GFDL SHiELD: A Unified System for Weather-to-Seasonal Prediction

Lucas Harris[1], Linjiong Zhou[1,2], Shian-Jiann Lin[1], Jan-Huey Chen[1,3], Xi Chen[1,2], Kun Gao[1,2], Matthew Morin[1,3], Shannon Rees[1,3], Yongqiang Sun[1,2], Mingjing Tong[1,4], Baoqiang Xiang[1,3], Morris Bender[1,2], Rusty Benson[1], Kai-Yuan Cheng[1,2], Spencer Clark[1,5], Oliver Elbert[1,5], Andrew Hazelton[1,2*], J. Jacob Huff[1,3], Alex Kaltenbaugh[1,3], Zhi Liang[1], Timothy Marchok[1], Hyeyum Hailey Shin[1,3], and William Stern[1]

[1]NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey.

[2]Cooperative Institute for Modeling the Earth System, Program in Oceanic and Atmospheric Sciences, Princeton University, Princeton, New Jersey.

[3]University Corporation for Atmospheric Research, Boulder, Colorado.

[4]SAIC, Princeton, New Jersey.

[5]Vulcan, Inc., Seattle, Washington.

Corresponding author: Lucas Harris (lucas.harris@noaa.gov)

*Current affiliation: NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida.

Key Points:

- A unified "one code, one executable, one workflow" global prediction modeling system is presented.
- SHiELD's multiple configurations show prediction skill and simulation fidelity matching or exceeding those of existing US models.
- The FV3 Dynamical Core provides a powerful foundation for unified prediction modeling.

1 Unified Modeling at GFDL

As computing power increases global atmosphere models are now capable of regular simulation at resolutions that had been the sole domain of regional atmospheric models. The Integrated Forecast System (IFS; ECMWF 2019a,b) of the European Center for Medium-Range Weather Forecasting (ECMWF) runs on a 9-km grid, and the Global Forecast System (GFS; Sela2010) of the US National Centers for Environmental Prediction (NCEP) runs on a 13-km grid. Some IPCC-class climate models now use grids with spacings as fine as 25 km (Chen and Lin 2013; Vecchi et al. 2019; Haarsma et al. 2017). Global atmosphere models lack the lateral boundary errors that contaminate the solutions of regional models after a few days of simulation. They thus allow us to extend mesoscale and storm-scale predictions into the medium range and beyond (Harris and Lin 2013, 2014; Zhou et al. 2019; Harris et al. 2019). Global modeling also brings many new challenges—one cannot “throw your garbage in the neighbor’s yard” in global modeling, so to speak. Biases and radiative imbalances must be minimized, as must errors *anywhere* in the atmosphere that could potentially grow and contaminate the entire domain.

A unified modeling system supports a variety of applications at a wide range of spatial and temporal scales within a single framework. These systems promise to simplify operational and research modeling suites and better exchange improvements and bug fixes between applications. The Unified Model of the United Kingdom Met Office (UKMO; Brown et al. 2012) is the most notable unified system. Variable-resolution models (Harris and Lin 2014, McGregor 2015) are particularly well-suited for unified modeling as they can efficiently reach very high resolutions over part of the earth, replacing the highest-resolution regional models (Hazelton et al. 2018a,b, Zhou et al. 2019) and potentially extending their lead times.

Here at GFDL a hierarchy of models has been developed for a variety of time and space scales, from centennial-scale earth-system simulations (Dunne et al. 2020) to very high-resolution weather prediction. The GFDL suite is unified around a single dynamical core, the GFDL Finite-Volume Cubed-Sphere Dynamical Core (FV3, or FV^3; Putman and Lin 2007), and a single framework, the Flexible Modeling System (FMS; Balaji 2012), and other shared components. We describe one part of this suite, the System for High Resolution Prediction on Earth-to-Local Domains, or SHiELD. This model, previously called fvGFS, was developed as a prototype of the Next-Generation Global Prediction System (NGGPS) of the National Weather Service, and of the broader Unified Forecast System (UFS). SHiELD continues GFDL’s high-resolution global modeling program previously established using the High-Resolution Atmosphere Model (HiRAM; Zhao et al. 2009; Chen and Lin 2013). SHiELD couples the nonhydrostatic FV3 dynamical core (Lin et al. 2017) to a physics suite originally from the GFS (Han et al. 2017, and references therein) and the Noah Land Surface Model (Ek et al. 2002). SHiELD can be used for a variety of timescales but has been designed with a particular focus on short-to-medium range weather (18 hours to 10 days) and into the subseasonal to seasonal (S2S; several weeks to several months) range. Seasonal to decadal predictions and centennial-scale climate projections coupled to a dynamical ocean are performed at GFDL using the Seamless System for Prediction and Earth System Research (SPEAR, Delworth et al. 2020), the Coupled Model version 4 (CM4; Held et al. 2020), and the Earth System Model version 4 (ESM4, Dunne et al. 2020).

cycled tracer advection and vertical remapping (cf. Lin 2004) are performed on an intermediate "remapping" timestep, in turn performed multiple times per physics timestep.

FV3's discretization along Lagrangian surfaces uses the piecewise-parabolic method, which previously used a monotonicity constraint to ensure positivity and to dissipate energy cascading to grid scale. In nonhydrostatic FV3 dynamical quantities (vorticity, potential temperature, and air mass) are advected by a non-monotonic scheme to reduce dissipation of resolved-scale modes. Previous work with nonhydrostatic FV3 had continued to use a monotonic advection scheme to avoid unphysical negative values. In this manuscript we present results using a new *positive-definite* but non-monotonic scheme to advect tracers, which greatly improves the representation of marginally-resolved and discontinuous features without creating computational noise at sharp gradients. This scheme is described in detail in Appendix A and applications to the representation of tropical cyclones in section 3d.

2.2 GFS/SHiELD Physics and Noah LSM

SHiELD inherits the GFS suite of physical parameterizations developed by the Environmental Modeling Center (EMC) of NCEP (2020). The initial 2016 version of SHiELD, implemented for dynamical core testing during Phase II of NGGPS, used physics largely identical to the then-operational GFSv13: The Simplified Arakawa-Schubert (SAS) shallow and deep convection schemes described in Han and Pan (2011); the hybrid Eddy-diffusivity Mass-flux (EDMF) scheme (Han et al. 2016); the Rapid Radiative Transfer Model (RRTM; Clough et al. 2005); the microphysics of Zhao and Carr (1997) and cloud-fraction scheme of Xu and Randall (1996); the Navy's simplified ozone scheme (McCormack et al. 2006); the GFS orographic gravity wave drag and mountain blocking schemes (Alpert 2002); and the convective gravity wave drag scheme of Chun and Baik (1998).

We have since made many changes to the physics to be able to support new applications, especially for convective scale prediction and marine phenomena, or to take advantage of new capabilities within the FV3 dynamical core. We first introduced the six-category GFDL microphysics and cloud fraction scheme (Zhou et al. 2019) with the fast microphysical processes split out of the physics driver and taking place on the shorter remapping timestep. Later, the GFDL microphysics was fully in-lined within FV3 (appendix B). Several new PBL schemes have also been used in SHiELD, including a modified hybrid EDMF PBL as per Zhang et al. (2015), and the Yonsei University scheme (YSU; Hong et al. 2006, Hong 2010, Wilson and Fovell 2018). We have also adopted the Scale-Aware SAS (Han et al. 2017) convection scheme in more recent versions of SHiELD.

The land surface model (LSM) is the Noah land-surface model (Ek et al. 2003), integrated within the physics and paired to the GFS surface-layer scheme. In 2017 Noah was upgraded to use the high-resolution land surface data (Wei et al. 2017), which greatly improves the appearance of land-surface fields in convective-scale simulations.

