Using GFDL C-SHiELD for the Prediction of Convective Storms during the 2021 Spring and Summer

Alex Kaltenbaugh
Lucas Harris
Kai-Yuan Cheng
Linjiong Zhou
Matthew Morin
William Stern

17 June 2022

GFDL Weather and Climate Dynamics Division
Technical Memorandum GFDL202201

1. Introduction

Advances in numerical modeling have improved prediction skill for both global convection-permitting and regional convection-allowing models. As severe storms continue to impact the United States each year, there has been a considerable need for accurate, extended-range forecasts of severe weather outbreaks. Global models have the advantage of being able to produce long-range (2–14 day) forecasts, but it is too expensive to run convective-scale models globally. Regional models allow for convective-scale prediction without the expense of global modeling. However, the extended-range prediction of convective storms requires accurate simulations across multiple spatial and temporal scales, including scale interactions. Previous studies (Harris & Lin, 2013; Harris et al., 2014, 2019; Zhou et al., 2019) have shown the benefits of using global models with variable resolutions for short-to-medium range mesoscale and storm-scale prediction.

At the Geophysical Fluid Dynamics Laboratory (GFDL), our research group recently introduced the System for High-resolution prediction on Earth-to-Local Domains (SHiELD), which couples the nonhydrostatic Finite-Volume Cubed-Sphere Dynamical Core (FV3; Putman & Lin, 2007) with model physics originating from the NCEP's operational weather forecast model, the Global Forecast System (GFS; Harris et al., 2020). SHiELD exists within a modeling suite at GFDL that is unified around the FV3 dynamical core, and it can be used on a variety of spatial and temporal scales with a focus on short-to-medium range and subseasonal-to-seasonal weather forecasting. Due to this flexibility, SHiELD has several configurations for specific forecasting purposes. Continental SHiELD (C-SHiELD) is the configuration designed for the explicit prediction of continental convection and severe weather, which uses a 3-km nest over the Contiguous United States (CONUS) within a uniform 13-km global grid (GFDL, 2021).

In recent years, FV3-based models including C-SHiELD have shown skill with convective-scale prediction of continental convection and severe weather (Potvin et al., 2019; Harris et al., 2019,2020; Zhang et al., 2019). In this report, we evaluate the current 2021 version of C-SHiELD and compare the results to the 2020 version and the global 13-km SHiELD. This report is organized as follows. Section 2 describes the model settings and configurations. Section 3 details the validation methods and all observational/reanalysis datasets utilized. Section 4 presents the results of the validations performed with C-SHiELD, and Section 5 summarizes the report and presents conclusions about the model's performance.

2. Model Description and Configuration

a. FV3 Dynamical Core

All SHiELD simulations are built around the FV3 dynamical core, which has been described in greater detail in previous papers (Harris & Lin, 2013; Lin, 2004; Putman & Lin, 2007; Harris et al. 2021; and references therein). FV3 is based on the finite-volume (FV) dynamical core used for many years by the National Oceanic and Atmospheric Administration (NOAA), the National Aeronautics and Space Administration (NASA), the National Center for Atmospheric Research (NCAR), and others. SHiELD uses a nonhydrostatic solver for the fully compressible Euler equations on a gnomonic cubed-sphere grid with a Lagrangian vertical coordinate (Harris et al., 2020). The algorithm used is horizontally explicit with fast vertically propagating sound and gravity waves solved by a semi-implicit method (Harris et al., 2019,2020). The piecewise-parabolic method is used in the discretization along Lagrangian surfaces, and dynamical quantities (e.g. vorticity, potential temperature, and air mass) are advected by a nonmonotonic scheme while tracers are advected using a positive-definite scheme (Harris et al., 2020). The horizontal discretization for C-SHiELD is on the gnomonic cubed-sphere grid and utilizes two-way grid nesting with a simplified two-way update (Harris & Lin, 2013, 2014; Harris et al., 2020).

b. Model Settings and Schemes

The physics used in SHiELD and C-SHiELD are similar to those used in the GFS version 16 (GFSv16). All simulations use the NCEP scale-aware Simplified Arakawa-Schubert (SA-SAS; Han et al., 2017) convection scheme, which includes shallow and deep convection. However, the deep convection is only applied to the global domain for C-SHiELD (GFDL, 2021). Both SHiELD and C-SHiELD use the five-category GFDL microphysics and cloud fraction scheme that is fully in-lined with the FV3 (Zhou et al., 2019; Harris et al., 2020), and both also use the scale-aware TKE-EDMF Planetary Boundary Layer (PBL) scheme described in Han and Bretherton (2019), which was also applied to convective-scale prediction by Snook et al. (2019) and Zhang et al. (2019). Both configurations use the Noah Land Surface model (Ek et al., 2003) that is integrated within the GFS physics and surface layer scheme for the global domain (Harris et al., 2020; GFDL, 2021). However, the 2021 version of C-SHiELD uses an updated Noah-Multiparameterization Land Surface Model with GFDL 2021 updates on its nested domain (GFDL, 2021). SHiELD and C-SHiELD both use a global cubed-sphere grid with

