Comparing the Forecasting Performance of the UFS and WRF Models during High-Impact Severe Weather Events

JUAN P. MANGUAL-PAGÁN*

National Weather Center Research Experiences for Undergraduates Program
Norman, Oklahoma

LARISSA J. REAMES
OU/CIMMS & NOAA/NSSL
Norman, Oklahoma

DAVID J. BODINE
OU/ARRC
Norman, Oklahoma

ABSTRACT

The numerical weather prediction community is seeking a common dynamical core to improve forecasts across a range of space and time scales, which motivated the development of the Unified Forecast System (UFS). The community-driven UFS uses the Finite Volume Cubed-Sphere (FV3) as its core. The FV3-based models are relatively untested at a convective-allowing scale. This study compares the performance of the UFS-FV3 and the WRF-ARW and verifies their respective simulations with radar data and verification metrics, such as initiation, timing, movement, and placement of convection. The case being examined is 28 April 2021 where large hail impacted the metropolitan areas of Norman, OK and San Antonio, TX. Both models had similar initiation, timing, and movement in both locations. Both models struggled to predict the hail event in Norman, OK due to the relative positioning of the storm and a cold front. While in San Antonio, the WRF predicted a strong storm with better placement, in regards to the radar, than the UFS. The WRF hailcast also predicted the severe hail threat to the west of San Antonio. The UFS hailcast, on the other hand, predicted abnormally small hail to the east and southeast of San Antonio, where no hail was reported. These findings highlight the need to further test the UFS hailcast scheme with the FV3-core.

1. Introduction

For decades, scientists have been developing Numerical Weather Prediction (NWP) models to be able to predict weather for days in advance with the aim to protect life and property, and anticipate long-term trends in climate. NWP models can benefit from using a common dynamical core across small to large space and time scales (e.g., convective storms to climate projections). By using a unified framework, community improvements to the model can be leveraged for a wide range of prediction scenarios. In 2016, the Finite Volume Cubed-Sphere (FV3) dynamical core (Putman and Lin 2007) was chosen to be implemented for the National Centers for Environmental Prediction (NCEP) Unified Forecast System (UFS). In 2019, the Global Forecast System (GFS) was upgraded from the spectral dynamical core to the FV3 dynamical core. The performance of the FV3 core at a global scale has been tested and well documented (Gallo et al. 2019). But, at a short-range (or convective-allowing) scale, the FV3 core is relatively untested.

The two NWP models this study will focus on are the UFS (referred as UFS-FV3) and the Weather Research and Forecasting (WRF) model (Skamarock et al. 2008). The FV3 core is capable of both hydrostatic and non-hydrostatic simulations. This study uses the Advanced Research WRF (ARW, referred as WRF-ARW) core, and it is utilized in non-hydrostatic mode to provide a baseline for comparison with the UFS-FV3. Why is the difference between hydrostatic and non-hydrostatic important in NWP models? Hydrostatic models use the hydrostatic approximation instead of the vertical momentum equation to estimate the vertical motion (American Meteorological Society 2012a). In non-hydrostatic models, the hydrostatic approximation is not taken into account and the vertical mo-

*Corresponding author address: Juan P. Mangual-Pagán, Department of Geological and Atmospheric Sciences - Iowa State University, 137 S Wilmoth Ave APT 201, Ames, Iowa 50014
E-mail: mangual@iastate.edu

mentum equation is used (American Meteorological Society 2012b). For global-scale prediction, models assume there is hydrostatic balance since strong upward or downward motions should average to near zero in large grid cells (e.g., tens of km). For small grid cells, it becomes more complicated, especially during severe weather conditions, the atmosphere may not always be in hydrostatic balance due to strong vertical motions. One example of the atmosphere not being in hydrostatic balance is inside of updrafts in supercells.

Tornadoes, high winds, heavy rainfalls, and hail are a threat to life and property, and these hazards are a high-priority to accurately predict. This study will primarily focus on the strengths and limitations of the UFS for hail prediction and will compare these predictions to the WRF. For the prediction of hail, models need to be able to accurately predict the evolution, movement, and intensity of the convective storms (Snook et al. 2016). But this is much harder than it seems and models do not have all of the information necessary to accurately predict hail. Forecast verification of hail is another challenge due to the dependence on reports from the public. These reports can suffer from size biases and under-reporting in low populated areas (Snook et al. 2016).

As previously mentioned, the FV3 core has been tested at a global scale, where the effects of non-hydrostatic effects are negligible. But at a short-range (or convective-allowing) scale, where non-hydrostatic effects are no longer negligible, the FV3 core has been relatively untested. One of the few instances that the FV3 core was studied at a short range level was during the 2018 Spring Forecast Experiment (SFE) (Gallo et al. 2019). With the use of statistics, the study compared the performance of the Geophysical Fluid Dynamics Laboratory FV3 (GFDL-FV3), the National Severe Storms Laboratory FV3 (NSSL-FV3), and the High Resolution Rapid Refresh v3 (HRRRv3) in various fields, including composite reflectivity. Fractional Skill Score (FSS, Roberts and Lean (2008)) and Critical Success Index (CSI, Schaefer (1990)) were used to see if the models had simulated reflectivity in the nearby areas. Both the FSS and CSI can tolerate position error when used in a neighborhood approach. The Gallo et al. (2019) study concluded that the HRRRv3 outperformed the GFDL-FV3 in composite reflectivity where there was a significant statistical difference. Other studies have looked at how FV3-based models predict precipitation at a short-range scale with similar statistics used in the 2018 SFE, such as Zhang et al. (2019).

