T. L. Black is retired.

Key Points:
- A limited area modeling capability has been developed for the Finite-Volume Cubed-Sphere dynamical core
- This capability was evaluated for a month-long period against a similarly configured two-way nest driven by a global model
- The limited area model is statistically comparable to the two-way nest for the first 24 h, with minor degradation by 48–60 h

Correspondence to:
J. R. Carley,
jacob.carley@noaa.gov

A Limited Area Modeling Capability for the Finite-Volume Cubed-Sphere (FV3) Dynamical Core and Comparison With a Global Two-Way Nest

T. L. Black[1], J. A. Abeles[2], B. T. Blake[2], D. Jovic[2], E. Rogers[1], X. Zhang[2], E. A. Aligo[2], L. C. Dawson[1], Y. Lin[3], E. Strobach[2], P. C. Shafran[2], and J. R. Carley[1]

[1]NOAA/NCEP/EMC, College Park, MD, USA, [2]IMSG at NOAA/NCEP/EMC, Rockville, MD, USA, [3]NOAA/NCEP/NCO, College Park, MD, USA

Abstract The development of a limited area model (LAM) capability for the nonhydrostatic Finite-Volume Cubed-Sphere (FV3) dynamical core is described and compared with a globally nested approach featuring two-way feedback. Comparisons of the computational performance of the LAM relative to the two-way nest reveal that the LAM configuration exhibits considerable improvement in efficiency. High-resolution (i.e., 3-km) LAM and nest configuration forecasts covering a 1-month period show statistically comparable results for most parameters. Forecast differences between the two configurations primarily arise in the upper air temperature and height fields, which show a statistically significant increase in the magnitude of negative biases in geopotential height and upper-air temperature using the LAM configuration relative to the nest at forecast lead times >24-h. Precipitation forecasts over the full 60-h forecast period are also evaluated and depict no statistically significant differences between the two configurations, with the nest configuration exhibiting slightly improved scores. Overall results suggest that while the FV3 LAM approach can introduce degradations into the forecast relative to the two-way interactive nest at lead times >24-h, these errors are generally small in magnitude and are accompanied by considerable improvement in computational efficiency.

Plain Language Summary In this study, we describe and evaluate a new limited area model (LAM) capability for the Finite-Volume Cubed-Sphere dynamical core. This capability provides a way to run the dynamical core over any region without the need for simultaneous integration of a global model counterpart, thus saving considerable computational resources. However, this framework is susceptible to errors from lateral boundaries, which are provided by an external model at a coarse temporal frequency compared to a global model employing a nest with two-way feedback. Short range forecasts at 3-km grid spacing over the contiguous United States show generally similar performance between the limited area and nest configurations for a month-long period in 2019, with the nest showing slightly better forecast scores near the end of the studied forecast length of 60 h. These results suggest that use of the LAM performs similarly to the more sophisticated, and computationally expensive, two-way nest configuration for convection-allowing, short range forecasts to 60 h.

1. Introduction

In 2016, the global Finite-Volume Cubed-Sphere (FV3) dynamical core (Harris & Lin, 2013, 2014; Lin, 2004; Putman & Lin, 2007) was selected as the atmospheric dynamical core to underpin the development of the National Centers for Environmental Prediction (NCEP) Unified Forecast System (UFS; https://ufscommunity.org/) through the Next Generation Global Prediction System project. In June 2019, a significant first step forward in the unification of the NCEP modeling suite took place when the FV3 dynamical core replaced the spectral dynamical core in the Global Forecast System (GFS) with the operational implementation of GFS version 15 (https://www.noaa.gov/media-release/noaa-upgrades-us-global-weather-forecast-model). Simultaneously, development work has been underway at NCEP to adopt the new FV3 dynamical core for regional modeling applications.

To provide regionally enhanced resolution, the FV3 provides a combination of two mesh refinement options: (a) stretching of the parent global domain (Harris et al., 2016) and (b) a two-way nest (Harris & Lin, 2013, 2014) recently extended to include nonhydrostatic motions (Harris, Chen, et al., 2020; Harris

et al., 2019). These options have worked demonstrably well in a variety of applications, including tropical and hazardous convective weather scenarios (Harris, Zhou, et al., 2020; Hazelton et al., 2018; Potvin et al., 2019; Zhang et al., 2019; Zhou et al., 2019). While this approach has been successful, it necessitates a simultaneous integration of a global model along with the higher resolution nest, which carries an additional computational cost.

Limited area model (LAM) applications have been in use for over 40 years in operational Numerical Weather Prediction (NWP) owing to their ability to provide regionally refined high resolution without requiring the costly integration of a high resolution global model (e.g., Benjamin et al., 2016; Black, 1994; Gerrity, 1977; Hoke et al., 1989; Janjić et al., 2001). There are a variety of approaches for the specification and application of lateral boundary conditions (LBCs) to ensure proper inflow/outflow of prognostic fields spanning multiple scales with high fidelity (e.g., T. Davies, 2014; Leps et al., 2019; Warner et al., 1997). Most approaches, and those employed in operational NWP, are straightforward and typically involve some blending (e.g., Black, 1994; Rogers et al., 1997) or relaxation (e.g., H. C. Davies, 1976; Skamarock et al., 2018). Blending involves a simple decaying weighted average between the specified lateral boundaries across a predefined buffer zone that spans a handful of grid points between the boundary edge, where the LBCs are fully specified, and the full interior of the domain where the state is fully determined by the model. Special treatments are often involved to avoid the problem of overspecification at outflow points. For example, in the case of Black (1994) only the velocity components tangential to the boundary are extrapolated from the interior of the integration. Alternatively, upwind differencing may also be applied to avoid overspecification, such as in Rogers et al. (1997). In relaxation, all prognostic variables are specified and a form of filtering, such as a diffusive relaxation term, is used to dampen any noise resulting from overspecification through a buffer zone. Both blending and relaxation approaches have been used with success for decades.

LAMs run at convection-allowing resolutions often feature data assimilation configurations with rapid updates and low latency for prediction of near-term, high-impact events (e.g., Benjamin et al., 2016; Gustafsson et al., 2018; Rogers et al., 2017; Wheatley et al., 2015). It is therefore customary for convection-allowing LAM configurations at NCEP to be executed prior to the GFS, which has a focus more toward the medium range (5–10 days), for the same initialization time. For example, the 0000 UTC High Resolution Rapid Refresh will begin running at ~0022 UTC while for the same cycle the GFS will start running at ~0245 UTC. Such a delayed start for the GFS allows it to ingest more observations, which are important for high-quality medium range forecasts (e.g., Kleist et al., 2009). In an operational setting, the simultaneous integration of a global and a convection-allowing nested domain would require that these systems have the same data cut-off for the same cycle. If both regional and global systems were to execute simultaneously, it would require either the high resolution application to be delivered several hours later, reducing the utility for near-term forecast applications, or a significant loss in observations for the global application to accommodate an earlier run time, resulting in degradation to the medium range forecast. A limited area capability is therefore required to satisfy the needs of operational NWP as well as provide flexibility in research settings, where the additional overhead of integrating and maintaining a global domain may be untenable or simply unnecessary for the research at hand. The focus of this work is to describe the LAM capability and compare it to a similarly configured two-way nest configuration using identical initial conditions for both configurations.

Here we describe the approach taken to introduce a LAM capability into the FV3 dynamical core framework. Owing to the existing nesting capability in the FV3, we are presented with the somewhat unique opportunity to evaluate the limited area capability against the more optimally configured nesting framework. To our knowledge this is the first such study to compare a convection-allowing LAM with a two-way interactive domain nested within a global model. In Section 2, we describe the limited area regional model approach. In Section 3, we evaluate the forecasts from the new limited area configuration for a near month-long period relative to a nest configuration. The study concludes with a discussion in Section 4.

2. Description of the LAM

The version of the fully compressible, nonhydrostatic FV3 that was enhanced for the LAM capability is the same as that which became operational in the NCEP GFS on June 12, 2019 (i.e., GFSv15). The dynamical core is based on the finite volume dynamics of Lin (1997, 2004); Lin and Rood (1996, 1997) along with

the nonhydrostatic extension described in Harris, Chen, et al. (2020) and later extended from the global latitude-longitude grid to a gnomonic cubed sphere grid by Putman and Lin (2007). The FV3 uses a C-D grid discretization where the horizontal wind components are solved for on the D-grid while the C-grid winds, which are determined at intermediate timesteps, are used to compute and advance the fluxes (Lin & Rood, 1997). The Lagrangian vertical coordinate (Lin, 2004) allows for straightforward extension toward nonhydrostatic motions as the deformation of the vertical layer constitutes the vertical motion. When the nonhydrostatic dynamic option is enabled, both the pressure depth and geometric depth of each vertical layer are considered prognostic variables. With the layer deformation comes the need to periodically perform a high-order, conservative remapping to an Eulerian reference vertical coordinate. In this work we use a hybrid-pressure reference coordinate. Remapping is invoked to avoid infinitesimally thin or folded layers (e.g., Griffies et al., 2020). All variables are layer-mean values; there is no vertical staggering. A semi-implicit solver is used to handle vertically propagating sound waves (Harris, Chen, et al., 2020). To date, the nonhydrostatic FV3 has been successfully applied in several convection-allowing applications with this approach (e.g., Harris, Zhou, et al., 2020; Harris et al., 2019; Potvin et al., 2019; Zhang et al., 2019; Zhou et al., 2019). A complete description of the dynamical core is beyond the scope of this manuscript.

The forecast integration in the LAM mode runs in the same way as in the global/nest version with the primary difference being the handling of the domain's boundaries. Conceptually, the FV3 LAM runs as a nest without a global parent, thus allowing the principles developed first for the nesting approach in the hydrostatic framework (Harris & Lin, 2013, 2014) and later in the nonhydrostatic framework (Harris et al., 2019) to be maintained for the LAM, with the notable omissions of both the two-way update and concurrent integration with a parent domain. The LBCs are comprised of prognostic variables from an external model into the "halo region" of the LAM domain. Here, the "halo region" refers to those grid cells forming a perimeter just outside the model integration domain that are necessary for the proper execution of the model dynamics. When applied during model integration the LBCs are linearly interpolated in both space and time to match the grid-spacing and timestepping of the LAM. In this section, we describe the details of how the LBCs are specified from an external model to accommodate the use of the existing infrastructure.

