Geophysical Research Letters

Advancements in Hurricane Prediction With NOAA's Next-Generation Forecast System

Jan-Huey Chen[1,2] (iD), Shian-Jiann Lin[1] (iD), Linus Magnusson[3], Morris Bender[1,4], Xi Chen[1,4] (iD), Linjiong Zhou[1,4] (iD), Baoqiang Xiang[1,2] (iD), Shannon Rees[1,2], Matthew Morin[1,2], and Lucas Harris[1] (iD)

[1]National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [2]University Corporation for Atmospheric Research, Boulder, CO, USA, [3]European Centre for Medium-Range Weather Forecasts, Reading, UK, [4]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA

Abstract We use the fvGFS model developed at the Geophysical Fluid Dynamics Laboratory to demonstrate the potential of the upcoming United States Next-Generation Global Prediction System for hurricane prediction. The fvGFS retrospective forecasts initialized with the European Centre for Medium-Range Weather Forecasts (ECMWF) data showed much-improved track forecasts for the 2017 Atlantic hurricane season compared to the best-performing ECMWF operational model. The fvGFS greatly improved the ECMWF's poor track forecast for Hurricane Maria (2017). For Hurricane Irma (2017), a well-predicted case by the ECMWF model, the fvGFS produced even lower five-day track forecast errors. The fvGFS also showed better intensity prediction than both the United States and the ECMWF operational models, indicating the robustness of its numerical algorithms.

Plain Language Summary When using European Centre for Medium-Range Weather Forecasts initial conditions, a new global weather model built at NOAA's Geophysical Fluid Dynamics Laboratory produces better hurricane forecast skill than the world-leading European model.

1. Introduction

Tropical cyclone (TC) prediction has long been an important mission for weather forecast agencies to help mitigate hazards along coastlines and inland. In the United States, the Global Forecast System (GFS) operated by the National Centers for Environmental Prediction provides the front-line guidance for TC forecasts. However, it is generally recognized that the European Centre for Medium-Range Weather Forecasts (ECMWF) provides the best hurricane/typhoon track guidance in the world (NOAA/NHC, 2017). Meanwhile, the forecast of TC intensity has remained a challenge for all global models including the GFS and the ECMWF Integrated Forecasting System (IFS; ECMWF, 2016).

In 2019, the Next-Generation Global Prediction System (NGGPS), an initiative of the National Weather Service (NWS), will become operational in the United States. The new GFS will be scientifically updated using the Finite-Volume Cubed-Sphere Dynamical Core (FV3; Lin, 2004; see also Harris & Lin, 2013; Putman & Lin, 2007) developed at the Geophysical Fluid Dynamics Laboratory (GFDL). This modern nonhydrostatic dynamical core demonstrated its accuracy, scalability, and computational efficiency during the NGGPS Phase II Dynamical Core Evaluation (NOAA/NWS, 2016). Benefiting from the high adaptability of FV3, the future GFS model is expected to provide a great opportunity for the unification of weather and climate prediction systems.

A new full physics global model, called fvGFS, was built at GFDL during the NGGPS Phase II, for testing the robustness of the dynamical core under a wide range of realistic atmospheric initial conditions. Different from the new GFS, the GFDL fvGFS is being developed mostly as a research model and includes more advanced dynamical and physical algorithms. Real-time 10-day fvGFS forecasts are run every 6 hr at GFDL and are publically available online (NOAA/GFDL, 2019). A sophisticated regridding tool (Chen et al., 2018) to ingest initial conditions (ICs) from two prominent global prediction systems, the current GFS and the ECMWF's IFS, has been developed to investigate the impact of the ICs on the forecast performance.