2.3 Mixed-layer Ocean

Initially sea-surface temperatures (SSTs) were prescribed as the climatological SST plus an SST anomaly from initial conditions which gradually decays to zero, without influence from the atmosphere. However, air-sea interactions are critical for several phenomena of interest to us, especially tropical cyclones and the Madden-Julian Oscillation (MJO) and may impact large-scale skill as well. To incorporate atmosphere-ocean interaction, we have implemented a configurations are usually refreshed every year with a new version, indicated by the year of the upgrade.

Our Tier-2 configurations address new challenges for numerical prediction and are still under development. Our 25-km (Subseasonal) S-SHiELD addresses the challenging domain of S2S prediction. Another configuration not discussed in this paper is the SHiELD global cloud-resolving model (GCRM) and addresses the frontier computational and data challenges of such simulations. This configuration was submitted to the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains (DYAMOND) intercomparison (Stevens et al. 2019, Satoh et al. 2019). Both configurations inspire the development of new functionality and capabilities within SHiELD and readily expose instabilities, climate drift, conservation issues, and other shortcomings. The advances driven by work on these frontier challenges help improve the Tier-1 configurations, demonstrating the value of a seamless prediction system. The domains for each of the four configurations plus the GCRM configuration are depicted schematically in Figure 1.

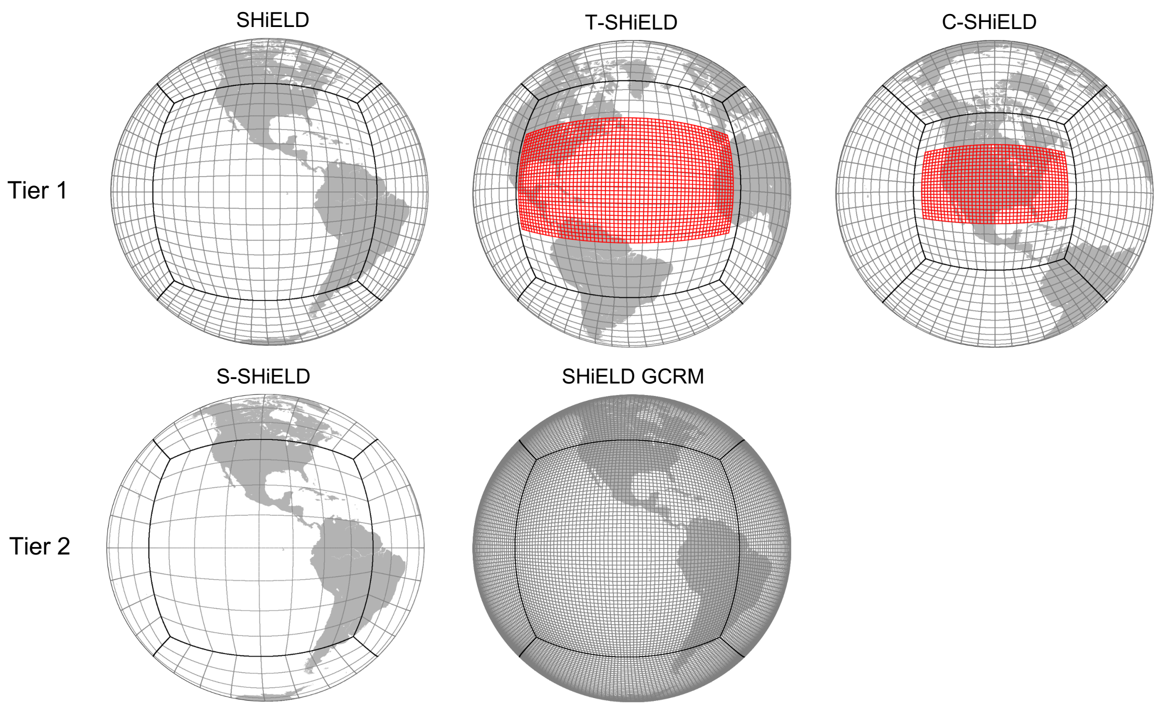

Figure 1. Current SHiELD configurations. Each plotted cell is 48x48 actual grid cells. Heavy black lines represent cubed-sphere edges; red lines represent nested grids. Note that the global domain of C-SHiELD (top center) is slightly stretched as per Harris et al. (2019).

Although all configurations follow the unified “one code, one executable, one workflow” structure of SHiELD, the configurations are not identical owing to the need to tailor each configuration for its specific application. Further, given the rapid pace of SHiELD development and the staggered development cycle for some of the configurations, we do not expect all of the Tier-1 configurations to always have the very latest developments. The development paths of the different SHiELD configurations can be seen in Table 1.

expensive owing to its nested grid but still completes a five-day simulation in under two hours on less than 3500 cores. T-SHiELD has a nested grid with twice as many columns as C-SHiELD but is only about 30% more expensive.

SHiELD is compiled with mixed-precision arithmetic: the dynamics (and the inlined components of the microphysics) use single-precision arithmetic while the physics uses double-precision. This differs from the practice used for most operational models (GFSv15 excluded) and for GFDL climate models, which use double-precision arithmetic throughout. Tests with the 2016 version of SHiELD had found no detectable difference in skill between predictions using mixed-precision and double-precision arithmetic, while leading to a cost reduction of about 40%.

3.1 SHiELD Medium-Range Weather Prediction

The flagship SHiELD configuration is designed for medium-range prediction with lead times of 24 hours to ten days. The design of SHiELD is similar to the operational GFS: a global c768 grid—a cubed-sphere with each face having 768 x 768 grid cells—with an average grid-cell width of about 13-km. The 2016 and 2017 versions of SHiELD used 63 vertical levels (Figure 2), the same as the hydrostatic GFSv14 but with the uppermost semi-infinite layer removed to permit nonhydrostatic simulation. SHiELD 2017 was then developed by NCEP and partners to become GFSv15 and its GFS Data Assimilation System (GDAS): specific implementation details can be seen at https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/implementations.php. Starting in 2018, SHiELD increased the number of vertical levels to 91, increasing the number of vertical levels below 700 mb from 19 to 23 and decreasing the depth of the lowest model layer from 45 to 33 m.

an ACC of 0.6 at 8.3–8.5 days globally and 8.5–8.7 days in the northern hemisphere, with some year-to-year and version-to-version variability.

The time series of day-5 global ACC and RMSE (Figure 4) shows that while there is a general secular improvement in both SHiELD and the GFS, there can be large seasonal and even interannual variability in forecast skill. Usually, predictions are more skillful in northern winter, as strong synoptic forcing dominates the large-scale weather patterns, but some northern summers see little to no forecast degradation. The implementation of GFSv13 on 11 May 2016, which included a major upgrade to the data assimilation cycling system of the GFS, significantly reduced RMSE in May and June 2016 compared to the preceding four months of the year. These results are worthy of further investigation. We do conclude that it may be misleading to use a short time period to evaluate or compare global prediction models.

The time-evolution of the large-scale forecast skill for both the GFS and SHiELD are very similar on monthly and shorter time-periods, which is expected as they use identical initial conditions, and SHiELD benefits from continual upgrades of the GFS initial conditions. As discussed in Chen et al. (2019b) the quality of the initial conditions is the preeminent factor in determining the forecast skill for the large-scale circulation as well as for metrics such as hurricane track forecasts that depend closely on the prediction skill of the large-scale flow.