Input data for the LAM domain is generated by two pre-processing steps. The first step creates the grid's horizontal specification file (i.e., the orography data file) and the static surface data files. The second step uses the output from the operational GFS to generate the atmosphere, surface, and boundary data files. This step runs for each cycle of the experiment featured in Section 3. The primary forecast variables contained in the boundary conditions are the pressure depth and geometric depth of model layers, virtual potential temperature (where the reference pressure is 1 Pa), vertical velocity (m s^{-1}), horizontal 2-D divergence (currently set to zero in the boundary), and the D-grid and C-grid u and v (horizontal) wind components. The nonhydrostatic pressure anomalies, which are needed in the calculation of the pressure gradient force, are obtained via calculation from already specified boundary condition data using the semi-implicit solver for the nonhydrostatic dynamics. Also, a general 4-D array holds all the tracers that are available in the boundary data. The LAM boundary for a hypothetical domain is depicted in Figure 1. The inner region and the points on that region's outline comprise the area of integration. Following the finite volume nature of the model, mass variables lie at the center of the grid cells while the wind components lie at the midpoints of the cells' edges. Specifically, the D-grid u and C-grid v winds lie at the midpoints of the upper and lower edges of the cells (blue dots in Figure 1) and the D-grid v and C-grid u winds lie at the midpoints of the right and left edges (red dots in Figure 1). FV3's dynamics algorithms reach three columns/rows outward from each integration point, therefore the full LAM domain must include at least three outer-boundary columns/rows surrounding the integration domain (i.e., the halo region). Following the reading of the input data derived from the external source, a key step in the forecast's initialization process is the remapping of all primary forecast variables from their vertical location in the input data to the levels used by the integration. For the LAM domain this must include the boundary variables as well. The third column/row of D-grid v and C-grid u boundary winds lies on the outer edge of the third column/row of boundary cells, and because remapping of the wind components requires adjacent pressure values, a fourth column/row of boundary grid cells is required to hold pressure. The fourth boundary columns/rows are the outermost seen in Figure 1. After the vertical remapping is done, the wind components are rotated to the orientation of the integration grid.
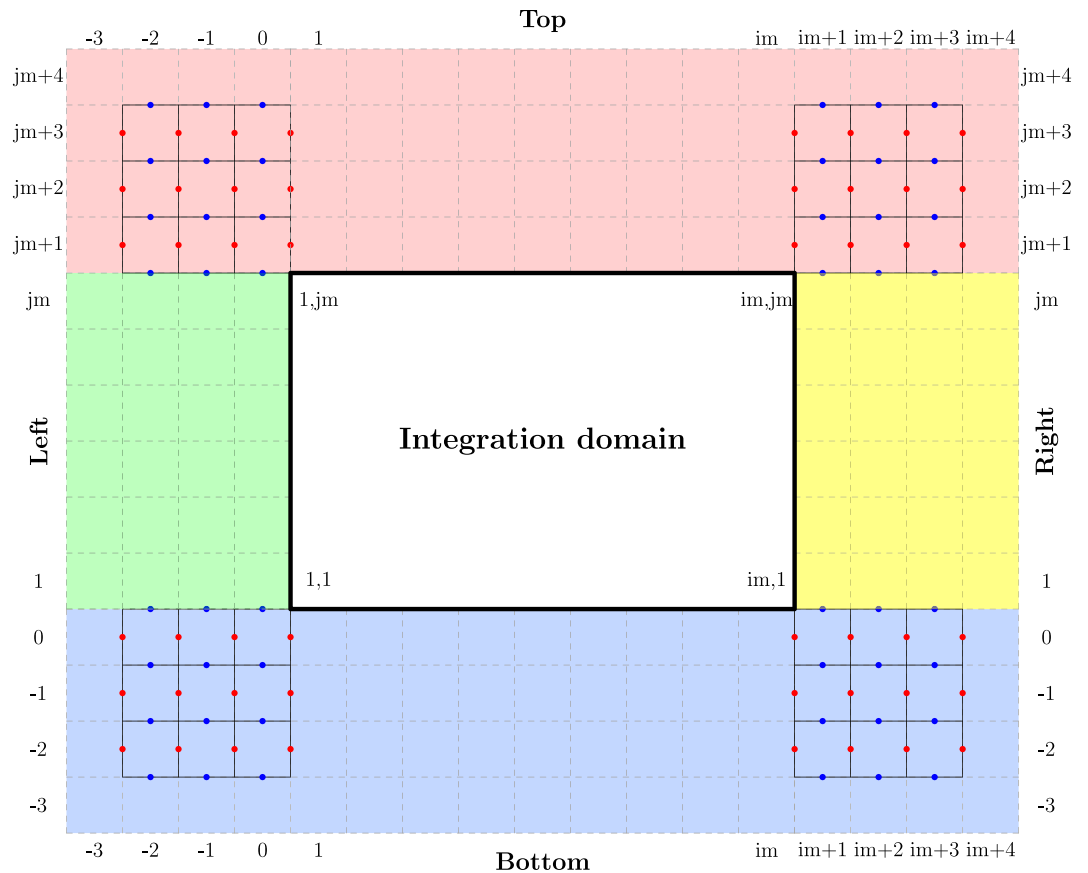
Figure 1. A depiction of the lateral boundary condition (LBC) configuration for the limited area model (LAM). The inner region and the points on that region's outline comprise the area of integration. The domain's boundary region (i.e., halo region) is depicted as a perimeter around the integration area via color shading and spans three cells, where each cell contains full prognostic variables and conforms to the D-grid and C-grid staggering of the FV3 dynamical core. A fourth line of boundary cells is also present for vertically remapping the outermost winds.

In order to have coherent arrays, the boundary data is organized according to which side of the domain it lies on. We refer to these sides as top, bottom, right, and left rather than geographic directions. The top and bottom sides of boundary data span the entire domain from left to right as depicted by the pink and blue strips in Figure 1. The right and left sides (yellow and green, respectively) span only the integration domain's sides and thus do not overlap the top and bottom. This necessitates careful handling of indices for the various mass and wind variables. In the distributed memory processing of the forecast, the Message Passing Interface (MPI) processing elements lying in the corners of the integration domain will naturally contain two different sides of boundary data (e.g., both the top and right sides for a task in the upper-right corner).

LBC data are generated from the external source (i.e., a global model) at regular intervals of time. As the forecast proceeds between two of these bracketing times, the values within the boundary are simply interpolated linearly in time between those two sets of data. When the later data time is reached, then data for the end of the following interval is read from the boundary file and the time interpolation within the boundary continues as the integration moves ahead. The bracketing times used in this study are at 3-h intervals.

2.1. Forecast Behavior Near Boundary Edges

Given that the boundary data is fully prescribed immediately adjacent to the integration values using a simple interpolation approach, it is worthwhile to inspect the behavior of the forecast near the boundaries relative to the two-way nesting configuration. In fact, prior work has demonstrated that two-way nesting can mitigate LBC-induced noise relative to a one-way approach in an idealized setting (Harris & Durran, 2010).
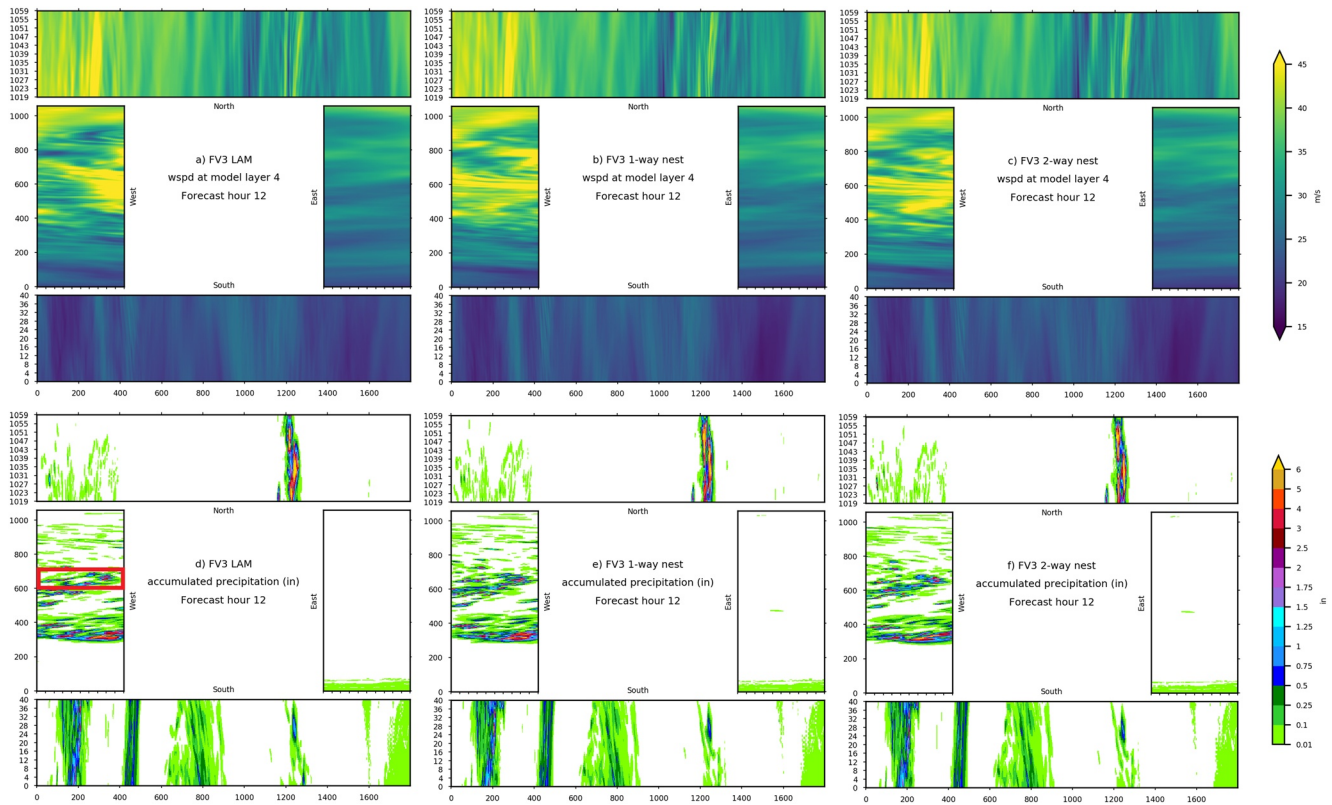
Figure 2. The first 40 rows and columns adjacent to the lateral boundaries showing values of wind speed (top) at model layer 4 (~267 Pa) and run total accumulated precipitation (bottom) from 12-h Limited area model (LAM), one-way NEST, and two-way NEST forecasts for an arbitrary case (valid 1200 UTC on August 8, 2019). The horizontal grid spacing is 3 km, and the physics and dynamics settings are identical in all configurations.