Table 1
Comparison Between the Current Operational NCEP/GFS, the GFDL fvGFS (2018 Version), and the ECMWF/IFS

	NCEP/GFS	GFDL fvGFS	ECMWF/IFS
Dynamical core	Hydrostatic spectral model with semiimplicit semi-Lagrangian horizontal discretization and three-dimensional Hermite interpolation[a]	Nonhydrostatic finite-volume cubed-sphere dynamical core (FV3)[b]	Hydrostatic spectral model with semiimplicit semi-Lagrangian horizontal discretization and finite element vertical discretization[f]
Physics suite	GFS[a]	Updated GFS with the GFDL microphysics[c], the YSU PBL[d], and a mixed-layer ocean model[e]	IFS[f]
Model resolution	13 km/64 levels	13 km/91 levels	9 km/137 levels
Initial conditions	GFS	GFS/IFS	IFS

[a]National Oceanic and Atmospheric Administration/National Weather Service/Environmental Modeling Center/The Global Forecast System (GFS)—Global Spectral Model (GSM; version 13.0.2), updated May 2016: http://www.emc.ncep.noaa.gov/GFS/doc.php. [b]Lin (2004). [c]Zhou et al. (2019). [d]Hong et al. (2006). [e]Polland et al. (1973). [f]ECMWF (2016).

In this study, we use the up-to-date version of the GFDL fvGFS model (Chen et al., 2018; Hazelton et al., 2018; Zhou et al., 2019) to investigate the impact of ICs on the 2017 Atlantic hurricane prediction. These results demonstrate the potential of the new FV3-based GFS to improve hurricane prediction which is expected to receive worldwide attention in the near future.

2. Model and Methodology

During the NGGPS Phase II, the GFS physics package provided by National Centers for Environmental Prediction/Environmental Modeling Center was implemented with the GFDL FV3 with a nonhydrostatic solver, called fvGFS. As discussed above, unlike the new GFS developed in the National Centers for Environmental Prediction/Environmental Modeling Center, this model has been continually developed as a research model by the GFDL FV3 team. In the latest (2018) model version, many advanced dynamical and physical algorithms have been tested and are included, that is, the GFDL microphysics scheme (Zhou et al., 2019), the YSU (Yonsei University) PBL scheme (Hong et al., 2006), and a mixed-layer-ocean model (Polland et al., 1973).

In addition to the intrinsic capability of the improved model, the quality of the initial conditions (ICs) also plays a major role in forecast skill for short- to medium-range weather prediction. In the absence of a native data-assimilation cycling system for fvGFS, we have developed a sophisticated regridding tool (Chen et al., 2018) to ingest ICs from two prominent global prediction systems, the current GFS and the ECMWF's IFS, which allows us to investigate the impact of the ICs on the forecast performance.

The three-month period from 1 August to 31 October 2017 was chosen to intercompare the tropical cyclone (TC) forecast skill of the operational GFS, the IFS, and the GFDL fvGFS. The fundamental differences between the three global models are summarized in Table 1. Model-predicted TCs in the twice-daily 10-day forecasts were made for each of the four models: the GFS, the IFS, the fvGFS initialized with GFS ICs (fvGFS_GFSIC), and the fvGFS initialized with IFS ICs (fvGFS_IFSIC). TCs are identified and tracked using the same TC tracking algorithm and criteria (Harris et al., 2016) for each model. Errors in the predicted TC tracks and intensities are computed relative to the Automated Tropical Cyclone Forecast best track data (Miller et al., 1990; Sampson & Schrader, 2000). For each observed TC, only the records after the first "TD (tropical depression)" record in the Automated Tropical Cyclone Forecast data set are used in this study. During the three months, there were 13, 11, and 18 named TCs in the North Atlantic, North East Pacific, and North West Pacific basins, respectively.
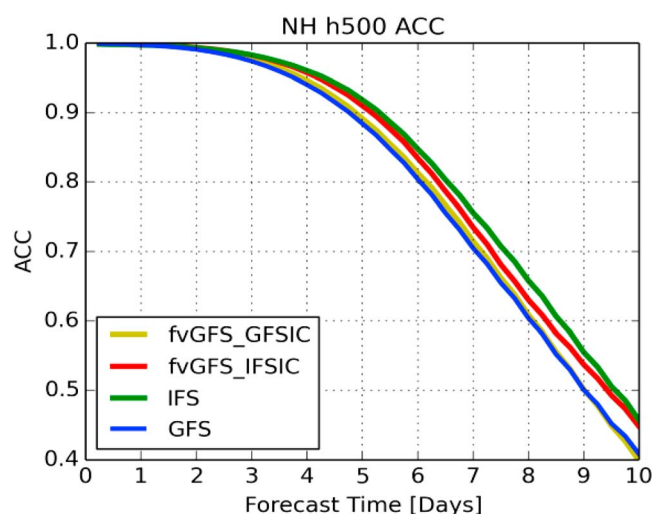