These results are for hindcasts but the ACC and RMSE for our real-time forecasts are nearly identical. An important caveat is that the operational GFS supports nearly the entire NCEP modeling suite, and so the GFS has many more demands and a much more stringent evaluation process imposed upon its development than does SHiELD. The development cycle of the GFS will therefore necessarily be less rapid and more methodological than that of SHiELD. Alternately, an experimental research model like SHiELD does have the freedom to pursue many different avenues for model development ("failure is always an option") so that the most successful new ideas can later be transitioned into operations, a major goal of the UFS.

00Z every five days, for a total of 144 cases per version. See Table 1 for the time periods being compared here.

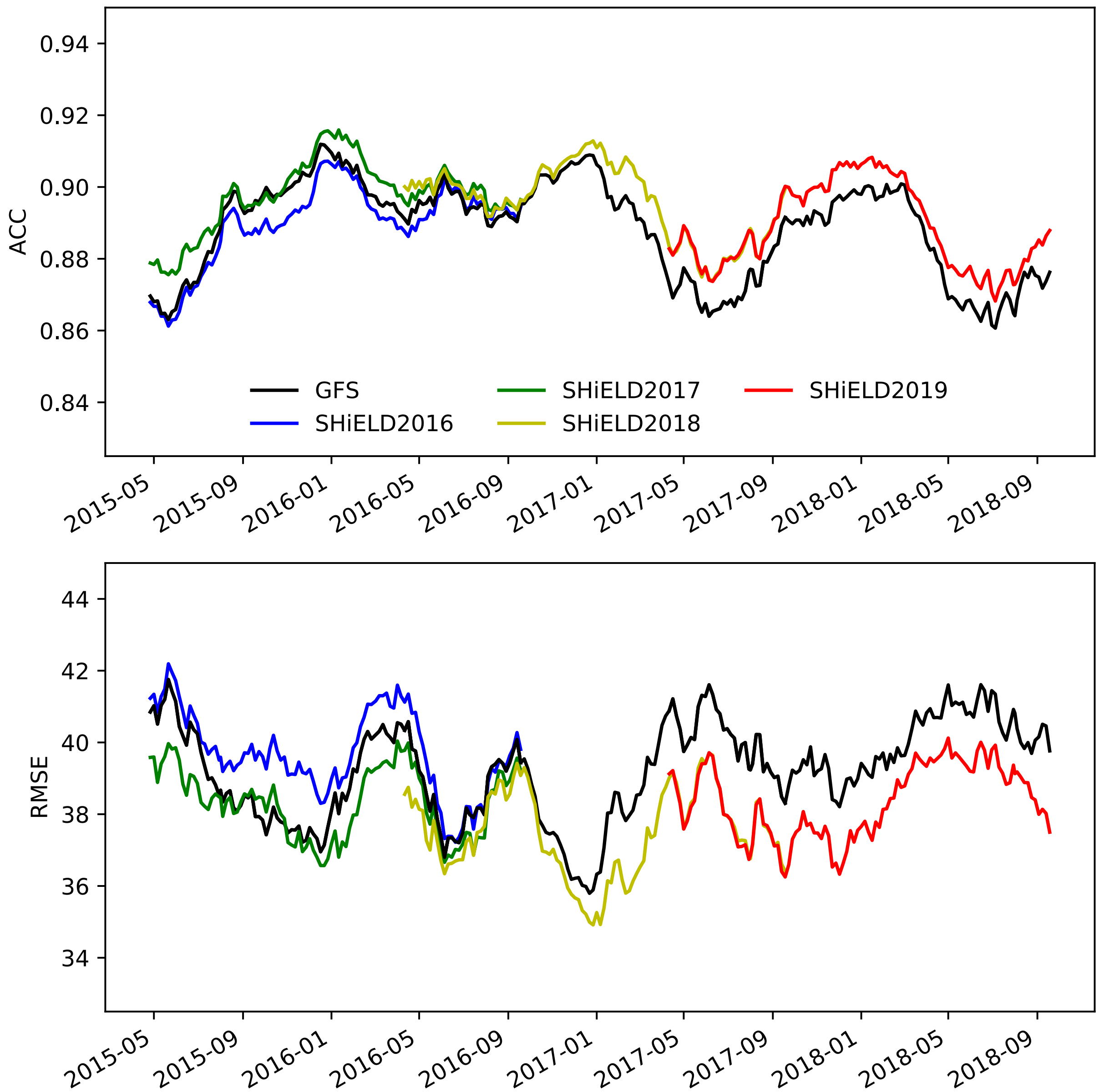

Figure 4. Six-month running-mean time series of global 500-mb geopotential height ACC (top) and RMSE (bottom, m) at day 5 for each version of the 13-km SHiELD and the contemporary operational GFS. Note that the operational GFS upgraded to v13 on 11 May 2016 and v14 on 19 July 2017.

Precipitation RMSE and biases have also improved during SHiELD development. The 2018 version significantly reduced both RMSE (Figure 5) and Bias (arithmetic difference between time-mean model and observed precipitation; Figure 6) at all lead times compared to earlier versions. Prediction of CONUS precipitation is more challenging given the smaller area dataset (regridded to 25 km); bottom row: CONUS verification vs. StageIV dataset (regridded to 13 km). Gray shading is the 95% confidence interval.

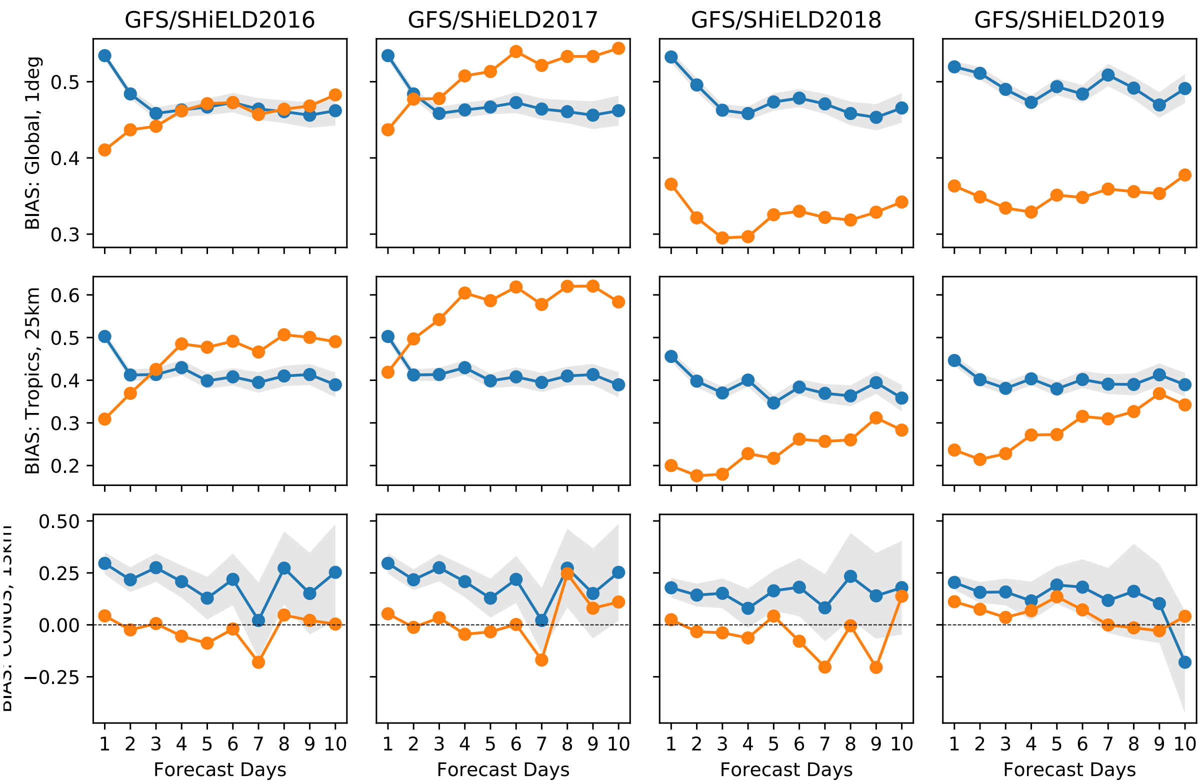

Figure 6. As in Figure 5 but for precipitation bias (mm d^{-1}), the arithmetic difference between means from the model and observations. Negative values imply too little mean precipitation.

