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.

Abstract

We present the System for High-resolution prediction on Earth-to-Local Domains (SHiELD), an atmosphere model coupling the nonhydrostatic FV3 Dynamical Core to a physics suite originally taken from the Global Forecast System. SHiELD is designed to demonstrate new capabilities within its components, explore new model applications, and to answer scientific questions through these new functionalities. A variety of configurations are presented, including short-to-medium-range and subseasonal-to-seasonal (S2S) prediction, global-to-regional convective-scale hurricane and contiguous US precipitation forecasts, and global cloud-resolving modeling. Advances within SHiELD can be seamlessly transitioned into other Unified Forecast System (UFS) or FV3-based models, including operational implementations of the UFS.

Continued development of SHiELD has shown improvement upon existing models. The flagship 13-km SHiELD demonstrates steadily improved large-scale prediction skill and precipitation prediction skill. SHiELD and the coarser-resolution S-SHiELD demonstrate a superior diurnal cycle compared to existing climate models; the latter also demonstrates 28 days of useful prediction skill for the Madden-Julian Oscillation. The global-to-regional nested configurations T-SHiELD (tropical Atlantic) and C-SHiELD (contiguous United States) shows significant improvement in hurricane structure from a new tracer advection scheme and promise for medium-range prediction of convective storms, respectively.

Plain Language Summary

At many weather forecasting centers where computer weather models are run, different models are run for different applications. However, each separate model multiplies the effort needed to maintain and upgrade each model and makes it difficult to move improvements between models.

We present a new “unified” weather modeling system, SHiELD, able to be configured for a variety of applications. This system uses a powerful computer code, FV3, to compute the fluid motion of the atmosphere at any scale, and also able to zoom in on areas of interest to better “see” severe storms or intense hurricanes. We show how we started from a quickly-assembled model for testing FV3 and then gradually improved the representation of different atmospheric processes and expanded into new uses for the system, including short-range severe thunderstorm prediction, hurricane forecasting, and forecasts out to as long as six weeks. We address some of the challenges that we faced and discuss prospects for future model improvements. Since many of the parts of SHiELD are used by models being developed by the National Weather Service for use by weather forecasters, the advances described here can be rapidly introduced into those models, eventually improving official forecasts.

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).

Since FV3 is designed to adapt to a variety of purposes and to any scale of atmospheric motion it is an ideal platform for a unified modeling system. All of the SHiELD configurations described here, as well as regional and doubly-periodic applications lying beyond the scope of this paper, use the same code base, the same executable, the same preprocessor, the same runscripts, and same post-processing tools, demonstrating a true unification for modeling on weather-to-S2S timescales. This approach also suggests how further unification with GFDL's climate models, which use a different atmospheric physics (Zhao et al. 2018), the MOM6 Dynamical Ocean (Adcroft et al. 2019), and the GFDL LM4 land model, may proceed. Advances in SHiELD can be seamlessly moved into other UFS models, including the 2019 upgraded GFSv15, and other FV3-based models. Most notably, advances in SHiELD can migrate into UFS models slated for operational implementation at NCEP, including the FV3-based GFSv15. NASA GEOS (Putman and Suarez 2017), NASA/Harvard GEOS-Chem High-Performance (GHCP), CESM-FV3, and the Chinese Academy of Sciences' F-GOALS all also use FV3 as their dynamical core and can benefit from the advances described below. This diversity of FV3-based models shows the advantages of using common components to leverage advances in the dynamical core but while still allowing centers to tailor their models to their own needs, the freedom to innovate new model designs, and to encourage the development of models as holistic integrated systems, rather than clumsily joining independent components.

SHiELD is designed for exploratory research into model design and development, with a focus on dynamics and physics-dynamics integration, and for research on prediction and atmospheric processes on timescales from a few hours to a few months. SHiELD is currently focused on deterministic prediction although effective S2S prediction will require the development of a simple ensemble (cf. Chen and Lin 2013). In this manuscript we use forecast skill as a principal means of establishing the scientific credibility of SHiELD as a research tool. Further research will more closely evaluate specific structures and processes within SHiELD, with some initial results described below (especially section 3.2) and in prior research (cf. Hazelton 2018a).

The design, evolution, configurations, and simulation characteristics of SHiELD are the subject of this paper. Section 2 describes the components of SHiELD and how they work together as a complete modeling system. Section 3 describes the four configurations of SHiELD for a variety of applications, including medium-range weather, continental convection, tropical meteorology and hurricanes, and S2S prediction. Section 4 summarizes the history of SHiELD development and discusses prospects for future work.

2 SHiELD Components

2.1 Nonhydrostatic FV3 Dynamical Core

All SHiELD simulations use the nonhydrostatic solver within the FV3 Dynamical Core. This core has been described in detail in other papers (Lin 2004, Putman and Lin 2007, Harris and Lin 2013, and references therein) and will only be summarized here. FV3 solves the fully-compressible Euler equations on the gnomonic cubed-sphere grid and a Lagrangian vertical coordinate. Fast vertically-propagating sound and gravity waves are solved by the semi-implicit method; otherwise the algorithm is fully explicit. FV3 advances sound and gravity wave processes and advects thermodynamic variables on the shortest "acoustic" timestep, while sub-

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

modification of the mixed-layer ocean (MLO) of Pollard et al. 1973. This simple ocean computes the mixed layer depth and heat within that mixed layer as prognostic variables, with tendencies computed from the net surface heat flux. The SST is nudged towards the NCEP Real-Time Global Sea Surface Temperature (RTGSST; Thiébaux et al. 2003) climatology plus a fixed initial anomaly which decays with a fixed timescale. The ocean mixed layer depth is also nudged toward observed climatology (de Boyer Montégut et al., 2004). While considerably simpler than the three-dimensional dynamical oceans in CM4 (Held et al. 2020) and in the GFDL Hurricane Model (Bender et al. 2019), the MLO still represents the thermodynamic and dynamic ocean interactions of greatest significance on the timescales for which SHiELD is used (Hazelton et al. 2018b), without incurring the complexity of a three-dimensional dynamical ocean.

2.4 Interoperability with other UFS models

SHiELD was designed to work with other models that use FV3, FMS, the GFS Physics Driver, and/or the Interoperable Physics Driver (IPD). The IPD is the interface between FV3 and the GFS Physics Driver, although it can support other physics suites. Innovations within SHiELD can then be seamlessly exchanged with other models using these same components. The UFS Atmosphere led by NCEP (https://github.com/NOAA-EMC/fv3atm/) analogous to SHiELD. For example, the transition of FV3 and the GFDL Microphysics into the operational GFSv15 was accelerated by the IPD. Conversely, schemes which have been introduced into the GFS Physics Driver by the broader community can then be integrated into SHiELD, including the numerous schemes implemented by Zhang et al. (2018).

3 SHiELD Configurations

SHiELD leverages the flexibility of FV3 to be able to make accurate and efficient simulations at a variety of spatial and temporal scales. Much of the development of SHiELD (and previously, of HiRAM) has been driven by a desire to improve the simulation quality at the convection-permitting resolutions covered by the range of SHiELD configurations.

We present four different configurations of SHiELD. All configurations are global domains using either a uniform grid or a locally refined grid using nesting or stretching (Harris and Lin 2013; Harris et al. 2016; Zhou et al. 2019). SHiELD can also run on FV3's doubly-periodic domain (Held and Zhou, 2006, Arnold and Putman, 2018) or on a regional domain using any regular quadrilateral grid (Dong et al., 2020), at spatial resolutions down to a few tens of meters (Jeevanjee 2017). These applications lie beyond the scope of this paper.

The four configurations can be fit within two "tiers"; Tier-1 configurations are the most well-tested, having originally been developed as prototypes to replace legacy NCEP models by FV3-based UFS systems, and having been run in near-real time for several years. These configurations demonstrate the capabilities of SHiELD, allow direct comparison to existing operational models, and provide robust tests of the forecast skill and reliability of SHiELD. Current real-time configurations are run twice daily and displayed at https://shield.gfdl.noaa.gov/.

The Tier-1 configurations are our flagship 13-km SHiELD, a prototype for the now-operational GFSv15 and for future upgrades of the GFS; (Tropical) T-SHiELD with a static, 3-km nest spanning the tropical North Atlantic, a prototype of the Hurricane Analysis and Forecast System (HAFS); and (Continental) C-SHiELD with a 3-km nest over the contiguous United States (CONUS), a prototype of the Regional Forecast System (RFS). Each of the Tier-1

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.

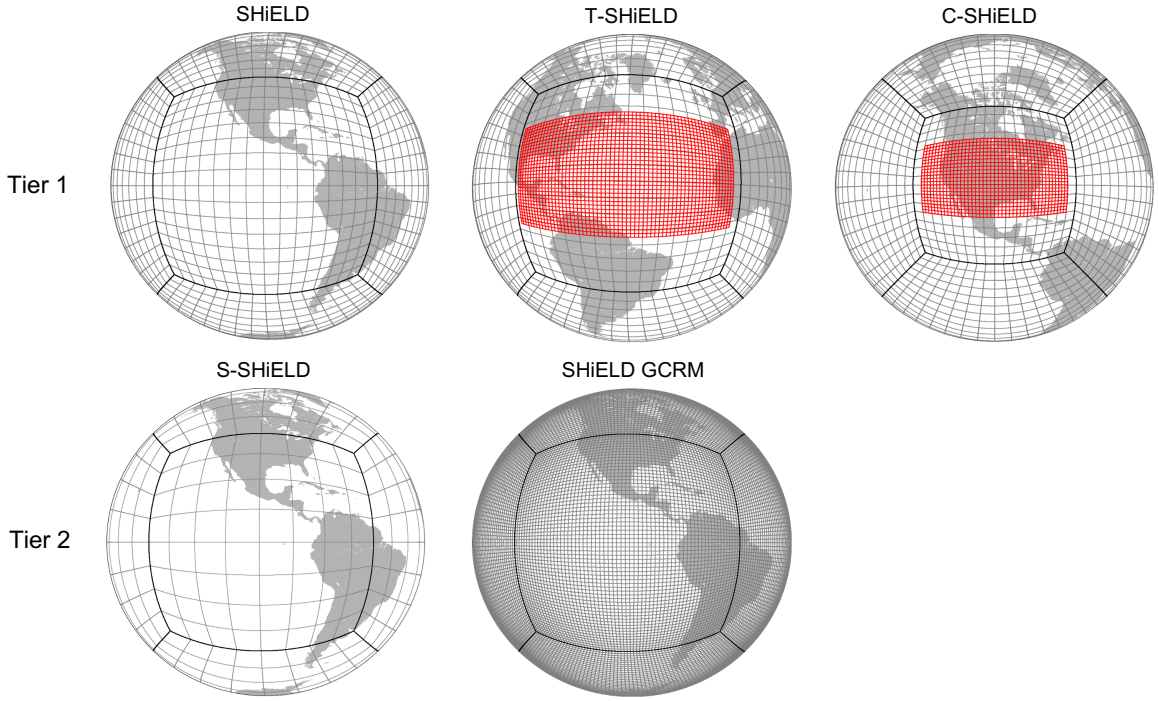


Figure 1. Current SHiELD configurations. Each plotted cell is 48x48 actual grid cells. Heavy black lines represent cubed-sphere edges; red lines represent nested grids. Note that the global domain of C-SHiELD (top center) is slightly stretched as per Harris et al. (2019).