768 by 768 grid cells on each face (C768 grid). For the 2020 version of C-SHiELD, the global grid is rotated and stretched to a width of about 9-km over the CONUS. By contrast, the global domain used in the 2021 C-SHiELD is unstretched and identical in grid resolution to the well-tuned 13-km SHiELD (GFDL, 2021). This allows for us to leverage the advantages of the unified modeling approach and focus on tuning the nested grid in C-SHiELD. The global and regional domains for each configuration are shown in Figure 1.

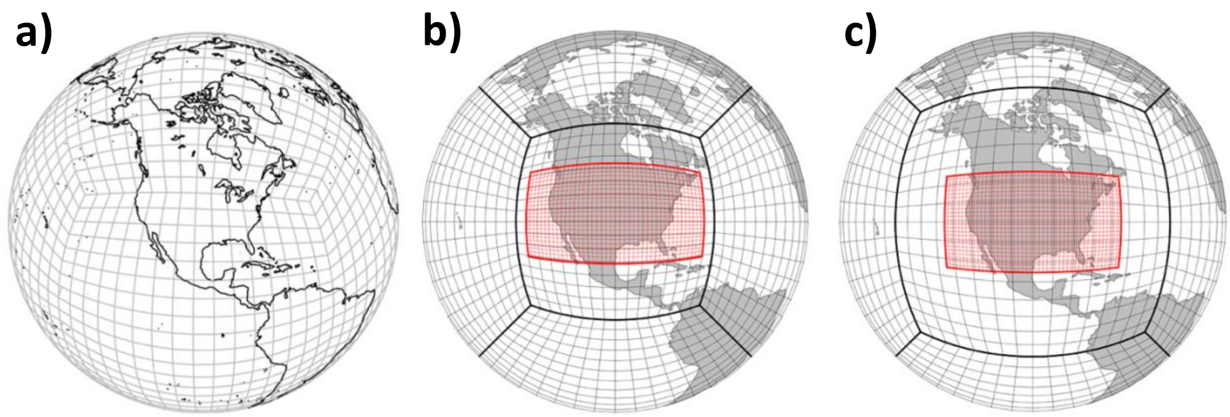


Figure 1: Global and nest domains for SHiELD (a), C-SHiELD 2020 (b), and C-SHiELD 2021 (c). Note that each grid cell in the figure represents many actual grid cells in the model configurations.

Despite the changes to the global domain, both versions of C-SHiELD have 3-km two-way nests that are placed over the CONUS (GFDL, 2021). SHiELD uses 91 Lagrangian vertical levels between ~12m and the model top around 40 Pa (Harris et al., 2020; GFDL, 2021) while C-SHiELD has 63 Lagrangian vertical levels between ~15 m and the model top at 64 Pa (GFDL, 2021). All C-SHiELD forecasts are integrated to 126 hours (5.25 days) while SHiELD forecasts are integrated to 240 hours (10 days) from the initialization time (Harris et al., 2020; GFDL, 2021). In terms of computational performance, SHiELD has a peak performance of about 80 min/10 days (or roughly 8.0 min/day) when run using 3024 cores on Gaea-C4. C-SHiELD has a peak performance of 139 min/5.25 days (or roughly 26.5 min/day) using 3024 cores on Gaea-C4 (GFDL, 2021). Table 1 summarizes the characteristics and settings of SHiELD and C-SHiELD.

2020 SHiELD	2020 C-SHiELD	2021 C-SHiELD
Medium-range weather prediction	Short-to-medium-range contiguous US severe weather prediction	Short-to-medium-range contiguous US severe weather prediction
Integrated to 10 days (240 hours)	Integrated to 5.25 days (126 hours)	Integrated to 5.25 days (126 hours)
Global 13-km (C768)	Stretched 13-km global domain (C768r), 3-km CONUS nested domain	Unstretched 13-km global domain (C768), 3-km CONUS nested domain
91 Levels; lowest mid-level at ~12 m, model top at 40 Pa	63 Levels; lowest mid-level at ~15 m, model top at 64 Pa	63 Levels; lowest mid-level at ~15 m, model top at 64 Pa
2020 In-Line GFDL MP	2020 In-Line GFDL MP with hail on nest	2021 In-Line GFDL MP with hail on nest
High-resolution Noah Land Model with GFDL 2019 updates	High-resolution Noah Land Model with GFDL 2019 updates	High-resolution Noah Land Model with GFDL 2020 updates (GFDL 2021 updates with Noah-MP on nest domain)
TKE-EDMF Planetary Boundary Layer with GFDL 2020 updates	TKE-EDMF Planetary Boundary Layer with GFDL 2020 updates	TKE-EDMF Planetary Boundary Layer with GFDL 2021 updates
NCEP 2017 Scale-Aware SAS (shallow and deep) Convection	NCEP 2017 Scale-Aware SAS (shallow and deep) Convection; deep convection on global domain only	NCEP 2017 Scale-Aware SAS (shallow and deep) Convection; deep convection on global domain only
Positive-definite Tracer Advection	Positive-definite Tracer Advection	Positive-definite Tracer Advection