Figure 1. Mean 500-hPa height anomaly correlation coefficients (ACCs) in the Northern Hemisphere. Mean ACCs from a total of 182 10-day forecasts (twice daily from 01 August to 31 October 2017) of the GFS (blue), the IFS (green), the fvGFS using GFS ICs (yellow; fvGFS_GFSIC), and the fvGFS using IFS ICs (red; fvGFS_IFSIC) verified against the NCEP analyses.
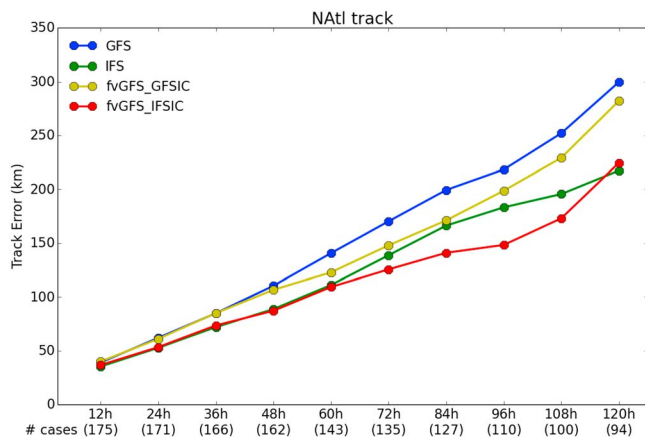
Figure 2. Mean TC track forecast errors (km) in the North Atlantic basin. Track errors as a function of forecast lead time for GFS (blue), IFS (green), fvGFS_GFSIC (yellow), and fvGFS_IFSIC (red). Numbers of homogeneous cases for each lead time are listed in the brackets at the bottom of each abscissa.

3. Results

We first examine the general forecast skill of the four sets of 10-day forecasts by computing the anomaly correlation coefficient (ACC) of the 500-hPa height field, one of the most widely used measures for the verification of global models. The three-month averaged ACCs for the Northern Hemisphere are shown in Figure 1. The results can be grouped into two pairs which are clearly related to the use of different ICs. The two pairs using the same ICs but different models start to diverge by day 4. The group using the IFS ICs (IFS and fvGFS_IFSIC) shows statistically significantly higher scores compared to the group using GFS ICs (GFS and fvGFS_GFSIC) up to day 10 (significant test not shown). The improvement shown by the fvGFS_IFSIC compared to the fvGFS_GFSIC demonstrates the superiority of the IFS ICs. One would ordinarily expect that the use of interpolated initial conditions in fvGFS should yield lower scores than those from the operational models which can take advantage of their native data assimilation cycling. While this is true for fvGFS_IFSIC compared to IFS, the ACC of the fvGFS_GFSIC is close to and at times higher than that of the GFS. We have found that the improvement shown by the fvGFS_GFSIC from days 3 to 8 can be attributed to the use of the modern FV3 dynamical core and the updated GFDL microphysics scheme (Chen et al., 2018).

Figure 2 shows homogeneous comparisons of the mean TC track forecast errors as a function of forecast lead time in the North Atlantic basin. Since the U.S. National Hurricane Center (NHC) issues TC track forecasts out to five days, track errors for only five days are shown. By 48 hr, we again see a striking difference using different ICs: the group using IFS ICs shows a 21% reduction in the track forecast error compared to the group using GFS ICs. This indicates that the quality of the ICs dominates the TC track forecast skill during the first two days, while the impact of using different numerical models is much less evident. After 48 hr, the fvGFS_GFSIC starts to show better track forecasts compared to the GFS, with the largest reduction in track error of 14.2% at 84 hr. For the IFS IC group, the TC track forecast errors are very similar through 60 hr. From 72- to 108-hr lead time, the fvGFS_IFSIC shows smaller TC track forecast errors than the IFS, with the largest error reduction at the 96-hr lead time of 19.1%. That the fvGFS_IFSIC shows better track forecasts than the IFS during days 3–5 is a very encouraging and important result, since this is the critical lead time for decision-makers to initiate plans for reducing the damage and loss of life from hurricanes (Goodwin & Donaho, 2010). Similar results can also be found in the western and eastern North Pacific basins, where the fvGFS_IFSIC has the best TC track forecast performance among the four sets of model predictions at most of the lead times (Figure S1). Based on the 95% significance level tests, the fvGFS_IFSIC shows statistically significantly better track forecasts than the operational GFS after the 48-hr lead time in the North Atlantic basin (not shown). For other regions, at most forecast lead times, none of the forecast shows statistically significantly better or worse results.

