectral	Non-hydrostatic NMM[2]
Model Type	Global	Global	Global	Regional
Horizontal Resolution	13 km	13 km	9 km	1.5-13.5 km
Vertical Layers	91 (top 55 km)	127 (top 80 km)	137 (top 80 km)	75 (top 31 km)
Data Assimilation (DA)	None (IC[3] from GFSv16)	GDAS[4] Hybrid 3DEnVar	4DVar	Hybrid and TDR-based EnKF/Var
Ocean Coupling	1D MLO[5]	None	NEMO[6]	MPIPOM-TC[7]
Microphysics	Inline GFDL Microphysics	Split GFDL Microphysics	EC Microphysics	Ferrier-Aligo Microphysics
Radiation	RRTM[8]	RRTM	RRTM-ECrad	RRTM
Boundary Layer Turbulence	SA-TKE-EDMF[9]	New SA-TKE-EDMF	EDMF	GFS Hybrid-EDMF
Convection	SA-SAS[10]	New SA-SAS	Tiedtke-Bechtold	SA-SAS

[1] FV3: Finite-Volume Cubed-Sphere Dynamical Core

[2] NMM: Non-hydrostatic Mesoscale Model

[3] IC: Initial Condition
[4] GDAS: Global Data Assimilation System
[5] MLO: Mixed Layer Ocean
[6] NEMO: Nucleus for European Modelling of the Ocean
[7] MPIPOM-TC: Message Passing Interface Princeton Ocean Model-Tropical Cyclone
[8] RRTM: Rapid Radiative Transfer Model
[9] SA-TKE-EDMF: Scale Aware Turbulent Kinetic Energy based Moist Eddy Diffusivity Mass Flux
[10] SA-SAS: Scale Aware Simplified Arakawa Schubert

Table 1. Summary of the key components of the four models evaluated in this study.

All of the SHiELD forecasts used in this study were initialized using analyses from the GFSv16 that went into operations at NCEP on 22 March 2021. All model forecasts were cold-started from the GFSv16 initial conditions with no further data assimilation. The SHiELD horizontal grid is identical to that of the GFSv16 (C768, ~13 km) and no horizontal interpolation of the atmosphere or surface analyses was necessary. Interpolation between the GFSv16's 127 levels and SHiELD's 91 levels was done by an accurate cubic-spline vertical remapping to maintain as much conservation and consistency between the GFSv16 analyses and FV3 dynamics as possible (see Section 10.1 of Harris et al. [2021] for more details). Although most of the analysis for this study focused on the 2021 tropical cyclone season starting from the 0z, 6z, 12z and 18z analysis cycles, forecasts were also made for the 2019 and 2020 seasons at 0z and 12z, in order to evaluate the robustness of the results for prior seasons. These retrospective SHiELD forecasts for the 2019 and 2020 seasons used the analyses from the retrospective pre-operational GFSv16 system, ensuring a homogeneous comparison between the GFSv16 and SHiELD. Although all of the SHiELD forecasts were run out to 10 days, all of the analysis of results presented in this study will focus on the 1 to 5 day forecast lead times which is consistent with the National Hurricane Center (NHC) and the Joint Typhoon Weather Center's (JTWC) official period of forecast responsibility for TCs in their respective basins of responsibility presented in this study (Atlantic and eastern Pacific for NHC, and western North Pacific for the JTWC). TC track forecasts for all of the models used in this study, are evaluated using the GFDL vortex tracker (Marchok 2021) verified against the widely used NHC's "best tracks" analyses (Landsea and Franklin 2013). Geopotential height is verified against the ERA5 reanalysis (Hersbach et al. 2020).

Finally, since comparison throughout this section will also be made to the HWRF and the IFS models, a summary of these four modeling systems is also presented in Table 1. Note that the vertical resolution of SHiELD is somewhat coarser (91 levels) compared to the GFSv16 (127 levels) and the IFS (137 levels) global models but slightly finer than the operational HWRF model (75 levels).

3. Analysis of SHiELD Track Performance

The focus of this section will be to quantify the superior track forecasting performance of the SHiELD model compared to the NWS's GFS, which, as mentioned previously, was upgraded by the NWS on 22 March 2021 to GFSv16 with an increase of vertical resolution, improved physics upgrades and major advancements in the data assimilation. Since improved vertical model resolution has been shown to be important for improved TC track skill in numerous studies (e.g., Feng and Wang 2021; Zhang et al. 2016; Zhang et al. 2015), the superior track performance in SHiELD was a somewhat surprising result which certainly warrants further investigation, particularly as SHiELD and GFSv16 were run with similar dynamical core and physical parameterizations. Evaluation of storm intensity was not a focus of this study as global models are still too coarse to adequately resolve the hurricane inner structure. In addition, global model data is typically not archived at native resolution, making intensity comparisons between the models unfair.

The 2021 Atlantic hurricane season, which will be the main focus of this study, was an active hurricane season with above average accumulated cyclone energy (ACE, units of 10^4 knot^2; Bell et al. 2000) of 145.1 and 21 storms that obtained the status of at least tropical storm or subtropical storm (Fig. 1). This activity is significantly greater than the long-term mean. Most noteworthy were the four hurricanes that obtained major hurricane status, two of which were exceptionally long-lived (Hurricanes Larry and Sam), and Hurricane Ida which had devastating impacts on the United States, making landfall in Louisiana on 29 August 2021, with winds of 130 knots (or 241 km/h). Hurricane Sam, the longest-lived storm of the 2021 Atlantic hurricane season, lasted 12 days with a total ACE value of 53.8. Despite the large number of named storms in 2021, only 7 obtained hurricane status with the bulk of the named storms characterized as weak systems with many of relatively short duration. Overall, since the TC genesis locations were distributed over a wide range of the Atlantic basin, the season should provide a robust and diverse sample of cases to evaluate model performance and skill.

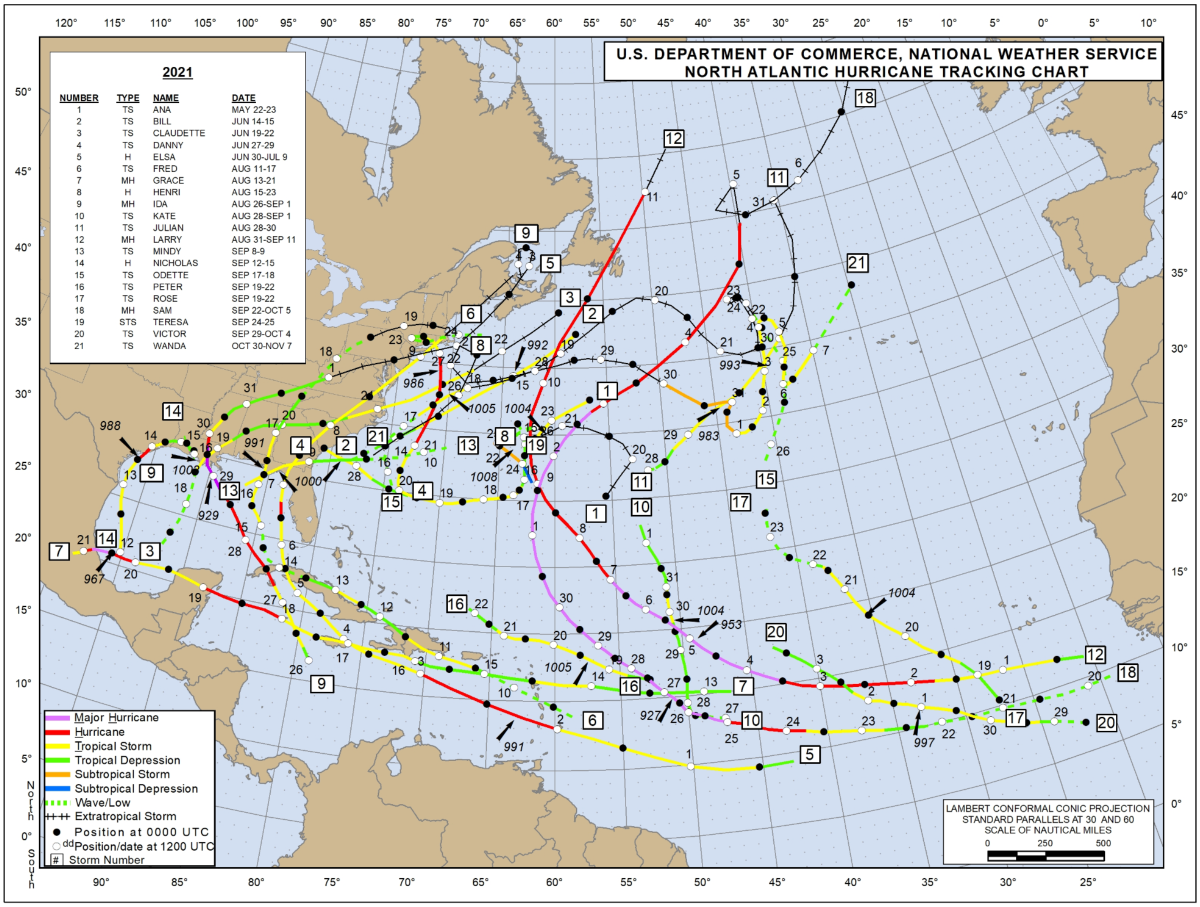

Fig. 1. Tracks of all TCs during the 2021 Atlantic hurricane season (from https://www.nhc.noaa.gov/data/tcr/index.php; courtesy of NHC).

As shown in Fig. 2a, SHiELD's mean track forecast errors were 10% and 15% less than those of the GFSv16 and HWRF, respectively for the average 3-5 day forecast lead times. The improvement of SHiELD compared to the GFSv16 was statistically significant at the 90% confidence level at 48h, and exceeded the 95% confidence level at days 3 and 4. Due to the small sample size at day 5, statistical significance was not found although average track error was reduced by about 8%.