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.

252

Configuration	SHiELD				T-SHiELD		C-SHiELD			S-SHiELD
Version	2016	2017	2018	2019	2017	2018	2017	2018	2019	2019
Evaluation Period	2015/01–2016/12	2015/01–2016/12	2016/01–2017/12	2017/01–2018/12	2017/08/18–2017/10/06	2017/08/18–2017/10/06	2017/05–2018/04	2018/05–2019/03	2019/04–2019/12	2011/10–2012/03 (MJO)
Simulation Frequency	00Z every 5 days (hindcast)				4x daily		00Z daily			00Z every 2 days
Simulation Length	10 days				126 hours		120 hours			40 days
Resolution	13 km (c768)				13 km (c768) + 3-km nest (2880 x 1536)		20-to-9 km stretched (c768r15) + 3-km 3x nest (2016 x 1080)			25-km
Grid Cells	3.54 M				3.54 M + 4.23 M		3.54 M + 2.18 M			885 K
Vertical Levels	63	63	91	91	63	63	63	63	63	91
Physics Timestep	225	150	150	150	90	90	90	90	90	450
Remapping, Tracer, and MP Timestep	112.5	150	150	150	90/22.5	90/22.5	90/22.5	90/22.5	90/22.5	225
Acoustic Timestep	18.75	18.75	18.75	18.75	12.8/4.5	12.8/4.5	12.8/4.5	12.8/4.5	12.8/4.5	28.125
Tracer Advection Scheme	Monotonic	Monotonic	Pos. Def	Pos. Def	Monotonic	Pos. Def	Monotonic	Pos. Def	Pos. Def	Pos. Def
Microphysics	Zhao-Carr	Split GFDL	Inline GFDL	Inline GFDL	Split GFDL	Split GFDL	Split GFDL	Inline GFDL	Inline GFDL	Inline GFDL
PBL Scheme	Hybrid EDMF	Hybrid EDMF	YSU	YSU	Mod. EDMF	YSU	Mod. EDMF	YSU	YSU	YSU
Deep Convection Scheme	SAS	SA-SAS	SA-SAS	SA-SAS	SAS	SA-SAS	None	None	None	SAS
Ocean Surface	Specified	Specified	MLO	MLO	Specified	MLO	Specified	Specified	MLO	MLO

253 **Table 1.** Development of the four SHiELD configurations and their yearly revisions described in
254 this paper. Timesteps are given in seconds; for nested simulations the format is global/nested
255 timesteps. All configurations and versions use the same Noah LSM and RRTM, and all use SAS
256 or SA-SAS shallow convection except 2017 and 2018 C-SHiELD.

257 All configurations are initialized using the real-time GFS analyses made available by
258 NCEP following Chen et al. (2018). This "cold starting" from the hydrostatic, spectral GFS
259 could potentially leave the convective-scale configurations (T-SHiELD, C-SHiELD) at a
260 comparative disadvantage to models with native, specialized convective-scale data assimilation.
261 This issue is minimized here due to the ability of FV3-based models to "spin up" their
262 convective scales within a few hours of initialization and experience little degradation thereafter
263 (Hazelton et al. 2018a,b; Marchok et al. 2018; Zhang et al. 2018; Harris et al. 2019).

264 Computational efficiency is crucial for useful simulation modeling, for both real-time and
265 experimental applications. We present the timings for the most recent iterations of SHiELD in
266 Table 2. The 13-km SHiELD needs only 3096 processor cores to complete one day in under 8.5
267 minutes, the threshold traditionally used for operational global prediction. The 25-km S-SHiELD
268 completes 1.5 years per day with just over 1700 cores; we are hoping to improve the
269 computational cost as part of further S-SHiELD development**.** C-SHiELD is necessarily more

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.

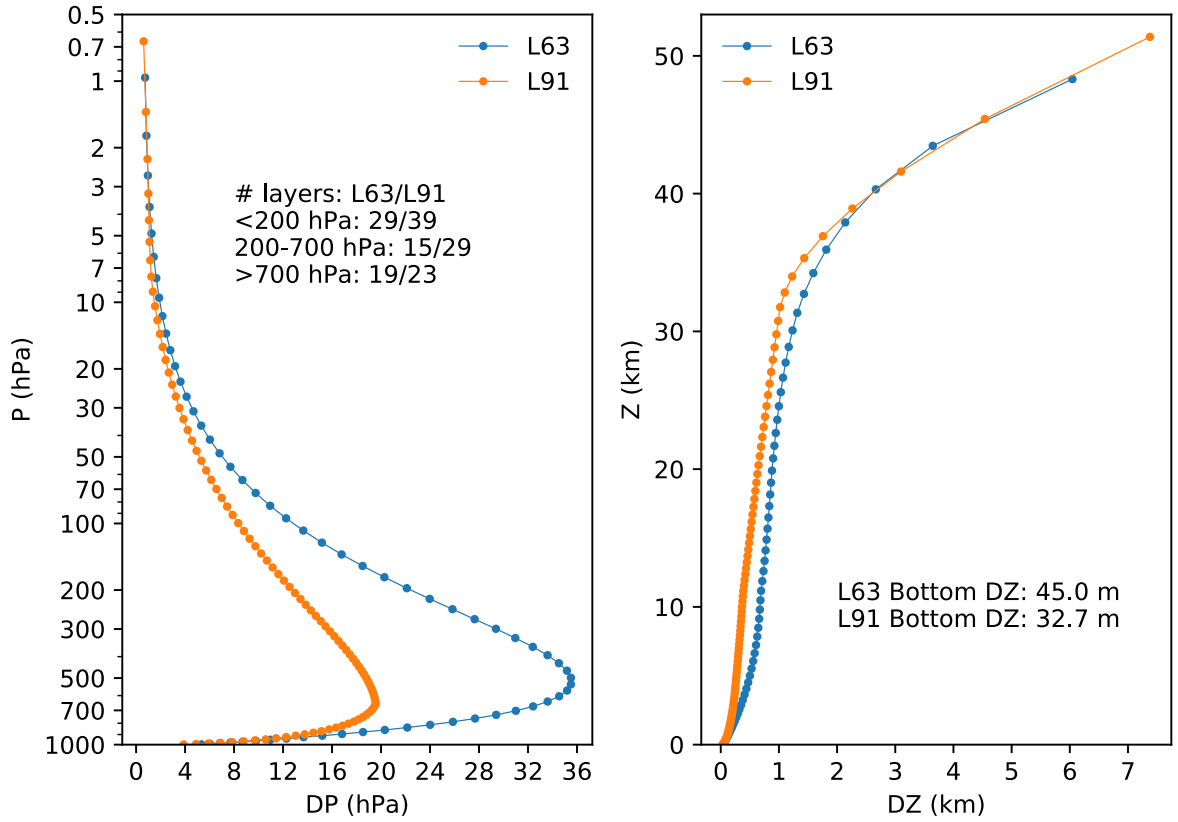


Figure 2. Distribution of vertical levels in various SHiELD configurations for a surface pressure of 1000 hPa and a standard atmospheric temperature structure.

The simulation characteristics and prediction skill of SHiELD have been previously discussed in several papers and will not be repeated here. Improving predictions of tropical cyclone track, intensity, and genesis has been a key driver of SHiELD development: Chen et al. (2019a) describes the 2016 and 2017 versions, while the considerably improved 2018 version is described in Chen et al. (2019b). Most notably SHiELD greatly improves upon other global models' ability to predict tropical cyclone intensity. The large-scale prediction skill, and CONUS precipitation and 2-m temperature skill, are briefly described for the 2016 and 2017 versions in LZhou2019 and Harris2019.

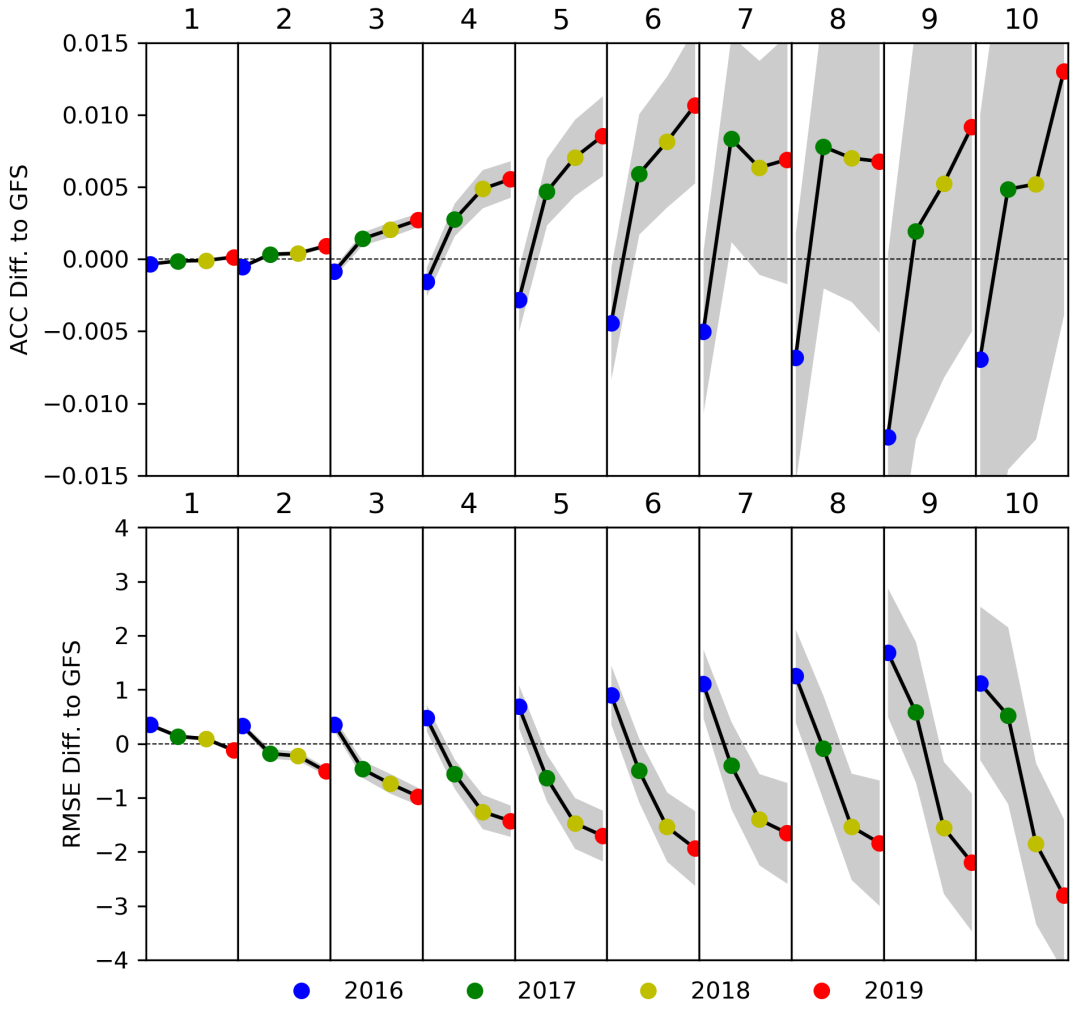
The anomaly correlation coefficient (ACC) of the 500-mb geopotential height field is the standard means for evaluating the large-scale prediction skill of medium-range prediction models. Figure 3 (top) shows that the global ACC of SHiELD has been better at all lead times than the contemporary GFS since the 2017 version, and significantly so on days 1–6. At all lead times except for days 7 and 8, each new version has improved upon the previous version. The result for root-mean square error (RMSE; Figure 3, bottom) is even more striking: every version is an improvement upon the previous at every lead time, and both the 2018 and 2019 versions are significantly better than the operational GFS. Results for just the northern hemisphere (20N--80N, Supplemental Figure S1) are less dramatic but SHiELD still shows statistically significant improvements in ACC and RMSE out to day 5. Both the GFS and all versions of SHiELD reach

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.