Such an advantage was further demonstrated in Harris and Lin (2014) where artifacts in accumulated precipitation along the boundaries of a one-way nested 10 years climate simulation were mostly absent from the two-way experiment. Since a LAM might be considered a temporally coarse variant of a one-way nest we might expect to detect such artifacts near the boundaries.

To evaluate the near-boundary forecast fields we compare identically configured 3 km LAM, one-way, and two-way nest simulations for a 12-h forecast (Figure 2) run over a domain encompassing the contiguous United States (CONUS; Figure 3). The one-way nest configuration is included as a means to subjectively examine any potential impact two-way feedback may have on near-boundary conditions. Initial and boundary conditions were from the GFSv15 with the LBCs provided at a 3-h interval for the LAM. The configurations are described more fully in Section 3, as they are identical to those used for the full suite of
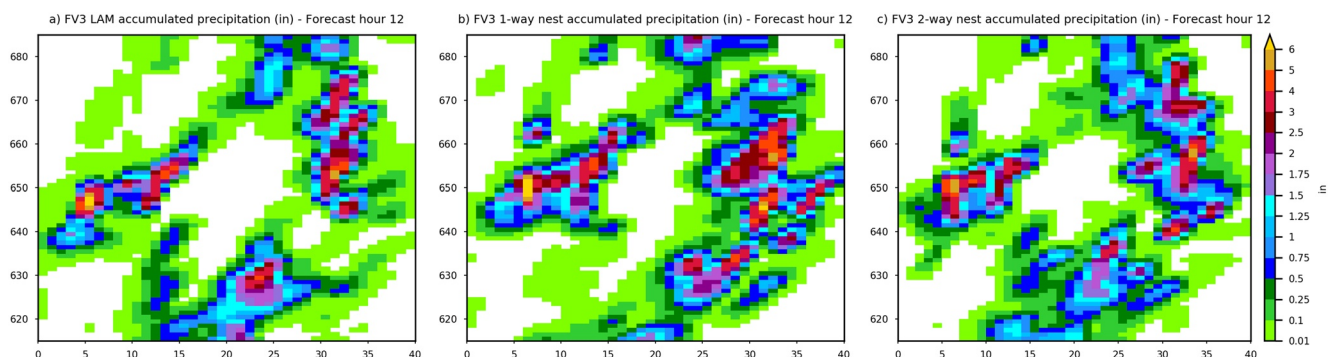

Figure 3. As in Figures 2d–2f except zoomed over the area outlined by the redbox depicted in Figure 2d.

forecast experiments. The 3 km grids are identical. Figure 2 depicts near-boundary forecast upper atmospheric winds (model layer 4, or ~267 hPa) and accumulated precipitation over regions spanning 40 cells, or ~120 km, into the model integration domains. We chose to focus on upper atmospheric winds, where air density is low, as it is a region and field susceptible to depicting spurious noise. The near-boundary winds (Figures 2a–2c) are meteorologically similar at forecast hour 12, with each configuration showing wind speeds approaching 40 m s^{-1} near the western and northern part of the domain. Differences are present between the LAM relative to the two nest configurations, especially along the western boundary near grid cell number 600. The one- and two-way nest configurations are nearly identical, though differences are apparent along the western boundary. Forecasts of 12-hr accumulated precipitation (Figures 2d–2f) are largely similar across all boundary regions with differences being most apparent along the western edge, associated with isolated heavy precipitation. The overall pattern of the precipitation is similar and further made apparent upon close examination. Figure 3 encompasses the area outlined denoted in Figure 2d located in the western edge of the domain. Here we see spatial patterns and intensity in heavy precipitation that are common across all configurations, with the LAM (Figure 3a) having the largest differences. This brief comparison shows that near-boundary forecast behavior is overall similar between the LAM and both nest configurations for this 12-hr forecast period. No discernible artifacts are present, however artifacts may be apparent for much longer integration periods, i.e., >60 h, as was described in Harris and Lin (2014).

Use of two-way relative to the one-way nest capability in FV3-based models for regional refinement has become standard practice (Harris, Zhou, et al., 2020; Zhang et al., 2019) as it offers potential benefits and costs little additional computational overhead relative to one-way (Harris & Lin, 2013). We now turn our focus toward a more comprehensive comparison of forecast performance between LAM and two-way nest configurations in the following Section.

3. Evaluation of FV3 LAM and FV3 Nest

Real time tests and forecast experiments using the FV3 LAM and FV3 nest were designed to investigate considerations for computational resources and model performance, respectively. Computational run time and forecast verification statistics were generated and compared for both the LAM and the nest within the global parent (hereafter NEST) configurations. To facilitate direct comparison, the studies described below all used identical initial conditions. The only difference is in regard to how local refinement is handled, i.e., limited area or two-way nest. There is no data assimilation featured in any of the experiments. The forecasts were conducted in real time to facilitate daily evaluation by developers and highly engaged stakeholders, a common practice in the convection-allowing NWP development community (e.g., Clark et al., 2018, 2020). This configuration may be considered an idealized research configuration and is not representative of the complexities of how such a system would be configured in an official operational environment, which are numerous (e.g., Benjamin et al., 2016; Gustafsson et al., 2018; Rogers et al., 2009). The simplifications here eliminate confounding factors that would otherwise obfuscate direct evaluation of LAM and NEST methods for regional refinement.

The physics and dynamics settings used were identical between the LAM and NEST configurations throughout the duration of the study. Both configurations used identical initial conditions from the GFSv15 model, which utilizes the Geophysical Fluid Dynamics Laboratory (GFDL) microphysics (Zhou et al., 2019), the hybrid eddy-diffusivity mass-flux planetary boundary layer scheme (Han et al., 2016), the GFS surface layer scheme (Long, 1986, 1989), the Rapid Radiative Transfer Model (Iacono et al., 2008; Mlawer et al., 1997) for both shortwave and longwave radiation, the scale-aware Simplified Arakawa-Schubert convection scheme (Han et al., 2017), and the Noah land surface model (Ek et al., 2003). A stretch factor of 1.5 was specified for the NEST's global parent, producing a grid spacing of ~9 km on the cube face where the nest was placed. A refinement ratio of 3 then yielded a grid spacing of ~3 km for the NEST. The NEST's boundary conditions were updated in each physics timestep by interpolation from the parent. In contrast, the LAM used 3-hourly LBCs provided by the GFSv15. NEST and LAM both used a 90 s physics timestep, 15 s vertical remapping timestep, and a 2.5 s acoustic timestep. The global model in the NEST simulation uses the same 90 s physics timestep but longer vertical remapping and acoustic timesteps of 45 and 7.5 s, respectively. Two-way feedback between the NEST and parent was used following Harris and Lin (2013) and Harris et al. (2019). Lateral boundary updates and two-way feedback between the NEST and its parent occurred every 90 s,
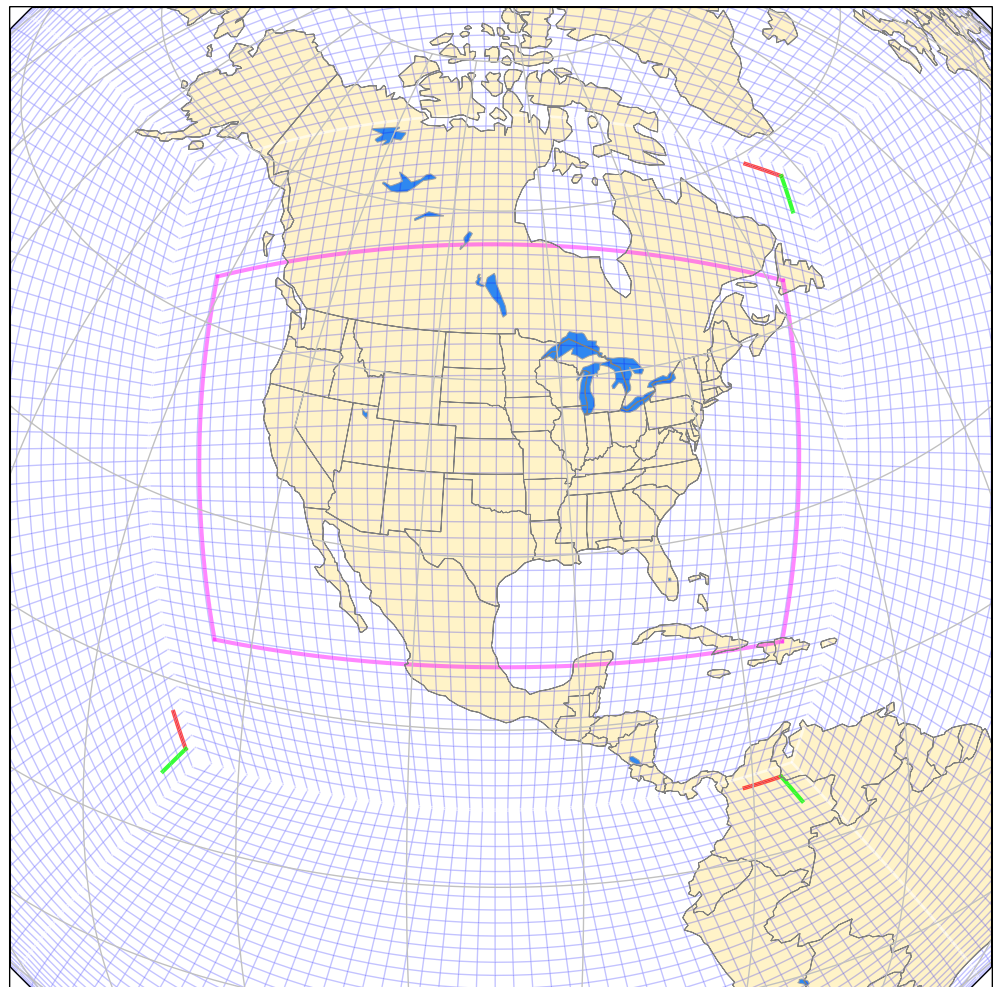
Figure 4. The computational domain for 3-km limited area model (LAM) and NEST experiments is depicted in pink. Note that the FV3 cubed-sphere grid is also depicted but is only relevant for the NEST experiment.

consistent with the physics timestep. In contrast to the LAM's use of 3-hourly LBCs, the NEST had its lateral boundaries updated 120 times more frequently. Precisely the same grid was used for the LAM as for the NEST. The NEST and LAM both used the GFSv15 physics suite, but without parameterized convection (i.e., convection-allowing). The global parent domain for the NEST configuration used the exact same convective parameterization employed in GFSv15. Both configurations used 64 vertical layers with a model top at 0.2 hPa, with identical topography over the same region of CONUS. The computational domain of the LAM and NEST experiments are shown in Figure 4.