In contrast to previous work, this study aims to look at the differences between how well the high-resolution models handled prediction of a high-impact hail event rather than a composite analysis across several weeks. Instead of using statistics to calculate the skill of the model, this study will assess the performance of both the UFS-FV3 and the WRF-ARW models through intercomparisons between models and verification of the model data with radar data and surface observations. A case study approach was chosen because it is more straightforward to understand what the models did and did not do well, and assess why those predictions succeeded or failed. High-impact hail events can cost billions of dollars in damages, hence accurate prediction of these events is particularly critical.

The case that will be examined is on 28 April 2021, when both Texas and Oklahoma were impacted by convective systems that produced high winds and large hail on or near large cities, such as Norman, Oklahoma, and San Antonio, Texas. The San Antonio storm is particularly interesting since the 6.4 inch (16.256 cm) hailstone in Hondo, Texas was the state record and was verified by the State Climate Extreme Committee (SCEC). Hondo is located 39.53 miles (63.62 km) to the west of San Antonio. Since the hailstone was verified at larger than 15 cm, it can be categorized as gargantuan (Kumjian et al. 2020). Reported hailstone sizes in both of these metropolitan areas ranged from 2 to 4+ inches (Storm Prediction Center (SPC) database).

The paper is organized as follows. Section 2 provides information about the UFS and WRF configurations, radar data, supplementary observations, and the process for forecast verification. Section 3 discuss the results of the models' simulations and the subsequent verification of the forecast. Section 4 provides a discussion and the implications for future research. Lastly, section 5 summarizes the results of the research.

2. Data and Methods

Model Configurations

The basic parametrizations of the UFS and WRF simulations both employ: Aerosol-aware Thompson microphysics (Thompson and Eidhammer 2014), Rapid Radiative Transfer Model (RRTMG) longwave and shortwave radiation (Mlawer et al. 1997; Iacono et al. 2000), Noah MP land surface model (Niu et al. 2011), and the Mellor-Yamada-Nakanishi-Niino (MYNN) scheme for the Boundary Layer (BL, Olson et al. (2019)). The UFS utilizes the Global Forecast System (GFS) surface layer scheme (Jiménez et al. 2012) and the WRF uses the modified MM5 Monin-Obukhov surface layer scheme (Jiménez et al. 2012). Both of these surface layer schemes are very similar and are based on the Monin-Obukhov similarity theory. For the UFS simulations, a Limited Area Model (LAM, Black et al. (2021)) was utilized. This allows the model to run on a regional domain, instead of the full globe. The domains of both the FV3-LAM and the WRF-ARW cover the Contiguous United States (CONUS). Both

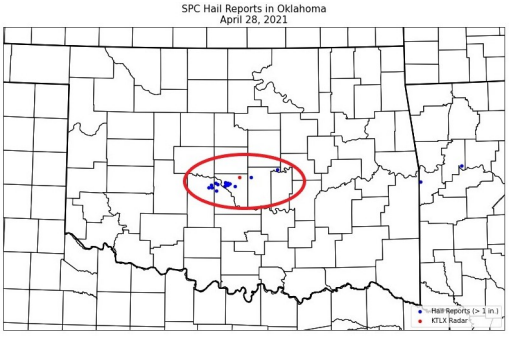

FIG. 1. SPC hail reports for 1 inch or larger hail across Oklahoma on 28 April 2021. Hail reports are shown in blue and the KTLX radar in red. Encapsulated in red is the general area of interest for hail.

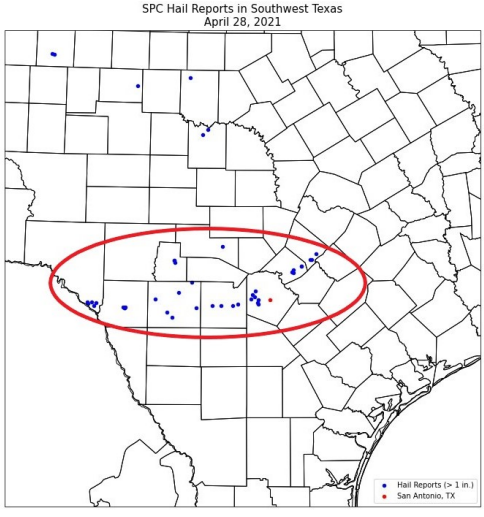

FIG. 2. SPC hail reports for 1 inch or larger hail across southwestern Texas on 28 April 2021. Hail reports are shown in blue and San Antonio, TX in red. Encapsulated in red is the general area of interest for hail.

models employ a 3 km grid spacing. The UFS uses an approximately 3 km grid spacing, but is fitted to a 3 km grid spacing in post-processing.

Both models were initialized at 18:00Z (1 p.m. CDT) on 28 April 2021 with the initial conditions given by GFS data. The forecast hours in the data were from 0 to 10 (F00-F10), so the model data spanned from 18Z on 28 April 2021 to 4Z on 29 April 2021. The model data output was every five minutes.

Radar Data

To verify the simulated reflectivity of the UFS-FV3 and the WRF-ARW, Next Generation Weather Radar (NEXRAD) archived data was used. The radar sites chosen were KTLX (Oklahoma City, OK) and KEWX (San Antonio, TX). The radar data was Level-II (or Base Data) at a 0.5° elevation angle, and the temporal resolution is approximately five minutes. The times of the radar scans are very similar to that of the model forecast hours, hence the radar data used was from 21Z 28 April 2021 to 4Z 29 April 2021. To create the figures of the radar data, a Python package called the Python ARM Toolkit (Py-Art) was utilized (Helmus and Collis 2016).