338

339 **Figure 3.** Global 500-mb geopotential height ACC (top) and RMSE (bottom, m) difference from
340 the contemporary GFS as a function of lead time (instantaneous at 00Z each day after
341 initialization) for each version of the 13-km SHiELD. Gray shading is the 95% confidence
342 interval. Each version of SHiELD is evaluated with two years of 10-day hindcasts initialized at

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

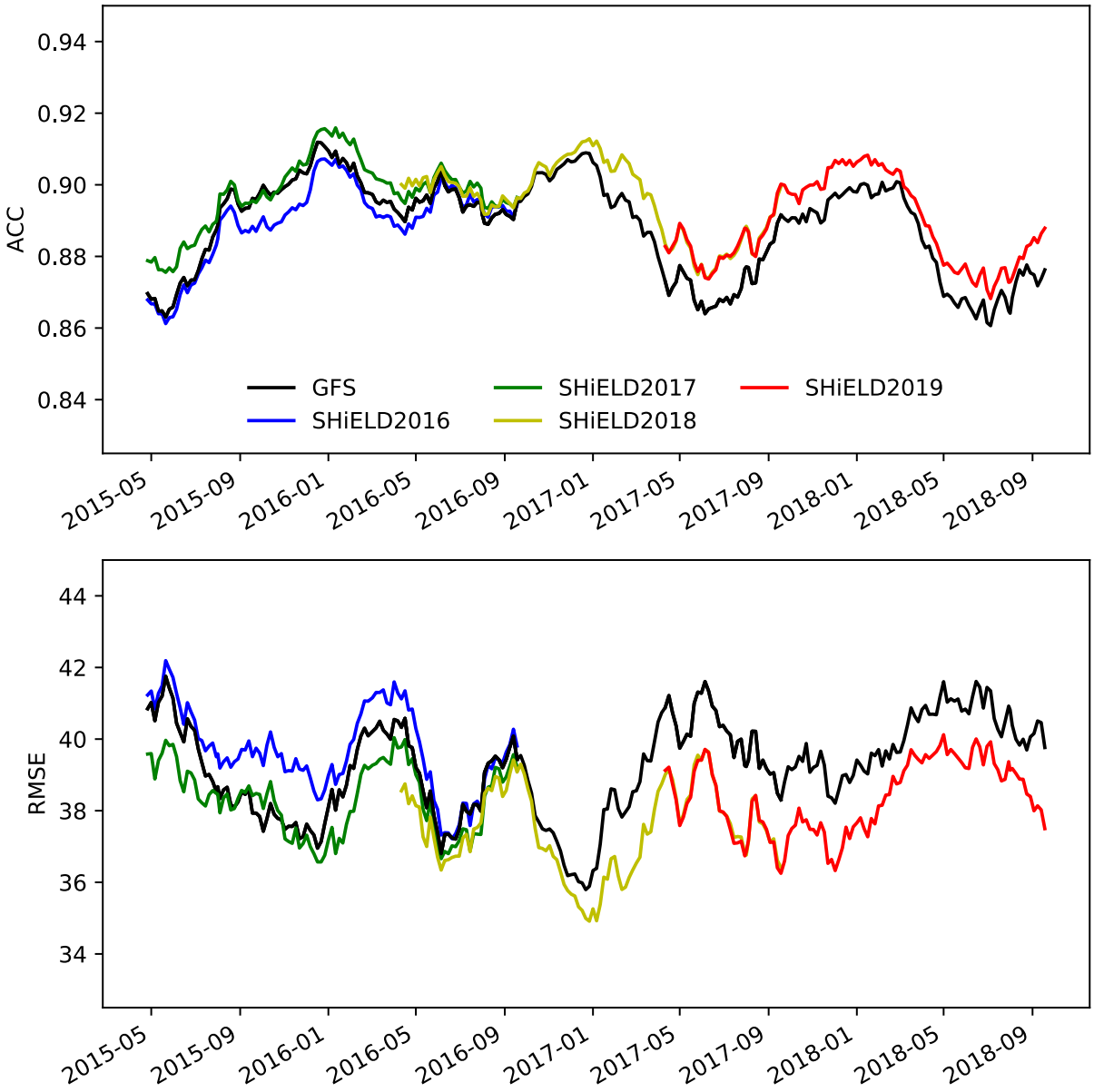


Figure 4. Six-month running-mean time series of global 500-mb geopotential height ACC (top) and RMSE (bottom, m) at day 5 for each version of the 13-km SHiELD and the contemporary operational GFS. Note that the operational GFS upgraded to v13 on 11 May 2016 and v14 on 19 July 2017.

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

and larger seasonal cycle but RMSE still improves every year and there is nearly no bias, especially in the 2019 version. Zhou et al (2019) give a more thorough description of precipitation forecast skill, including other metrics. Probability distribution functions (PDFs) of precipitation (Figure 7) show that all of the versions depicted here have a low bias in the frequency of moderate precipitation and a high bias of both light and heavy precipitation rates compared to TRMM, although versions of SHiELD using the GFDL microphysics (2017 and later) modestly alleviate these biases. Both the GFS and all versions of 13-km SHiELD capture the observed CONUS PDF very well.

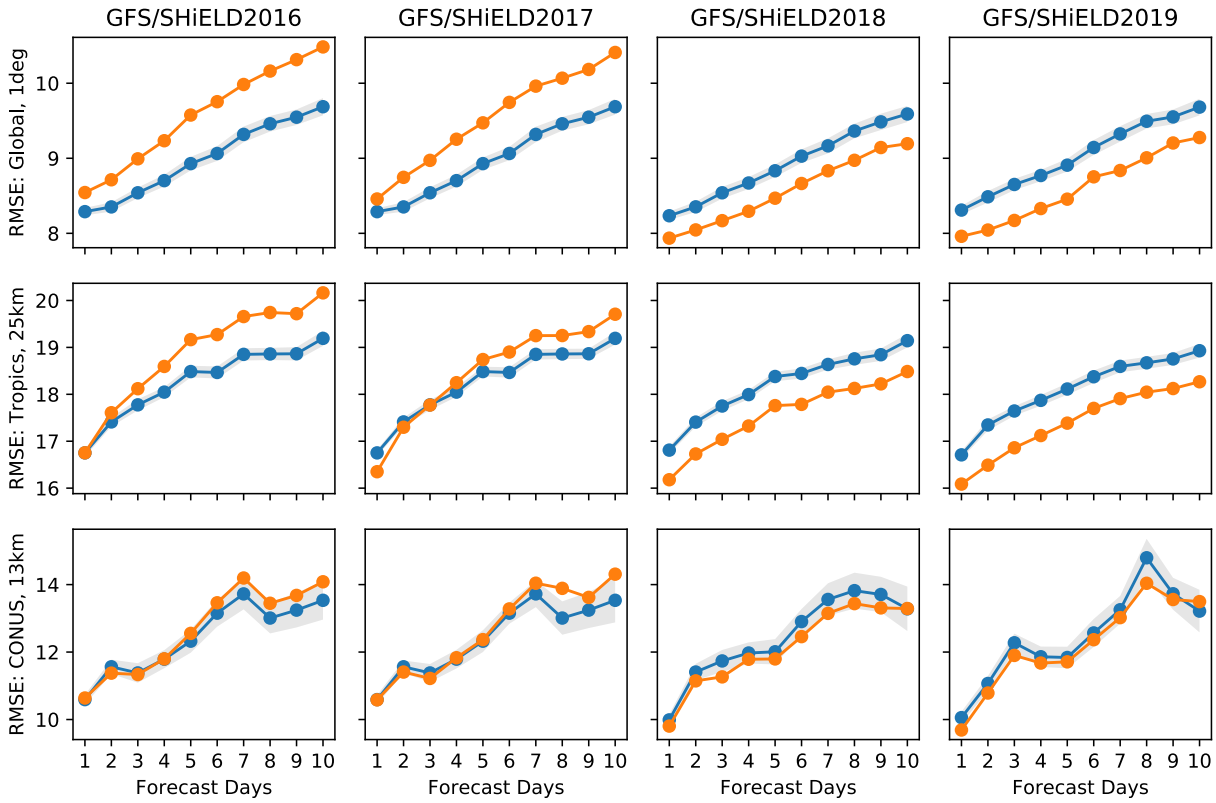


Figure 5. RMSE of 24-hour precipitation (mm) for different versions of 13-km SHiELD (orange) compared to contemporary GFS (blue). Each version's results are aggregated over the same two-year of hindcast periods plotted in Figures 3 and 4. Top row: Global verification vs. GPCP dataset (regridded to 1 degree); middle row: tropics (30S--30N) verification vs. TRMM

dataset (regridded to 25 km); bottom row: CONUS verification vs. StageIV dataset (regridded to 13 km). Gray shading is the 95% confidence interval.