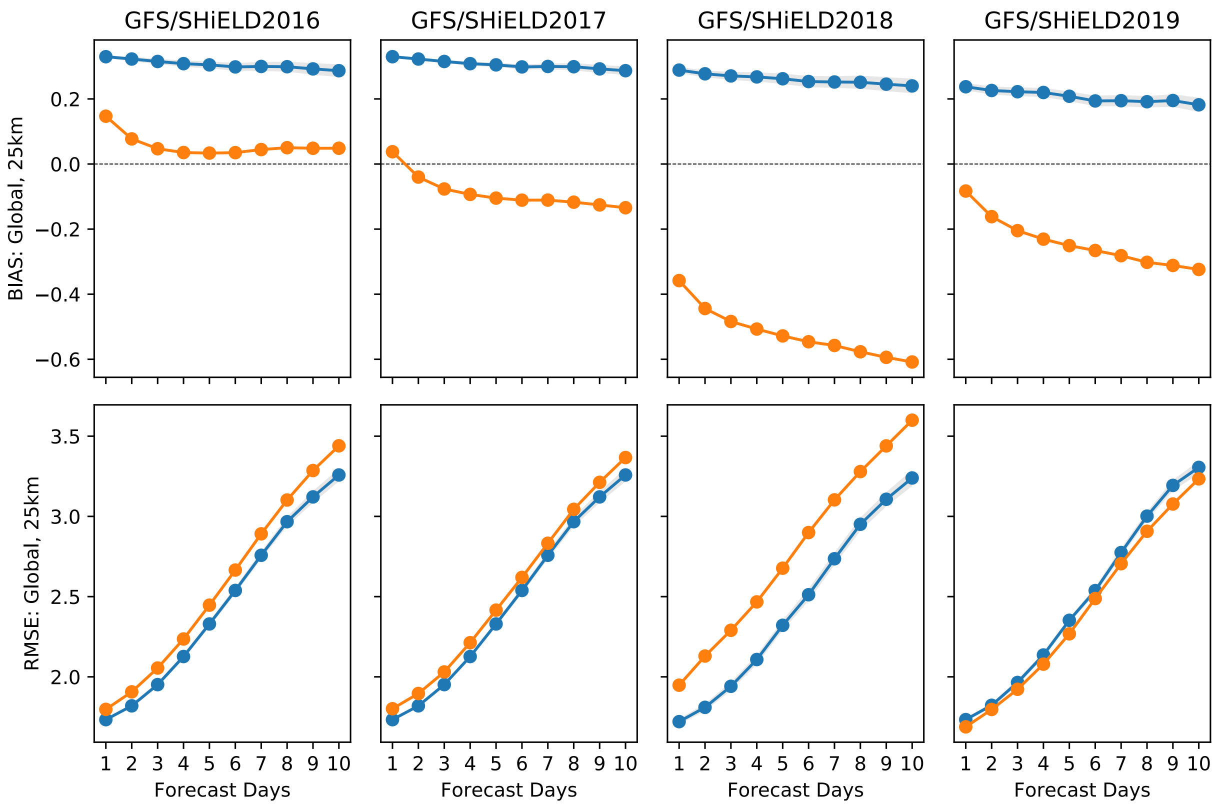

Figure 8. Global 2-m temperature (deg K) bias (top) and RMSE (bottom) for 13-km SHiELD (orange) compared to contemporary GFS (blue), both validated against ERA5 Reanalysis (Hersbach et al. 2020).

3.2 T-SHiELD North Atlantic Nest for Tropical Cyclone Prediction

T-SHiELD uses the variable-resolution capabilities of FV3 to replicate the tropical cyclone track skill of global models and the intensity skill of convective-scale regional hurricane models. This configuration uses the 13-km SHiELD grid and then places a large factor-of-four two-way nest over the tropical North Atlantic (Figure 1). The resulting nested domain has grid cells of about 3-km width and interacts with its parent global domain. Earlier experiments and a comprehensive evaluation of T-SHiELD 2017 were described in Hazelton et al. (2018a, 2018b). T-SHiELD has been used as the initial prototype for the Hurricane Analysis and Forecast System (HAFS; Hazelton et al., 2020). Here we will describe further evolution of T-SHiELD, including progress towards rectifying two forecast issues in T-SHiELD 2017: an under-intensification bias for rapidly intensifying storms, and storms with a radius of maximum winds (RMW) that is too large. Note that there is no 2019 version of T-SHiELD.

Hazelton et al. (2018b) found that the RMW in T-SHiELD 2017 was often larger than observed and in particular larger than that in HWRF simulations from the same set of cases. Zhang et al. (2015) found that reducing the parameterized mixing in the PBL scheme reduced the size of the RMW in HWRF. While reducing the parameterized mixing in the hybrid EDMF scheme gave modest improvement to hurricane structure in T-SHiELD, there was no appreciable reduction in the size of the eyewall. A dramatic and immediate impact was instead found by prototype of T-SHiELD 2018 with the monotonic (CTRL) and positive-definite tracer advection schemes (PD). The RMW is denoted as a dashed black line. Note that a localized extremum (left panels) may not be visible in the azimuthal averages (right panel), especially during rapid intensification.

A more systematic comparison of wind radii between the 2017 and 2018 T-SHiELD versions (Figure 10, d) shows that the effect of the PD scheme is not limited to a single storm. Noting that the difference between the two T-SHiELD versions is more than just the PD scheme, we do see a systematic and substantial decrease in the radius of the 64-kt (33 m s^{-1}, hurricane force) winds in the 2018 version. The 2018 version spins up the vortex such that within 36 hours of initialization, the 64-kt radii reduce to and then remain a consistent 20-25 nautical miles (nm; 37–46 km) for the rest of the forecast period. This represents a reduction of more than half at 120-h lead time compared to the 2017 version, which steadily widens the 64-kt radii during the simulation. There is also a reduction in radii forecast errors compared to Best Track estimates in T-SHiELD 2018, with the qualification that there is considerable (potentially 40% for 64-kt: Landsea and Franklin 2013) uncertainty in estimates of wind radii. This uncertainty can impact the initialization of tropical cyclones using real-time storm message files (Bender et al. 2017), and thereby of estimates of size-related impacts like precipitation and extreme winds.