TCs are primarily steered by the large-scale flow, the skill of which is partially represented by the 500-hPa ACC scores. While the IFS attained the best ACC score (Figure 1), it did not show the best TC track performance in the first five days (Figure 2). To investigate the cause of this disparity, analyses of the track forecasts of Hurricanes Irma and Maria between the IFS and the fvGFS_IFSIC are examined. Both hurricanes occurred in September 2017 and reached Category 5 intensity. According to the NHC TC reports (NOAA/NHC, 2018a), the IFS made the best track forecasts for Irma compared to other operational model guidance, including both global and regional models, such as the GFS, the Met Office Unified Model, and the Hurricane Weather Research and Forecasting model. In addition, the IFS was the only operational model to consistently outperform the NHC official track forecasts for Irma. On the other hand, the ECMWF exhibited poor track skill for Maria, with the largest track forecast error among all operational models at three- to five-day lead times (NOAA/NHC, 2018b).

The five-day track forecast errors for Irma are shown in Figure 3a. The IFS track forecasts are statistically significantly better than those of the GFS with a 160-km track forecast error difference at 120 hr. Strikingly, using the same ICs, the fvGFS produced better track forecasts for Irma in both IC groups. The
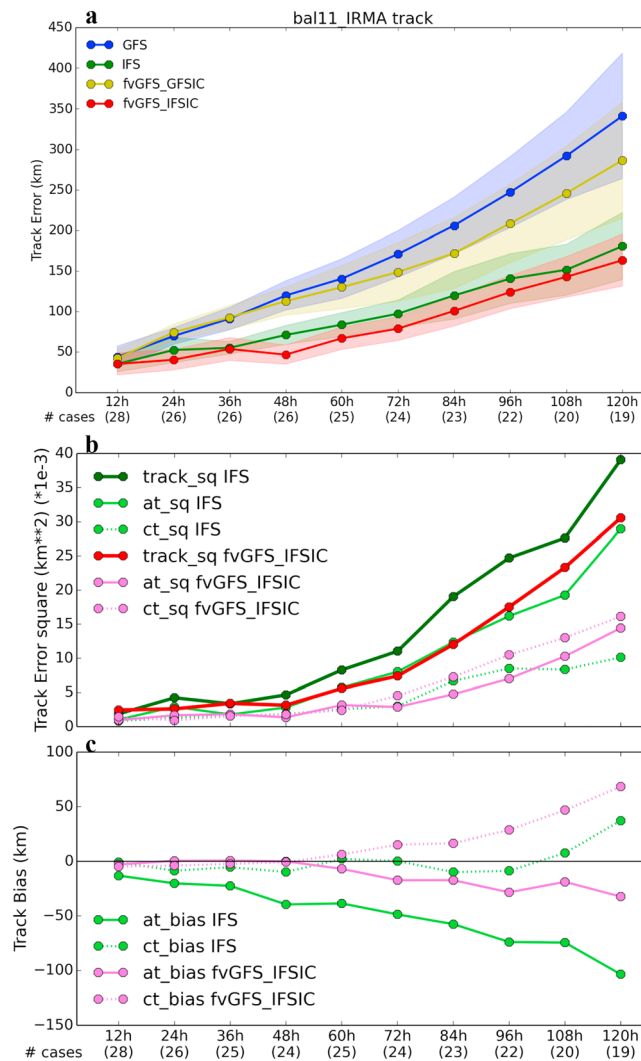
Figure 3. Analyses of track forecast errors for Hurricane Irma (2017). (a) Mean track forecast errors (km) as the function of forecast lead time for the GFS (blue), IFS (green), fvGFS_GFSIC (yellow), and fvGFS_IFSIC (red). The 95% confidence levels for each model are indicated by the same transparent color shading. (b) The squares of total track errors (solid dark green and red lines), along-track errors (solid light green and purple lines) and cross-track errors (dashed light green and purple lines) of the IFS (dark and light green) and the fvGFS_IFSIC (red and purple) forecasts. (c) The biases of along-track (solid lines) and cross-track (dashed lines) errors for the IFS (light green) and the fvGFS_IFSIC (purple) forecasts. Numbers of homogeneous cases for each lead time are listed at the bottom of (a) and (c).