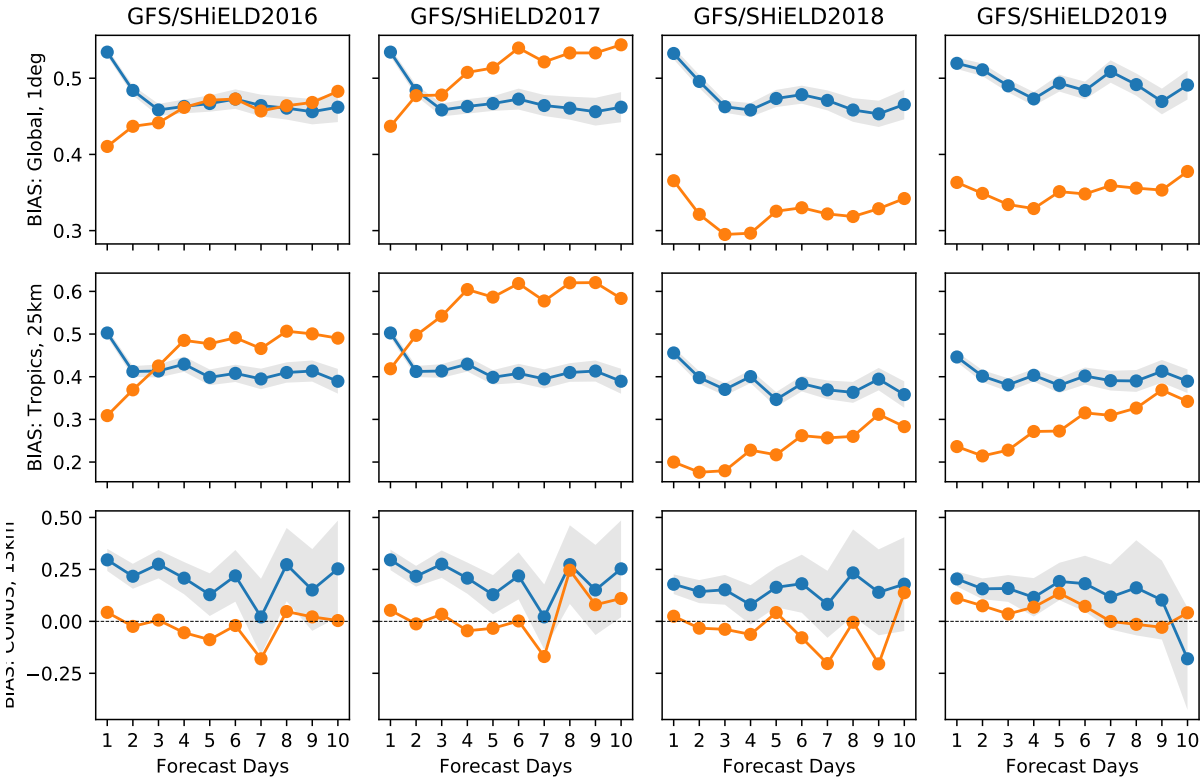
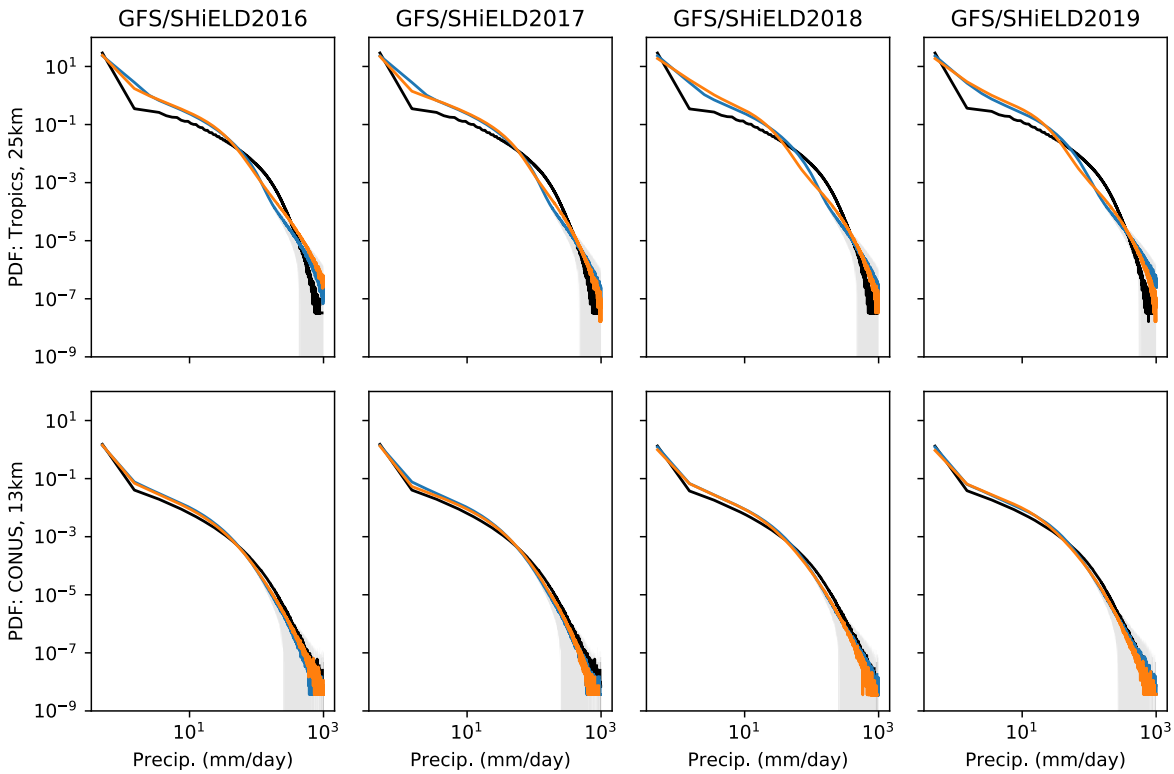


Figure 6. As in Figure 5 but for precipitation bias (mm d^{-1}), the arithmetic difference between means from the model and observations. Negative values imply too little mean precipitation.



372

373 **Figure 7.** Precipitation PDF for 13-km SHiELD (orange) compared to contemporary GFS (blue).
374 Top: Tropical (30S--30N) precipitation vs. TRMM (black). Bottom: CONUS precipitation vs.
375 StageIV (black).

376 Another sensible weather metric is 2-m temperature, which has an interesting
377 development history (Figure 8). The initial 2016 version of SHiELD had a very small warm bias,
378 significantly less than the small (0.3 K) warm bias of the operational GFS. The 2018 version of
379 SHiELD, which otherwise had significant improvements in other skill metrics, developed a cool
380 bias which increased to 0.6 K by day 10. Investigation traced the cool bias to two sources: the
381 switch from the hybrid EDMF PBL to YSU, which by default has significantly less near-surface
382 mixing and thereby allows the surface to cool too much, and the change in how cloud droplets
383 absorb radiation when the Inline GFDL Microphysics was introduced. In 2019 the cloud-
384 radiation interactions were significantly revised, and the background diffusion in the YSU PBL
385 was increased, which significantly reduced both the cold bias and the error in 2-m temperature.
386 The cold bias in SHiELD 2019 ranges from 0.1 K on the first day to 0.35 K on day 10, which is
387 approximately equal to the positive bias of the operational GFS.

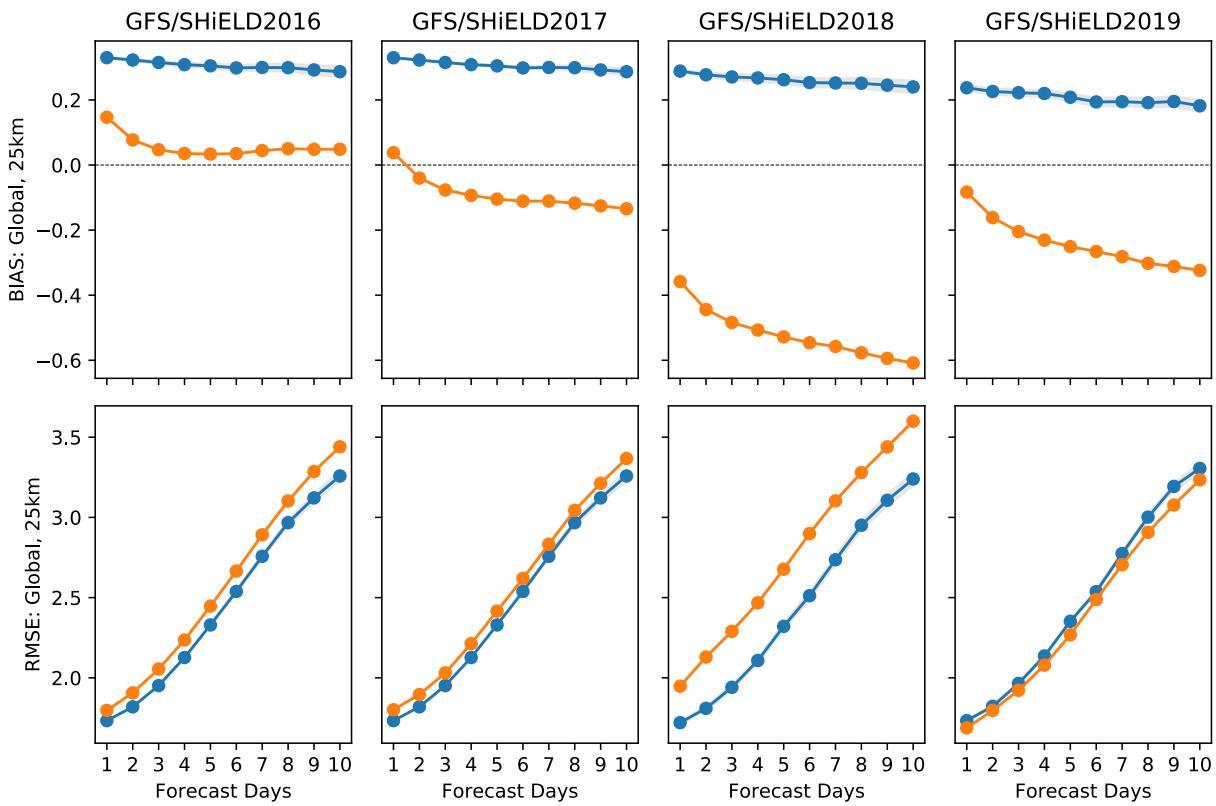


Figure 8. Global 2-m temperature (deg K) bias (top) and RMSE (bottom) for 13-km SHiELD (orange) compared to contemporary GFS (blue), both validated against ERA5 Reanalysis (Hersbach et al. 2020).

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

using the positive-definite (PD) advection scheme for water vapor and microphysical tracers. Results from T-SHiELD 2018 simulations of Major Hurricane Irma, initialized prior to its rapid intensification, show that a simulation using the older monotonic advection scheme (Figure 9) produces a gradually expanding vortex that does not intensify. Meanwhile, the simulation with the new PD scheme and *no other changes to the physics or dynamics, including advection of dynamical quantities,* produces an intensifying storm with a contracting eyewall. Notably, the vertical velocity within the eyewall is much more coherent with the PD scheme and is continually displaced within the eyewall, which we suspect may be driving both the intensification of Irma and a continued contraction of the eye, as well as contributing to enhanced precipitation within the eyewall. For this reason, the positive-definite advection scheme was selected for T-SHiELD 2018.

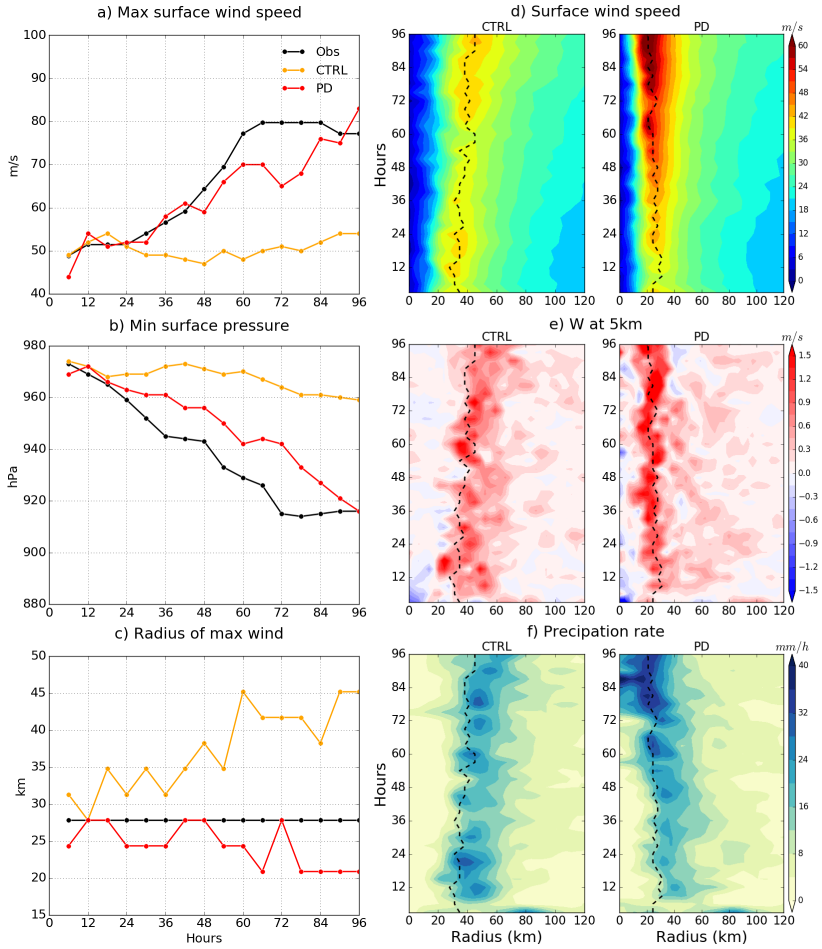


Figure 9. Hurricane Irma (2017) forecast initialized at 00 UTC 03 September 2017. Left column shows time-series plots of maximum 10-m winds (a), minimum central pressure (b), and RMW (c) compared against extended Best Track observations (Demuth et al. 2006). Right column shows time-radius plots of azimuthally averaged (d) 10-m winds (e) 5-km vertical velocity and (f) precipitation rate from forecasts of Hurricane Irma initialized 03 September 2017, from a

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.