Table 1: Summary of model settings and schemes for SHiELD (left), C-SHiELD 2020 (middle), and C-SHiELD 2021 (right). For C-SHiELD 2020, the "C768r" label is used to distinguish the stretched global domain from the unstretched domain used in SHiELD 2020 and C-SHiELD 2021.

3. Validation Methods and Description

In this report, the validations and analyses were performed using the Model Evaluation Tools (MET) Version 9.1.1 developed by NCAR in 2020 (MET, 2020). We compared C-SHiELD 2021 to C-SHiELD 2020 as well as the 13-km SHiELD and operational GFS (version 16). Note that SHiELD is labeled as “RT2020_C768” in all figures shown in this report. For the precipitation and surface analyses, we also compared C-SHiELD to the 3-km nest of the operational North American Mesoscale model (NAM; NCEP, 2017) and the High-Resolution Rapid Refresh model (HRRR; GSL, 2020). The 3-km nest of the NAM has been designed specifically for convective-scale prediction, so it serves as a baseline against which C-SHiELD can be evaluated. For the large-scale fields, we validated the models against the fifth-generation reanalysis dataset (ERA5) from the European Centre for Medium-Range Weather Forecasts (ECMWF) with a 30-km horizontal resolution (ECMWF, 2021). Surface fields (e.g. 2-meter temperature, dewpoint, specific humidity, and 10-meter winds) were validated against the 2.5-km Real-Time Mesoscale Analysis (RTMA; NCEP, 2021), and precipitation was validated against the 6-hourly 4-km multi-sensor STAGE IV quantitative precipitation dataset (NCAR, 2022). Table 2 summarizes all atmospheric fields that were validated in this analysis.

Abbreviated Name(s)	Full Name(s)	Units
h500, h200, h850, h700	Geopotential height at 500, 200, 850, and 700 hPa	Meters (m)
t500, t200, t850, t700	Temperature at 500, 200, 850, and 700 hPa	Kelvin (K)
u500, u200, u850, u700	U-wind at 500, 200, 850, and 700 hPa	Meters per second (m/s)

v500, v200, v850, v700	V-wind at 500, 200, 850, and 700 hPa	Meters per second (m/s)
q500, q850, q700	Specific humidity at 500, 850, and 700 hPa	Grams per kilogram (kg/kg)
TMP2m	2-meter temperature	Kelvin (K)
u10, v10	10-meter u-wind and v-wind	Meters per second (m/s)
lcc, hcc, tcc	Low, high, and total cloud fractions	Percent (%)
sshf, slhf	Surface sensible and latent heat fluxes	Watts per meter squared (W/m^2)
Plt	Total precipitation	Millimeters per day (mm/day)
Fsds	Surface incoming shortwave radiation	Watts per meter squared (W/m^2)
Fsus	Surface net shortwave radiation	Watts per meter squared (W/m^2)
Fsdt	TOA incoming shortwave radiation	Watts per meter squared (W/m^2)
Fsut	TOA net shortwave radiation	Watts per meter squared (W/m^2)
Flut	TOA net longwave radiation	Watts per meter squared (W/m^2)

Table 2: Full names and units of fields that were validated.

We analyzed 143 forecasts run daily at 00Z between April 11 and August 31, 2021. All forecasts were validated every 6 hours out to 120 hours (5 days) from the initialization time. For the large-scale and surface analyses, root-mean-squared error (RMSE), bias, and anomaly correlation coefficient (ACC) were computed by MET at each grid point and then averaged over three regions: CONUS, the Eastern U.S. (EAST), and the Western U.S. (WEST), which are shown in Figure 2.

Figure 2: Map showing the CONUS, EAST, and WEST domains used in the model validations.

For the precipitation analyses, we analyzed the frequency bias and forecast skill. The frequency bias is a ratio computed by dividing the number of forecast events by the number of observation events within the grid domain that exceed a specific precipitation threshold (mm/6hr). A frequency bias of 1 would indicate an unbiased forecast. The precipitation skill was evaluated using two different metrics: the Fractional Skill Score (FSS; Roberts & Lean, 2008) and the Gilbert Skill Score (GSS), which is also known as the Equitable Threat Score (ETS; Hogan & Mason, 2012). The FSS compares the fractional coverage of forecast and observed events that exceed a certain rain threshold within a spatial domain. This is done in spatial windows of increasing size until the window covers the entire domain of the forecast and/or observations. By contrast, the GSS measures the fraction of observed and forecast events that were correctly predicted within a certain precipitation threshold, but it is adjusted for hits that are associated with random chance. Both the FSS and GSS are measured between 0 and 1, with 0 representing no skill and 1 representing perfect skill. Examining both metrics allows us to account for differences in forecast skill across various spatial scales while also obtaining a measure of skill that eliminates the success rate due to random chance.