3.1. Computational Performance

Tests were performed on NOAA's research and development supercomputer, known as Hera, to evaluate the computational efficiency of the LAM and NEST configurations. Hera has 40 cores per node consisting of 2.4 GHz Intel Skylake processors, 96 GB memory per node, and an HDR-100 Infiniband interconnect. Comparisons were conducted for 24-h forecasts, with model history writes turned off to eliminate any overhead associated with input/output (I/O) as well as performance interference from I/O contention on the shared computing system. Each test used the same node configuration, with 20 MPI tasks with two threads each per node. Figure 5 reveals the total amount of clock time to complete a 24-h forecast as a function of task count. For a given number of tasks, the LAM completed the 24-h forecast in approximately half the time of the NEST. Further, we can also see that the LAM uses less than half the tasks that the NEST needs
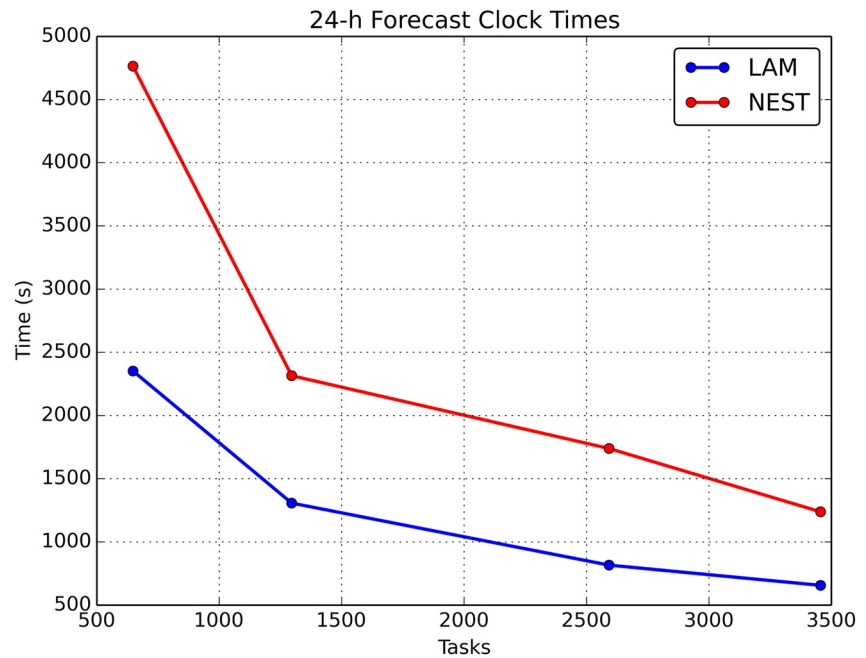
Figure 5. Total amount of clock time elapsed to complete a 24-h forecast as a function of various task counts (i.e., Message Passing Interface (MPI) processes) comparing the limited area model (LAM) (blue) and NEST (red).

for completing in a given amount of time. For example, the LAM needs only about 1/3 as many tasks as the NEST to complete in a clock time of 1200 s. Based on these results, the LAM is considerably more efficient than the NEST, as neither the integration of a global model nor the global model's interaction with it are required. However, such performance gains noted with the LAM may not be worthwhile if the quality of the resulting forecasts are significantly degraded. The accuracy of LAM and NEST forecasts will be compared in Section 3.2.

3.2. Forecast Verification

Forecast experiments were run in real-time during a study period which began on March 15, 2019 and ended on April 16, 2019 to determine whether the LAM could match the performance of the NEST. LAM and NEST forecasts were initialized daily at 0000 UTC and integrated forward 60 h on domains encompassing the CONUS (Figure 4). While both configurations ran in real-time during this period, they were subject to machine outages and maintenance windows. A total of 29 complete forecast cycles were examined in this study.

Comprehensive verification was conducted to compare LAM and NEST performance for upper air and precipitation forecasts. Upper air variables (i.e., geopotential height, temperature, and specific humidity) were verified every 12 h using radiosonde observations valid at 0000 UTC and 1200 UTC. Metrics of interest included bias (i.e., mean error of the forecast) and bias-corrected root-mean-squared error (BCRMSE).

Quantitative precipitation forecasts (QPFs) were verified for 6- and 24-h accumulation intervals using the 4.76-km Climatology-Calibrated Precipitation Analysis (CCPA; Hou et al., 2014) data for validation. The 24-h accumulations were regridded to a common 12-km grid using budget interpolation (Accadia et al., 2003). Metrics of interest include contingency table statistics such as equitable threat score (ETS) and frequency bias. Verification was performed in 24-h (i.e., daily) periods valid from 1200 UTC to 1200 UTC. Owing to the 0000 UTC initialization of the LAM and NEST, 24-h QPFs were only verified for 36- and 60-h forecast leads. Neighborhood-based QPF verification was conducted for 6-h accumulations valid at 0000, 0600, 1200, and 1800 UTC using the 4.76-km CCPA grid with neighborhood sizes ranging from roughly 5 km (i.e., grid-scale) to ~150 km. Fractions skill score (FSS; Roberts & Lean, 2008) was used as the primary metric to assess the 6-h QPFs. We chose to use a finer common grid for FSS verification since it is applied

LAM vs. NEST Scorecard – Upper Air
for LAM and NEST
2019-03-15 00:00:00 – 2019-04-16 00:00:00

			CONUS				
			F12	F24	F36	F48	F60
Bias	Geopotential Height	250 hPa	-0.256	▼ -1.112	▼ -1.204	▼ -2.205	▼ -3.273
		500 hPa	-0.190	▼ -0.712	▼ -0.426	▼ -1.083	▼ -1.679
		850 hPa	-0.085	▼ -0.536	-0.234	▼ -0.937	▼ -1.517
	Temperature	250 hPa	-0.006	▼ -0.030	▲ -0.060	▼ -0.105	▲ -0.166
		500 hPa	-0.011	▼ -0.021	-0.019	▼ -0.029	-0.025
		850 hPa	-0.008	-0.010	-0.002	0.003	0.007
	Specific Humidity	300 hPa	0.000	0.002	0.000	-0.002	-0.001
		500 hPa	0.001	0.000	-0.006	0.004	0.006
		850 hPa	0.003	-0.007	0.017	-0.001	0.013
BCRMSE	Geopotential Height	250 hPa	-0.025	0.110	0.130	-0.028	-0.050
		500 hPa	-0.066	0.036	0.130	0.354	0.583
		850 hPa	-0.095	-0.012	0.022	▼ 0.515	0.395
	Temperature	250 hPa	0.002	-0.002	-0.006	0.010	0.034
		500 hPa	0.004	0.006	0.011	0.003	-0.004
		850 hPa	0.004	0.002	0.003	-0.010	-0.001
	Specific Humidity	300 hPa	0.000	0.000	0.000	0.000	0.001
		500 hPa	-0.007	-0.001	0.003	0.007	0.008
		850 hPa	-0.012	0.013	0.022	0.012	0.004

▲	LAM is better than NEST at the 99% significance level
	LAM is better than NEST at the 95% significance level
	Difference between LAM and NEST is not statistically significant
	LAM is worse than NEST at the 95% significance level
▼	LAM is worse than NEST at the 99% significance level
	Null or Missing Data
	Statistic for symbols: DIFF_SIG , Statistic for values: DIFF

Figure 6. Upper air scorecard. Units for geopotential height, temperature, and specific humidity pairwise differences (limited area model (LAM) minus NEST) are m, K, and kg kg^{-1}, respectively.

across multiple spatial scales and is therefore useful to begin with the finest grid resolution (e.g., Wolff et al., 2014). This choice, as well as the use of 6-h QPFs for evaluation, is also consistent with recommendations from the World Meteorological Organization (WMO, 2013).

Scorecards have been demonstrated to be a useful means of comparing the performance of two modeling experiments by summarizing differences in verification metrics for fields of interest (Gallo et al., 2020). This allows for the consolidation of a large set of relatively comprehensive statistics to be examined at the same time. On each scorecard, statistics were aggregated for the appropriate combinations of variable, forecast lead, and threshold. The scorecards depict numerical values, denoting the pairwise differences between the LAM and NEST verification statistics, and a combination of colors and symbols that indicate the level of statistical significance of the pairwise differences (Figure 6). The statistical significance of the pairwise differences between experiments was computed using bootstrap resampling using 1,000 replicates with replacement. This approach is used for all instances where statistical significance testing is employed in this study.

The upper air scorecard shown in Figure 6 summarizes differences in bias and BCRMSE for geopotential height, temperature, and specific humidity at the 250-, 500-, and 850-hPa pressure levels. Forecast bias (i.e., mean error) differences indicate statistically significant degradation in the middle and upper tropospheric temperature and geopotential height forecasts from the LAM beginning at forecast hour 24 and growing in magnitude through the forecast period. The LAM and NEST both had negative (i.e., low) biases for their respective geopotential height forecasts (not shown), but the negative pairwise differences indicate that the LAM biases were more negative than those of the NEST. BCRMSE differences, however, were generally found to be statistically insignificant for all three variables and pressure levels. LAM 500- and 850-hPa geopotential height forecasts had increased BCRMSE relative to the NEST for forecast leads greater than 24 h, but these increases (indicative of degradation) were only statistically significant during the last 12 h of the forecast period. These BCRMSE and bias results indicate that a growing negative temperature and height bias was the greatest contributor to the error in the LAM forecasts after the first day of the model integration. While the NEST and LAM both displayed negative geopotential height biases that grew in magnitude throughout the forecast period, the mean error was greater in the LAM than in the NEST.