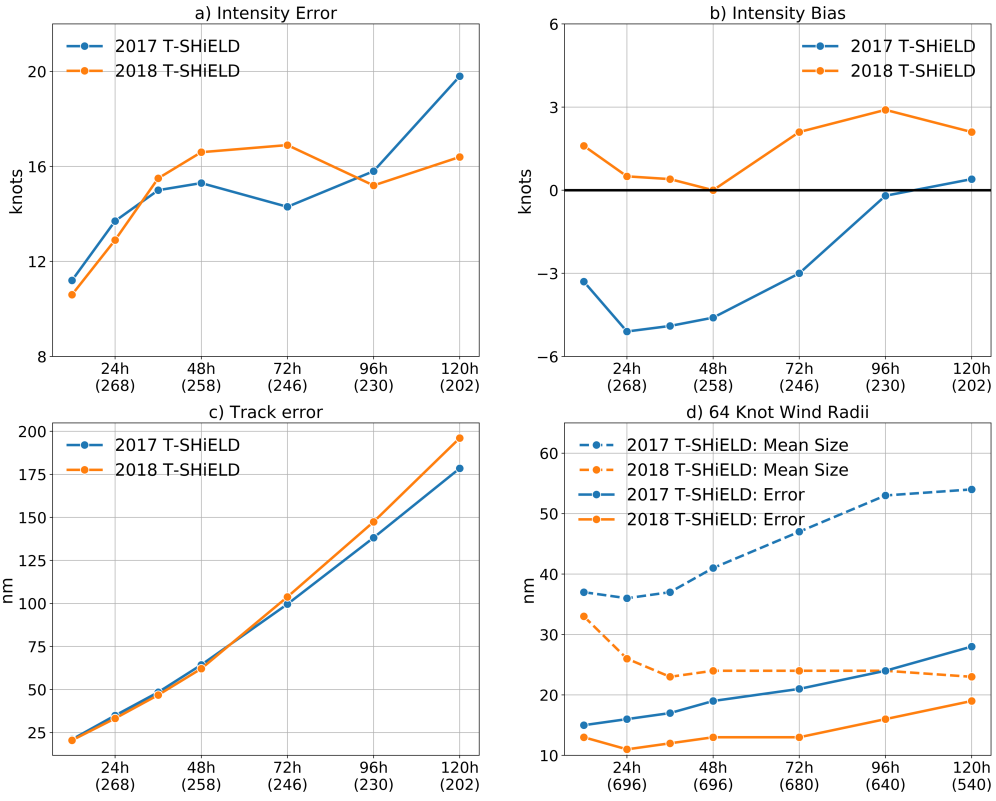


Figure 10. Verification of T-SHiELD 2017 and 2018 during the 2017 Atlantic hurricane season against the Best Track Dataset: intensity (a) error and (b) bias; (c) track error; (d) 64-kt (33 m s^{-1}, hurricane-force) radii. Units shown (kt, nautical miles) are standard for US operational prediction. In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

parentheses; in (d) the number in parentheses is the number of storm quadrants available for validation.

The multiple changes in the 2018 version of T-SHiELD combined to create tropical cyclones which are stronger overall (Figure 10a,b), with little to no bias towards more intense storms at all lead times. There is a minor degradation in track error in the 2018 version at longer lead times (Figure 10c). The adoption of the PD scheme and YSU PBL scheme likely created forecasts of more intense storms mitigated by the introduction of the interactive mixed-layer ocean. While the weak bias of the 2017 version was alleviated, intensity predictions were not appreciably improved except at 120-h lead time, and in fact were degraded between 36 and 72 hours after initialization. These results show once again the great challenge of improving intensity prediction. The reduction in RMWs in simulations using the PD scheme will be discussed in more detail in a forthcoming manuscript.

3.3 C-SHiELD Nest for Continental US Convection

C-SHiELD was designed to efficiently reach convective-scale resolutions in a global domain, in this case to replicate the capability of regional convective-scale models for continental convection such as the 3-km NAM Nest and the members of the High-Resolution Ensemble Forecast (HREF). C-SHiELD also is designed to extend convective-scale forecasts beyond the 18-to 60-hour ranges of existing US operational CONUS models into the medium-range timescales and beyond. The nested domain of C-SHiELD serves as a prototype for the Regional Forecast System (RFS; Carley et al. 2020) and the Rapid-Refresh Forecast System (RRFS; Alexander et al. 2020), both using the regional domain capability being developed within FV3.

The 2017 version of C-SHiELD is described in Harris et al. (2019). Modified versions of C-SHiELD with different microphysics and PBL schemes are described in Zhang et al. (2018) and Snook et al. (2019). C-SHiELD 2018 saw considerable updates as shown in Table 2; C-SHiELD 2019 added incremental updates, including re-configuration of the numerical diffusion and GFDL microphysics. We will limit our discussion to the evolution of broad forecast characteristics, but we will perform year-round validation instead of restricting the analysis to a single season. The time periods evaluated are given in Table 1. The exception is for the Surrogate Severe verification below, which is only verified for peak severe weather season of April to August of each year.

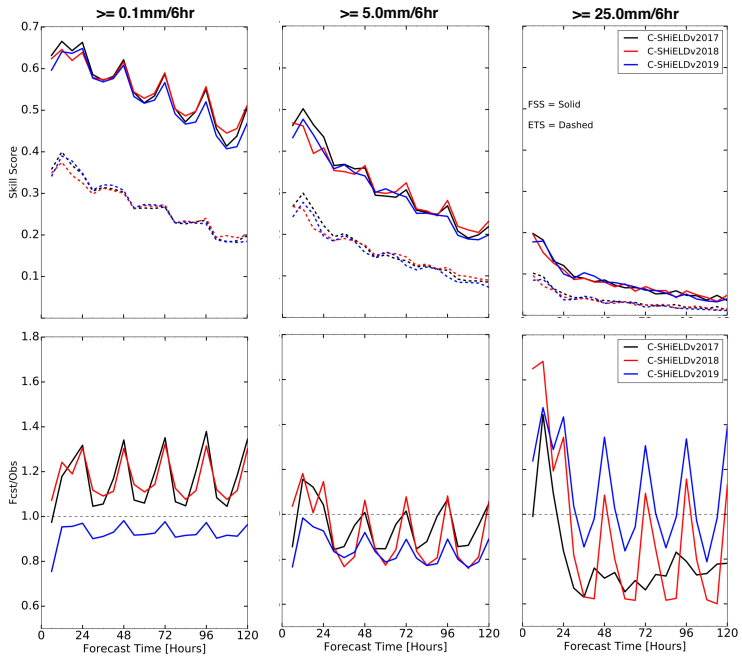


Figure 11. Precipitation skill scores (top) and bias score (bottom) vs. StageIV for 6-hr CONUS precipitation in three versions of C-SHiELD, given for precipitation events greater than three six-hourly accumulation thresholds (0.1, 5.0, and 25.0 mm). Skill scores are given for both Equitable Threat Score (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

the significantly higher updraft helicities in FV3-based models (Potvin et al. 2019). This in turn is likely due to the emphasis on vorticity in the horizontal discretization as described in Harris2019.

The maximum FSS in the 2018 and 2019 versions is about 0.8, on par with operational and research convective-scale models (cf. Sobash et al. 2019) and significantly higher than the 2017 version. There is a uniform over-prediction bias for all but the highest UH thresholds (Figure 12, bottom row). This bias was significant in the 2017 version but is decreased every year for most threshold-radius combinations, and for the highest-FSS combination decreases from 0.47 in 2017 to 0.22 in 2019. C-SHiELD 2019 still has a high frequency bias except for the very highest UH thresholds, as it is still too aggressive at creating strong storms.

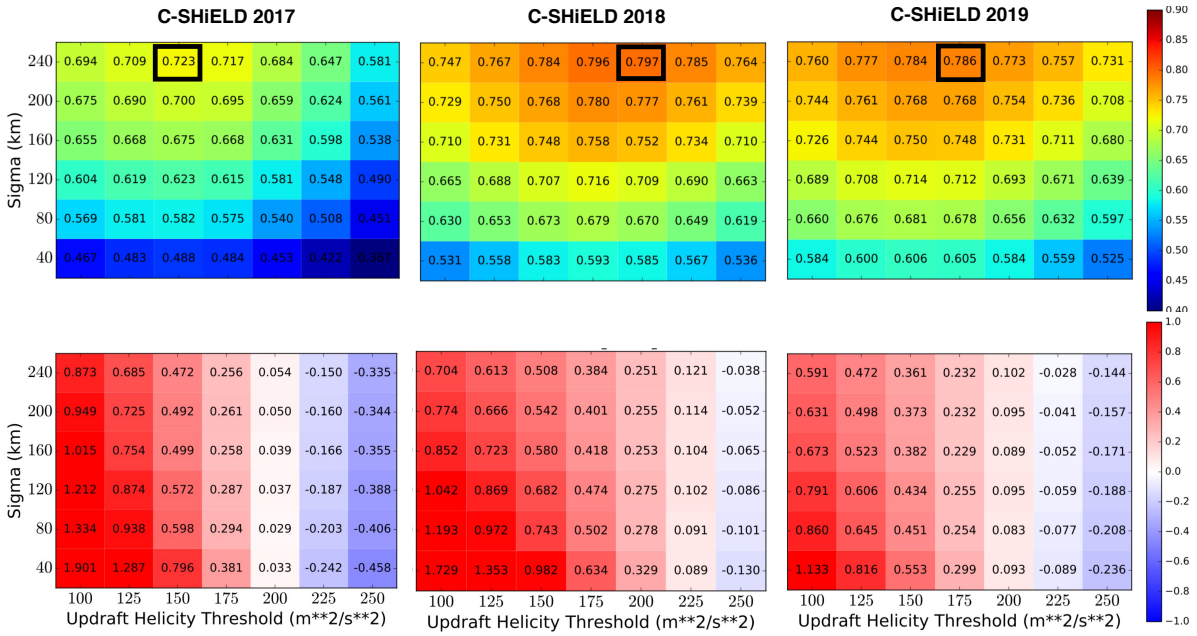


Figure 12. FSS (top) and Bias score minus 1 (bottom) for surrogate severe predictions with 12–36 hour lead times for three versions of C-SHiELD initialized at 00Z. Heavy black outline corresponds to the combination of UH threshold (m^2s^{-2}) and smoothing radius (sigma, km) giving the highest FSS.