Supplementary Observations

To supplement radar data, additional observations, reports, and mesoanalysis discussions were used to have a better representation on what was happening near the surface and in the upper air. SPC archived surface observations plotted surface temperature and dew point temperature, mean sea level pressure and two hour pressure change, sky cover, wind speed and direction, visibility, and present weather. These observations can show if hail was reported at the station, or it can also aid in showing the passage of a front. SPC storm reports have preliminary data on tornado, hail, and wind reports on a particular day. For the purpose of this study, only the hail reports were taken into consideration. For the areas of KEWX (San Antonio, TX) and KTLX (Oklahoma City, OK) there was a combined total of 57 hail reports greater than 1 inch. All of the SPC hail reports for Oklahoma and southwestern Texas are shown in Fig. 1 and Fig. 2, along with a highlighted radius for the majority of the reports for both locations. In the San Antonio metropolitan area there were a total of 38 hail reports of greater than 1 inch and 18 reports of 1.75 inch or greater hail. As for the Oklahoma City metropolitan area, there were a total of 19 hail reports of greater than 1 inch and 17 reports of 1.75 inch or greater hail (SPC). These reports do not represent the entirety of the hailstorm at both places, but it does give a good sense of the hail size and the timing of the hail.

Forecast Verification

In order to verify the forecasts made by the UFS and the WRF, this study used 4 main metrics for the hailstorms in both locations. These were the storm initiation, timing, placement, and movement. Storm initiation in the models was defined as the instance that the storm consistently had values above 30 dBZ. The storm placement was defined as the center of the area of the highest reflectivity. Along with these metrics, the hailcast of both models is used to determine where the models produced hail within a time period.

3. Results

San Antonio, Texas

Overall, both model provided a reasonably accurate forecast of the convection west of San Antonio. In particular, the WRF had some storms to the west of San An-

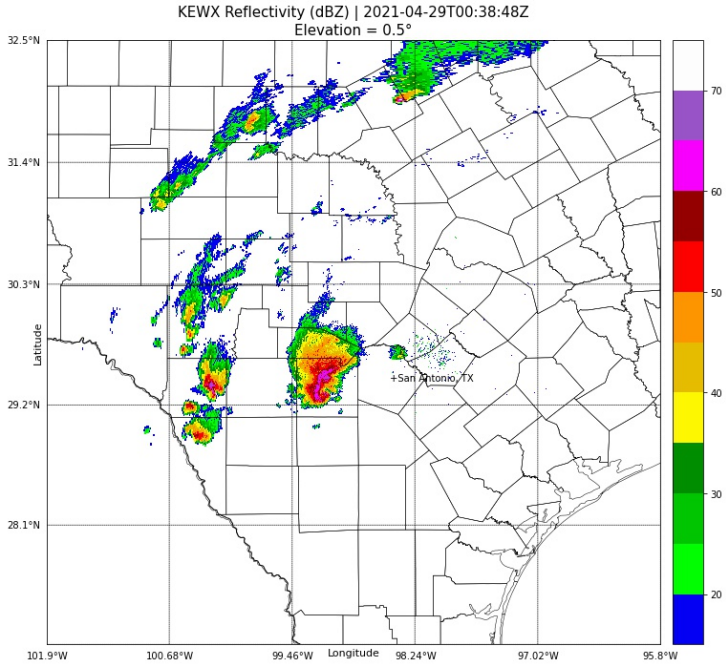

FIG. 3. KEWX reflectivity (dBZ), at a 0.5° elevation angle, across southwestern Texas during 0:38:48Z on 29 April 2021.

tonio with similar size and strength as what the KEWX radar showed, although it did under-estimate the number of storms. The UFS, on the other hand, overconvected towards the west and south Texas. In west Texas, the UFS overestimated the number of storms, but it did have the cells in the right location with a good approximate size and strength. As for the cells in southern Texas, these storms were not seen in the WRF simulation or radar reflectivity, but the UFS did have these cells moving northeast and reaching reflectivity values of 50-60 dBZ.

Both of the models initiated the storm in almost the same area and with very little difference in timing (5 minutes). The UFS initiates the storm at 20:50Z (F02:50) at 29.2°N 101.137°W, while the WRF initiates the storm at 20:55Z (F02:55) at 29.185°N 101.137°W. Both models initiated this storm approximately an hour later than what radar reflectivity showed. The models' initiation of the storm is very close to where the radar showed the storm at their respective times. The models' initiations were slightly off to the east, but this displacement was minimal. Comparing the models, the initial placement differed by 1.68 km (or 1.05 miles). Radar reflectivity shows the storm actually initiating at approximately 19:35Z. By the time that the models were initiating the storm, radar showed high enough reflectivity values for hail near the Texas-Mexico border. The UFS does have a stronger initiation than the WRF, with the highest reflectivity values at initiation being around 35-45 dBZ. The highest reflectivity values at the WRF's initiation was around 30-35 dBZ. Both models had similar initial sizes, with the UFS's storm being slightly larger.

Radar showed the San Antonio storm moving east towards San Antonio, with a secondary storm forming behind the initial storm that later produces hail. Both models had the storm moving northeast, eventually placing the storm north of what radar showed. At 00:38:48Z on 29 April 2021, the radar showed the highest reflectivity values at 29.475°N 99.155°W, shown in Fig. 3. At 00:40Z, the WRF's storm was placed at 29.888°N 99.38°W and 50.08 km (31.57 miles) from the actual storm seen by the radar. At the same time, the UFS's storm was placed at 30.025°N 98.85°W and 67.88 km away from the actual storm. Fig. 4-5 show the placement of the storms by the WRF and the UFS, respectively. The difference between storm placement in the models was 53.16 km (33.03 miles). At this time, the WRF had a small area of 60+ dBZ in the southern tip of the storm. These reflectivity values indicated a potential of large hail. The UFS, on the other hand, did not have the storm nearly as strong, but the UFS had some areas to the south of the storm reaching 50-60 dBZ. The UFS also had a better grasp of the secondary storm than the WRF, with the UFS's storm reaching moderate intensity (45-55 dBZ) at times. The WRF did not have the secondary storm well developed, but it does hint at the storm developing in the right location.