NEST and LAM precipitation forecasts were found to be largely similar, but the impact of poorer upper air forecasts was evident in trends for LAM QPF statistics. Both models were considerably wetter than the verifying analysis throughout the entire diurnal cycle (not shown). This wet bias is quite evident based on all 24-h QPF thresholds having a frequency bias greater than 1 as depicted via a performance diagram (Roebber, 2009) shown in Figure 7. Furthermore, 24-h QPFs were found to be markedly similar in terms of the other performance diagram metrics: critical success index, probability of detection, and success ratio (Figure 7). NEST precipitation forecasts were found to have a slight advantage over LAM precipitation forecasts at longer lead times (Figure 8), but these differences were not determined to be statistically significant. Verification of 6-h QPF found NEST FSSs to be greater than LAM FSSs (Figure 9) at the 5 mm per 6-h threshold, but the differences were on the order of 0.01. The small magnitudes suggest the pairwise differences were not practically significant, and confidence testing at the 95% significance level found that the differences were not
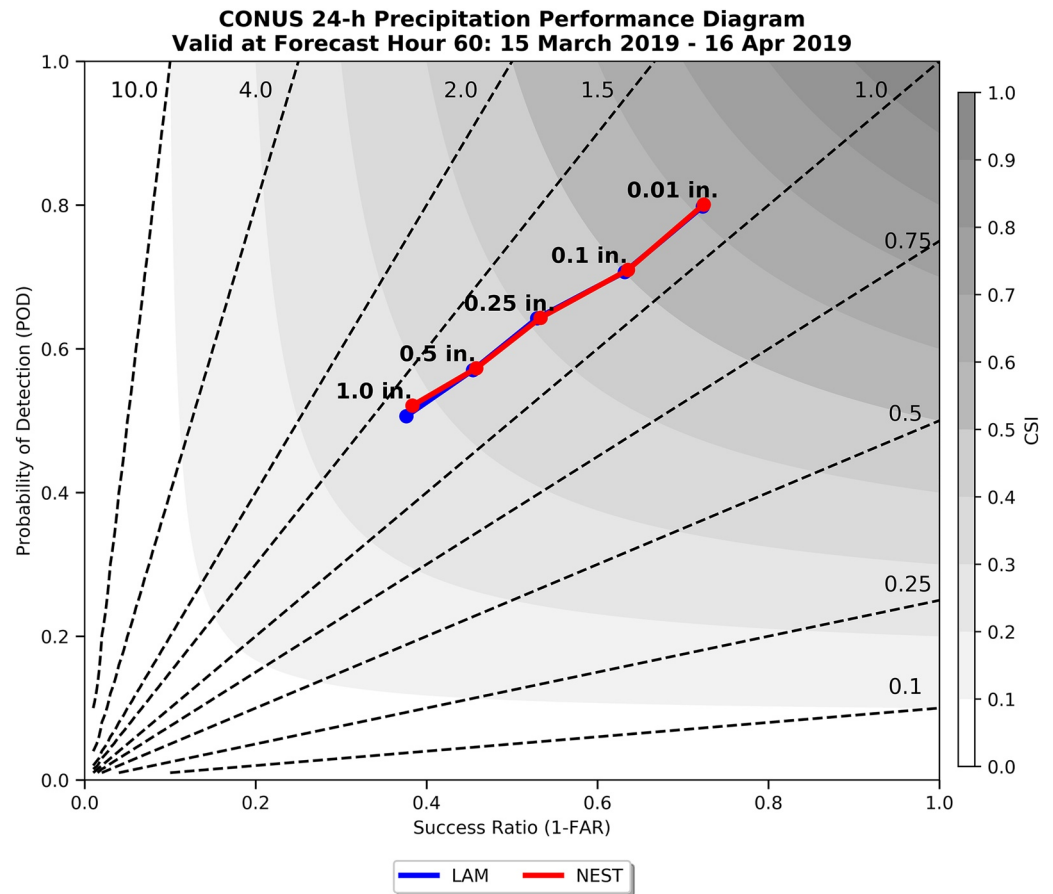
Figure 7. Performance diagram for 24-h Quantitative precipitation forecasts (QPF) forecasts valid from forecast hour 36 to 60. Dots mark 24-h QPF thresholds (>0.01, 0.1, 0.25, 0.5, and 1.0 inches).

statistically significant. A 6-h QPF scorecard found consistent results at additional accumulation thresholds and forecast leads (not shown).

Overall, objective verification found that LAM and NEST forecasts were generally statistically comparable. The most noteworthy statistical differences were found in the bias of the upper air geopotential height and temperature forecasts, particularly at forecast leads greater than 24 h. Both configurations were found to have negative geopotential height and upper-air temperature biases that grew in magnitude throughout the forecast integration. These biases are attributable to the physics being configured following the experimental version of GFSv15, which had a well-documented negative height and temperature biases during the study period of this experiment (Bentley & Manikin, 2019). While these biases were seen in both configurations, they were exacerbated in the LAM. Precipitation verification found NEST forecasts to be superior to LAM forecasts, especially at longer forecast ranges, but the differences were not statistically significant.

3.3. Case Study

LAM and NEST forecasts were also subjectively evaluated throughout the duration of the study period. Qualitative comparisons found the forecasts to be largely similar with no systematic differences in the sensible weather forecasts generated from each configuration. This subsection reviews a case that provides a representative example of the aforementioned statistical verification differences and the overall similarity of the sensible weather forecast guidance. The case of interest features a convective outbreak that caused widespread severe weather over the southeastern United States between 1200 UTC April 13, 2019 and 1200 UTC April 14, 2019 (https://www.spc.noaa.gov/climo/reports/190413_rpts.html). Comparisons focus on 500-hPa geopotential height and composite reflectivity forecasts and analyses valid in the middle
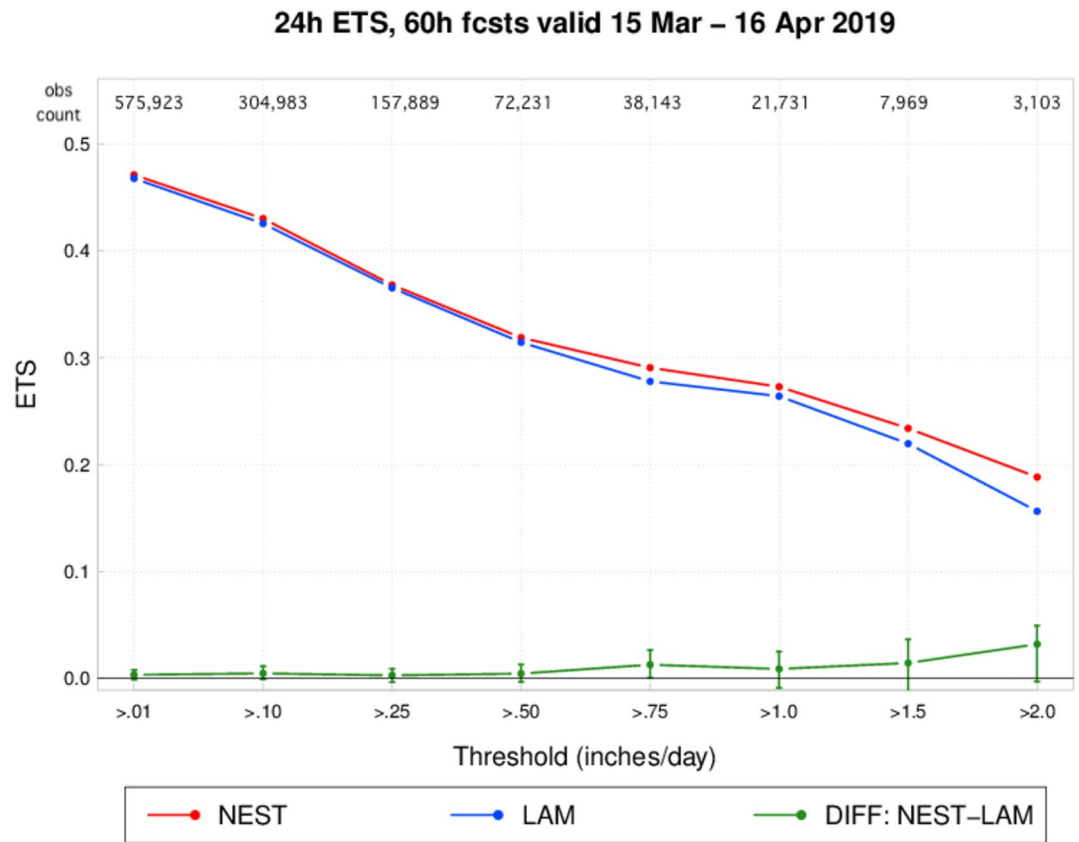
Figure 8. Equitable threat score (ETS) versus daily precipitation thresholds, for 36–60-h forecasts. The confidence interval for the pairwise difference line is shown at 95%.

of this time period (i.e., 0000 UTC April 14, 2019). LAM and NEST forecasts initialized 24- and 48-h prior to this valid time (i.e., 0000 UTC April 13 and 0000 UTC April 12, 2019) will be referred to as the Day 1 and Day 2 forecasts, respectively. The LAM and NEST upper air forecasts were validated using the operational GFS analysis interpolated to a 0.25 deg lat-lon grid. Composite reflectivity forecasts were validated using reflectivity analyses from the Multi-Radar Multi-Sensor system (MRMS; Smith et al., 2016).

Differences between the LAM and NEST 500-hPa geopotential height forecasts are evident in the Day 1 (Figure 10a) and Day 2 forecasts (Figure 10b). For both initialization times, the LAM generally forecasted lower 500-hPa geopotential heights than the NEST. This was nearly ubiquitous across the entire domain and was most pronounced near the center of the closed low. Close inspection of the height contours reveals a southward shift in the LAM and NEST contours relative to those of the GFS analysis. This is consistent with the negative geopotential height bias that was noted in the statistical verification of both forecast configurations. The differences between the LAM and NEST forecasts and the negative height bias relative to the GFS analysis were both larger in magnitude on Day 2 compared to Day 1, again consistent with the geopotential height bias shown in Figure 6. In the Day 2 forecast, both configurations displayed a slight lag in the trough that was centered over the southern Great Plains. The largest differences in geopotential height between the LAM and NEST occur over convective regions such as the mesoscale convective system located in northeast Texas (Figure 11). While these differences near convection appear large, they are more representative of differences between small-scale, discrete phenomena.