We also investigate if skillful prediction of severe weather is possible beyond the first forecast day. Figure 13 shows surrogate severe FSS for days 1 through 4 (hours 12–36, 36–60, 60–84, and 84–108, respectively). The FSS value is not as high on later days as on the first, but even on day 4 the FSS is still a respectable 0.74, indicating that there is skill in predicting severe weather multiple days in advance. These high skill scores may be partially due to the relatively large smoothing radius of 240 km.

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.

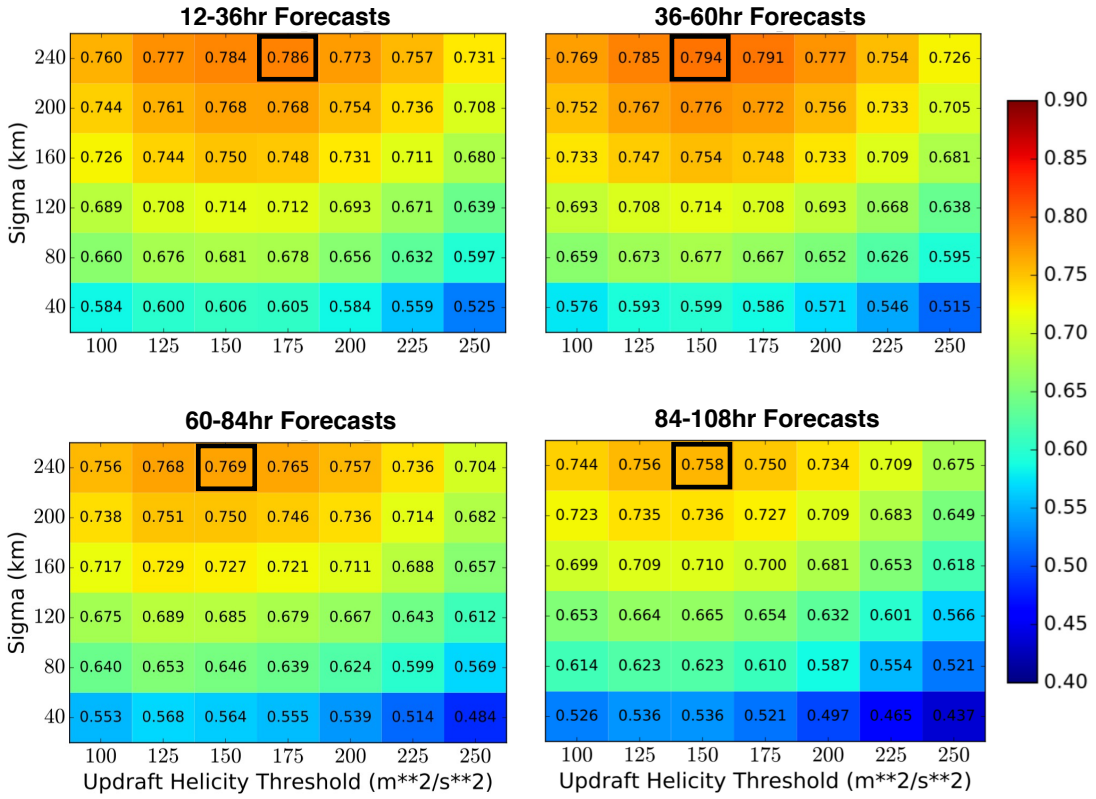


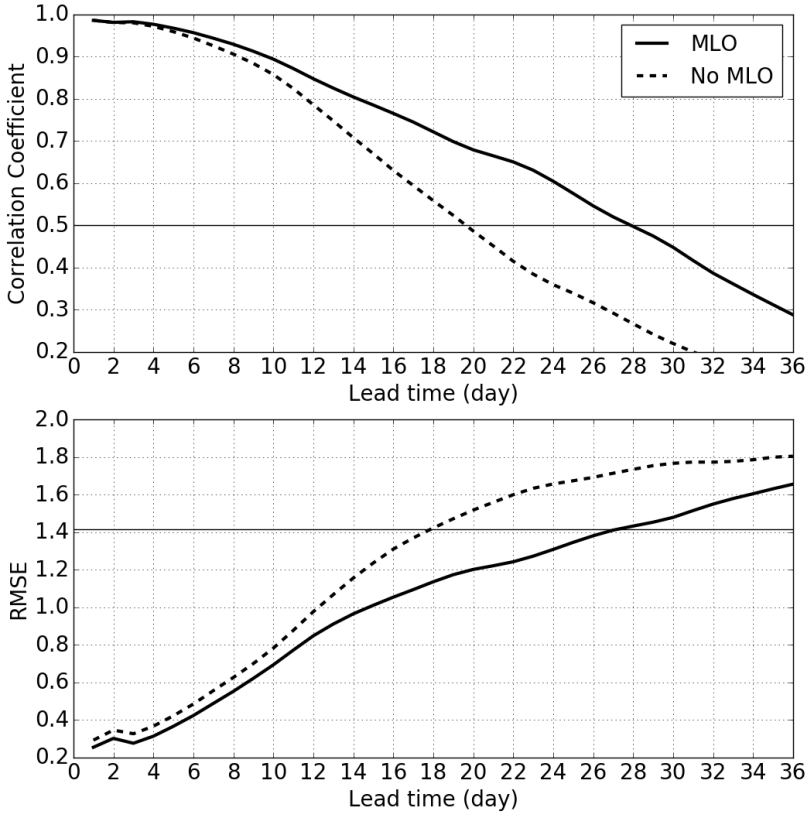
Figure 13. FSS for surrogate severe predictions at different lead times for 00Z initializations of C-SHiELD 2019.

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.

549 The MJO plays a major role in subseasonal variability but has been a challenge for many
550 models to predict or even simulate reasonably (Kim et al. 2018). To explore the MJO prediction
551 skill of S-SHiELD we performed 92 40-day predictions, one initialized at 00Z every two days
552 from 1 October 2011 to 31 March 2012, covering the active Dynamics of the MJO (DYNAMO;
553 Yoneyama et al. 2013) observation period. The Real-time Multivariate MJO Index (RMM;
554 Wheeler and Hendon 2004) is calculated for the hindcasts following the methodology of Xiang
555 et al. (2015) and Vitart et al (2017). For each hindcast we compute daily-mean anomalies of
556 outgoing longwave radiation (OLR) and zonal wind at 200mb (U200) and 850mb (U850),
557 averaged between 15S-15N. These forecast anomalies are not bias-corrected since we use
558 observed climatology as reference instead of model climatology. We then subtract the averaged
559 anomalies of the previous 120 days from the total anomalies to remove the signals of interannual
560 and longer time-scale variability; observed anomalies are appended to the anomalies in the
561 hindcast. The normalized U200, U850 and OLR anomalies are then projected onto the pre-
562 computed Empirical Orthogonal Functions (EOFs) from Wheeler and Hendon (2004) to obtain
563 the two RMM indices.

564 We find that S-SHiELD with the MLO (Figure 14) has good skill (correlation > 0.7) out
565 to 19 days and useful skill (correlation > 0.5) out to 28 days. The RMSE likewise shows similar
566 skill (RMSE < $\sqrt{2}$ out to 27 days). This skill may not be representative of other time periods
567 given that skill is known to be higher during strong events (cf. Xiang et al 2015) and the period
568 of evaluation is relatively short. However, this result does give us confidence that S-SHiELD
569 simulates the MJO well enough for useful S2S prediction. We plan to expand our evaluation of
570 the MJO in forthcoming work.



571
572 **Figure 14.** Prediction skill of the MJO's RMM indices in S-SHiELD with (solid) and without

(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.

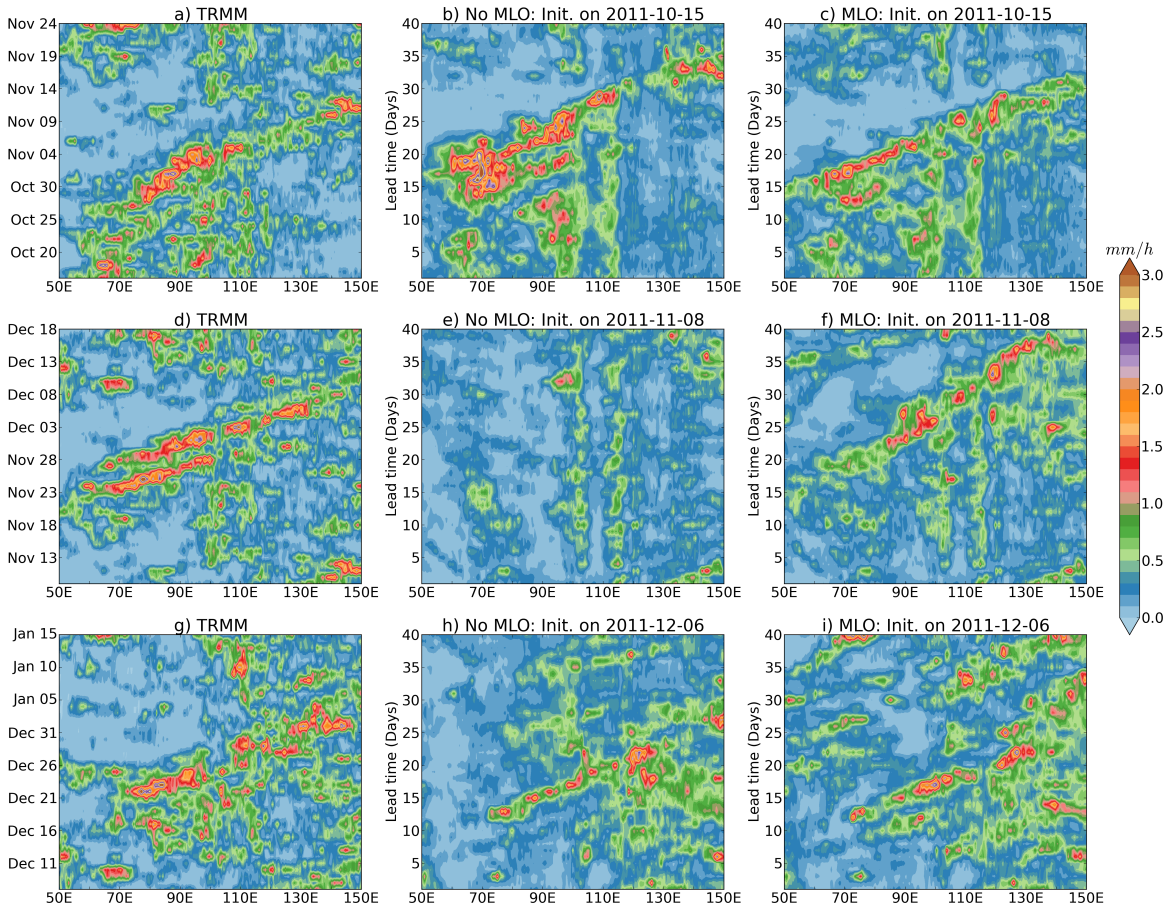


Figure 15. Precipitation (averaged from 5S--5N) from TRMM (left), S-SHiELD without MLO (center), and S-SHiELD with MLO (right), for initializations at (top) 15 October (middle) 8 November and (bottom) 6 December 2011.