The WRF had a very good forecast of the potential hail threat to the west of San Antonio during its 18Z run. The WRF hailcast predicted hail ranging from 0.25 to 2.0

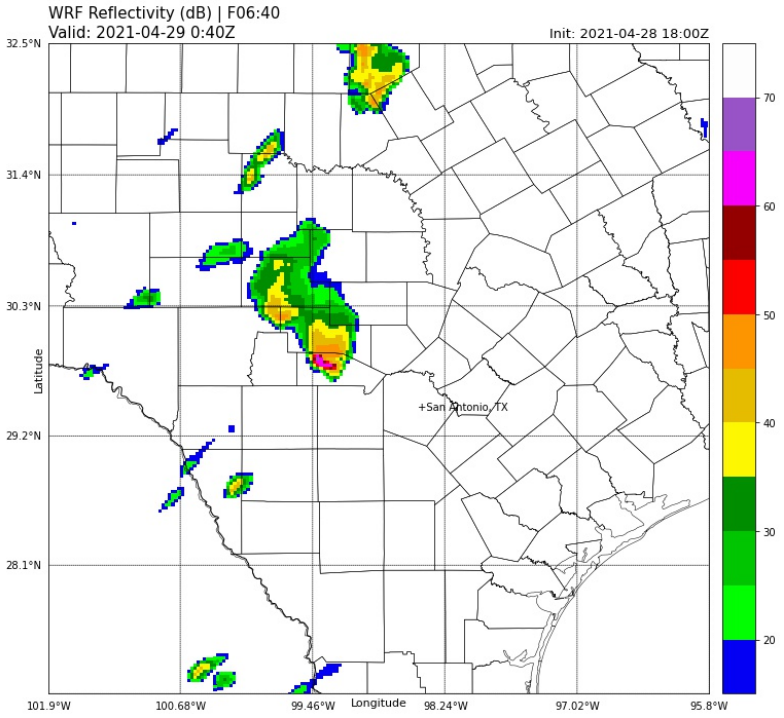

FIG. 4. WRF simulated reflectivity (dBZ) across southwestern Texas during 0:40Z on 29 April 2021. Reflectivity values under 15 dBZ are not plotted.

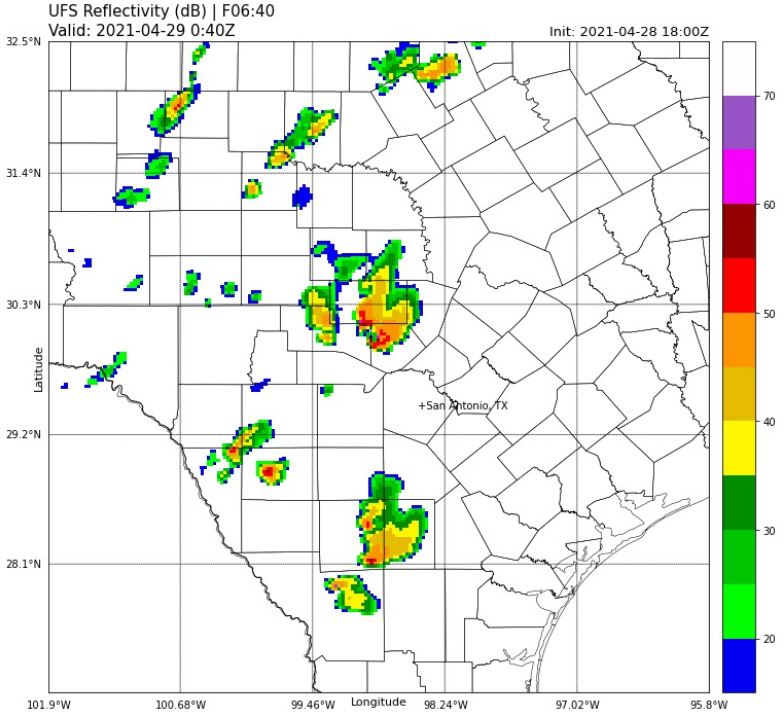

FIG. 5. UFS simulated reflectivity (dBZ) across southwestern Texas during 0:40Z on 29 April 2021. Reflectivity values under 15 dBZ are not plotted.

inches to the west of San Antonio. Something important to note is that the WRF did not predict any hail closer to the Texas-Mexico border in southwest Texas, as shown in Fig. 6 (a). A combination of SPC hail reports and radar reflectivity showed hail near the Texas-Mexico border, as the storm was intensifying and subsequently moving towards San Antonio. The WRF did not have hail in that area due to its late initiation. As for the UFS, the UFS hailcast variable showed irregular hail sizes across southern Texas, as shown in Fig. 6 (b). The UFS was predicting abnormally minuscule hail of 0.0005 to 0.0025 inches across a large area of southern Texas, where there was no hail threat. The UFS hailcast was not expected to give such irregular results, since when looking at the storm during its peak at 00:40Z, 55-60 dBZ were observed.. These reflectivity values are high enough for, at least, some hail to be present in the model. More information about the UFS hailcast will be examined in the discussion section.

Norman, Oklahoma

Overall, both models had a grasp of the general convection around eastern and southeastern Oklahoma, but overestimated the convection. The WRF did show reflectivity values upwards of 60+ dBZ in eastern Oklahoma, which hinted at the possibility of hail. The WRF had an accurate general depiction of the locations and intensity of convection. As for northern Oklahoma, both models overconvected, but the storms were generally weak.