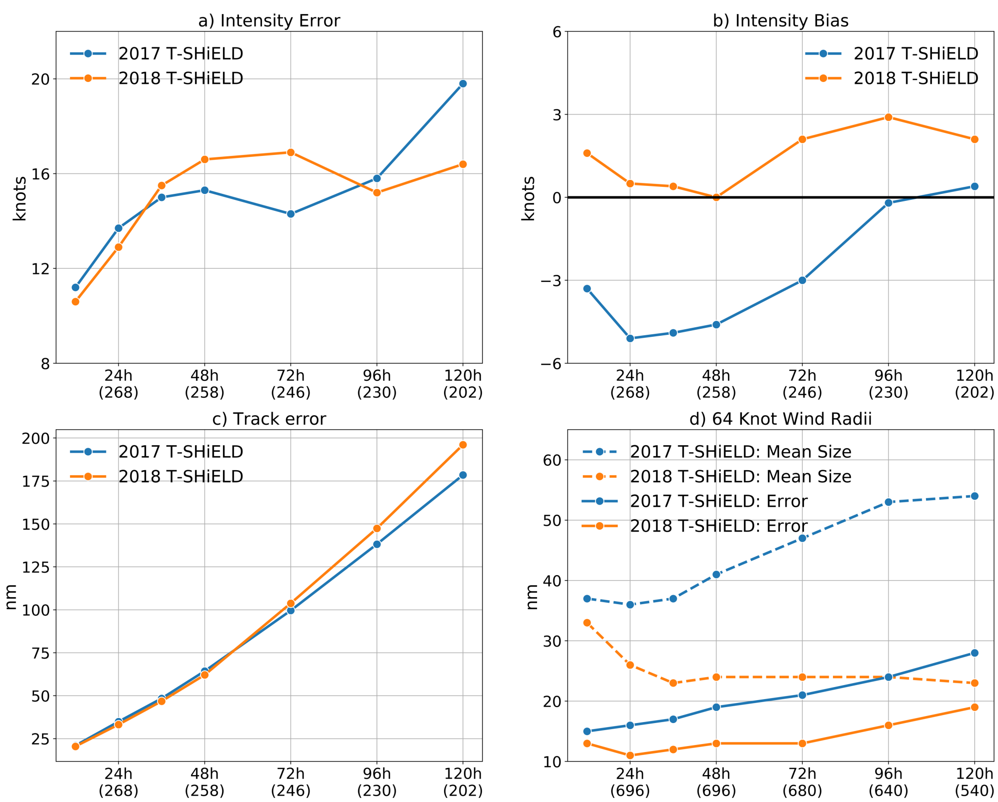

Figure 10. Verification of T-SHiELD 2017 and 2018 during the 2017 Atlantic hurricane season against the Best Track Dataset: intensity (a) error and (b) bias; (c) track error; (d) 64-kt (33 m s^{-1}, hurricane-force) radii. Units shown (kt, nautical miles) are standard for US operational prediction. In a–c the number of cases (individual storms) available at each lead time is shown in

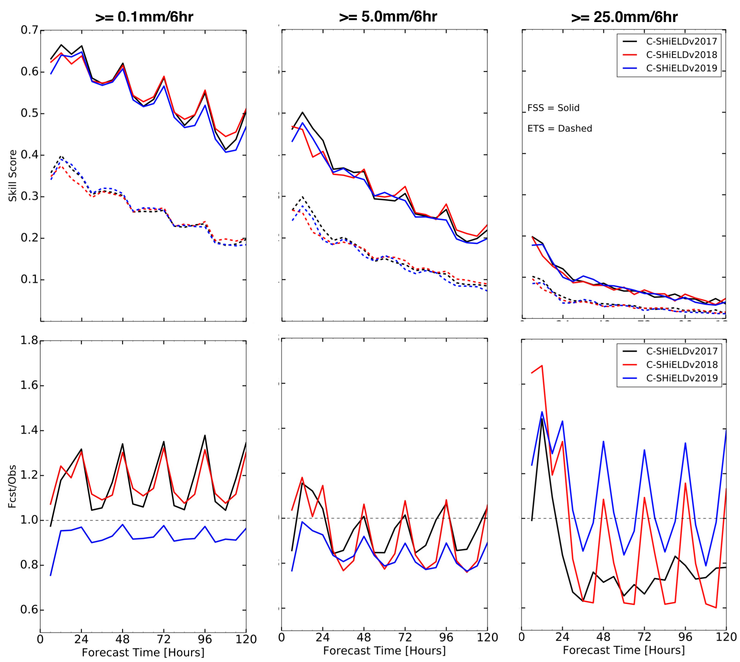

Figure 11. Precipitation skill scores (top) and bias score (bottom) vs. StageIV for 6-hr CONUS precipitation in three versions of C-SHiELD, given for precipitation events greater than three six-hourly accumulation thresholds (0.1, 5.0, and 25.0 mm). Skill scores are given for both Equitable Threat Score (ETS; Hogan and Mason 2012) and Fractions Skill Score (FSS; Roberts and Lean 2008). C-SHiELD 2017 is validated from May 2017 to May 2018; C-SHiELD 2018 is validated from April 2018 to May 2019; C-SHiELD 2019 is validated from January to December 2019. Validation is performed on the 4-km StageIV grid using 3x3 neighborhoods, corresponding to a 12-km radius.

Precipitation forecast skill (Figure 11, top panels) is similar among all three versions of C-SHiELD. The 2019 version has the least overall bias (Figure 11, bottom panels) as earlier versions had too much light and too little heavy precipitation. The 2019 version reduced the diurnal cycle in the bias of light and moderate precipitation, although this was still apparent in the bias score for heavy precipitation and still had a prominent high bias of heavy precipitation during the first 30 hours. We speculate that the re-configuration of the numerical diffusion, which improved storm placement, and the revised settings for the GFDL microphysics, which improved structure and evolution of the storms, combined to improve the biases in the 2019 version.

We use the surrogate severe technique of Sobash et al. (2011) to validate our 2–5 km updraft helicity (UH) fields against storm reports from the Storm Prediction Center. This is a well-established method used for evaluation of convective-scale prediction models (cf. https://hwt.nssl.noaa.gov/sfe/2018/docs/HWT_SFE_2018_Prelim_Findings_v1.pdf). We create surrogate severe fields and validate against observed severe fields to compute FSS and Bias scores in C-SHiELD and plot the results as a function of UH threshold and smoothing radius (Figure 12), similar to Figure 17 in Sobash et al. (2016). For all versions of C-SHiELD the highest FSS is found from the largest smoothing radius of 240 km and for UH thresholds of 150–200 m^2 s^{-2}, with slightly higher or lower thresholds giving similar skill scores. The UH threshold giving the best score for C-SHiELD is higher than in many other convective-scale models due to

These multiple-day severe weather forecasts are in the spirit of the convective outlooks issued by the Storm Prediction Center (www.spc.noaa.gov/products/outlook; Edwards 2015) based on predictions of synoptic-scale environments favorable for severe weather. The advantage of using a dynamical convective-scale prediction model on medium-range timescales is that explicit prediction of storms, instead of just environments, potentially can give forecasts of convective modes and specific hazards.