It is interesting that the SHiELD's track forecast skill during the 2021 Atlantic hurricane season was even greater than the GFDL high-resolution T-SHiELD model (not shown), which also provided outstanding track prediction. Note from the x-axis labels of Fig. 2b that only 12 of the 21 Atlantic TCs in 2021 survived for long enough to have at least one forecast case that extended to 96 hours. This was due to the fact that the 2021 Atlantic hurricane season, particularly in the early part of the season, was dominated by weak and short-lived storms which SHiELD did very well in predicting. The improved track prediction of SHiELD compared to the GFSv16 was consistent with improved prediction of the 500 hPa geopotential height ACC (Fig. 3) which is one of the most widely used large-scale metrics to evaluate model skill in NWP models. Note that the improved ACC in SHiELD averaged for the entire 2021 Atlantic hurricane season, exceeded the 95% confidence interval for 3 to 5 day forecast lead times.

Two storms which produced a much-improved performance of SHiELD compared to the GFSv16, were Hurricanes Grace (Fig. 4) and Tropical Storm (TS) Victor (Fig. 5) as can be seen in Fig. 2b. Much of the improved track forecasting performance for these two storms came from a significantly reduced north bias in SHiELD compared to the GFSv16, particularly for the early forecast cycles of TS Victor, where the GFSv16 erroneously accelerated the storm quickly northward. Analysis of the 500 hPa geopotential height field (Fig. 6) during the period of TS Victor indicated that the predicted height fields were too low in both models compared to the ERA5 reanalysis over much of the eastern Atlantic in the deep tropics of the Main Development Region (MDR), although the negative height anomaly was worse in the GFSv16 near this storm. It is uncertain how much of an impact this had on the hurricane track. However, it is likely that the weaker ridge that was apparent in the GFSv16 in the eastern Atlantic deep tropics (Fig. 6f) during the period of TS Victor, likely contributed to the north bias in many of the TS Victor forecasts.

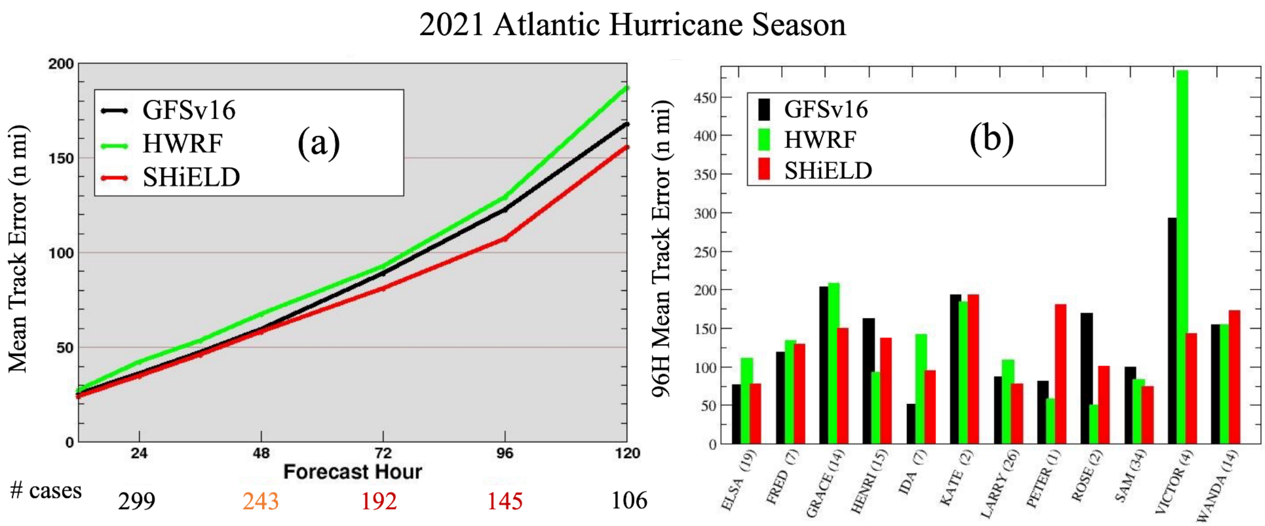

Fig. 2. 2021 Atlantic hurricane season (a) mean track forecast errors (n mi) for the GFSv16 (black), HWRF (green) models compared to SHiELD (red), and (b) 96h mean track forecast errors (n mi) for all storms that had at least one forecast case that lasted for 96 hours. Number of verifying cases are shown at the bottom of panel (a), with forecast lead times showing statistically significant improvement at 90% and 95% confidence intervals between the SHiELD and the GFSv16, indicated by orange and red colors respectively. Number of cases that are verified at 96h is shown at the bottom of panel (b) identified with the storm names.

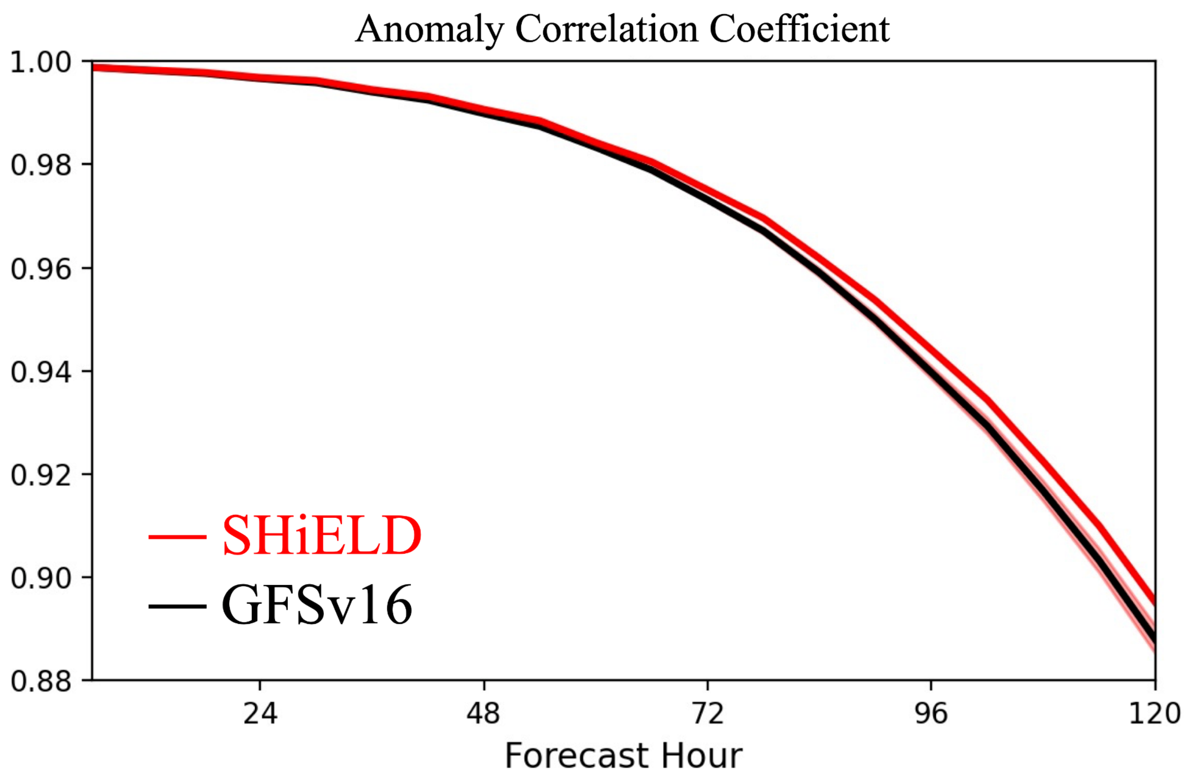

Fig. 3. Mean 500 hPa geopotential height ACC in the northern hemisphere for both SHiELD and the GFSv16. The. Mean ACC was computed from a total of 183 forecasts (1 June to 10 November 2021) of the SHiELD (red) and the GFSv16 (black) models verified against the ERA5 reanalysis. Pink shaded area is the 95% significance interval of their difference.

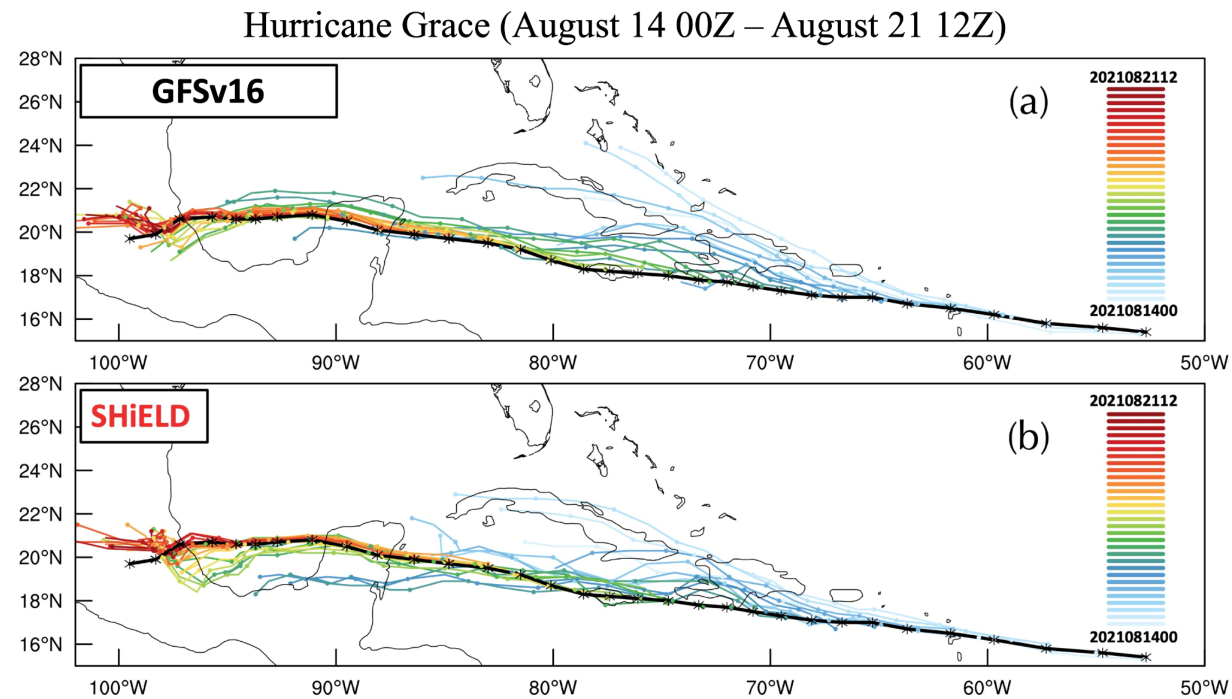

Fig. 4. Composite 5-day forecast tracks of Hurricane Grace for cases initiated at 0z and 12z synoptic times, for the (a) GFSv16 and (b) SHiELD. Black dashed line is from the "best tracks" analyses. Color lines represent different initial dates and times.