Klingaman and DeMott (2020) found that climate models exaggerate the effect of ocean coupling on the MJO by over-intensifying the MJO in El Niño years. S-SHiELD does not have a coupled dynamical ocean and nudges towards climatology, and so can only represent the El Niño-Southern Oscillation (ENSO) state at initialization; indeed, the DYNAMO period was during a La Niña event (see https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php). Hence, this ENSO contamination of the link between ocean coupling and the MJO is not present in S-SHiELD.

The diurnal cycle of precipitation is another challenge for climate models. Covey et al (2016) found nearly all climate models, even the 30-km resolution GFDL HiRAM, struggle with both the phase and amplitude of the diurnal cycle, especially over land and during boreal summer. Figure 16 presents the JJA diurnal cycle from a 10-year S-SHiELD simulation with MLO SSTs nudged towards climatology, with results from 13-km SHiELD hindcasts shown for reference. We find that the observed phase of the diurnal cycle is beautifully matched by S-SHiELD, over both land and ocean. Most notably the CONUS evening maximum of precipitation is reproduced. However, the amplitude of the cycle is biased low over land areas, possibly due to the inability of S-SHiELD's 25-km grid to produce the propagating mesoscale convective systems characteristic of heavier warm-season precipitation events. This appears to

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).

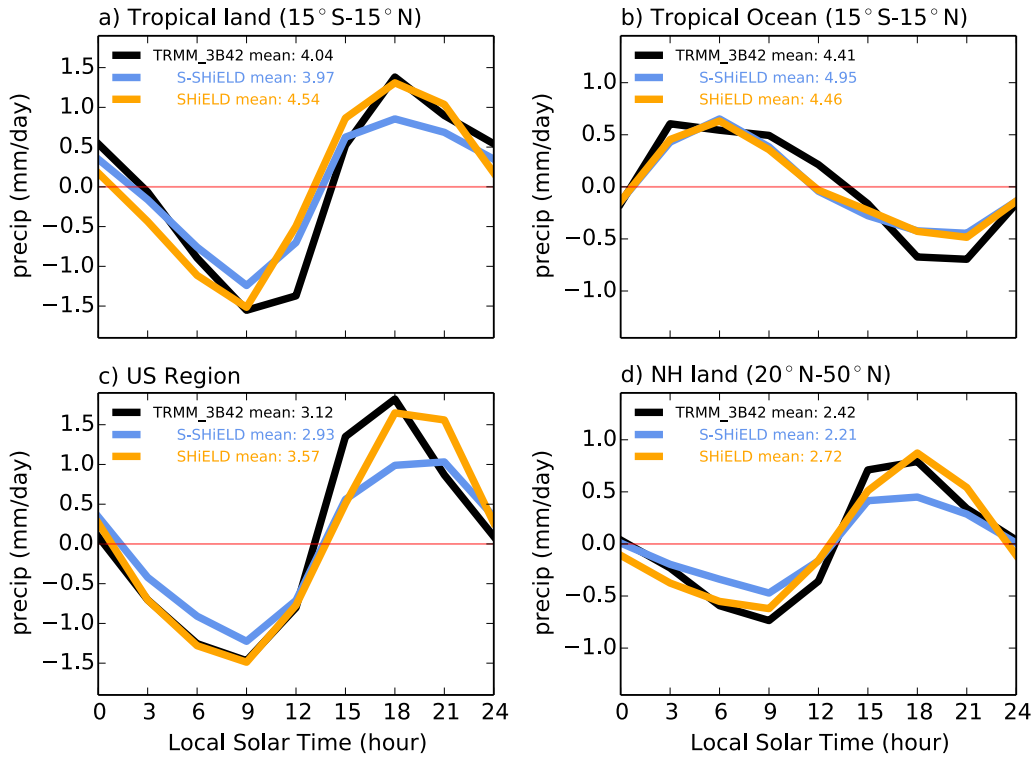


Figure 16. JJA diurnal cycle of precipitation as a function of local solar time in a 10-year S-SHiELD climate integration with the MLO nudged towards climatological SSTs, and from days 6--10 of three years of 13-km SHiELD hindcasts (initialized every five days), compared to TRMM 2011-2018 observations.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

of unified modeling in which there is a single modeling system with one codebase, one executable, one preprocessor, one set of runscripts, and one set of post-processing tools. This greatly simplifies the modeling suite and allows improvements to be exchanged between configurations.

The fundamental characteristics of SHiELD compared to previous-generation and existing operational models are documented in this and other publications. For some applications we have previously demonstrated capabilities similar to that of existing modeling systems, such as severe-storm prediction in C-SHiELD (Harris et al 2019) and tropical cyclone intensity prediction in T-SHiELD (Hazelton et al 2018a,b). We have shown significant improvements over existing models, especially over existing global models, for large-scale and hurricane prediction skill in 13-km SHiELD (Zhou et al 2019, Chen et al 2019a), and the diurnal cycle and MJO prediction in S-SHiELD. We have even shown entirely new possibilities for prediction modeling, such as skillful hurricane intensity forecasts in 13-km SHiELD (Chen et al. 2019b), and the possibility of medium-range convective-scale prediction in C-SHiELD. *Ultimately, the true strength of SHiELD is that all of these characteristics are demonstrated in the same modeling system.*

SHiELD is designed to be an experimental research modeling system, with a particular set of scientific goals set by its developers, and thereby is more restricted in scope than the GFS, HAFS, RFS, and other general-purpose models intended for operational weather forecasting and to support broad audiences of users. While improved prediction skill is a major scientific goal and an important "vital sign" of model development, we also develop SHiELD as a means to demonstrate new modeling capabilities. SHiELD is also intended to be principally a physical atmosphere modeling system and is not intended for research into oceanic dynamics, decadal-to-centennial projection, biogeochemistry, or other topics taking place at either longer timescales or greater complexity than SHiELD is designed for. Improvements within SHiELD can be seamlessly transitioned into other FV3-based models that do address these topics, including other Unified Forecast System models and the FV3-based coupled earth-system models at GFDL, within NASA, NCAR, and elsewhere. As such SHiELD's progress will continue to contribute to the development and improvement of these modeling systems. SHiELD is a part of GFDL's fourth-generation modeling suite (GFDL 2019, Figures 1 and 2) and shares common infrastructure with CM4, ESM4, and SPEAR. SHiELD uses a different physics suite and land model from the other GFDL configurations, but otherwise is constructed similarly. Advances can then be exchanged between the configurations, allowing for mutual improvement, seamless cross-timescale modeling, and potentially unification of GFDL's weather and climate modeling efforts. A significant two-way interaction between SHiELD and other UFS configurations (GFS, HAFS, RFS, etc.) is taking and promises to continue driving furthered improvement of all UFS applications.

Further development of SHiELD, including both FV3 and the SHiELD physics, will continue to improve the prediction skill of the configurations, address issues which have been identified, and broaden the scope towards new applications. As computing power allows, models will be pushed to higher horizontal and vertical resolution, physical processes developed to improve simulation quality and prediction skill, and to address emerging scientific questions. New capabilities within FV3, including regional and doubly periodic domains, will permit efficient simulation of processes at kilometer and sub-kilometer scales for basic science and for process studies to improve physical parameterizations. We are also working on a native SHiELD

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:

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

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

$B_{oi} = Q_i^- + Q_i^+$

$\Delta A_i = Q_i^+ - Q_i^-$

$A_{4i} = -3\ B_{oi}$

If abs($\hat{q}_{i+1/2} - \hat{q}_{i-1/2}$) > $-A_{4i}$ and $\overline{q_i} + \Delta A_i{}^2/(4A_{4i}) + \frac{1}{12}A_{4i} < 0$ then

    If $Q_i^- Q_i^+ > 0$ then

        $Q_i^- = Q_i^+ = B_{oi} = 0$

    Elseif dAi > 0 then

        $Q_i^+ = -2 * Q_i^-$

        $B_{oi} = -Q_i^-$

    Else

        $Q_i^- = -2 * Q_i^+$

        $B_{oi} = -Q_i^+$

Appendix B: Split and In-line GFDL Microphysics

The GFDL microphysics, a single-moment six-category microphysics, has its origin in the microphysics of Lin et al. (1983) as implemented within GFDL ZETAC (Pauluis and Garner, 2006; Knutson et al., 2007, 2008) with further developments from Lord et al. (1984) and Krueger et al. (1995). It was later substantially revised for use in HiRAM (Chen and Lin, 2011, 2013; Harris et al., 2016; Gao et al., 2017, 2019) by adding the following updates:

1. Time-splitting is applied between warm-rain and ice-phase processes, with the warm-rain processes called twice per invocation.
2. PPM is applied for sedimentation of all condensate species except cloud water, ensuring shape preservation and stability.
3. The heat content of condensates is included when heating/cooling grid cells.
4. Scale awareness is achieved by assuming a horizontal subgrid distribution and a second-order vertical reconstruction for autoconversion processes with a slope which increases with grid-cell width.
5. Additional microphysical processes, including ice nucleation and cloud ice sedimentation, were introduced.

In the Split GFDL Microphysics first implemented within SHiELD, microphysical processes were divided into fast and (relatively) slow processes, where the fast processes (primarily phase changes and latent heating/cooling) are updated after the vertical remapping in FV3, while the slower processes remain in the physical driver. More recently, the entire GFDL microphysics was Inlined within the dynamical core. The advantages of Inlining are 1) to separate the physical processes based on different time scales to better interact with dynamics processes; and 2) to be able to make the physical parameterization thermodynamically consistent

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).

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. In *33rd Conf. on Hurricane and Tropical Meteorology*.

Chen, J. H., Lin, S. J., Magnusson, L., Bender, M., Chen, X., Zhou, L., ... & Harris, L. (2019). Advancements in hurricane prediction with NOAA's next-generation forecast system. *Geophysical Research Letters*, *46*(8), 4495-4501.

Chen, J. H., Lin, S. J., Zhou, L., Chen, X., Rees, S., Bender, M., & Morin, M. (2019). Evaluation of Tropical Cyclone Forecasts in the Next Generation Global Prediction System. *Monthly Weather Review*, *147*(9), 3409-3428.

Chun, H. and J. Baik. (1994). Weakly Nonlinear Response of a Stably Stratified Atmosphere to Diabatic Forcing in a Uniform Flow. *J. Atmos. Sci., 51*, 3109–3121, https://doi.org/10.1175/1520-0469(1994)051<3109:WNROAS>2.0.CO;2

Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., Boukabara, S., & Brown, P. (2005). Atmospheric radiative transfer modeling: A summary of the AER codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *91*(2), 233–244.

Colella, P., & Woodward, P. R. (1984). The piecewise parabolic method (PPM) for gas-dynamical simulations. *Journal of computational physics*, *54*(1), 174-201.