track forecast errors for fvGFS_GFSIC were less than those of the GFS at two- through five-day lead times, with 13% and 16% error reductions on day 3 and day 5, respectively. With the IFS ICs, a consistent improvement by the fvGFS_IFSIC over the IFS is shown after 24 hr. At 48 hr, track forecast errors of the IFS and the fvGFS_IFSIC were 70.8 and 46.5 km, respectively. This represented a 34% reduction in track error achieved by the fvGFS, which was statistically significant at the 95% confidence level.

Track errors can be due to biases either in forecasts of the TC translational speed or the TC direction of motion. To identify the types of errors, the along-track error (ATE) and cross-track error (CTE; perpendicular to the track) are computed. Both ATE and CTE are calculated as great circle distances. By the Pythagorean Theorem the square of the total error equals the squares of the ATE and CTE. In Figure 3b, the squares of total track errors, CTEs, and ATEs for the IFS and the fvGFS_IFSIC forecasts are shown for Irma. The squared CTEs of the two models are very similar. However, the squared ATEs of the IFS are much larger than those of in the fvGFS_IFSIC after 48 hr, indicating that the fvGFS more accurately predicted the
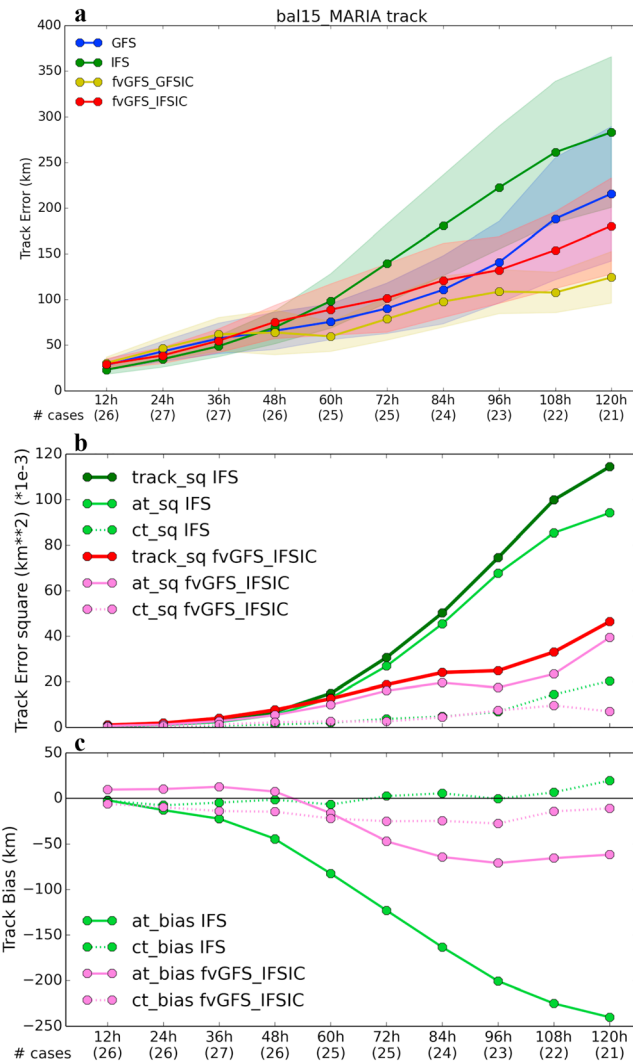
Figure 4. Same as in Figure 3 but for Hurricane Maria (2017).