Despite the aforementioned differences in the LAM and NEST geopotential height forecasts, both configurations produced qualitatively similar forecasts of the organized convection over the southeastern United States (Figure 11). The spatial extent and position of the composite reflectivity fields from the LAM and NEST compare well with MRMS observations for both Day 1 (Figures 11a and 11b) and Day 2 (Figures 11c and 11d). The most notable difference between the forecasts and the MRMS observations are the relatively
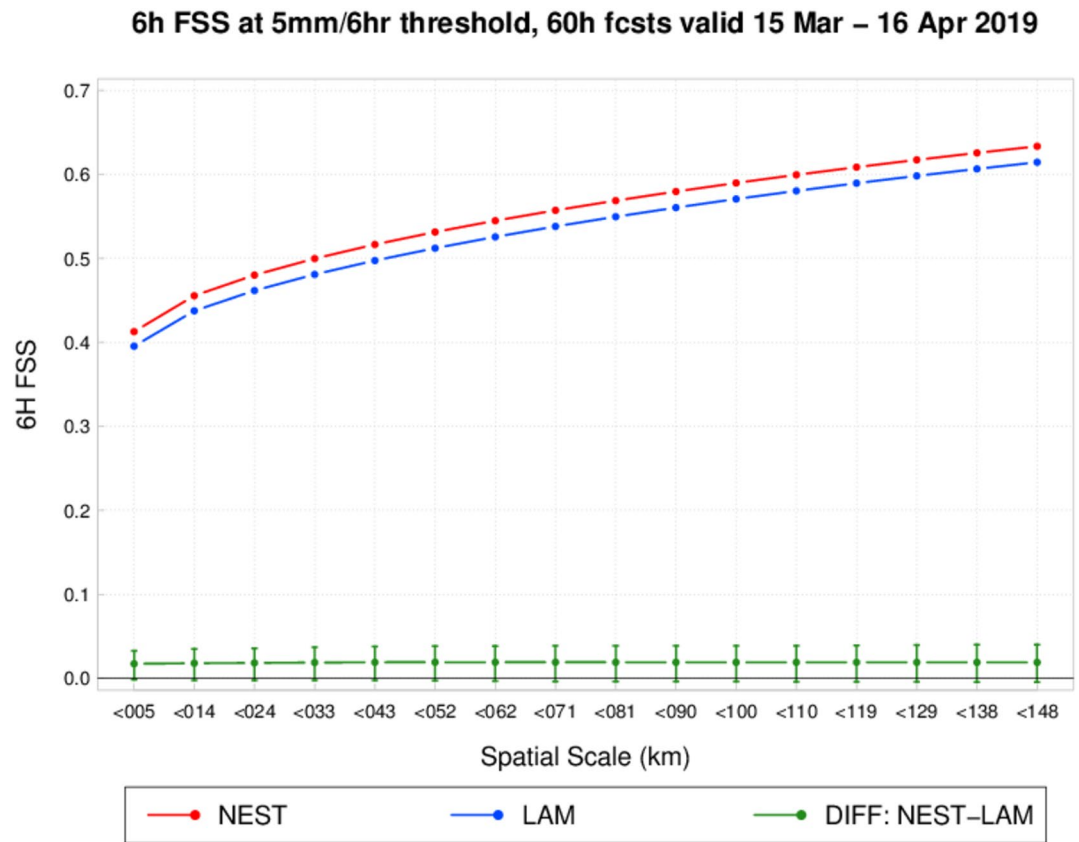
Figure 9. Fractions skill score (FSS) versus horizontal spatial scales, for 54–60-h forecasts, at the 5 mm per 6-h threshold. The confidence interval for the pairwise difference line is shown at 95%.

high magnitudes of simulated composite reflectivity, which is a feature common between both forecasts and thus not an artifact inherent to the LAM configuration.

Qualitatively, it is difficult to say whether one configuration performed better than the other because differences were minor and small-scale in nature. The overall similarity of these convective-scale forecasts agrees with the precipitation verification, which found that the differences between the LAM and NEST precipitation forecasts were not statistically significant. Furthermore, despite the Day 2 forecasts predicting a deeper and slower upper trough than was observed (Figure 10b), the LAM and NEST produced convective-scale forecasts that were comparable to the Day 1 forecasts and compared well with observations. These results suggest that the negative geopotential height bias (relative to observations) seen in the LAM and NEST forecasts did not have a practical impact on forecasts of organized convection or precipitation. While this negative geopotential height bias was exacerbated in the LAM configuration relative to the NEST, the precipitation and convective forecasts were not significantly impacted by the differences in the upper air forecasts.

4. Discussion

This study describes the development of a LAM capability within the FV3 dynamical core framework and the subsequent validation of the LAM. The FV3 LAM runs as a stand-alone, high-resolution configuration without a global parent domain, where the primary differences are the omission of the concurrent integration with the external parent and the handling of the LBCs. The LBCs consist of a linear interpolation in space and time of prognostic variables from an external model into the halo region of the LAM domain, which includes the grid cells that form a perimeter just outside the integration domain. The forecast initialization process involves reading in the input data from the external source, remapping all primary forecast
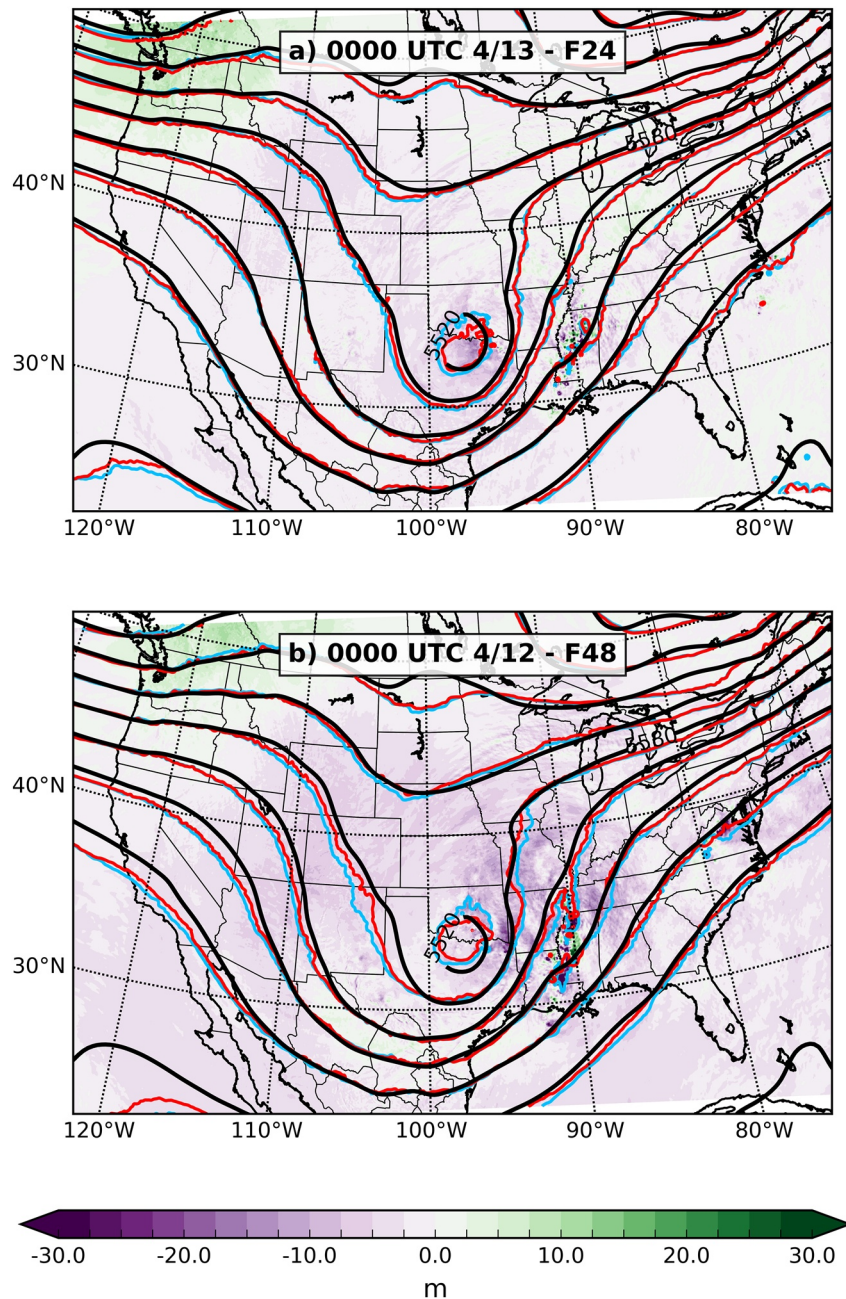
Figure 10. Limited area model (LAM) and NEST 500-hPa geopotential height forecasts valid at 0000 UTC April 14, 2019 are contoured in the light blue and red, respectively. Color shading shows the difference (LAM – NEST) between the two 500-hPa geopotential height forecasts. The forecasts are also compared with the Global Forecast System (GFS) analysis (black contours) valid at the same time. The 24-h LAM and NEST forecasts initialized at 0000 UTC April 13, 2019 are shown in (a). The 48-h forecasts initialized at 0000 UTC April 12, 2019 are shown in (b).

variables from their vertical location in the input data to the levels used by the integration, and rotating the wind components to the orientation of the integration grid.