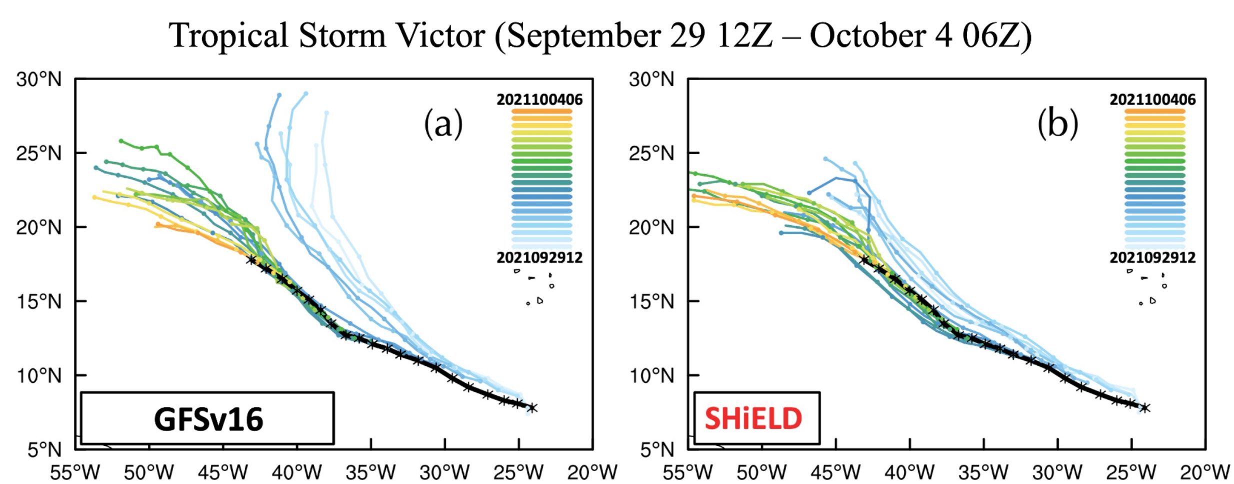

Fig. 5. The same as Fig. 4, but for TS Victor.

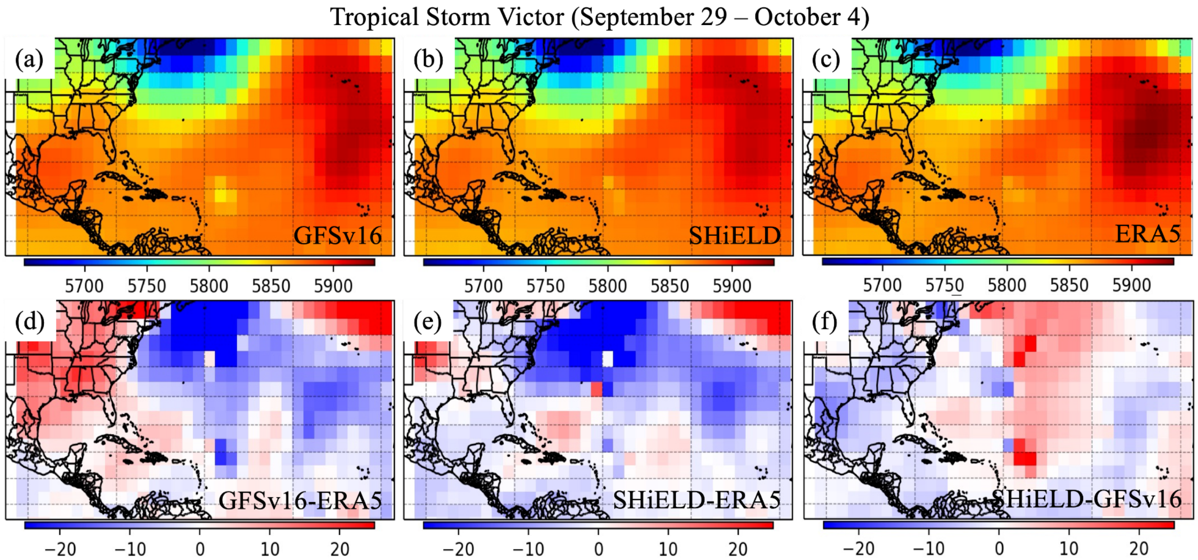

Fig. 6. The 500 hPa geopotential height fields (m) for the (a) GFSv16 and (b) SHiELD averaged for the life-cycle of TS Victor, and compared to the (c) ERA5 reanalysis. Difference fields (d, e) between the models and the ERA5 reanalysis as well as (f) between the two models are also shown.

The track forecast errors for Hurricane Sam, the longest-lived storm of the 2021 Atlantic hurricane season, were also significantly reduced for SHiELD at all forecast lead times. At 3 and 4 days, SHiELD's mean track forecast errors of 52 and 74 n mi were 25% less than the operational GFSv16 (Fig. 2b). A prominent slow bias particularly during recurvature likely contributed to the larger errors in the GFSv16 at the longer forecast lead times (Fig. 7d). This appears to be consistent with the higher geopotential heights in SHiELD particularly in the region traversed by Hurricane Sam (Fig. 7c). Also, in the early period of Hurricane Sam, the weaker ridging predicted by the GFSv16 east of the Caribbean likely accounted for the premature recurvature in the GFSv16 compared to SHiELD (Fig. 7e). Overall, the forecast errors for all three models were extremely low for Hurricane Sam (Fig. 2b), which was one of the better forecasted TCs of 2021.

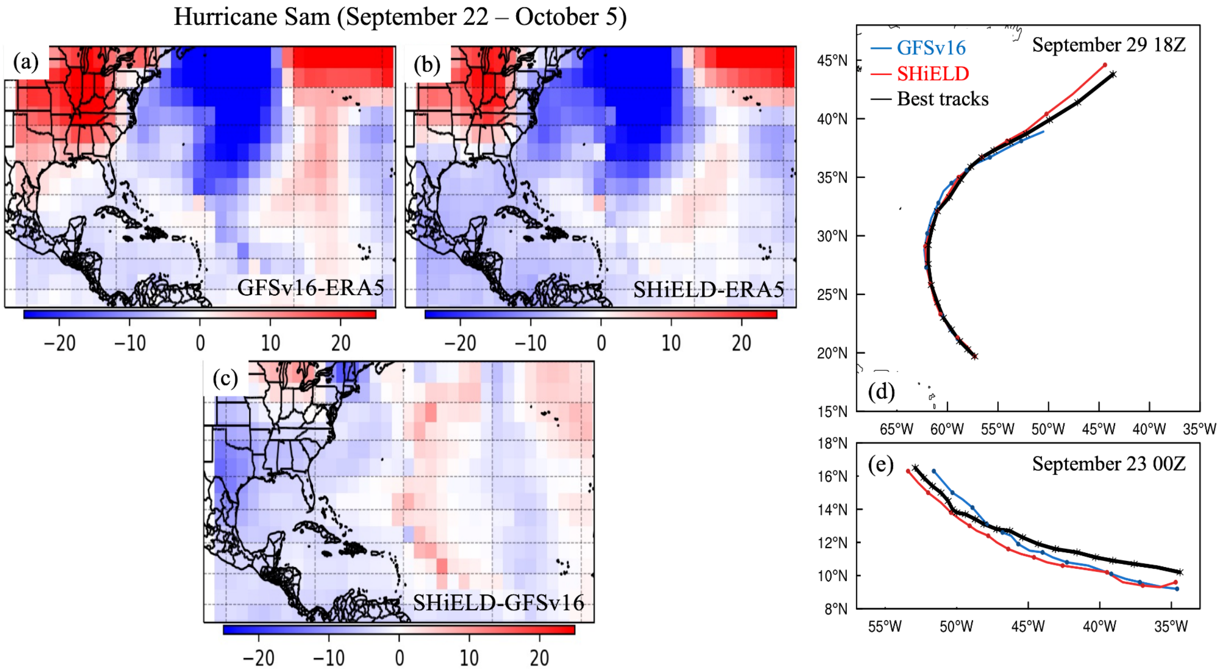

Fig. 7. 500 hPa geopotential height difference fields (m) between the (a) GFSv16 and ERA5 reanalysis, (b) SHiELD and ERA5 reanalysis, and (c) SHiELD and GFSv16, averaged for the lifecycle of Hurricane Sam. Hurricane tracks are compared between the SHiELD (red) and the GFSv16 (blue), for 5-day forecasts initialized at (d) September 29 18z and (e) September 23 0z. Black line is from the "best tracks" analyses.