4. Results

a. Large-scale Analysis

When evaluating the performance of models, it is useful to first evaluate forecast skill with large-scale fields, notably the 500-mb geopotential height. Poor 500-mb skill can be indicative of forecast degradation, which can degrade even weather forecasts at shorter lead times and of severe weather (Harris et al., 2019, 2020). Thus, it is important to validate the 500-mb height forecasts before examining specific weather events. Figure 3 shows the time series of the April-August averaged 500-mb height root-mean-square error (RMSE: m) and anomaly correlation coefficient (ACC). The RMSE and ACC in C-SHiELD 2021 are improved from the 2020 version. In addition, C-SHiELD 2021 shows a very similar 500-mb performance to the 13-km SHiELD, which is designed specifically for short- to medium-range weather prediction. Both SHiELD and C-SHiELD 2021 perform similarly to (and occasionally better than) the operational GFS.

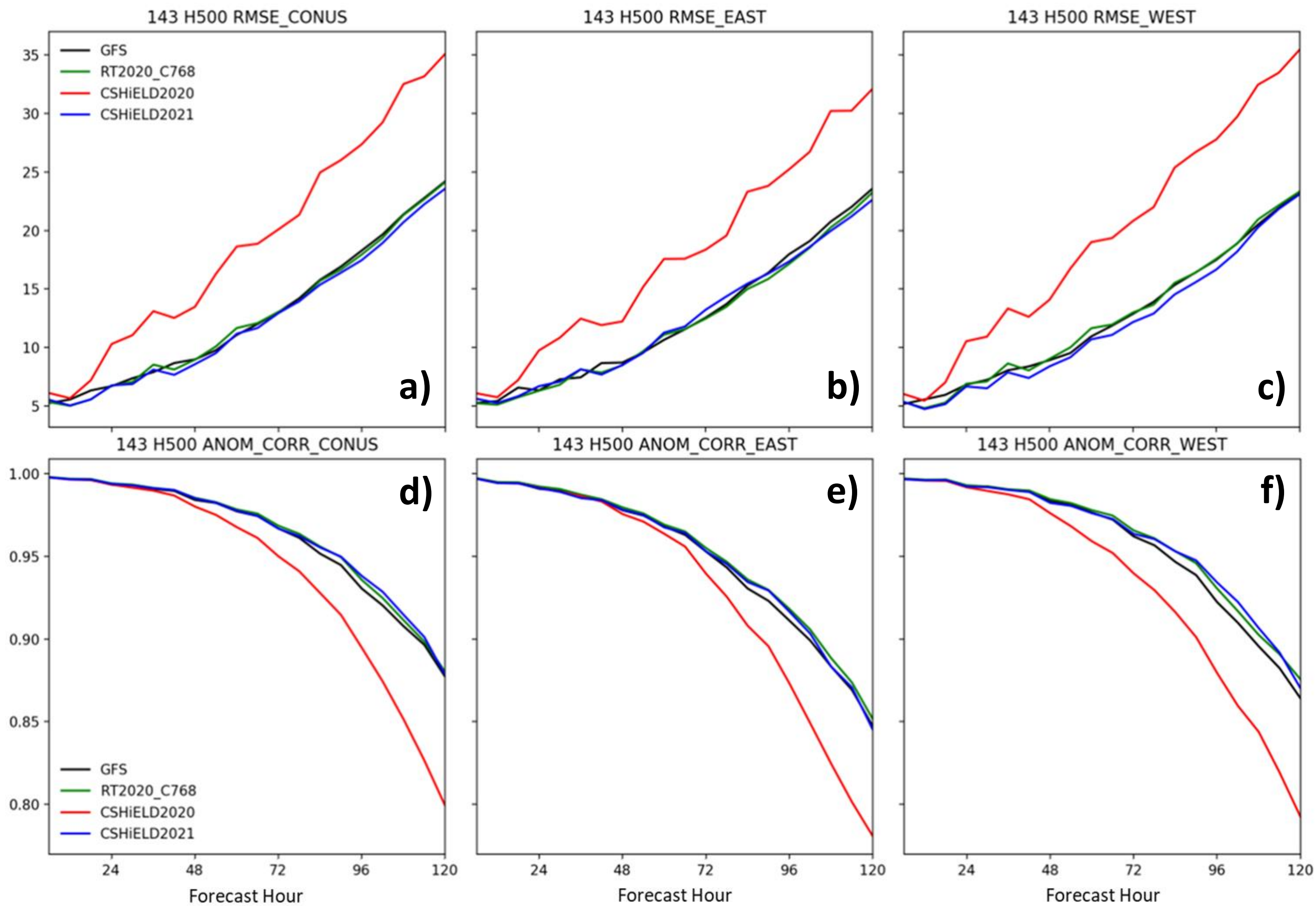


Figure 3: April-August averaged 500-mb height RMSE (m; a-c) and ACC (d-f). RT2020_C768 refers to SHiELD. Column 1 is the Contiguous U.S. (CONUS), column 2 is the Eastern U.S. (EAST), and column 3 is the Western U.S. (WEST).