Both the UFS and WRF initiated the storm at 23:30Z (F05:30), but radar reflectivity shows the storm initiating at 21:03:02Z. Both models initiated the storm approximately 2.5 hours later than the observed storm. Radar shows the initial placement of the storm at 35.07°N 100.07°W. The UFS's initial placement of the storm was at 35.558°N 99.43°W and the WRF's was at 35.556°N 99.22°W. The UFS's storm initiation was displaced approximately 93.9 km (58.35 miles) northeast of the actual storm initiation. Similarly, the WRF's storm initiation was approximately 94.9 km (or 58.97 miles) northeast. This shows a difference of 1 km (0.62 miles) between the initial storm placements by the UFS and the WRF. The WRF does have a similar structure to the radar reflectivity at initiation, where the UFS has a more compact cell with slightly higher reflectivity values. By the time that the models initiated the storm, radar showed the storm placed at 35.07°N 98.8°W with reflectivity values reaching 45-55 dBZ.

In contrast to San Antonio, both models struggled with the movement and subsequent intensification of the storm, when comparing radar and model reflectivity. This was due to the placement of a weak cold frontal boundary across central Oklahoma. When plotting derived reflectivity and wind direction (not shown), it can clearly be seen that the Norman storm was north of the boundary in both models. Since, both models initiated at the same time and under the same conditions, this is the most likely reason as to why both models struggled with the movement and placement of the Norman cell. At 1:55:53Z, the radar reflectivity showed the highest reflectivity values for the Norman hailstorm, (Fig. 7). This would have happened at F07:55 in the models, but during that time both models had the Norman cell placed northwest of Norman with moderate intensity (40-50 dBZ) at almost the same place. The WRF had the storm placed at 35.799°N 98.38°W (Fig. 8) and the UFS had it at 35.81°N 98.41°W (Fig. 9). Norman is situated at 35.223°N 97.44°W, hence the WRF had the storm 107 km (66.49 miles) away and the UFS placed the same storm 109.18 km (67.84 miles). The difference between the placement by the models was only 2.18 km (1.35 miles). Afterwards, both models have the storm moving eastward towards Norman, but dissipate before reaching the city. The WRF had the storm starting to dissipate and values lower than 30 dBZ during 2:00Z (F09:00) and the UFS shortly after.

Since both models did not have the storm over Norman and producing hail at any point of the storm's lifespan, the focus will shift to what caused this movement/placement. As previously mentioned, this was caused by the models' placement of the Norman cell with respect to the cold frontal boundary in central Oklahoma. This is further reaffirmed with SPC Mesoscale Discussion #0463 (SPC). This discussion mentions how along this boundary there were developing supercells that produced hail in the Oklahoma City area. As previously stated, both models initiated the storm north of this boundary. This meant that the storm could not intensify and move eastward with the boundary towards Norman. Shifting the attention back to the hailcast, Fig.10 and Fig.11 show the southeast displacement for hail after 10 forecast hours. When referring to the radar reflectivity, the storms to the east were significant enough at some times to produce hail, but not intense enough for hail larger than an inch. Although the WRF underestimated the size of hail in these storms, it still had a grasp over the potential threat for hail in southeastern Oklahoma. Similarly to San Antonio, TX, the UFS continued to produce hail in abnormally small sizes and across a large area.

4. Discussion

Both of the models had similar initiation, timing, and placement. This was mostly due to the storms occurring close to 18Z. Since the model data initializes at 18Z and is fed the same initial conditions, this is what causes the similarity between the models. Had the model data been from an earlier time, such as 12Z, the results would most likely been entirely different. As seen in the results section, the UFS hailcast was seen to display irregular hail sizes in both Norman, OK and San Antonio, TX. Fig. 10

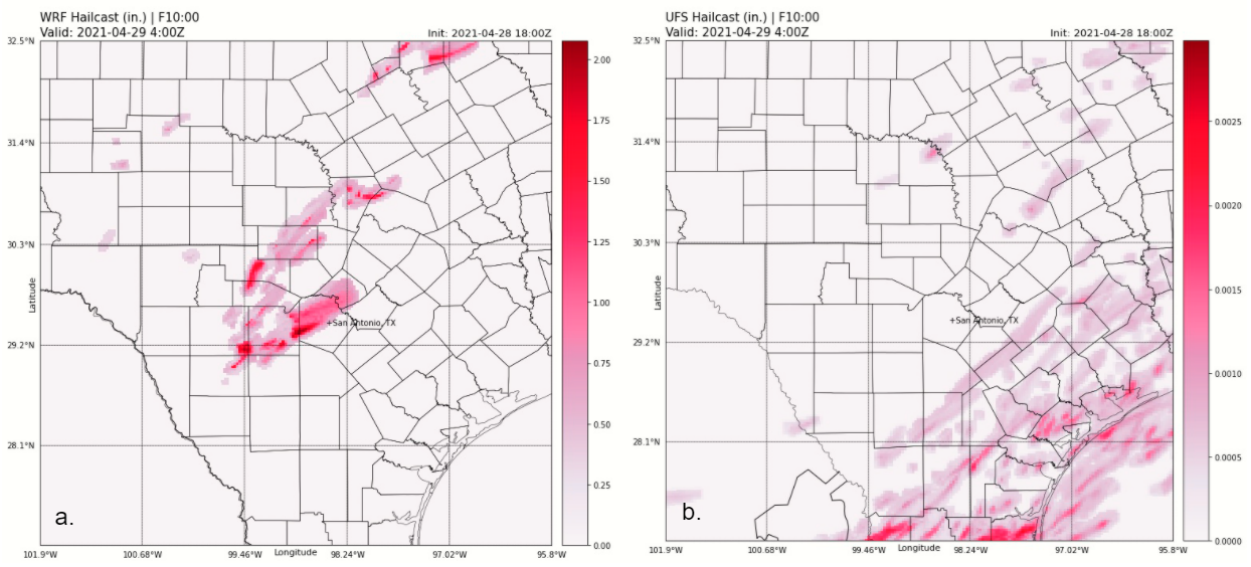

FIG. 6. WRF (a) and UFS (b) hailcast from F00 to F10 across southwestern Texas. WRF hailcast ranged from 0.0 to 2.0 inches. UFS hailcast ranged from 0.0 to 0.0025 inches