The newly developed LAM capability was compared against the nesting capability that was previously available within the FV3 dynamical core framework. Computational performance testing found that the LAM configuration was nearly twice as efficient as the nested configuration, owing to the removal of the dependency on a global parent domain's simultaneous integration. Objective verification of the LAM and NEST experiments found the forecasts to be statistically comparable, with the statistically significant differences
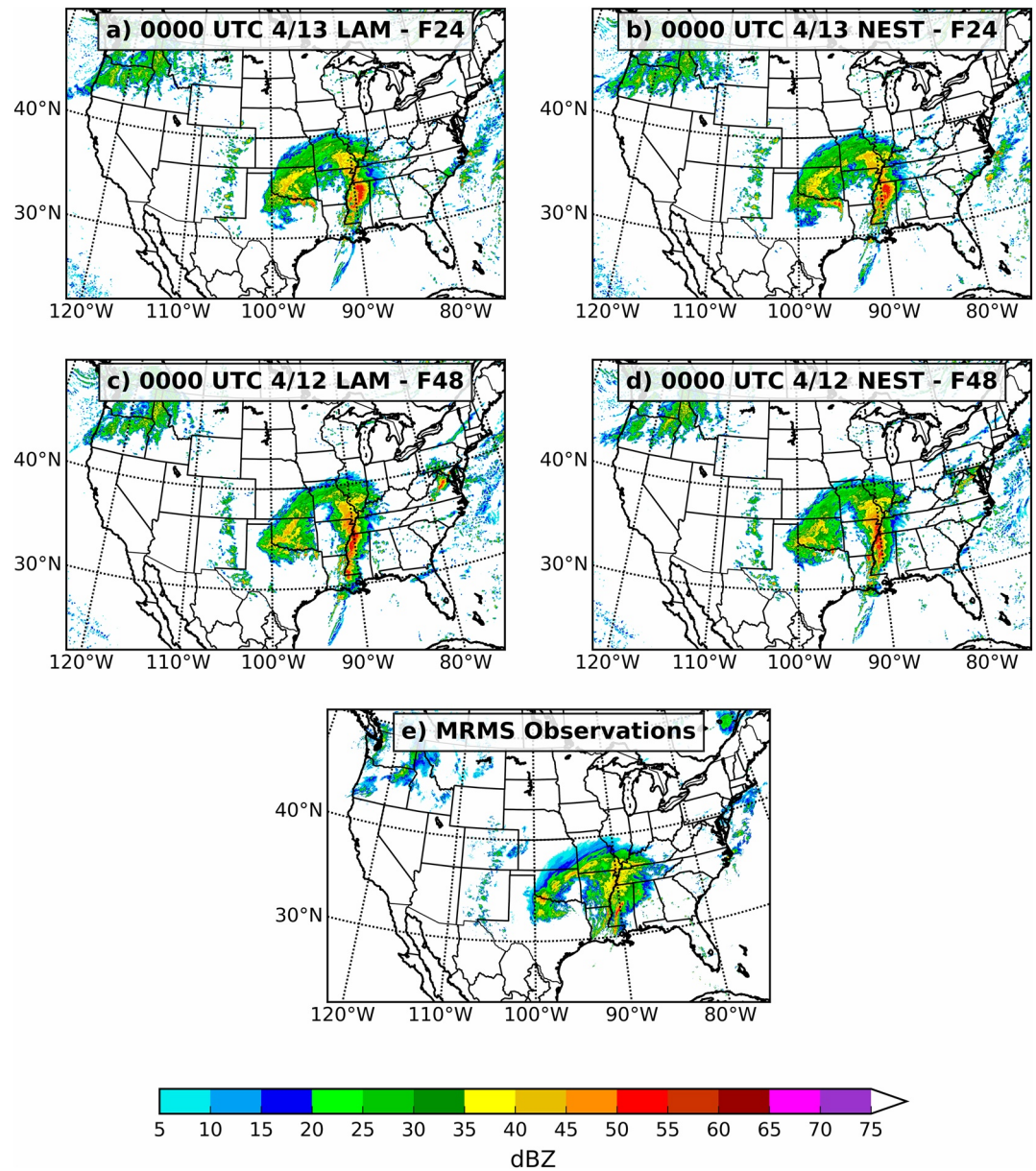
Figure 11. Composite reflectivity forecasts (a–d) and Multi-Radar Multi-Sensor system (MRMS) observations (e) valid at 0000 UTC April 14, 2019. The 24-h limited area model (LAM) and NEST forecasts initialized at 0000 UTC April 13, 2019 are shown in (a) and (b), respectively. The 48-h LAM and NEST forecasts initialized at 0000 UTC April 12, 2019 are shown in (c) and (d), respectively.

largely being limited to the bias of the geopotential height and temperature forecasts in the middle and upper troposphere. The noteworthy differences between the LAM and NEST generally grew in magnitude and became more statistically significant throughout the forecast integration, consistent with prior studies that indicate error growth is typically linear before reaching domain-wide saturation (e.g., Warner et al., 1997). Subjective evaluations of day-to-day forecasts matched the findings of the statistical verification, and overall qualitative differences were small.

Differences between NEST and LAM configurations lie principally within three areas: (1) feedback between the high resolution and the lower resolution model. (2) the source model providing the LBCs, and (3) the frequency of boundary updates. The LAM configuration lacked the two-way updates that occurred between the NEST and its global parent. We hypothesize that impact of two-way feedback is relatively modest in

comparison to the influence of less frequent boundary updates. Figures 2 and 3 both demonstrate that one- and two-way nests are almost identical near the boundaries for short forecast lead times. Further, our experiments mitigated potential differences in the LBC source model by configuring the global parent domain in a similar manner to that of the GFSv15. Therefore forecast differences in these experiments are likely dominated by the frequency of boundary updates. In NEST, boundary updates occurred at every physics timestep of 90 s while the LAM's boundary conditions were linearly interpolated in time between states specified every 3-h. The reduced frequency of boundary updates likely explains the faster growth in upper-air biases in the LAM. These findings are not unexpected as the effects of boundary forcing can have ill effects on the forecast fields, however this is balanced by the considerable improvement in efficiency and flexibility allowed by the LAM. In addition, more frequent LBC updates, e.g., hourly, in the LAM could reduce errors (T. Davies, 2014). For forecast lengths appreciably longer than those featured in this study, such as several days to weeks, it is preferable to exercise the two-way nested configuration. Recent advances with the System for High-resolution prediction on Earth-to-Local Domain (SHiELD), which is underpinned by the nonhydrostatic FV3 dynamical core, have shown promising results in the application of a globally nested configuration with two-way feedback for medium range (3–5 days) prediction of convective storms over CONUS (Harris, Zhou, et al., 2020).

Testing of the LAM capability in this study benefited from utilizing boundary conditions specified from an external model having the same fundamental dynamics, physics, and distribution of vertical levels. In an operational setting, the LAM configuration would be somewhat less ideal than that tested here. Since high resolution regional NWP systems have earlier data cut-offs than their global counterparts for a given forecast cycle, these systems are forced to use LBCs generated from an older cycle of the global model. Therefore, inconsistencies are more likely to arise and become problematic along the boundary edges. This challenge is exacerbated when one introduces the data assimilation procedure, where the interior model integration domain is updated to reflect the best estimate of the atmospheric state, yet the lateral boundaries still reflect an estimate that is from a model cycle that is usually at least 6-h older.

The LAM approach developed here uses the same gnomonic cubed-sphere projection as the global cubed-sphere model (e.g., Figure 4). Technically, there is no real need for such a choice as the LAM could be constructed using an orthogonal projection. In this work we chose to keep the projections identical since it allowed for a simpler implementation and was essential for a direct comparison between the NEST and LAM configurations. This was a necessary first step in developing the initial LAM capability. A technique has since been developed to minimize the variance of the grid cell sizes across the LAM domain and thus produce a more uniform grid. This technique is the subject of a future manuscript.

We note that the LBC approach taken here differs from the "rim zone" strategy described in T. Davies (2014), where the model operates directly on the LBC values within a small rim of the integration domain and replaces the prognosed fields with the LBCs at the end of each timestep. Such an approach could be adopted by including a similar rim region, immediately adjacent to the boundary halo (Figure 1), yet within the computational domain. At the end of each physics timestep the prognostic state within the rim would be replaced by the external LBCs. This method is being considered for future application as an option to further reduce LBC related error.

The introduction of an FV3 LAM capability is a necessary step forward in the advancement of the NCEP UFS, as it will formulate the basis of the Rapid Refresh Forecast System (RRFS), a convection-allowing data assimilation and forecast ensemble featuring at least an hourly update cadence. To support development toward the RRFS future work will involve, but is not limited to: testing under the context of rapidly updated data assimilation, examination of forecast behavior spanning multiple seasons and phenomena. The introduction of near boundary stretching, which makes the lateral boundaries more remote, may also improve the LAM (T. Davies, 2017). In addition, recent work by Dong et al. (2020) has shown the LAM capability to be effective in the early phases of the development of Hurricane Analysis and Forecast System (HAFS). As an integral part of the UFS, the RRFS will facilitate the consolidation of many of the current regional applications in the NCEP production suite.

Data Availability Statement

In-situ observations used for verification may be obtained from NCEI, though not all observations are available to the public owing to data restrictions. Precipitation data are archived at the NCAR Earth Observing Laboratory: https://doi.org/10.5065/D6PG1QDD. Radar observations may be obtained through NCEI and the NOAA Big Data Project: https://www.ncdc.noaa.gov/data-access/radar-data/noaa-big-data-project. Initial and lateral boundary conditions used in the simulations in this study are from publicly available GFS forecasts archived at the National Centers for Environmental Information (NCEI).

References

Accadia, C., Mariani, S., Casaioli, M., Lavagnini, A., & Speranza, A. (2003). Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Weather and Forecasting*, *18*(5), 918–932. https://doi.org/10.1175/1520-0434(2003)018<0918:SOPFSS>2.0.CO;2

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., et al. (2016). A North American hourly assimilation and model forecast cycle: The rapid refresh. *Monthly Weather Review*, *144*(4), 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Bentley, A., & Manikin, G. (2019). *FV3GFS (GFSv15.1) Status Update: Addressing excessive snowfall and low-level cold bias (EMC Model Evaluation Group)*. Retrieved from https://www.emc.ncep.noaa.gov/users/Alicia.Bentley/fv3gfs/updates/MEG_4-04-19_FV3GFS_COLD.pptx

Black, T. L. (1994). The new NMC Mesoscale Eta Model: Description and forecast examples. *Weather and Forecasting*, *9*(2), 265–278. https://doi.org/10.1175/1520-0434(1994)009<0265:Tnnmem>2.0.Co;2

Clark, A. J., Jirak, I. L., Dembek, S. R., Creager, G. J., Kong, F., Thomas, K. W., et al. (2018). The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous weather testbed spring forecasting experiment. *Bulletin of the American Meteorological Society*, *99*(7), 1433–1448. https://doi.org/10.1175/BAMS-D-16-0309.1

Clark, A. J., Jirak, I. L., Gallo, B. T., Roberts, B., Knopfmeier, K. H., Clark, R. A., et al. (2020). A real-time, simulated forecasting experiment for advancing the prediction of hazardous convective weather. *Bulletin of the American Meteorological Society*, *101*(11), E2022–E2024. https://doi.org/10.1175/bams-d-19-0298.1

Davies, H. C. (1976). A lateral boundary formulation for multi-level prediction models. *Quarterly Journal of the Royal Meteorological Society*, *102*(432), 405–418. https://doi.org/10.1002/qj.49710243210

Davies, T. (2014). Lateral boundary conditions for limited area models. *Quarterly Journal of the Royal Meteorological Society*, *140*(678), 185–196. https://doi.org/10.1002/qj.2127

Davies, T. (2017). Dynamical downscaling and variable resolution in limited-area models. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 209–222. https://doi.org/10.1002/qj.2913

Dong, J., Liu, B., Zhang, Z., Wang, W., Mehra, A., Hazelton, A. T., et al. (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

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al. (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

Gallo, B. T., Kalb, C. P., Gotway, J. H., Fisher, H. H., Roberts, B., Jirak, I. L., et al. (2020). Initial development and testing of a convection-allowing model scorecard. *Bulletin of the American Meteorological Society*, *100*(12), ES367–ES384. https://doi.org/10.1175/bams-d-18-0218.1

Gerrity, J. P. (1977). *The LFM model-1976: A documentation* (p. 68). NOAA technical memorandum NWS-NMC 60.