TC translational speed than the IFS. In Figure 3c, the comparison of ATE biases also indicates that the IFS has more of a slow bias than the fvGFS_IFSIC. The IFS shows smaller biases as well as smaller squared CTEs than the fvGFS after 60 hr. However, the degradation in track error from the slow TC bias dominates the total track errors.

As previously mentioned, the IFS performed poorly for Maria (Figure 4, green line) compared to the GFS (blue line). Before day 2, the differences among the four forecasts are small. However, after 60 hr, the IFS started to show larger track forecast errors than the other three forecasts. When using the same IFS ICs, the track forecast errors of the fvGFS_IFSIC are only slightly larger than those of the GFS at 72 and 84 hr. The errors of the fvGFS_IFSIC do not increase as rapidly as the IFS errors. During the 96–120-hr lead times, the fvGFS_IFSIC shows even smaller errors than the GFS. The reduction of error by the fvGFS_IFSIC compared to the IFS and the GFS are 36–41% and 6–18%, respectively. Furthermore, the fvGFS_GFSIC consistently showed the best forecast performance after 48 hr. By day 5, the fvGFS_GFSIC track forecast error is reduced by 42% compared to the GFS. From Figures 4b and 4c we see that the total track errors in both the IFS and the fvGFS_IFSIC are mainly from the slow track speed bias, while the negative ATEs of the IFS are much larger than that of the fvGFS_IFSIC at all lead times.

Prediction of TC intensity remains a challenge for global models, including the GFS and the IFS. The weaknesses in TC intensity forecasts from these two operational models can be seen from the relationship
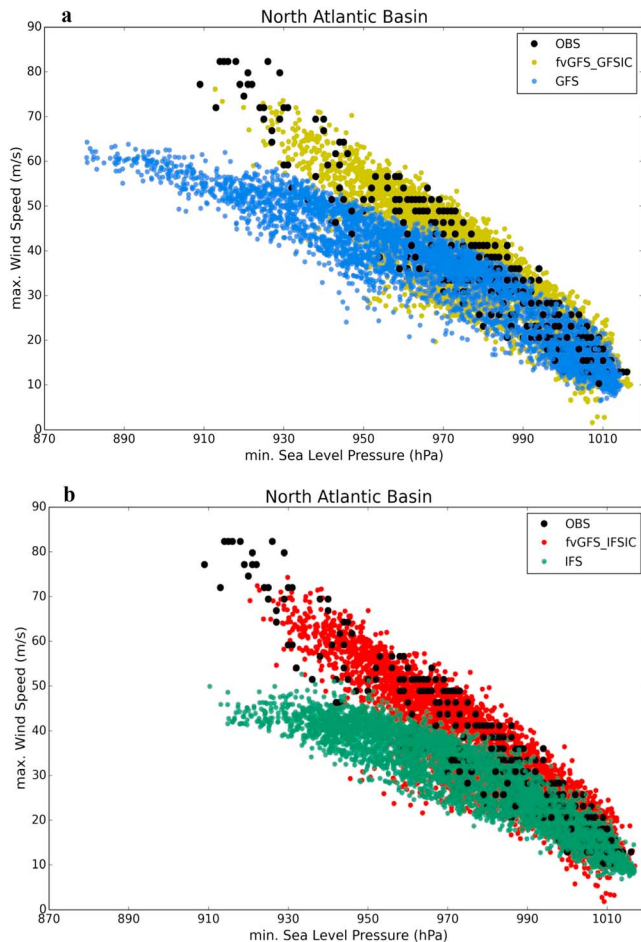
Figure 5. The relationship of maximum 10-m wind (m/s) to minimum sea level pressure (hPa) for TCs in the North Atlantic basin during 2017. The observations at every 6 hr from ATCF best-track data are denoted with black dots. (a) Forecasts of GFS TCs are plotted with blue dots, and fvGFS_GFSIC TCs are denoted with yellow dots. (b) Forecasts of IFS TCs are denoted with green dots, and of fvGFS_IFSIC TCs using red dots.