Covey, C., P.J. Gleckler, C. Doutriaux, D.N. Williams, A. Dai, J. Fasullo, K. Trenberth, and A. Berg. (2016). Metrics for the Diurnal Cycle of Precipitation: Toward Routine Benchmarks for Climate Models. J. Climate, 29, 4461–4471. https://doi.org/10.1175/JCLI-D-15-0664.1

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research: Oceans*, *109*(C12).

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J. H., Dunne, K. A., ... & Harrison, M. J. (2020). SPEAR–the next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *Journal of Advances in Modeling Earth Systems*.

Demaria, E. M. C., Palmer, R. N., and Roundy, J. K. (2016). Regional climate change projections of streamflow characteristics in the Northeast and Midwest U.S. *Journal of Hydrology: Regional Studies, 5*, March 2016, 309-323. https://doi.org/10.1016/j.ejrh.2015.11.007.

DeMott, C. A., Klingaman, N. P., Tseng, W.-L., Burt, M. A., Gao, Y., & Randall, D.A. (2019) The convection connection: How ocean feedbacks affect tropical mean moisture and MJO propagation. *Journal of Geophysical Research Atmospheres, 124*, 11910– 11931. https://doi.org/10.1029/2019JD031015

Demuth, J. L., DeMaria, M., & Knaff, J. A. (2006). Improvement of advanced microwave sounding unit tropical cyclone intensity and size estimation algorithms. *Journal of applied meteorology and climatology*, *45*(11), 1573-1581.

839 Dong, J., and coauthors. (2020). The Evaluation of Real-Time Hurricane Analysis and Forecast
840 System (HAFS) Stand-Alone Regional (SAR) Model Performance for the 2019 Atlantic
841 Hurricane Season. *Atmosphere, 11*(6)*,* 617. https://doi.org/10.3390/atmos11060617

842 Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., et al.
843 (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model
844 description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, 12,
845 e2019MS002015. https://doi.org/10.1029/2019MS002015

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

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

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

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

853 Gao, K., Chen, J. H., Harris, L. M., Lin, S. J., Xiang, B., & Zhao, M. (2017). Impact of
854 intraseasonal oscillations on the tropical cyclone activity over the Gulf of Mexico and western
855 Caribbean Sea in GFDL HiRAM. *Journal of Geophysical Research: Atmospheres*, *122*(24), 13-
856 125.

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

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

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

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

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

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

Han, J., Wang, W., Kwon, Y. C., Hong, S.-Y., Tallapragada, V., & Yang, F. (2017). Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. *Weather and Forecasting*, *32*, 2005–eeting2017.

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.

Harris, L., Lin, S. J., & Chen, J. H. (2014, May). Great Plains warm-season precipitation in a two-way nested high-resolution GCM. In *EGU General Assembly Conference Abstracts* (Vol. 16).

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*(11), 4293-4314.

Harris, L. M., Rees, S. L., Morin, M., Zhou, L., & Stern, W. F. (2019). Explicit Prediction of Continental Convection in a Skillful Variable-Resolution Global Model. *Journal of Advances in Modeling Earth Systems, 11(*6), 1847-1869.

Hazelton, A. T., Bender, M., Morin, M., 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 and Forecasting*, *33*(5), 1317-1337.

Hazelton, A. T., Harris, L., & Lin, S. J. (2018). Evaluation of tropical cyclone structure forecasts in a high-resolution version of the multiscale GFDL fvGFS model. *Weather and Forecasting*, *33*(2), 419-442.

Hazelton, A., et al. (2020). The Global-Nested hurricane analysis and forecast system (HAFS): results from the 2019 Atlantic hurricane season. *In 100th Annual Meeting.* AMS.

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., ... & Wittenberg, A. T. (2019). Structure and performance of GFDL's CM4. 0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691-3727.

Held, I. M., Zhao, M., & Wyman, B. (2007). Dynamic radiative-convective equilibria using GCM column physics. *Journal of the Atmospheric Sciences, 64*(1), 228-238.

Hersbach, H, Bell, B, Berrisford, P, et al. The ERA5 global reanalysis. *Q J R Meteorol Soc*. 2020; 146: 1999– 2049. https://doi.org/10.1002/qj.3803

Hogan, R. J., & Mason, I. B. (2011). Deterministic forecasts of binary events. *Forecast verification: A practitioner's guide in atmospheric science*, 31-59.

Hong, S. Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly weather review*, *134*(9), 2318-2341.

Hong, S.-Y. (2010). A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon. *Quarterly Journal of the Royal Meteorological Society, 136*, 1481–1496.

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

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

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

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

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

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

925 Landsea, C.W. and J.L. Franklin. (2013). Atlantic Hurricane Database Uncertainty and
926 Presentation of a New Database Format. *Mon. Wea. Rev., 141*, 3576–
927 3592. https://doi.org/10.1175/MWR-D-12-00254.1

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

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

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

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

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

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

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

946 McCormack, J. P., Eckermann, S. D., Siskind, D. E., and McGee, T. J. (2006). CHEM2D-OPP:
947 A new linearized gas-phase ozone photochemistry parameterization for high-altitude NWP and
948 climate models, *Atmos. Chem. Phys., 6*, 4943–4972. https://doi.org/10.5194/acp-6-4943-2006.

949 McGregor, J.L. (2015). Recent developments in variable-resolution global climate
950 modelling. *Climatic Change*, *129*, 369–380. https://doi.org/10.1007/s10584-013-0866-5

951 Moum, J. N., de Szoeke, S. P., Smyth, W. D., Edson, J. B., DeWitt, H. L., Moulin, A. J., ... &
952 Fairall, C. W. (2014). Air–sea interactions from westerly wind bursts during the November 2011
953 MJO in the Indian Ocean. *Bull. Amer. Meteorol. Soc.*, *95*(8), 1185-1199.

954 Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., Mayer, M., and Keeley, S. P. E. (2018).
955 Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS
956 cycle 43r1) for HighResMIP, *Geosci. Model Dev.*, 11, 3681-3712, https://doi.org/10.5194/gmd-
957 11-3681-2018

958 NCEP (2020). List of GFS Implementations. Available at
959 https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/implementations.ph
960 p . Last accessed 22 May 2020.

961 Pauluis, O., & Garner, S. (2006). Sensitivity of radiative–convective equilibrium simulations to
962 horizontal resolution. *Journal of the atmospheric sciences*, *63*(7), 1910-1923.

963 Pollard, R. T., Rhines, P. B., & Thompson, R. (1973). The deepening of the mixed layer.
964 *Geophysical Fluid Dynamics*, *3*, 381-404.

965 Potvin, C.K., J.R. Carley, A.J. Clark, L.J. Wicker, P.S. Skinner, A.E. Reinhart, B.T. Gallo, J.S.
966 Kain, G.S. Romine, E.A. Aligo, K.A. Brewster, D.C. Dowell, L.M. Harris, I.L. Jirak, F. Kong,
967 T.A. Supinie, K.W. Thomas, X. Wang, Y. Wang, and M. Xue. (2019). Systematic Comparison
968 of Convection-Allowing Models during the 2017 NOAA HWT Spring Forecasting Experiment.
969 *Wea. Forecasting, 34*, 1395–1416. https://doi.org/10.1175/WAF-D-19-0056.1

970 Putman, W M., and Shian-Jiann Lin (2007). Finite-volume transport on various cubed-sphere
971 grids. *Journal of Computational Physics*, 227(1), 55-78.

972 Putman, W. M., & Suarez, M. (2011). Cloud-system resolving simulations with the NASA
973 Goddard Earth Observing System global atmospheric model (GEOS-5). *Geophysical Research*
974 *Letters*, *38*(16).

975 Putman, W. M., & Suárez, M. J. (2017). GEOS Atmospheric Model: Challenges at Exascale.

976 Roberts, N.M. and H.W. Lean. (2008). Scale-Selective Verification of Rainfall Accumulations
977 from High-Resolution Forecasts of Convective Events. *Mon. Wea. Rev., 136*, 78–97.
978 https://doi.org/10.1175/2007MWR2123.1

979 Satoh, M. (2007). Global Cloud-Resolving Model Development and its seamless Applications to
980 Weather & Climate Researches. In *Abstract Collection of the Third China-Korea-Japan Joint*
981 *Conference on Meteorology*.

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.

1019 Wilson, T. H, and R. G. Fovell. (2018). Modeling the evolution and life cycle of radiative cold
1020 pools and fog. *Weather and Forecasting, 33*, 2031–220.

1021 Xiang, B., Zhao, M., Jiang, X., Lin, S. J., Li, T., Fu, X., & Vecchi, G. (2015). The 3–4-week
1022 MJO prediction skill in a GFDL coupled model. *Journal of Climate*, *28*(13), 5351-5364.

1023 Xu, K.-M., & Randall, D. A. (1996). A semiempirical cloudiness parameterization for use in
1024 climate models. *Journal of the atmospheric sciences*, *53*(21), 3084–3102.

1025 Yoneyama, K., C. Zhang, and C.N. Long. (2013). Tracking Pulses of the Madden–Julian
1026 Oscillation. Bull. *Amer. Meteor. Soc., 94*, 1871–1891. https://doi.org/10.1175/BAMS-D-12-
1027 00157.1

1028 Zhang, C., Xue, M., Supinie, T. A., Kong, F., Snook, N., Thomas, K. W., et al. (2019). How well
1029 does an FV3-based model predict precipitation at a convection-allowing resolution? Results from
1030 CAPS forecasts for the 2018 NOAA hazardous weather test bed with different physics
1031 combinations. *Geophysical Research*
1032 *Letters*, 46, 3523– 3531. https://doi.org/10.1029/2018GL081702

1033 Zhang, J.A., D.S. Nolan, R.F. Rogers, and V. Tallapragada. (2015). Evaluating the Impact of
1034 Improvements in the Boundary Layer Parameterization on Hurricane Intensity and Structure
1035 Forecasts in HWRF. *Mon. Wea. Rev., 143*, 3136–3155. https://doi.org/10.1175/MWR-D-14-
1036 00339.1

1037 Zhao, M., Golaz, J. C., Held, I. M., Ramaswamy, V., Lin, S. J., Ming, Y., ... & Guo, H. (2016).
1038 Uncertainty in model climate sensitivity traced to representations of cumulus precipitation
1039 microphysics. *Journal of Climate*, *29*(2), 543-560.

1040 Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018). The GFDL
1041 global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed
1042 SSTs. *Journal of Advances in Modeling Earth Systems*, 10, 691–734.
1043 https://doi.org/10.1002/2017MS001208

1044 Zhao, M., Held, I. M., Lin, S. J., & Vecchi, G. A. (2009). Simulations of global hurricane
1045 climatology, interannual variability, and response to global warming using a 50-km resolution
1046 GCM. *Journal of Climate*, *22*(24), 6653-6678.

1047 Zhao, Q., & Carr, F. H. (1997). A prognostic cloud scheme for operational NWP
1048 models. *Monthly Weather Review*, *125*(8), 1931–1953.

1049 Zhou, L., Lin, S.-J., Chen, J.-H., Harris, L. M., Chen, X., & Rees, S. L. (2019). Toward
1050 convective-scale prediction within the next generation global prediction system. *Bulletin of the*
1051 *American Meteorological Society*. https://doi.org/10.1175/BAMS-D-17-0246.1

1052