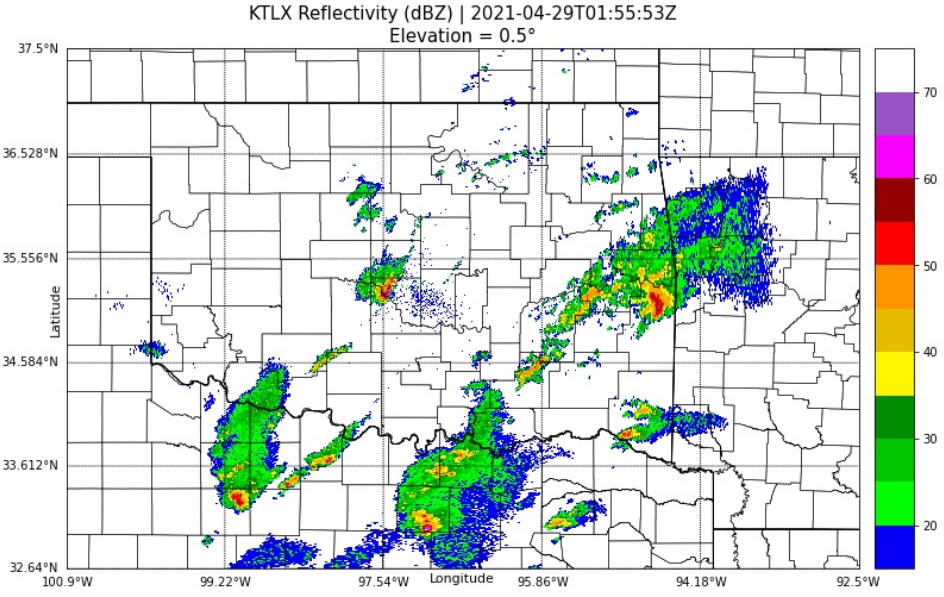

FIG. 7. KTLX reflectivity (dBZ), at a 0.5° elevation angle, across Oklahoma during 1:55:53Z on 29 April 2021. Reflectivity values under 15 dBZ are not plotted.

and Fig. **??** displayed the data values that were in the raw data. Hence, the values plotted from the hailcast was not an error. The UFS and WRF also use the same microphysics scheme, thus the hail prediction should not be too different between the models. Further research can be done looking into what type of precipitation the UFS predicting with the San Antonio storms. Variables such as rain and graupel would be of interest. For the Norman storm, further research could be done in manipulating the initial conditions of the 18z run, so the storms could initiate closer to the boundary and see if both of the models could then predict the hail threat. In terms of predicting the San Antonio hail event, both models had significant storms in their simulations capable of producing hail, even if the UFS's hailcast failed to do so. An ensemble of Convective-Allowing Models (CAMs) could have been of benefit when forecasting for this event before it occurred. An ensemble such as the experimental Warn-on-Forecast

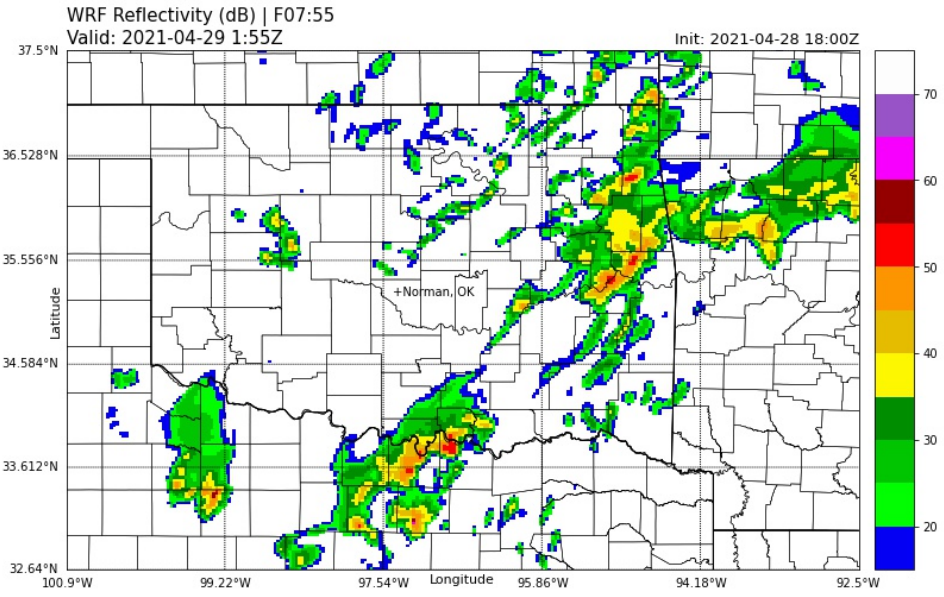

FIG. 8. WRF simulated reflectivity (dBZ) across Oklahoma during 1:55Z on 29 April 2021. Reflectivity values under 15 dBZ are not plotted.