Griffies, S. M., Adcroft, A., & Hallberg, R. W. (2020). A primer on the vertical lagrangian-remap method in ocean models based on finite volume generalized vertical coordinates. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2019MS001954. https://doi.org/10.1029/2019MS001954

Gustafsson, N., Janjic, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., et al. (2018). Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1218–1256. https://doi.org/10.1002/qj.3179

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*(5), 2005–2017. https://doi.org/10.1175/waf-d-17-0046.1

Han, J., Witek, M. L., Teixeira, J., Sun, R., Pan, H.-L., Fletcher, J. K., & Bretherton, C. S. (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. *Weather and Forecasting*, *31*(1), 341–352. https://doi.org/10.1175/waf-d-15-0053.1

Harris, L., Chen, X., Zhou, L., & Chen, J.-H. (2020). *The nonhydrostatic solver of the GFDL finite-volume cubed-sphere dynamical core*. NOAA Technical Memorandum OAR GFDL, 2020-003. https://doi.org/10.25923/9wdt-4895

Harris, L., Zhou, L., Lin, S.-J., Chen, J.-H., Chen, X., Gao, K., et al. (2020). GFDL SHiELD: A unified system for weather-to-seasonal prediction. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2020MS002223. https://doi.org/10.1029/2020MS002223

Harris, L. M., & Durran, D. R. (2010). An idealized comparison of one-way and two-way grid nesting. *Monthly Weather Review*, *138*(6), 2174–2187. https://doi.org/10.1175/2010mwr3080.1

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

Harris, L. M., & Lin, S.-J. (2014). Global-to-regional nested grid climate simulations in the GFDL high resolution atmospheric model. *Journal of Climate*, *27*(13), 4890–4910. https://doi.org/10.1175/Jcli-D-13-00596.1

Harris, L. M., Lin, S.-J., & Tu, C. (2016). High-resolution climate simulations using GFDL HiRAM with a stretched global grid. *Journal of Climate*, *29*(11), 4293–4314. https://doi.org/10.1175/JCLI-D-15-0389.1

Harris, L. M., Rees, S. L., Morin, M., Zhou, L. J., & 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. https://doi.org/10.1029/2018ms001542

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. https://doi.org/10.1175/Waf-D-18-0056.1

Hoke, J. E., Phillips, N. A., Dimego, G. J., Tuccillo, J. J., & Sela, J. G. (1989). The regional analysis and forecast system of the national meteorological center. *Weather and Forecasting*, *4*(3), 323–334. https://doi.org/10.1175/1520-0434(1989)004<0323:TRAAFS>2.0.CO;2

Hou, D., Charles, M., Luo, Y., Toth, Z., Zhu, Y., Krzysztofowicz, R., et al. (2014). Climatology-calibrated precipitation analysis at fine scales: Statistical adjustment of Stage IV toward CPC gauge-based analysis. *Journal of Hydrometeorology*, *15*(6), 2542–2557. https://doi.org/10.1175/jhm-d-11-0140.1

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*(D13), D13103. https://doi.org/10.1029/2008JD009944

Janjić, Z. I., Gerrity, J. P., & Nickovic, S. (2001). An alternative approach to nonhydrostatic modeling. *Monthly Weather Review*, *129*(5), 1164–1178. https://doi.org/10.1175/1520-0493(2001)129<1164:AAATNM>2.0.CO;2

Kleist, D. T., Parrish, D. F., Derber, J. C., Treadon, R., Wu, W.-S., & Lord, S. (2009). Introduction of the GSI into the NCEP Global Data Assimilation System. *Weather and Forecasting*, *24*(6), 1691–1705. https://doi.org/10.1175/2009waf2222201.1

Leps, N., Brauch, J., & Ahrens, B. (2019). Sensitivity of limited area atmospheric simulations to lateral boundary conditions in idealized experiments. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2694–2707. https://doi.org/10.1029/2019ms001625

Lin, S.-J. (1997). A finite-volume integration method for computing pressure gradient force in general vertical coordinates. *Quarterly Journal of the Royal Meteorological Society*, *123*(542), 1749–1762. https://doi.org/10.1002/qj.49712354214

Lin, S.-J. (2004). A "Vertically Lagrangian" finite-volume dynamical core for global models. *Monthly Weather Review*, *132*(10), 2293–2307. https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2

Lin, S.-J., & Rood, R. B. (1996). Multidimensional flux-form semi-lagrangian transport schemes. *Monthly Weather Review*, *124*(9), 2046–2070. https://doi.org/10.1175/1520-0493(1996)124<2046:Mffslt>2.0.Co;2

Lin, S.-J., & Rood, R. B. (1997). An explicit flux-form semi-lagrangian shallow-water model on the sphere. *Quarterly Journal of the Royal Meteorological Society*, *123*(544), 2477–2498. https://doi.org/10.1002/qj.49712354416

Long, P. J. (1986). *An economical and compatible scheme for parameterizing the stable surface layer in the medium range forecast model (NCEP office note 321)*. Retrieved from https://www.ncep.noaa.gov/officenotes/NOAA-NPM-NCEPON-0004/01408602.pdf

Long, P. J. (1989). *Derivation and suggested method of the application of simplified relations for surface fluxes in the medium-range forecast model: Unstable case (NCEP office note 356)*. Retrieved from https://www.ncep.noaa.gov/officenotes/NOAA-NPM-NCEPON-0005/0140893E.pdf

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16663–16682. https://doi.org/10.1029/97JD00237

Potvin, C. K., Carley, J. R., Clark, A. J., Wicker, L. J., Skinner, P. S., Reinhart, A. E., et al. (2019). Systematic comparison of convection-allowing models during the 2017 NOAA HWT spring forecasting experiment. *Weather and Forecasting*, *34*(5), 1395–1416. https://doi.org/10.1175/WAF-D-19-0056.1

Putman, W. M., & Lin, S.-J. (2007). Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, *227*(1), 55–78. https://doi.org/10.1016/j.jcp.2007.07.022

Roberts, N. M., & Lean, H. W. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, *136*(1), 78–97. https://doi.org/10.1175/2007mwr2123.1

Roebber, P. J. (2009). Visualizing multiple measures of forecast quality. *Weather and Forecasting*, *24*(2), 601–608. https://doi.org/10.1175/2008waf2222159.1

Rogers, E., Baldwin, M., Black, T., Brill, K., Chen, F., DiMego, G., et al. (1997). Changes to the NCEP Operational "Early Eta" Analysis/Forecast System. *NOAA/NWS Technical Procedures Bulletin*, *447*, 36.

Rogers, E., Carley, J., Ferrier, B., Aligo, E., Gayno, G., Janjic, Z., et al. (2017). Upgrades to the NCEP North American Mesoscale (NAM) System. In *Research activities in atmospheric and oceanic modelling, CAS/JSC Working Group on Numerical Experimentation*. Report No. 47. WCRP Report No.12/2017. Geneva. WMO.

Rogers, E., DiMego, G., Black, T., Ek, M., Ferrier, B., Gayno, G., et al. (2009). The NCEP North American Mesoscale Modeling System: Recent changes and future plans. In *Paper presented at the 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction*. Omaha, NE. American Meteorological Society.

Skamarock, W. C., Duda, M. G., Ha, S., & Park, S.-H. (2018). Limited-area atmospheric modeling using an unstructured mesh. *Monthly Weather Review*, *146*(10), 3445–3460. https://doi.org/10.1175/mwr-d-18-0155.1

Smith, T. M., Lakshmanan, V., Stumpf, G. J., Ortega, K. L., Hondl, K., Cooper, K., et al. (2016). Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bulletin of the American Meteorological Society*, *97*(9), 1617–1630. https://doi.org/10.1175/bams-d-14-00173.1

Warner, T. T., Peterson, R. A., & Treadon, R. E. (1997). A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bulletin of the American Meteorological Society*, *78*(11), 2599–2618. https://doi.org/10.1175/1520-0477(1997)078<2599:Atolbc>2.0.Co;2

Wheatley, D. M., Knopfmeier, K. H., Jones, T. A., & Creager, G. J. (2015). Storm-scale data assimilation and ensemble forecasting with the NSSL experimental warn-on-forecast system. Part I: Radar data experiments. *Weather and Forecasting*, *30*(6), 1795–1817. https://doi.org/10.1175/waf-d-15-0043.1

WMO. (2013). Suggested methods for the verification of precipitation forecasts against high resolution limited area observations. Joint Working Group on Forecast verification Research.

Wolff, J. K., Harrold, M., Fowler, T., Gotway, J. H., Nance, L., & Brown, B. G. (2014). Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Weather and Forecasting*, *29*(6), 1451–1472. https://doi.org/10.1175/waf-d-13-00135.1

Zhang, C. X., Xue, M., Supinie, T. A., Kong, F. Y., Snook, N., Thomas, K. W., et al. (2019). How well does an FV3-based model predict precipitation at a convection-allowing resolution? Results from CAPS forecasts for the 2018 noaa hazardous weather test bed with different physics combinations. *Geophysical Research Letters*, *46*, 3523–3531. https://doi.org/10.1029/2018gl081702

Zhou, L., Lin, S.-J., Chen, J.-H., Harris, L. M., Chen, X., & Rees, S. L. (2019). Toward convective-scale prediction within the next generation global prediction system. *Bulletin of the American Meteorological Society*, *100*(7), 1225–1243. https://doi.org/10.1175/BAMS-D-17-0246.1