(Figure 5) between the predicted maximum 10-m wind and the minimum sea level pressure (SLP) of North Atlantic TCs in the four model forecasts compared to observations (black dots). The GFS underpredicts the maximum wind speed yet overpredicts the minimum SLP of TCs (Figure 5a): with numerous TCs exhibiting minimum SLP lower than 910 hPa, but almost none reaching wind speeds of over 65 m/s. The underprediction of TC maximum wind speed is even worse in the IFS (Figure 5b), also shown by Magnusson et al. (2019). However, the wind-pressure relationship is significantly improved by the fvGFS in both the fvGFS_GFSIC and the fvGFS_IFSIC forecasts (Figures 5a and 5b). The fvGFS predicted storm intensities correlate well with observations across almost the full range of observed intensities. Note that the nominal horizontal resolution of the fvGFS is similar to the GFS but coarser than the IFS (Table 1). Homogeneous comparisons of mean TC intensity forecast errors are shown in Figure S3. The two fvGFS forecasts show smaller absolute intensity errors of maximum 10-m wind and minimum SLP compared to either the GFS or IFS forecasts in the North Atlantic basin, as well as in the East and the West Pacific basins, during almost the entire five days. Both Figures 5 and S3 show that the differences in the fvGFS_GFSIC and fvGFS_IFSIC intensity performance are small, indicating that the improvement of TC intensity forecasts and pressure-wind relationship with fvGFS is primarily due to the capabilities of the model itself and not the ICs.

4. Summary and Discussion

Our results, based on TC forecasts from the active 2017 Atlantic hurricane season, demonstrate that both the model (dynamics and physics) and the ICs are important for hurricane prediction. Compared to the current operational GFS, the FV3-powered fvGFS demonstrates improvement to TC track forecasts when using the same ICs. However, when using superior ECMWF ICs, the fvGFS produces not only much improved track forecasts compared to the GFS but also improvement over the IFS. The slow track speed bias of the TCs in the IFS has been present for many years (Haiden et al., 2017). From the analyses for two high-impact hurricanes (Maria and Irma), the fvGFS_IFSIC improved the TC translational speed predicted by the IFS. This encouraging result demonstrates the capability of fvGFS to successfully capture the large-scale steering flow impacting TC motion. We found that the ICs were not the dominant cause for the poor IFS track forecasts of Maria. Moreover, for a successful IFS prediction, like Irma, the track forecasts are further improved using fvGFS by reducing the slow bias in the IFS. The difference in track speed biases between fvGFS and IFS was also found for the rest of the 2017 North Atlantic season (Figure S2).

To improve the TC track forecasts of the GFS, the quality of the ICs and the capability of the numerical model are both critical. This can be demonstrated by comparing the forecasts of fvGFS_GFSIC and fvGFS_IFSIC for these two hurricanes. For Irma, using the same fvGFS model, the track forecast with IFS ICs is significantly better than with the GFS ICs after the 24-hr lead time (Figure 3a), indicating the deficiency of the GFS ICs. For Maria, the best track forecast is achieved by fvGFS_GFSIC, which could imply that the quality of the GFS ICs was better than that of the IFS ICs for this particular case (Figure 4a). However, in addition to the large improvements shown by the fvGFS_GFSIC compared to the GFS, the fvGFS_IFSIC still showed better track forecasts compared to the operational GFS. This suggests that the upcoming upgrade of the operational GFS with the new FV3-based GFS should yield promising track forecast improvements, especially when accompanied by improvements in the physical parameterizations. This is also suggested by Figure 2, comparing the GFS and fvGFS_GFSIC as well as the IFS and fvGFS_IFSIC track performance for this larger sample size.

To develop the world's best global model for hurricane prediction, our results suggest that the United States also needs to invest in improving the data assimilation system to produce higher-quality ICs for model initialization. As for hurricane intensity prediction, which has been a challenge for both the GFS and the IFS, the fvGFS also demonstrates promising improvement. Therefore, we believe the fvGFS could provide important and significant improvements to hurricane predictions with a coherently designed native data assimilation system that can produce comparable ICs to the ECMWF.

References

Chen, J.-H., Chen, X., Lin, S.-J., Magnusson, L., Bender, M., Zhou, L., & Rees, S. (2018). Tropical cyclones in GFDL fvGFS—Impacts of dycore, physics and initial conditions. 33rd Conference on Hurricane and Tropical Meteorology, Ponte Vedra, FL, Amer. Meteor. Soc., 9B.4.