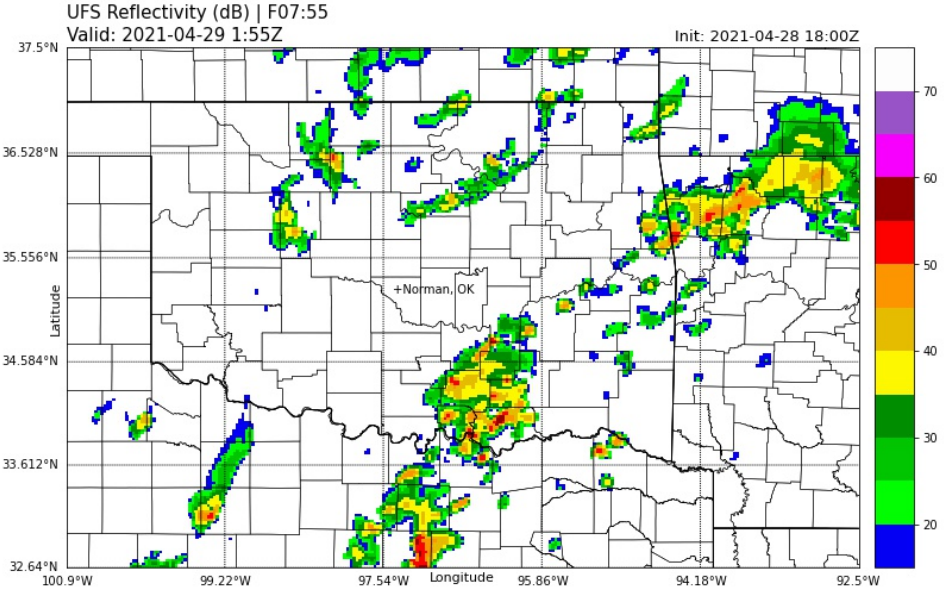

FIG. 9. UFS simulated reflectivity (dBZ) across Oklahoma during 1:55Z on 29 April 2021. Reflectivity values under 15 dBZ are not plotted.

System (WoFS) would have aided in determining a more accurate storm placement in comparison to the WRF or the UFS. For Norman, the placement of the storm relative to the boundary is what both of the models predicted incorrectly. An ensemble, such as WoFS, could have had one of its members predicting the storm along with the boundary and subsequently riding the boundary towards Norman. Whether or not the ensemble member eventually had the storm producing hail is another matter.

5. Conclusions

There has been few documented instances of the FV3-core being tested at a short-range scale. This study focused on comparing the performance of the UFS-FV3 and WRF-

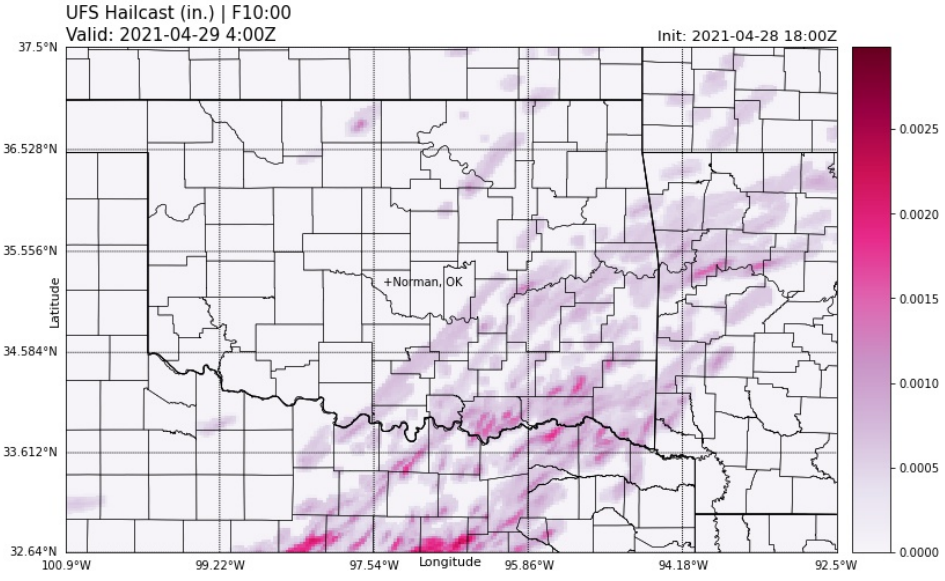

FIG. 10. UFS hailcast from F00 to F10 across Oklahoma. Hail sizes ranged from 0.0 to 0.0025 inches.

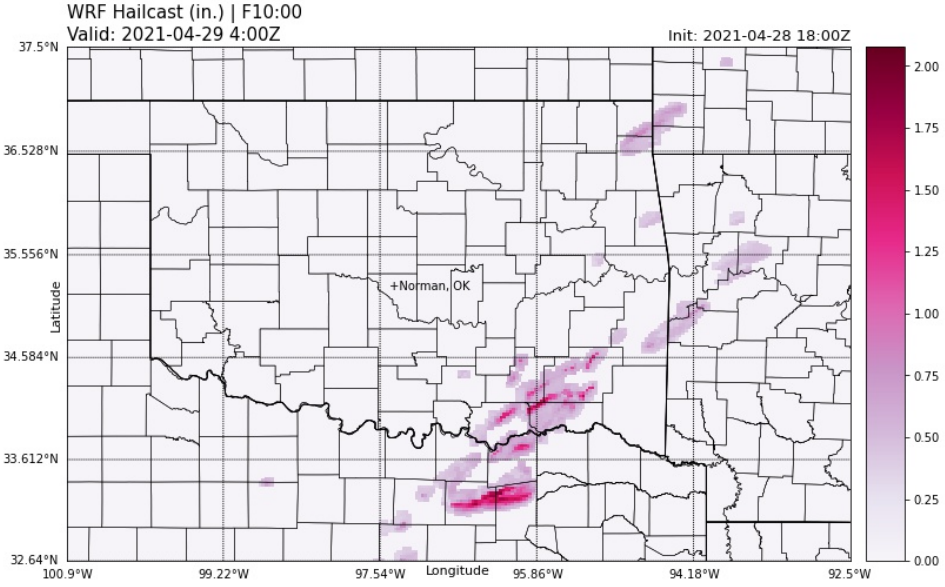

FIG. 11. WRF hailcast from F00 to F10 across Oklahoma. Hail sizes ranged from 0.0 to 2.0 inches.