Another metric used to evaluate track performance for operational guidance is the track skill normalized with respect to the Climatology and Persistence (CLIPER) model, which typically serves as a baseline for model skill (Sampson and Schrader 2000). Track skill refers to the average percent of reduced track forecast error of each model relative to the forecast track from a reference model which is based entirely on climatology and persistence (Neumann 1973). In Fig. 8 the "early guidance" is presented which employs a time-interpolation technique that produces model guidance which can be made available to the operational centers to produce their official forecast (e.g., National Hurricane Center Model Error Trends, 2021, https://www.nhc.noaa.gov/verification/verify6.shtml). The presentation of the "early guidance" is necessary for proper comparison with the official forecast and is standard operational procedure at NHC and other operational centers. The IFS model's performance is also included, since it has had the highest track skill in the Atlantic over the past several years, as mentioned previously. In order to maximize the sample size of cases, in the previous figures the IFS forecasts were not included since this model is only available twice daily, compared to the GFSv16, HWRF and SHiELD, which are run four times daily. Also, in many of the forecasts in the early portion of the Atlantic hurricane season, the IFS was unable to follow many of the weaker storms to 5-days, which also reduced the sample size since all of these model comparisons involve a perfectly homogeneous set of model forecasts (model forecasts are only included in the verifications if the forecast being verified at that forecast time is available from all models).

Following operational procedures, the IFS forecasts are interpolated 12 hours in time to produce the "early guidance" at 0z and 12z in contrast to the GFSv16 and HWRF, which only have to be interpolated 6 hours in time. Despite this obvious disadvantage, the IFS still has been the most skillful model in the Atlantic in most northern hemisphere basins. However, in 2021 at lead times beyond 48h, the IFS performed worse than the GFSv16 while SHiELD showed superior track skill compared to all operational guidance (Fig. 8). Note that at 96h the SHiELD was actually comparable in skill to the official forecast and approached 70% skill relative to CLIPER at 3 and 4 days.

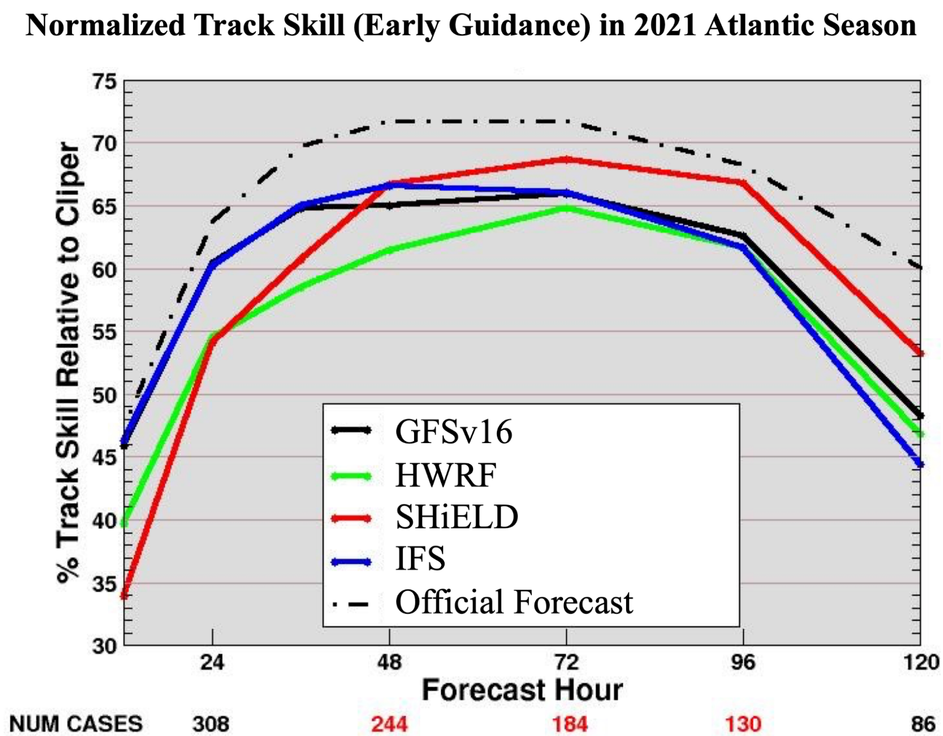

Fig. 8. Early model track guidance of the GFSv16 (black), HWRF (green), SHiELD (red), IFS (blue), and the NHC official forecast (black dotted dashed) normalized relative to the CLIPER (Climatology and Persistence) model for the 2021 Atlantic hurricane season. The percent (%) track skill refers to percent of reduced averaged track error compared to CLIPER. Number of cases are shown at the bottom, with forecast lead times showing statistically significant improvement at 90% and 95% confidence intervals between the SHiELD and the GFSv16, indicated by orange and red colors respectively.

A spatial analysis of track forecast errors and biases was performed to help identify differences in model forecast performance across subregions of the Atlantic basin. The analysis was evaluated for each lead time on a one-degree latitude-longitude grid by employing a technique that applies a Gaussian smoothing to the forecast errors and biases and then averages them for each point on the grid. This Gaussian smoothing was accomplished using the same Barnes analysis technique that is utilized in the GFDL vortex tracker (Marchok 2021). For the current application, an e-folding radius of 450 km and a radius of influence of 1200 km were used. A minimum of five forecast data points within the radius of influence at each analysis grid point must exist to provide a spatial analysis estimate at the analysis grid point.

The 96h spatial distribution of track forecast errors and biases is presented in Fig. 9 for SHiELD, GFSv16, and HWRF, averaged for the entire 2021 Atlantic hurricane season. Both the SHiELD and GFSv16 models produced extremely low forecast errors and biases in the central Atlantic (65W to 50W) with a modest north bias in the subtropical western Atlantic, as was evident for Hurricane Grace. All three models had their largest track forecast errors and a distinct northeast bias in the eastern Atlantic, particularly in the sub-tropics. However, the degraded performance of the GFSv16 compared to the SHiELD is clearly evident in this region, as seen previously in the track errors of TS Victor, where the distinct north bias occurred in the GFSv16. Also, the better performance of SHiELD in the Western Caribbean region is evident, compared to both the HWRF and the GFSv16.

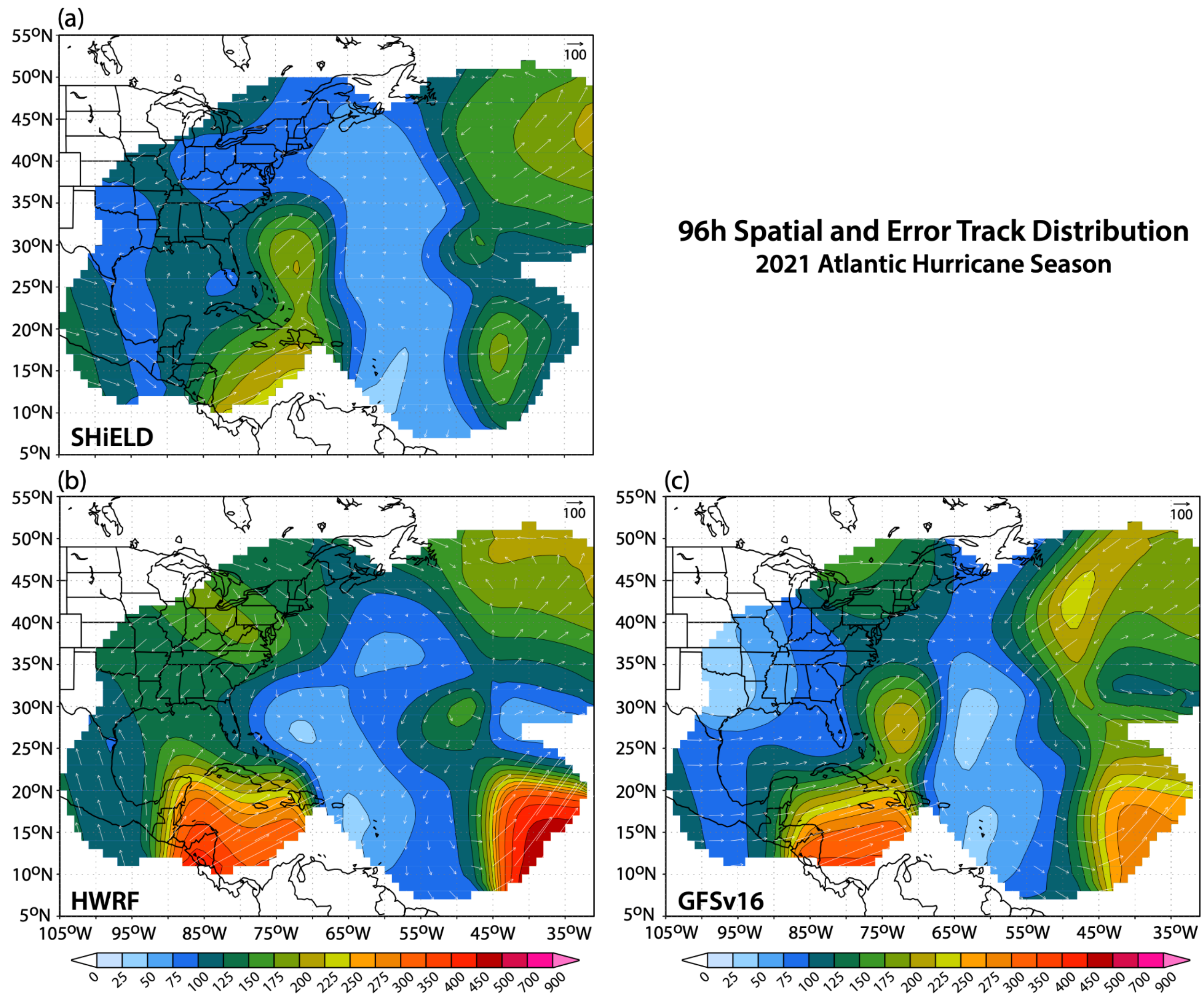

Fig. 9. Spatial distribution of the 96h model track forecast error (color contour) and bias (arrows) averaged for the entire 2021 Atlantic hurricane season for (a) SHiELD, (b) HWRF, and (c) GFSv16. The length of the vector arrows corresponds to a 100 n mi of track forecast bias.