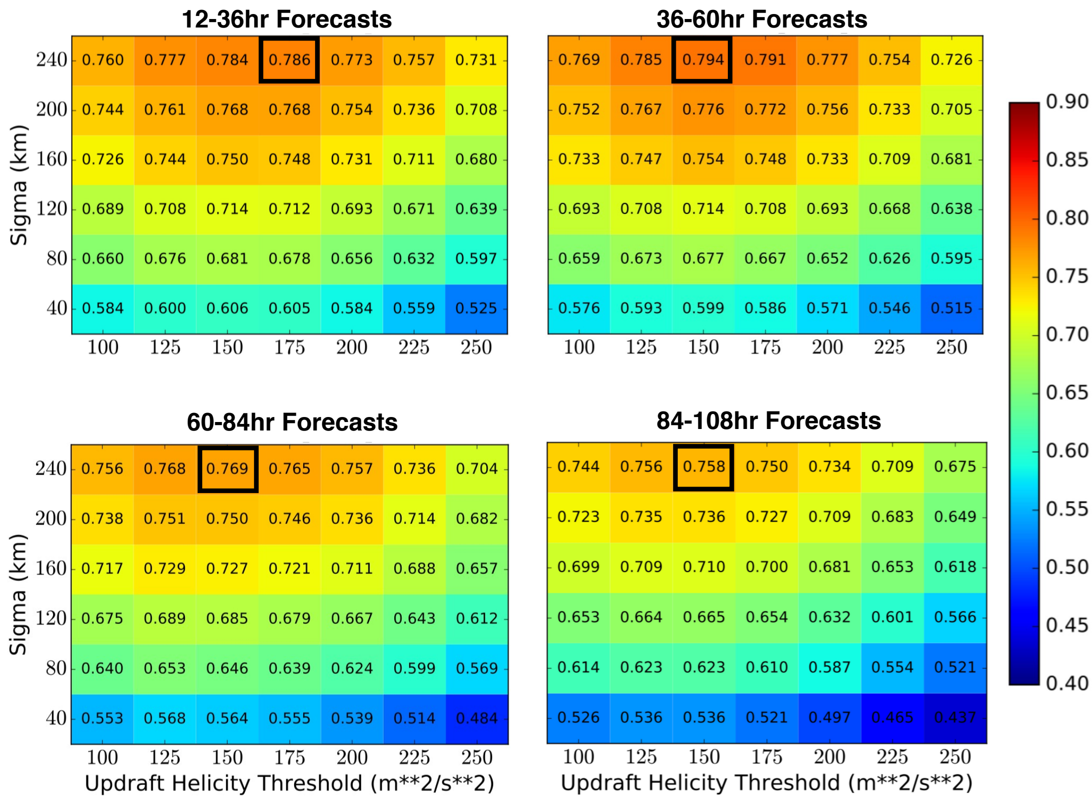

Figure 13. FSS for surrogate severe predictions at different lead times for 00Z initializations of C-SHiELD 2019.

3.4 S-SHiELD Subseasonal-to-Seasonal Prediction

We briefly describe the characteristics of the Tier-2 S-SHiELD configuration, using a 25-km grid designed for climate integrations and for subseasonal and seasonal predictions. S-SHiELD is configured similarly to the 13-km SHiELD, although SHiELD's two-day relaxation timescale of SSTs in the MLO towards the "frozen anomalies" is extended to 15 days in S-SHiELD. Unlike the vast majority of climate models, S-SHiELD is nonhydrostatic and uses a more sophisticated microphysics which is updated much more frequently. While these features do make S-SHiELD more expensive than analogous 25-km hydrostatic climate models (cf. Murakami et al 2016; Roberts et al. 2018), previous experience with HiRAM (Chen and Lin 2012, 2013; Gao et al 2018) has shown that nonhydrostatic dynamics and better microphysical-dynamical coupling yields a better representation of mesoscale convective systems and in particular of tropical cyclones, a major emphasis of our group's research.

(dotted) the interactive MLO for 92 40-day predictions initialized during the 2011-2012 DYNAMO period. Top: Correlation coefficient; bottom: RMSE.

The behavior of the MJO in GFDL's CMIP6-generation climate models (Zhao et al. 2018) suggests that the two keys for a good MJO simulation are an appropriate convection scheme and some form of interactive ocean, a result found also by DeMott et al. (2019) and others. A second set of S-SHiELD experiments was performed using specified climatological SSTs plus frozen anomalies. These simulations without the interactive MLO had much smaller RMM correlations, with predictions no longer useful after day 20, and larger errors. The effect of the interactive ocean is made clear in Figure 15, in which S-SHiELD with the MLO correctly predicted the formation of all three strong MJO events during this period 10–15 days in advance, and correctly propagated all events through the Maritime Continent (near 120 E longitude), although the propagation speed is slower than observed and there is some disruption near the Maritime Continent. However, S-SHiELD with prescribed SSTs has difficulty propagating the MJO through the Maritime Continent and for the November event creates no MJO whatsoever. The November event proves particularly challenging for S-SHiELD without the MLO as it performs poorly at a range of lead times (Supplemental Figure S2) but poses no problem for S-SHiELD with the MLO. It is clear that the simple, inexpensive MLO used in S-SHiELD is sufficient to significantly extend the predictability of the MJO.

DeMott et al. (2019) did not describe any deficiencies of the MJO from models using a 1D column ocean instead of a 3D dynamical ocean, which suggests a limited role for direct feedbacks between ocean circulations and the MJO. However they did not rule out indirect effects of the MJO on ocean circulation that could impact other S2S-timescale phenomena or MJO teleconnections. Other investigators have found that the MJO does alter ocean circulations on intraseasonal timescales, notably the result of Moum et al. (2014) found during DYNAMO. It remains to be seen whether these ocean dynamical effects of the MJO are of sufficient impact to affect S2S prediction skill. One advantage of the MLO is that we can nudge to climatological SSTs and so do not have climate drifts that challenge fully coupled models.

be a resolution effect as 13-km SHiELD reproduces both the correct phase and amplitude of precipitation. We also find that the majority of precipitation in S-SHiELD (55% globally and 80% between 20S and 20N) is from the SAS convective scheme, although this does not adversely affect the phase of the diurnal cycle. S-SHiELD does have the correct phase and amplitude (albeit slightly too high) of the diurnal cycle of 2-m temperature over land (Supplemental Figure S3).

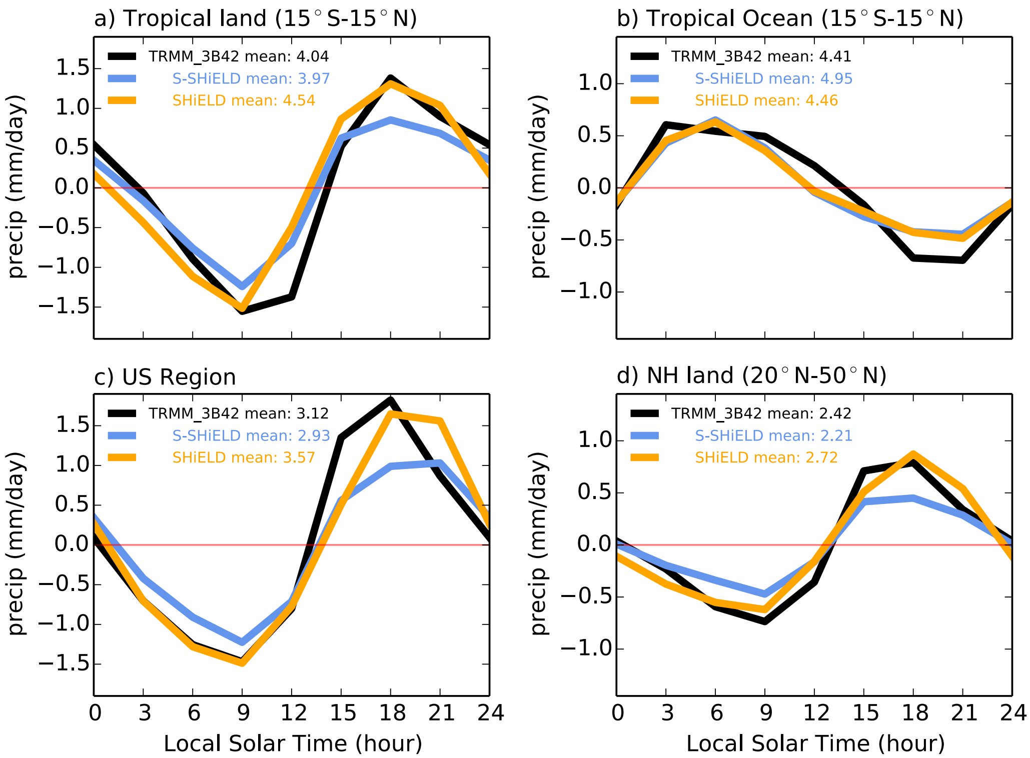

Figure 16. JJA diurnal cycle of precipitation as a function of local solar time in a 10-year S-SHiELD climate integration with the MLO nudged towards climatological SSTs, and from days 6--10 of three years of 13-km SHiELD hindcasts (initialized every five days), compared to TRMM 2011-2018 observations. Means are given in the legends as mm/day.