ARW. This study compared the UFS-FV3 and WRF-ARW to radar reflectivity on the San Antonio, TX and Norman, OK hailstorms of 28 April 2021. Both models were in agreement with the initiation and timing of the San Antonio hailstorm, but less so when it came to the placement of the storm at its peak. Both models initiated the storm approximately 2.5 hours later than what the radar showed. The WRF had the better placement and higher intensity of the storm than the UFS. The WRF hailcast also produced large hail, with some areas west of San Antonio reaching 2.0+ inches. The UFS's hailcast produced extremely small hail to the east and southeast of San Antonio and no hail to the west of San Antonio. For Norman, OK, both models were in consensus for the initiation, timing, and placement of the hailstorm, with minor differences. Similar to San Antonio, both models initiated the Norman storms late by approximately an hour. Both models struggled with producing hail over Norman, due to their placement of the storm north of a cold frontal boundary in western Oklahoma and it caused the storms to be left stagnant northwest

of Norman. In conclusion, in this particular case, the models were mostly in agreement in terms of initiation, timing, and subsequent movement of the storms in both locations. For San Antonio, the WRF had a stronger storm than the UFS, and the WRF's hailcast showed areas of severe hail, while UFS's hailcast did not. For Norman, both models struggled with the placement of the storm with respect to the boundary and did not produce hail over Norman.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant No. AGS-2050267. This work was prepared by the authors with funding provided by National Science Foundation Grant No. AGS-2050267, and NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA11OAR4320072, U.S. Department of Commerce. David Bodine is supported by NSF Grant AGS-1823478. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, NOAA, or the U.S. Department of Commerce. The authors would like to thank Steve Nesbitt, professor of Atmospheric Sciences at the University of Illinois at Urbana-Champaign, for providing a python script to plot radar data over counties.

References

American Meteorological Society, 2012a: *Glossary of Meteorology*. Http://glossary.ametsoc.org/wiki/Hydrostatic_model.

American Meteorological Society, 2012b: *Glossary of Meteorology*. Http://glossary.ametsoc.org/wiki/nonhydrostatic_model.

Black, T. L., and Coauthors, 2021: A limited area modeling capability for the finite-volume cubed-sphere (fv3) dynamical core and comparison with a global two-way nest. *J. Adv. Modeling Earth System*, **13 (6)**, doi:https://doi.org/10.1029/2021MS002483.

Gallo, B., and Coauthors, 2019: Initial development and testing of a convection-allowing model scorecard. *Bull. Amer. Meteor. Soc.*, **100**, doi:10.1175/BAMS-D-18-0218.1.

Helmus, J. J., and S. M. Collis, 2016: The python arm radar toolkit (py-art), a library for working with weather radar data in the python programming language. *J. Open Res. Software*, **4**.

Iacono, M. J., E. J. Mlawer, S. A. Clough, and J.-J. Morcrette, 2000: Impact of an improved longwave radiation model, rrtm, on the energy budget and thermodynamic properties of the ncar community climate model, ccm3. *J. Geophys. Res.*, **105 (D11)**, 14 873–14 890, doi:https://doi.org/10.1029/2000JD900091.

Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, 2012: A revised scheme for the wrf surface layer formulation. *Mon. Wea. Rev.*, **140 (3)**, 898–918.

Kumjian, M. R., and Coauthors, 2020: Gargantuan hail in argentina. *Bull. Amer. Meteor. Soc.*, **101 (8)**, 1241 – 1258, doi:10.1175/BAMS-D-19-0012.1.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: Rrtm, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102 (D14)**, 16 663–16 682, doi:https://doi.org/10.1029/97JD00237.

Niu, G.-Y., and Coauthors, 2011: The community noah land surface model with multiparameterization options (noah-mp): 1. model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116 (D12)**, doi:https://doi.org/10.1029/2010JD015139.

Olson, J. B., J. S. Kenyon, W. Angevine, J. M. Brown, M. Pagowski, K. Sušelj, and Coauthors, 2019: A description of the mynn-edmf scheme and the coupling to other components in wrf-arw. doi:https://doi.org/10.25923/n9wm-be49.

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

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136 (1)**, 78–97.

Schaefer, M. G., 1990: Regional analyses of precipitation annual maxima in washington state. *Wat. Resources Res.*, **26 (1)**, 119–131, doi:https://doi.org/10.1029/WR026i001p00119.

Skamarock, W., J. Klemp, J. Dudhia, D. Gill, D. Barker, W. Wang, and J. Powers, 2008: A description of the advanced research wrf version 3. **27**, 3–27.

Snook, N., Y. Jung, J. Brotzge, B. Putnam, and M. Xue, 2016: Prediction and ensemble forecast verification of hail in the supercell storms of 20 may 2013. *Wea. Forecasting*, **31 (3)**, 811–825.

Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636, doi:10.1175/JAS-D-13-0305.1.

Zhang, C., and Coauthors, 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. *J. Geophys. Res.*, **46**, doi:10.1029/2018GL081702.