Aside from the 500-mb height, C-SHiELD 2021 shows improvements in forecast skill over C-SHiELD 2020 in many other fields, including the geopotential heights, winds, and temperatures between 850 hPa and 200 hPa (Figure 4). Indeed, for nearly all fields the RMSE is improved in C-SHiELD 2021, the principal exceptions being the surface shortwave and turbulent fluxes.

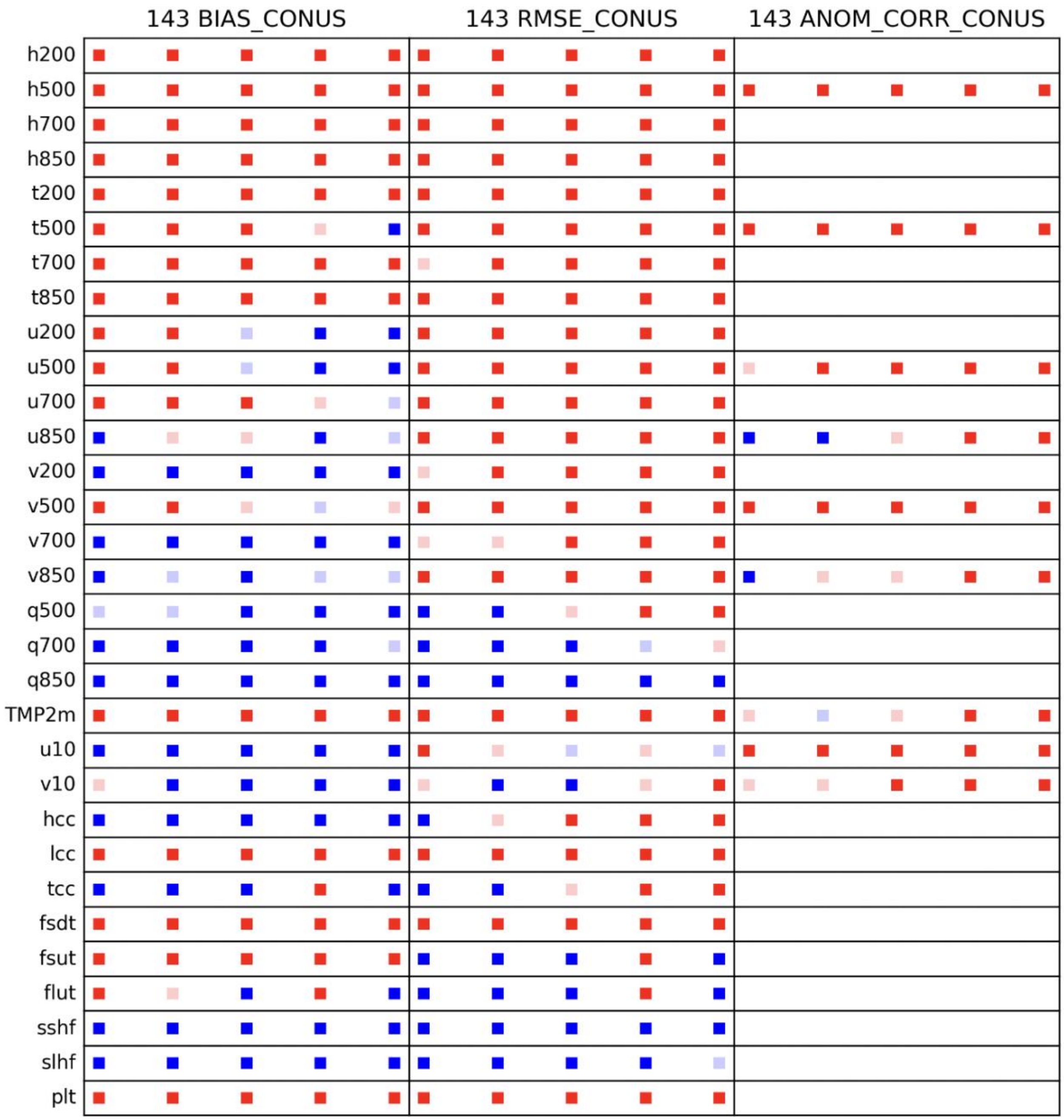
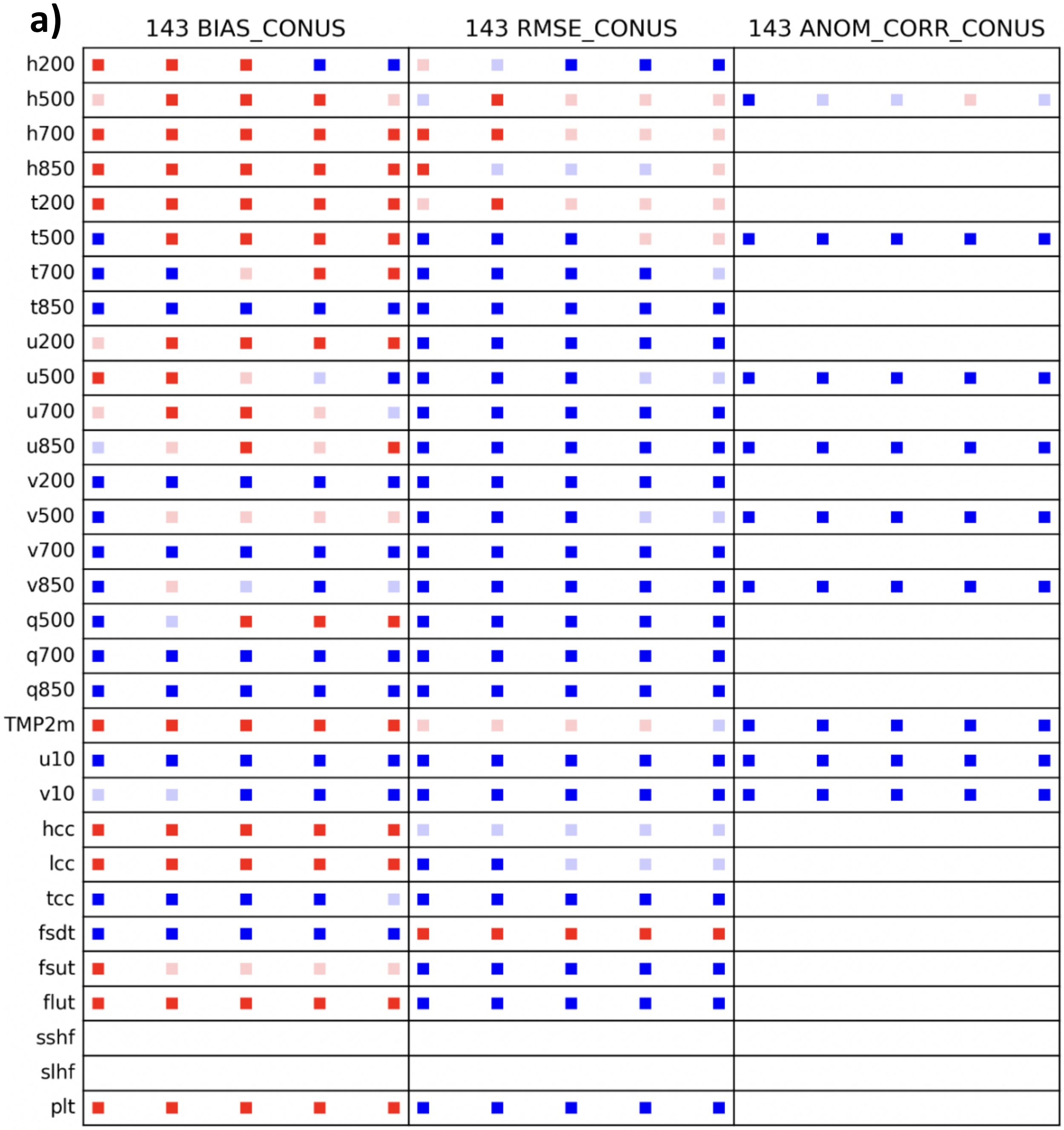


Figure 4: Score card comparing C-SHiELD 2021 to C-SHiELD 2020. The red color indicates when C-SHiELD 2021 performed better than C-SHiELD 2020 while the blue color indicated where it performed worse. Darker colors represent statistically significant changes while lighter colors do not.