Hagos et al. (2016) found that the diurnal cycle of cloudiness and precipitation plays a key role in the propagation of the MJO through the Maritime Continent. Since S-SHiELD has considerably better diurnal cycles of precipitation and temperature over land, especially over tropical land, than do most climate models, we might expect that this improved representation of the diurnal cycle may be contributing to the improved representation of the MJO seen above.

4 Conclusion and Prospects

We have developed the SHiELD modeling system as a research tool to demonstrate new capabilities of the FV3 Dynamical Core and of our physical parameterizations, develop new ideas in atmospheric prediction modeling, and to explore processes and phenomena within the atmosphere. Since late 2015 when FV3 was first coupled to the then-operational GFS Physics Driver we have developed SHiELD into a promising vehicle for improving the prediction and understanding of atmospheric phenomena. SHiELD also demonstrates the potential and viability data assimilation cycling system to take advantage of the new advances and to create initial conditions most consistent with the forward prediction model configurations. Finally, development will continue of our Tier-2 configurations, with near real-time S2S predictions being made using S-SHiELD, and continued extension into the global cloud-resolving regime (cf. Stevens2019) towards new scientific problems not adequately addressed by existing regional models or by coarse-resolution global models.

Acknowledgments

SHiELD grew out of a major collaboration between GFDL and EMC and would not have been possible without the physical parameterization suite, software, data, and especially input initial conditions and baseline forecasts made freely available by EMC and the National Weather Service. We thank Jongil Han for providing SA-SAS, and George Gayno and Helin Wei for providing EMC pre-processing tools and land model inputs and for significant assistance with these tools and datasets. We also thank James Franklin (NHC, retired) for advice on the accuracy of the wind radii in the best-track dataset. Kate Zhou and Tom Delworth provided reviews of this manuscript. Xi Chen, Linjiong Zhou, Kun Gao, Yongqiang Sun, Kai-Yuan Cheng, and Morris Bender are funded under award NA18OAR4320123 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. Xi Chen, Zhou, and Cheng were additionally funded by the Next-Generation Global Prediction System project of the National Weather Service. The National Oceanic and Atmospheric Administration's Hurricane Supplemental Program Office partially funded Zhou, Gao, and Bender under award NA19OAR0220147; Sun under NA19OAR0220145; and Cheng under NA19OAR0220146. Supporting data can be found at doi:10.5281/zenodo.3997344. We thank two anonymous reviewers for their insightful comments.

Appendix A: Positive-Definite Advection Scheme

The Lagrangian dynamics in FV3 uses 1D advection operators to build the 2D advection scheme of Lin and Rood (1996). In hydrostatic FV3 these operators are typically monotonic (Lin 2004), in that no new extrema are created by the advection; however monotonic advection can be overly diffusive for some applications. In nonhydrostatic FV3 the monotonicity constraint is not used for advection of dynamical quantities (vorticity, heat, air mass), but positivity still needs to be enforced for scalar tracers. We introduce a positive-definite scheme, which uses a weaker constraint than monotonicity which only prevents the appearance of negative values.

This positivity constraint can be applied to any scheme similar to VanLeer (1974) or PPM (Collella and Woodward 1984) in which first-guess continuous edge values $\hat{q}_{i+1/2}$ and $\hat{q}_{i-1/2}$ are interpolated from the cell-averaged values $\overline{q_i}$ where *i* is a grid index. As with a standard monotonicity constraint we break the continuity of the sub-grid reconstructions across grid-cell interfaces, creating left-edge and right-edge values, Q_i^- and Q_i^+, respectively, as well as a curvature value B_{oi} for each grid cell, which are then used to compute the flux as in Putman and Lin (2007), Appendix B.

To adjust the edge values to ensure positivity, we use the algorithm below on cell *i*, where notation is as in Lin (2004), Appendix A:

$$Q_i^- = \hat{q}_{i-1/2} - \overline{q_i}$$

$$Q_i^+ = \hat{q}_{i+1/2} - \overline{q_i}$$

with the dynamical core. Other updates in the Inline microphysics include a time-implicit monotonic scheme for sedimentation to ensure stability without needing to subcycle; precise conservation of the total moist energy; and transportation of heat and momentum carried by condensates during sedimentation.

Appendix C: A Note on Terminology

The term “model” means many different things in many contexts, and can be confusing. In this paper, we use the term “model” only in the abstract (“other general-purpose models”, “NCEP Modeling Suite”) or as part of the name of another system (“Noah Land Surface Model”, “GFDL Hurricane Model”). For concreteness, we refer to SHiELD as a “modeling system” which can be used in a variety of “configurations” (13-km SHiELD, C-SHiELD, T-SHiELD, S-SHiELD), each upgraded to new yearly versions (SHiELD 2016, SHiELD 2017, etc.).

References

Alexander, C., Carley, J., Heinselman, P. L., & Harris, L. (2020). Advancements of the FV3 Stand-Alone Regional Model. In *100th Annual Meeting.* AMS.

Alpert, J. C. (2004, January). Sub-grid scale mountain blocking at NCEP. In *Proceedings of 20th Conference on WAF, 16th conference on NWP*.

Arnold, N. P., & Putman, W. M. (2018). Nonrotating convective self-aggregation in a limited area AGCM. *Journal of advances in modeling earth systems*, *10*(4), 1029-1046.

Balaji, V. (2012). The flexible modeling system. In *Earth System Modelling-Volume 3* (pp. 33-41). Springer, Berlin, Heidelberg.

Bender, M.A., T.P. Marchok, C.R. Sampson, J.A. Knaff, and M.J. Morin. (2017). Impact of Storm Size on Prediction of Storm Track and Intensity Using the 2016 Operational GFDL Hurricane Model. *Wea. Forecasting, 32*, 1491–1508. https://doi.org/10.1175/WAF-D-16-0220.1

Bender, M. A., Marchok, T., Tuleya, R. E., Ginis, I., Tallapragada, V., & Lord, S. J. (2019). Hurricane Model Development at GFDL: A Collaborative Success Story from a Historical Perspective. *Bulletin of the American Meteorological Society*, *100*(9), 1725-1736.

Brown, A., et al. (2012). "Unified Modeling and Prediction of Weather and Climate: A 25-Year Journey." *Bulletin of the American Meteorological Society*, *93*(12): 1865-1877.

Carley, J. R., et al. (2020). Advances toward an Operational Convection-Allowing Ensemble Prediction System in the Unified Forecast System at NOAA. In *100th Annual Meeting.* AMS.

Chen, J. H., & Lin, S. J. (2013). Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *Journal of Climate*, *26*(2), 380-398.

Chen, J. H., & Lin, S. J. (2011). The remarkable predictability of inter-annual variability of Atlantic hurricanes during the past decade. *Geophysical Research Letters*, *38*(11).

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*(6)*,* 617. https://doi.org/10.3390/atmos11060617

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., et al. (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS002015. https://doi.org/10.1029/2019MS002015

Edwards, R. (2015). Overview of the Storm Prediction Center. In *13th History Symposium*.

ECMWF. (2019). Part III: Dynamics and Numerical Procedures. *IFS Documentation CY46R1*.

ECMWF. (2019). Part IV: Physical Processes. *IFS Documentation CY46R1*.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., & Tarpley, J. D. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research*, *108*(D22), 8851. https://doi.org/10.1029/2002JD003296

Gao, K., Chen, J. H., Harris, L. M., Lin, S. J., Xiang, B., & Zhao, M. (2017). Impact of intraseasonal oscillations on the tropical cyclone activity over the Gulf of Mexico and western Caribbean Sea in GFDL HiRAM. *Journal of Geophysical Research: Atmospheres*, *122*(24), 13-125.