Evaluation of the Atlantic spatial forecast errors and biases at 96h for the combined 2019, 2020 and 2021 hurricane seasons (Fig. 10) shows a similar pattern. The reduction in average 72h and 96h forecast error between SHiELD and the GFSv16 (not shown) averaged 10% and 9% (96 n mi vs. 106 n mi and 131 n mi vs. 145 n mi, respectively) for the combined three-year sample. A pronounced northeast track bias in the eastern Atlantic was also evident in the combined three seasons, similar to just 2021 alone (cf. Figs. 9, 10), contributing to the excessive track errors for that part of the basin, particularly for the GFSv16. It is interesting to note the very small track forecast bias both in SHiELD and the GFSv16 in the central Atlantic compared to the HWRF, which exhibited a pronounced south bias in the three-year mean. Note that in both the 2021 Atlantic hurricane season as well as the combined three-year sample SHiELD produced better track forecast performance in the western Caribbean compared to the GFSv16, with both models showing comparable track performance in the Gulf of Mexico. Despite large year-to-year variability in model track forecast performance (to be discussed later), both of the global models have similar bias and error distributions in the combined three-year sample as well as 2021 contrasted to the HWRF, which had a somewhat different spatial distribution in the three-year sample compared to 2021 in both the central and eastern Atlantic.

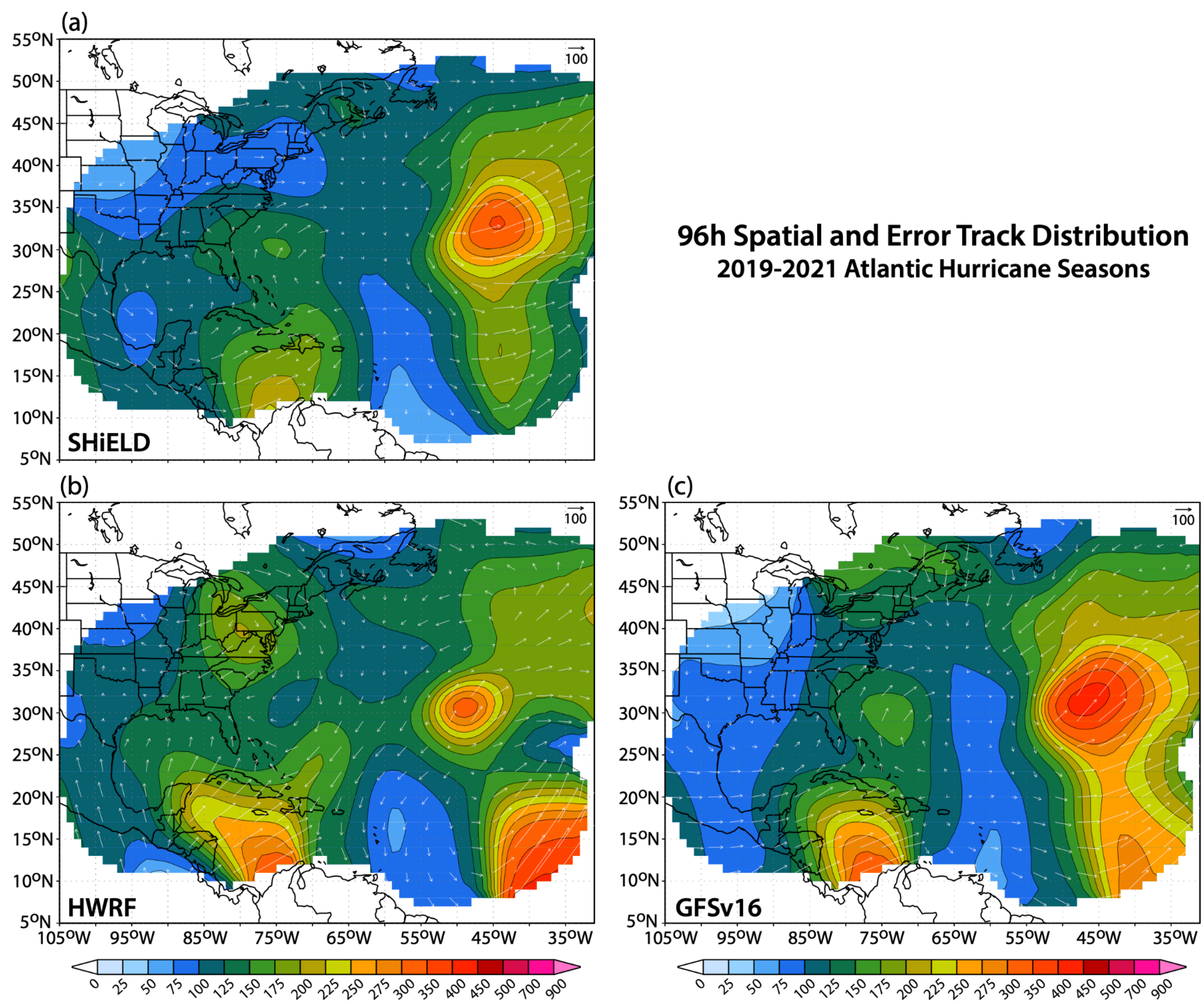

Fig. 10. The same as Fig. 9, but for the entire 2019, 2020 and 2021 Atlantic hurricane seasons.

The analysis of the 2021 mean track forecast error for the two major Pacific basins is presented in Fig. 11 to demonstrate the robustness of the improved track performance of SHiELD in the other major northern hemisphere basins in 2021 compared to the GFSv16. In order to include the official forecast in the comparisons, the early guidance is presented, and both the HWRF and IFS model results are included. In the eastern Pacific, the mean forecast error was comparable between SHiELD and GFSv16 through the first 3 days, with SHiELD exhibiting about 7% to 10% reduced track forecast error at the 4 to 5 day lead times. However, the IFS was considerably more skillful at all forecast lead times (e.g., 12% and 17% reduced track error at 3-5 days compared to the SHiELD and GFSv16, respectively). Nevertheless the three global models showed superior performance compared to the regional HWRF model in this basin at all forecast times. In contrast, in the western Pacific in 2021, the HWRF and GFSv16 exhibited very similar track forecast errors beyond 2 days and SHiELD showed about 7% reduced track error at 3 to 5 days compared to these two models which was statistically significant at the 95% interval. In contrast to the Atlantic, the IFS was the top performer for track in the western Pacific for the operational models while the SHiELD performance was very comparable to the IFS except at 96h.

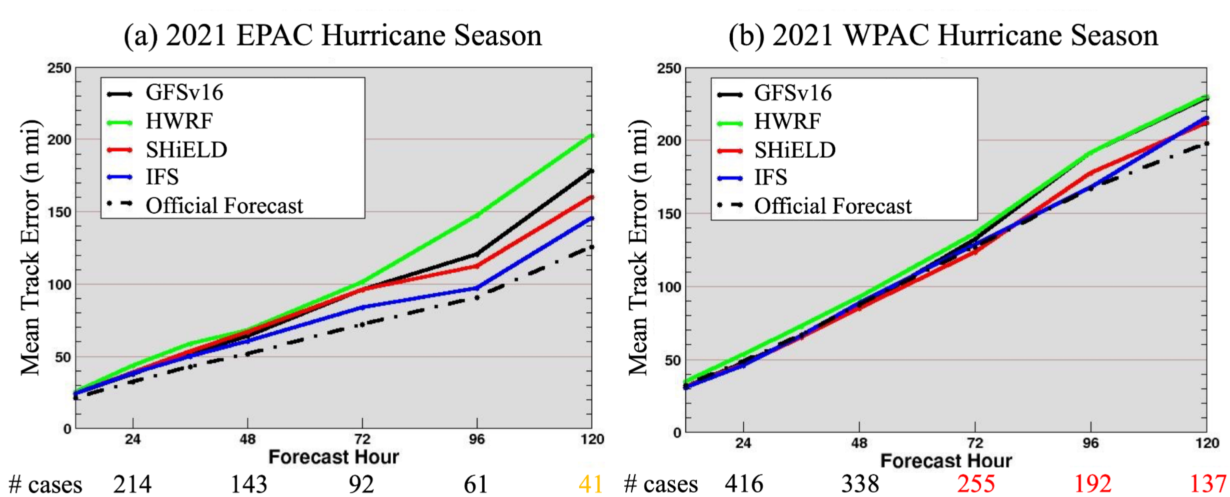

Fig. 11. Mean track forecast errors (n mi) for the GFSv16 (black), HWRF (green), SHiELD (red), IFS (blue), and the NHC's official forecast (black dotted dashed) for the 2021 (a) eastern Pacific and (b) western Pacific hurricane seasons. Number of cases are shown at the bottom of

each panel, with forecast lead times showing statistically significant improvement at 90% and 95% confidence intervals between SHiELD and the GFSv16, indicated by orange and red colors respectively.

Finally, an evaluation of mean track forecast error was made for the 2019 and 2020 Atlantic hurricane seasons (Fig. 12) to also evaluate the robustness of the Atlantic SHiELD track performance compared to the GFSv16 model performance during the two prior Atlantic hurricane seasons. While those hurricane seasons occurred prior to the 2021 implementation of GFSv16, the version of the GFS that created the analyses used to generate initial conditions for these retrospective SHiELD runs is the GFSv16, making these comparisons completely valid. In order to maximize the sample size, HWRF is excluded from this analysis since some gaps occurred in the availability of this model in these retrospective runs. The IFS forecasts were also excluded due to the lack of availability of the 2021 version of this model in these prior two years. In 2019, a much-improved performance of SHiELD was evident at all forecast lead times, particularly at days 2-4 where the track error was decreased 15% to 23%. In contrast, in the 2020 Atlantic hurricane season, the model track performance of the SHiELD and GFSv16 was comparable between the two models except at 5 days, where the SHiELD track errors were marginally degraded by 5%. The year-to-year variability in model performance is not surprising as the long term synoptic patterns and environmental conditions that often dominate during a given year tend to vary from one season to another (McBride and Zehr 1981; Landsea and Gray 1992; Knaff 1997; Klotzbach 2011). This likely contributes to the stronger model performance for one season compared to another. However, the strong performance of SHiELD in 2019, a mostly neutral performance in 2020 and a strong performance again in 2021, increases our confidence that the SHiELD model is producing superior model TC skill compared to the already strong performing GFSv16.