ECMWF (2016). IFS Documentation CY41R2. Technical report, ECMWF. Retrieved from https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation

Goodwin, B. S. Jr., & Donaho, J. C. (2010). Tropical storm and hurricane recovery and preparedness strategies. *ILAR Journal*, *51*, 104–119.

Haiden, T., Janousek, M., Bidlot, J., Ferranti, L., Prates, F., Vitart, F., et al. (2017). Evaluation of ECMWF forecasts, including 2016–2017 upgrades, ECMWF Tech Memo, https://www.ecmwf.int/en/elibrary/17913-evaluation-ecmwf-forecasts-including-2016-2017-upgrades

Harris, L. M., & Lin, S.-J. (2013). A two-way nested global-regional dynamical core on the cubed-sphere grid. *Monthly Weather Review*, *141*(1), 283–306. https://doi.org/10.1175/MWR-D-11-00201.1

Harris, L. M., Lin, S.-J., & Tu, C. (2016). High-resolution climate simulations using GFDL HiRAM with a stretched global grid. *Journal of Climate*, *29*, 4293–4314. https://doi.org/10.1175/JCLI-D-15-0389.1

Hazelton, A. T., Bender, M. A., Morin, M. J., Harris, L., & Lin, S.-J. (2018). 2017 Atlantic hurricane forecasts from a high-resolution version of the GFDL fvGFS model: Evaluation of track, intensity, and structure. *Weather Forecasting*, *33*, 1317–1337. https://doi.org/10.1175/WAF-D-18-0056.1

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*, 2318–2341. https://doi.org/10.1175/MWR3199.1

Lin, S.-J. (2004). A "vertically Lagrangian" finite-volume dynamical core for global models. *Monthly Weather Review*, *132*, 2293–2307.

Magnusson, L., Bidlot, J.-R., Bonavita, M., Brown, A., Browne, P., De Chiara, G., et al. (2019). ECMWF activities for improved hurricane forecasts. *Bulletin of the American Meteorological Society*, *100*, 445–458. https://doi.org/10.1175/BAMS-D-18-0044.1

Miller, R. J., Schrader, A. J., Sampson, C. R., & Tsui, T. L. (1990). The Automated Tropical Cyclone Forecast System (ATCF). *Weather Forecasting*, *5*, 653–660.

NOAA/GFDL (2019). Real-time-10-day fvGFS forecasts. Retrieved from https://data1.gfdl.noaa.gov/fvGFS/fvGFS_products.php

NOAA/NHC (2017). National Hurricane Center Forecast Verification. Retrieved from https://www.nhc.noaa.gov/verification/verify6.shtml

NOAA/NHC (2018a). National Hurricane Center Tropical Cyclone Report—Hurricane Irma (AL112017). Retrieved from https://www.nhc.noaa.gov/data/tcr/AL112017_Irma.pdf

NOAA/NHC (2018b). National Hurricane Center Tropical Cyclone Report—Hurricane Maria (AL152017). Retrieved from https://www.nhc.noaa.gov/data/tcr/AL152017_Maria.pdf

NOAA/NWS (2016). The verification report for the NGGPS Phase II Dynamical Core Evaluation. Retrieved from https://www.weather.gov/sti/stimodeling_nggps_implementation_atmdynamics

Polland, R. T., Rhines, P. B., & Thompson, R. O. R. Y. (1973). The deepening of the wind-mixed layer. *Geophysical Fluid Dynamics*, *3*, 381–404.

Putman, W. M., & Lin, S.-J. (2007). Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, *227*, 55–78. https://doi.org/10.1016/j.jcp.2007.07.022

Sampson, C. R., & Schrader, A. J. (2000). The Automated Tropical Cyclone Forecasting System (version 3.2). *Bulletin of the American Meteorological Society*, *81*, 1231–1240.

Zhou, L., Lin, S.-J., Chen, J.-H., Harris, L., Chen, X., & Rees, S. L. (2019). Toward Convective-Scale Prediction Within the Next Generation Global Prediction System. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-17-0246.1