Improvement upon SHiELD and the GFS (Figure 5) is more difficult since these models are optimized for large-scale skill, and indeed for most quantities C-SHiELD does not perform as well. However, in some fields, especially radiative fluxes, 2-m temperature, and geopotential height, C-SHiELD's performance over the CONUS is better than the global models, sometimes significantly so.

a)



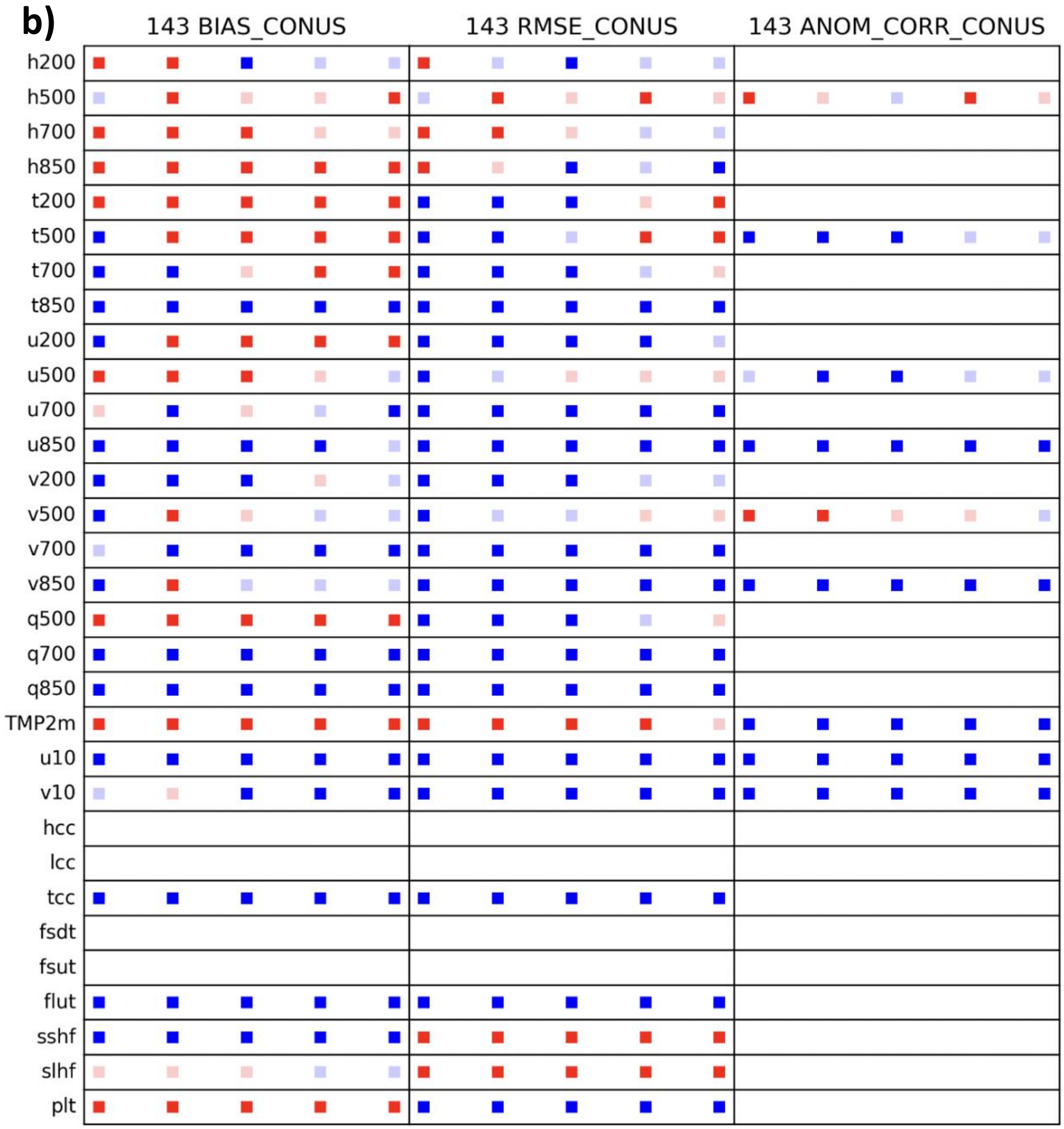


Figure 5: Score cards comparing C-SHiELD 2021 to SHiELD (a) and the GFS (b). The red color indicates when C-SHiELD 2021 performed better while the blue color indicated where it performed worse. Darker colors represent statistically significant changes while lighter colors do not.

b. Surface Analysis

After analyzing the forecast skill at multiple pressure levels, we validated the models near the surface at a 3-km resolution. The regional convective-scale models (NAM and HRRR) are heavily optimized to reduce biases and errors in these fields, making them challenging baselines for C-SHiELD. Figure 6 shows the time series of the April-August averaged 2-meter temperature (TMP2m) bias (K) and RMSE (K) over the three regions (CONUS, EAST, and WEST).