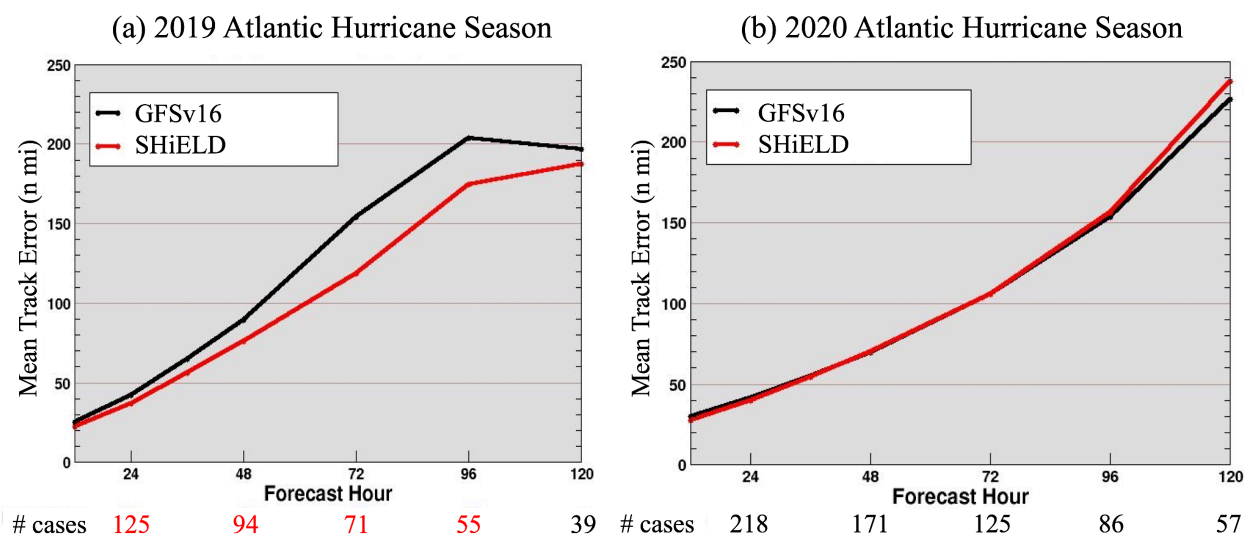

Fig. 12. Mean track forecast error (n mi) for the GFSv16 (black) compared to SHiELD (red) for the (a) 2019 and (b) 2020 Atlantic hurricane seasons. Number of cases are shown at the bottom of each panel, with forecast lead times showing statistically significant improvement at 90% and 95% confidence intervals between SHiELD and the GFSv16, indicated by orange and red colors respectively.

Since it is evident that SHiELD significantly outperforms GFSv16 in TC track forecast beyond day 2 particularly for the 2021 hurricane seasons, in the following section, we will dig into the reason why SHiELD exhibits an outstanding track performance. As noted in Table 1, in the 22 March 2021 upgrade of the GFS to version 16, a new scale-aware turbulent kinetic energy-based moist eddy-diffusivity mass-flux (SA-TKE-EDMF) vertical turbulence mixing scheme was implemented in the GFSv16 to better represent the planetary boundary layer (PBL) processes (Han et al. 2021). Prior to the version 16 upgrade, modifications to the scale-aware simplified Arakawa-Schubert (SA-SAS) convection were also made in 2021 in the GFS, to address issues with a model cold bias (Han et al. 2021). Here, the SA-TKE-EDMF and SA-SAS schemes in GFSv16 are referred to as “new”, while those in SHiELD are referred to as “old”. Evaluation of each of these two separate physics upgrades on the 2021 Atlantic track performance in the SHiELD model was made (Fig. 13) for the 2021 0z cases. Here, experiment s1 is the control (i.e., the version of the convection and PBL schemes used in this study). Note that the upgrades to the PBL scheme had small impact on track (comparing experiment s1 vs. s2) except at day 5 where the number of cases was relatively small. With further upgrade of the convection scheme, the impact on TC track from the combined upgraded convection and PBL schemes (comparing experiment s1 vs. s3) was statistically significant in the shorter lead times (between 5% and 10% at forecast lead times of 2 to 3 days), indicating the new convection scheme degraded the SHiELD track performance (actually the statistical significance exceeded 99% at day 3).

The new convection and PBL schemes did not produce a similar degradation in GFSv16 compared to GFSv15 but rather resulted in small positive improvements in track forecasting performance (Yang 2020). This difference is possibly due to the impact of significantly greater vertical resolution in the GFSv16 compared to SHiELD (Fig. 14). As noted in Fig. 14, the enhancement of vertical resolution in the GFSv16 is large (70% increment below 700 hPa, 14% increment between 700 hPa and 200 hPa, and 41% increment above 200 hPa). These results point to the importance of the vertical model resolution to possibly impact model track, as noted in previous studies (e.g., Feng and Wang 2021; Zhang et al. 2016; Zhang et al. 2015). This also suggests the care that should be given in the tuning of a model and then careful evaluation of a particular modelits performance, when implementing new physics packages, particularly in regard to complex model interactions involving the convection and PBL. Nevertheless, as previously stated, the consistent superior performance of SHiELD is surprising given the significantly enhanced vertical resolution in the GFSv16. It remains to be seen if the improved model skill will even be greater when the vertical resolution in SHiELD is increased further. Although this question remains unanswered these results again point out the need of careful retuning of the convection and PBL schemes in order to optimize the benefit of the enhanced vertical resolution.

Another important difference between the GFSv16 and the SHiELD is the cloud microphysics scheme. Although both models use the GFDL cloud microphysics (Zhou et al. 2019; Harris et al. 2020), GFSv16 uses the split version and SHiELD uses the up-to-date inline version. The differences in these two versions are described in Harris et al. (2020), and are further compared in Zhou and Harris (2022). Since very significant updates have been made to SHiELD’s inline GFDL cloud microphysics scheme since the implementation of GFSv15, and later GFSv16, to pinpoint the major changes that have lead to the significant impact on hurricane track prediction shown in this study is difficult. Further investigation is needed in the future.

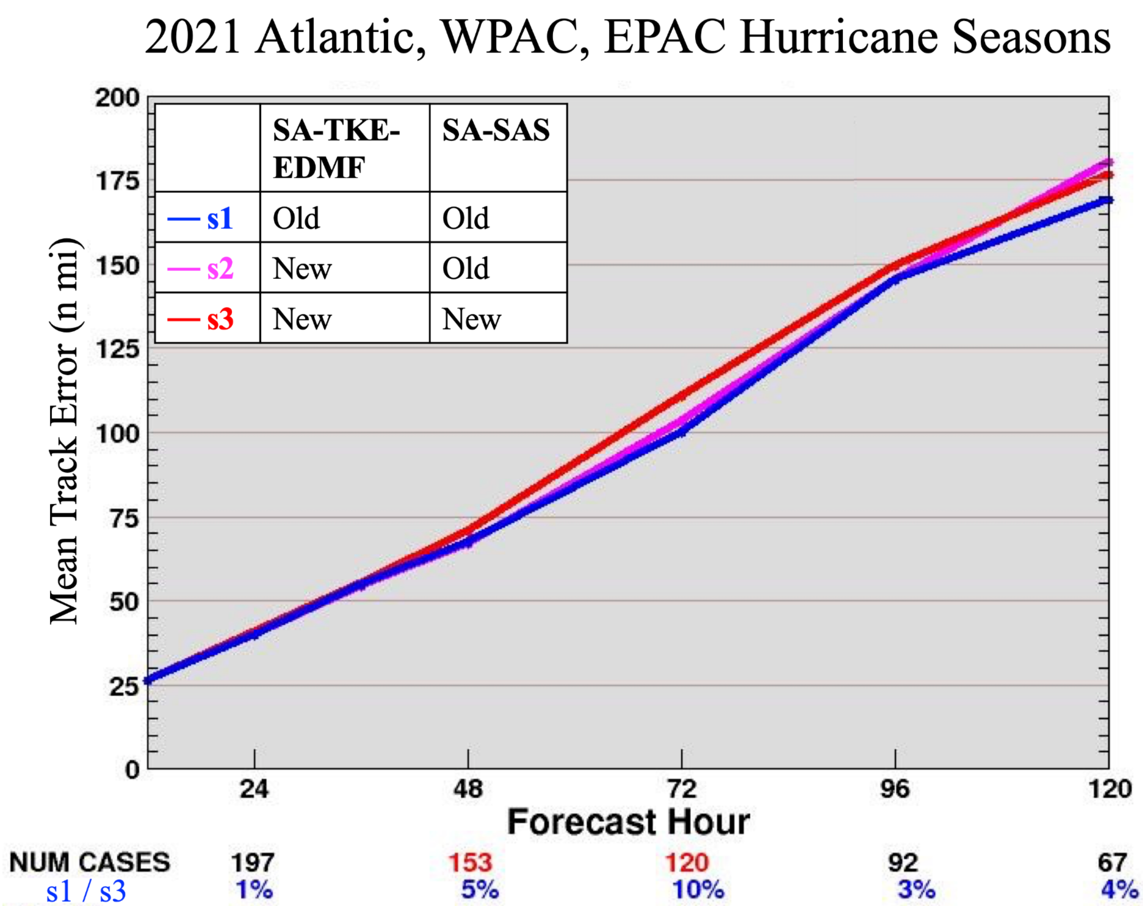

Fig. 13. Mean track forecast error (n mi) for three configurations of SHiELD (s1, s2, and s3) run for the 2021 Atlantic, western Pacific, and eastern Pacific hurricane seasons at 0z synoptic times, using two different versions of the SA-SAS convection and the SA-TKE-EDMF PBL schemes. Experiment s1 employed the same version of physics for the SHiELD model used in this study (i.e., control). In the embedded table, “new” represents the upgraded SAS-TKE-EDMF or SA-SAS scheme in GFSv16; “old” represents the current scheme in SHiELD. Number of cases are shown at the bottom, with forecast lead times showing statistically significant improvement at 90% and 95% confidence intervals between s1 and s3, indicated by orange and red colors respectively. Percentage of improvement in s1 upon s3 are also shown at the bottom.