Gao, K., Chen, J. H., Harris, L., Sun, Y., & Lin, S. J. (2019). Skillful Prediction of Monthly Major Hurricane Activity in the North Atlantic with Two-way Nesting. *Geophysical Research Letters*, *46*(15), 9222-9230.

GFDL. (2019). The 5–10 Year Strategic Plan. Available at https://www.gfdl.noaa.gov/wp-content/uploads/2019/10/2019_GFDL_External_Review_Strategic_Plan.pdf

Haarsma, R., van der Linden, E. C., Selten, F., & van der Schrier, G. (2017). Extreme future central European droughts in a high-resolution global climate model. In *EGU General Assembly Conference Abstracts* (Vol. 19, p. 14128).

Hagos, S. M., Zhang, C., Feng, Z., Burleyson, C. D., De Mott, C., Kerns, B., Benedict, J. J., and Martini, M. N. (2016). The impact of the diurnal cycle on the propagation of Madden-Julian Oscillation convection across the Maritime Continent. *J. Adv. Model. Earth Syst., 8*, 1552– 1564. doi:10.1002/2016MS000725.

Han, J., M.L. Witek, J. Teixeira, R. Sun, H. Pan, J.K. Fletcher, and C.S. Bretherton. (2016). Implementation in the NCEP GFS of a hybrid eddy-diffusivity mass-flux (EDMF) boundary layer parameterization with dissipative heating and modified stable boundary layer mixing. *Wea. Forecasting, 31*, 341–352. https://doi.org/10.1175/WAF-D-15-0053.1

Han, J., & Pan, H.-L. (2011). Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Weather and Forecasting*, *26*(4), 520–533.

Jeevanjee, N. (2017). Vertical velocity in the gray zone. *Journal of Advances in Modeling Earth Systems*, *9*(6), 2304-2316.

Kim, H., F. Vitart, and D.E. Waliser. (2018). Prediction of the Madden–Julian Oscillation: *A Review. J. Climate, 31*, 9425–9443. https://doi.org/10.1175/JCLI-D-18-0210.1

Klingaman, N. P., & Demott, C. A. (2020). Mean-state biases and interannual variability affect perceived sensitivities of the Madden–Julian oscillation to air–sea coupling. *Journal of Advances in Modeling Earth Systems, 12*, e2019MS001799. https://doi.org/10.1029/2019MS001799

Knutson, T. R., Sirutis, J. J., Garner, S. T., Held, I. M., & Tuleya, R. E. (2007). Simulation of the recent multidecadal increase of Atlantic hurricane activity using an 18-km-grid regional model. *Bulletin of the American Meteorological Society*, *88*(10), 1549-1565.

Knutson, T. R., & Tuleya, R. E. (2008). *Tropical cyclones and climate change: revisiting recent studies at GFDL* (pp. 120-144). Cambridge, UK: Cambridge University Press.

Krueger, S. K., Fu, Q., Liou, K. N., & Chin, H. N. S. (1995). Improvements of an ice-phase microphysics parameterization for use in numerical simulations of tropical convection. *Journal of Applied Meteorology*, *34*(1), 281-287.

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

Lin, Y. L., Farley, R. D., & Orville, H. D. (1983). Bulk parameterization of the snow field in a cloud model. *Journal of climate and applied meteorology*, *22*(6), 1065-1092.

Lin, S. J. (2004). A "vertically Lagrangian" finite-volume dynamical core for global models. *Monthly Weather Review*, 132(10), 2293-2307.

Lin, S. J. (2013). From Large-Eddy-Simulation to climate modeling: GFDL's unified global-regional non-hydrostatic modeling framework. In *Tropical Cyclone Research Forum: 67th IHC Presentations*.

Lin, S.-J., Harris, L. M., Benson, R., Zhou, L., Chen, J.-H., & Chen, X. (2017). Towards a unified prediction system from weather to climate scale. Second Symposium on Multi-Scale Atmospheric Predictability, Seattle, WA, Paper 3.1.

Lord, S., Willoughby, H. E., & Piotrowicz, J. M. (1984). Role of a parameterized ice-phase microphysics in an axisymmetric, nonhydrostatic tropical cyclone model. *Journal of the Atmospheric Sciences*, *41*(19), 2836-2848.

Marchok, T. (2018, April). Factors Important for Tropical Cyclone Tracking in NWP Output. In *33rd Conference on Hurricanes and Tropical Meteorology*. AMS.

Marchok, T., Morin, M. J., Knaff, J., Sampson, C. R., Hazelton, A., & Lin, S. J. (2018, April). An Evaluation of Surface Wind Structure Forecasts from the fvGFS and Operational Dynamical Models. In *33rd Conference on Hurricanes and Tropical Meteorology*. AMS.

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S. J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports, 5(3),* 172-184.

Sela, J. G. (2010). The derivation of the sigma pressure hybrid coordinate Semi-Lagrangian model equations for the GFS.

Snook, N., Kong, F., Brewster, K. A., Xue, M., Thomas, K. W., Supinie, T. A., ... & Albright, B. (2019). Evaluation of convection-permitting precipitation forecast products using WRF, NMMB, and FV3 for the 2016–17 NOAA Hydrometeorology Testbed Flash Flood and Intense Rainfall Experiments. *Weather and Forecasting*, *34*(3), 781-804.

Sobash, R.A., J.S. Kain, D.R. Bright, A.R. Dean, M.C. Coniglio, and S.J. Weiss. (2011). Probabilistic Forecast Guidance for Severe Thunderstorms Based on the Identification of Extreme Phenomena in Convection-Allowing Model Forecasts. *Wea. Forecasting, 26*, 714–728. https://doi.org/10.1175/WAF-D-10-05046.1

Sobash, R.A., G.S. Romine, C.S. Schwartz, D.J. Gagne, and M.L. Weisman. (2016). Explicit Forecasts of Low-Level Rotation from Convection-Allowing Models for Next-Day Tornado Prediction. *Wea. Forecasting, 31*, 1591–1614. https://doi.org/10.1175/WAF-D-16-0073.1

Sobash, R.A., C.S. Schwartz, G.S. Romine, and M.L. Weisman. (2019). Next-Day Prediction of Tornadoes Using Convection-Allowing Models with 1-km Horizontal Grid Spacing. *Wea. Forecasting, 34*, 1117–1135. https://doi.org/10.1175/WAF-D-19-0044.1

Thiébaux, J., E. Rogers, W. Wang, and B. Katz. (2003). A New High-Resolution Blended Real-Time Global Sea Surface Temperature Analysis. *Bull. Amer. Meteor. Soc., 84*, 645–656. https://doi.org/10.1175/BAMS-84-5-645

Van Leer, B. (1974). Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second-order scheme. *Journal of computational physics*, *14*(4), 361-370.

Vecchi, G. A., Murakami, H., Delworth, T. L., Underwood, S., Wittenberg, A. T., Zeng, F. J., ... & Kapnick, S. B. (2019). Tropical cyclone sensitivity to global forcing: seeds and probability. *AGUFM*, *2019*, A32F-06.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., ... & Hendon, H. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, *98*(1), 163-173.

Wei, H., Zheng, W., Meng, J., Gayno, G., Hou, Y., & Ek, M. (2017). Planned land surface changes for the next NEMS implementation. In *28th Conf. on Weather Analysis and Forecasting/ 24th Conf. on Numerical Weather Prediction, American Meteorological Society, Seattle, WA*, pp. 600.

Wheeler, M. C., & Hendon, H. H. (2004). An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly weather review*, *132*(8), 1917-1932.