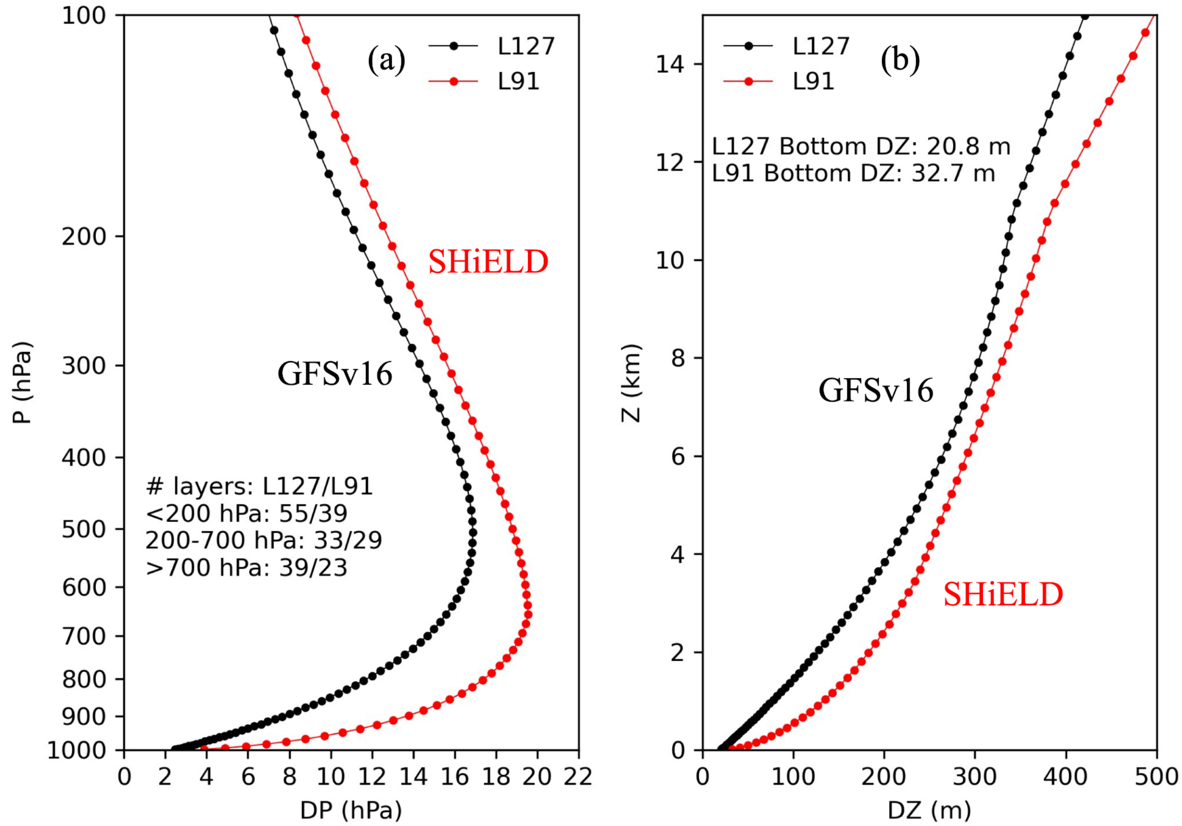

Fig. 14. Comparison of the distribution of model levels between the GFSv16 (black, 127 levels) and SHiELD (red, 91 levels), presented on (a) pressure levels for the surface to 100 hPa and on (b) height levels for the surface to 15 km. Values along the x-axis indicate the stepping increment in between vertical levels.

4. Summary and Discussion

The purpose of this study was to quantify the outstanding track performance during the 2021 Atlantic hurricane season of the SHiELD 13 km global model that is under development at GFDL and based on the FV3 dynamical core that was transitioned into operations by the NWS on 12 June 2019 (i.e., GFSv15) as a replacement for the spectral based GFSv14. On 22 March 2021 the GFSv15 was upgraded by the NWS to GFSv16 with improved model physics and the vertical resolution was doubled from 63 to 127 vertical levels. This upgraded GFSv16 model performed exceptionally well in 2021 and produced smaller TC track errors compared to any operational model in the Atlantic basin. However, despite much coarser vertical model resolution, the GFDL SHiELD model demonstrated superior track forecasting performance in the Atlantic basin compared to the GFSv16, when run from identical initial conditions. This improvement was found to be statistically significant at days 2, 3 and 4. Superior performance compared to GFSv16, which was also statistically significant, was found in the western Pacific basin beyond 48h, where the SHiELD forecast errors were very comparable to the ECMWF's IFS, which was the top performing operational model in that basin in 2021. In the eastern Pacific, where the IFS significantly outperformed all other operational models, SHiELD still performed better than the GFSv16 at 4 and 5 days and was comparable at the earlier forecast lead times. In this study it was shown that similar superior Atlantic track forecast skill compared to the GFSv16 was also seen in 2019 when retrospective forecasts were performed using the GFSv16 initial condition, with mostly neutral impacts in 2020.

The IFS model, which has been a top performing track model in the Atlantic over the previous 5 years, did not perform as well in 2021 relative to other models, with the upgraded GFSv16 the top performing operational track model for track skill in the Atlantic. However, the SHiELD model track skill was shown to be superior to all operational models beyond 48h and was even comparable to the official forecast at days 3 and 4. Analysis of the spatial distribution of the forecast error for the Atlantic showed that the largest errors from both SHiELD and GFSv16 track forecasts occurred in the subtropical eastern Atlantic, associated with a distinct northeast bias that was somewhat reduced in the SHiELD forecasts. The overall smaller spatial forecast error in SHiELD compared to GFSv16 in the Atlantic basin significantly contributed to the better overall TC track forecast performance for the season. Analysis of the three-year spatial distribution of track forecast bias and error showed that this pattern was present to some extent in all three years in the GFSv16 model. This appeared to be partly related to a tendency for premature recurvature of systems into the westerlies. For example, analysis of TS Victor in the 2021 Atlantic hurricane season showed a pronounced northeast bias existed in the GFSv16 although the observed storm did not recurve. A negative 500 hPa geopotential height anomaly persisted in both the GFSv16 and SHiELD in the eastern Atlantic for much of the season, however it was somewhat worse in the GFSv16 during the passage of TS Victor, likely contributing to the excessive north bias as the subtropical ridge weakened too quickly.

Since SHiELD was run with an older version of both the convection and PBL schemes, a subset of cases (0z only) was rerun in SHiELD with the new convection and PBL packages used in GFSv16. Although the impact on hurricane track was minimal with the newer PBL scheme before day 5, the new convection scheme did negatively impact the hurricane track forecasts in the shorter forecast lead times, particularly at days 2 and 3. However, in contrast to a negative impact on track skill in SHiELD the new convection and PBL packages had a positive impact on the GFSv16. It is interesting to note that the impact on the track forecast performance from either physics package was minimal in SHiELD for the case of TS Victor (not shown), so it is unlikely that the new convection and PBL schemes were a contributor to the poor performance of the GFSv16 on this storm compared to SHiELD. However, it is not surprising that the impact of the newer convection and PBL schemes is significantly different between the two models since the number of vertical levels was largely increased in the GFSv16 compared to SHiELD. Since previous studies have shown that increased vertical resolution in NWP models does consistently lead to better model performance, it is indeed likely that the SHiELD track forecasting skill may be further improved with increased vertical resolution if the model physics is properly retuned to the new vertical resolution. This will be soon investigated in future model upgrades.

Consistent with the superior track skill, the SHiELD produced more skillful values for the mean 500 hPa geopotential height anomaly correlation coefficient (ACC), which is one of the most widely used large-scale metrics to evaluate model skill in NWP. As was shown, the improved ACC was statistically significant at forecast lead times beyond 2 days. Thus, based on the overall improved track skill of SHiELD compared to the GFSv16, we have increased confidence that the SHiELD model does provide more reliable TC track prediction at least in the northern hemisphere primarily due to better prediction of the large-scale steering flow. (A robust comparison of track performance in the southern hemisphere is yet to be done). In addition to factors already explicitly mentioned in this study, possible reasons for the improvements involve upgrades GFDL has made to the inline GFDL microphysics (Harris et al. 2020; Zhou et al. 2022) and refinements to the FV3 dynamical core (Harris et al. 2020; Gao et al. 2021). However, due to the complicated impacts and interactions of these changes to other components of the model, it was very difficult to pinpoint precise reasons for the improved overall model skill.

As the GFDL SHiELD development team continues to investigate the impacts of these changes on the SHiELD improved performance, efforts are ongoing to make further improvements to the model such as increased resolution in both the vertical and horizontal, and testing of new model physics. As the model skill continues to improve, it is hoped that some of these model upgrades and refinements could be transitioned into operations. However, the first important step is to quantify that the improved model skill is real and robust on a significantly large and robust sample with identical initial conditions, and in the case of tropical cyclone track prediction, over multiple seasons and forecast basins. This has been clearly established by our results.

Acknowledgments.

The authors would like to thank Timothy Marchok of GFDL and Kun Gao of Princeton University, for their very valuable comments and suggestions on an earlier version of this paper. The authors greatly benefited from numerous discussions with Timothy Marchok, and his valuable assistance in the use of the spatial error and bias software used in this study is appreciated. Special thanks are also given to Vijay Tallapragada and Jongil Han of NCEP's Environmental Modeling Center, for their helpful suggestions and the ongoing collaboration between our 2 respective agencies. The authors thank Catherine Raphael for assisting on the preparation of several of the figures. Lastly the authors are indebted to Shian-Jiann Lin, the former head of the Weather and Climate Dynamics Division at GFDL, and the developer of the FV3 dynamic core, whose inspiring leadership over many years played a key role in the successful development of the SHiLED model.

Data Availability Statement.

All model data produced during this study have been archived locally and are available upon request to the corresponding author.

REFERENCES

Bell, G. D., and Coauthors, 2000: Climate Assessment for 1999. *Bulletin of the American Meteorological Society*, **81,** S1-S50. Doi: https://doi.org/10.1175/1520-0477(2000)81[s1:CAF]2.0.CO;2

Bender, M. A., and I. Ginis, 2000: Real-Case Simulations of Hurricane–Ocean Interaction Using A High-Resolution Coupled Model: Effects on Hurricane Intensity. *Mon Weather Rev*, **128,** 917-946. Doi: https://doi.org/10.1175/1520-0493(2000)128<0917:RCSOHO>2.0.CO;2

Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The Operational GFDL Coupled Hurricane–Ocean Prediction System and a Summary of Its Performance. *Mon Weather Rev*, **135,** 3965-3989. Doi: https://doi.org/10.1175/2007MWR2032.1

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. *Bulletin of the American Meteorological Society*, **100,** 1725-1736. Doi: https://doi.org/10.1175/BAMS-D-18-0197.1

Biswas M. K., S. Abarca, L. Bernardet, I. Ginis, E. Grell, M. Iacono, E. Kalina, B. Liu, Q. Liu, T. Marchok, A. Mehra, K. Newman, J. Sippel, V. Tallapragada, B. Thomas, W. Wang, H. Winterbottom, and Z. Zhang, 2018: Hurricane Weather Research and Forecasting (HWRF) Model: 2018 Scientific Documentation, Available at https://dtcenter.org/HurrWRF/users/docs/index.php

Cangialosi, J. P., 2018: 2017 Hurricane season. National Hurricane Center Forecast Verification Report, 73 pp. Available at https://www.nhc.noaa.gov/verification/pdfs/Verification_2017.pdf

Cangialosi, J. P., 2019: 2018 Hurricane season. National Hurricane Center Forecast Verification Report, 73 pp. Available at https://www.nhc.noaa.gov/verification/pdfs/Verification_2018.pdf

Cangialosi, J. P., 2020: 2019 Hurricane season. National Hurricane Center Forecast Verification Report, 75 pp. Available at https://www.nhc.noaa.gov/verification/pdfs/Verification_2019.pdf

Cangialosi, J. P., 2021: 2020 Hurricane season. National Hurricane Center Forecast Verification Report, 77 pp. Available at https://www.nhc.noaa.gov/verification/pdfs/Verification_2020.pdf

Dunne, J. P., and Coauthors, 2020: The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall Coupled Model Description and Simulation Characteristics. *J Adv Model Earth Sy*, **12,** e2019MS002015. Doi: https://doi.org/10.1029/2019MS002015

Feng, J., and X. Wang, 2021: Impact of Increasing Horizontal and Vertical Resolution during the HWRF Hybrid EnVar Data Assimilation on the Analysis and Prediction of Hurricane Patricia (2015). *Mon Weather Rev*, **149,** 419-441. Doi: https://doi.org/10.1175/MWR-D-20-0144.1

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. *Journal of the Atmospheric Sciences*, **78,** 3007-3021. Doi: https://doi.org/10.1175/jas-d-20-0331.1

Han, J, W. Li, F. Yang, E. Strobach, W. Zheng, and R. Sun, 2021: Updates in the NCEP GFS Cumulus Convection, Vertical Turbulent Mixing, and Surface Layer Physics, Office Note 505, 18 pp. Doi: https://doi.org/10.25923/cybh-w893

Han, J, F. Yang, R. Montuoro, W. Li, and R. Sun, 2022: Implementation of a positive definite mass-flux scheme and a method for removing the negative tracers in the NCEP GFS planetary boundary layer and cumulus convection schemes, Office Note 506, 14 pp. doi: https://doi.org/10.25923/5051-3r70

Harris, L., X. Chen, W. Putman, L. Zhou, and J.-H. Chen, 2021: A Scientific Description of the GFDL Finite-Volume Cubed-Sphere Dynamical Core, NOAA Technical Memorandum OAR GFDL, 2021-001, 109 pp. doi: http://dx.doi.org/10.25923/6nhs-5897

Harris, L., and Coauthors, 2020: GFDL SHiELD: A Unified System for Weather-to-Seasonal Prediction. *J Adv Model Earth Sy*, **12,** e2020MS002223. Doi: https://doi.org/10.1029/2020ms002223

Hazelton, A., and Coauthors, 2022: Performance of 2020 Real-Time Atlantic Hurricane Forecasts from High-Resolution Global-Nested Hurricane Models: HAFS-globalnest and GFDL

T-SHiELD. *Weather and Forecasting*, **37,** 143-161. Doi: https://doi.org/10.1175/WAF-D-21-0102.1

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146,** 1999-2049. Doi: https://doi.org/10.1002/qj.3803

Klotzbach, P. J., 2011: El Niño–Southern Oscillation's Impact on Atlantic Basin Hurricanes and U.S. Landfalls. *Journal of Climate*, **24,** 1252-1263. Doi: https://doi.org/10.1175/2010JCLI3799.1

Knaff, J. A., 1997: Implications of Summertime Sea Level Pressure Anomalies in the Tropical Atlantic Region. *Journal of Climate*, **10,** 789-804. Doi: https://doi.org/10.1175/1520-0442(1997)010<0789:IOSSLP>2.0.CO;2

Landsea, C. W., and J. L. Franklin, 2013: Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format. *Mon Weather Rev*, **141,** 3576-3592. Doi: https://doi.org/10.1175/MWR-D-12-00254.1

Landsea, C. W., and W. M. Gray, 1992: The Strong Association between Western Sahelian Monsoon Rainfall and Intense Atlantic Hurricanes. *Journal of Climate*, **5,** 435-453. Doi: https://doi.org/10.1175/1520-0442(1992)005<0435:TSABWS>2.0.CO;2

Marchok, T., 2021: Important Factors in the Tracking of Tropical Cyclones in Operational Models. *J Appl Meteorol Clim*, **60,** 1265-1284. Doi: https://doi.org/10.1175/jamc-d-20-0175.1

McBride, J. L., and R. Zehr, 1981: Observational Analysis of Tropical Cyclone Formation. Part II: Comparison of Non-Developing versus Developing Systems. *Journal of Atmospheric Sciences*, **38,** 1132-1151. Doi: https://doi.org/10.1175/1520-0469(1981)038<1132:OAOTCF>2.0.CO;2

Neumann, C. J, 1972: An Alternate to the HURRAN (Hurricane Analog) Tropical Cyclone Forecast System. NOAA Technical Memorandum. NWS SR-62, 24 pp. url: https://repository.library.noaa.gov/view/noaa/3605

Pollard, R. T., P. B. Rhines, and R. O. R. Y. Thompson, 2006: The deepening of the wind-Mixed layer. *Geophysical Fluid Dynamics*, **4,** 381-404. Doi: https://doi.org/10.1080/03091927208236105

Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, **227,** 55-78. Doi: 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). *Bulletin of the American Meteorological Society*, **81,** 1231-1240. Doi: https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2

Tallapragada, V., and Coauthors, 2016: Hurricane Weather Research and Forecasting (HWRF) Model: 2015 Scientific Documentation, *No. NCAR/TN-522+STR*, 123 pp. Doi:http://dx.doi.org/10.5065/D6ZP44B5

Yang, F., 2020: Development and Evaluation of NCEP's Global Forecast System Version 16, UFS Webinar. Url: https://ufscommunity.org/wp-content/uploads/2020/10/UFS_Webnair_GFSv16_20201022_FanglinYang.pdf

Zhang, B., and Coauthors, 2016: Increasing vertical resolution in US models to improve track forecasts of Hurricane Joaquin with HWRF as an example. *Proceedings of the National Academy of Sciences*, **113,** 11765-11769. Doi: https://doi.org/10.1073/pnas.1613800113

Zhang, D.-L., L. Zhu, X. Zhang, and V. Tallapragada, 2015: Sensitivity of Idealized Hurricane Intensity and Structures under Varying Background Flows and Initial Vortex Intensities to Different Vertical Resolutions in HWRF. *Mon Weather Rev*, **143,** 914-932. Doi: https://doi.org/10.1175/MWR-D-14-00102.1

Zhou, L., and L. Harris, 2022: Integrated Dynamics-Physics Coupling for Weather to Climate Models: GFDL SHiELD with In-Line Microphysics. To be submitted.

Zhou, L., S.-J. Lin, J.-H. Chen, L. M. Harris, X. Chen, and S. L. Rees, 2019: Toward Convective-Scale Prediction within the Next Generation Global Prediction System. *Bulletin of the American Meteorological Society*, **100,** 1225-1243. Doi: https://doi.org/10.1175/bams-d-17-0246.1

Zhou, L., and Coauthors, 2022: Weather Prediction in SHiELD: Effect from GFDL Cloud Microphysics Scheme Upgrade. *Earth and Space Science Open Archive and submitted to J. Adv. Model. Earth Syst., in revision.* Doi: http://dx.doi.org/10.1002/essoar.10510017.1