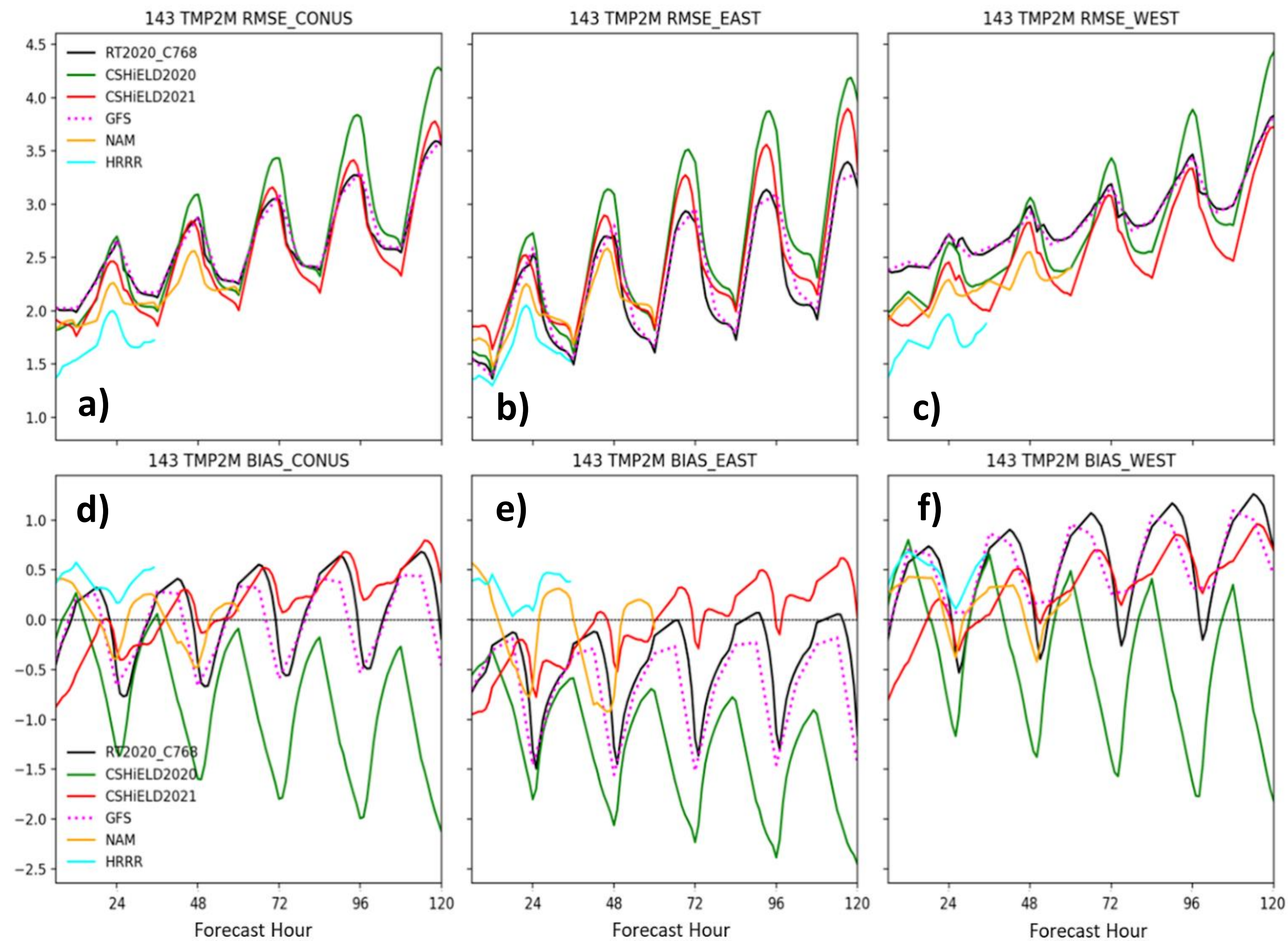


Figure 6: April-August averaged 2-meter temperature RMSE (K; a-c) and BIAS (K; d-f). RT2020_C768 refers to SHiELD. Column 1 is the Contiguous U.S. (CONUS), column 2 is the Eastern U.S. (EAST), and column 3 is the Western U.S. (WEST).

The TMP2m bias for C-SHiELD 2021 is improved over C-SHiELD 2020 and SHiELD, especially during the first three days (72 hours) of the forecasts. Beyond three days, C-SHiELD 2021 begins to drift towards a positive temperature (or warm) bias while C-SHiELD 2020 has a rather cold bias. The RMSE of C-SHiELD 2021 is also noticeably lower than C-SHiELD 2020 and closer to that of SHiELD over the CONUS. It is also worth noting that the diurnal cycles of the BIAS and RMSE, noted in earlier versions of C-SHiELD (Harris et al. 2019), are significantly reduced in C-SHiELD 2021, which is indicative of year-over-year improvements. The TMP2m bias and RMSE for C-SHiELD 2021 are roughly on par with the 3km NAM while the HRRR has a smaller RMSE and a stronger warm bias in the first 24-36 hours. The changes in TMP2m biases between C-SHiELD 2020 and 2021 over the CONUS can be visualized in Figures 7b and 7c, which show that the cold bias over the Eastern U.S. is substantially reduced in C-SHiELD 2021.

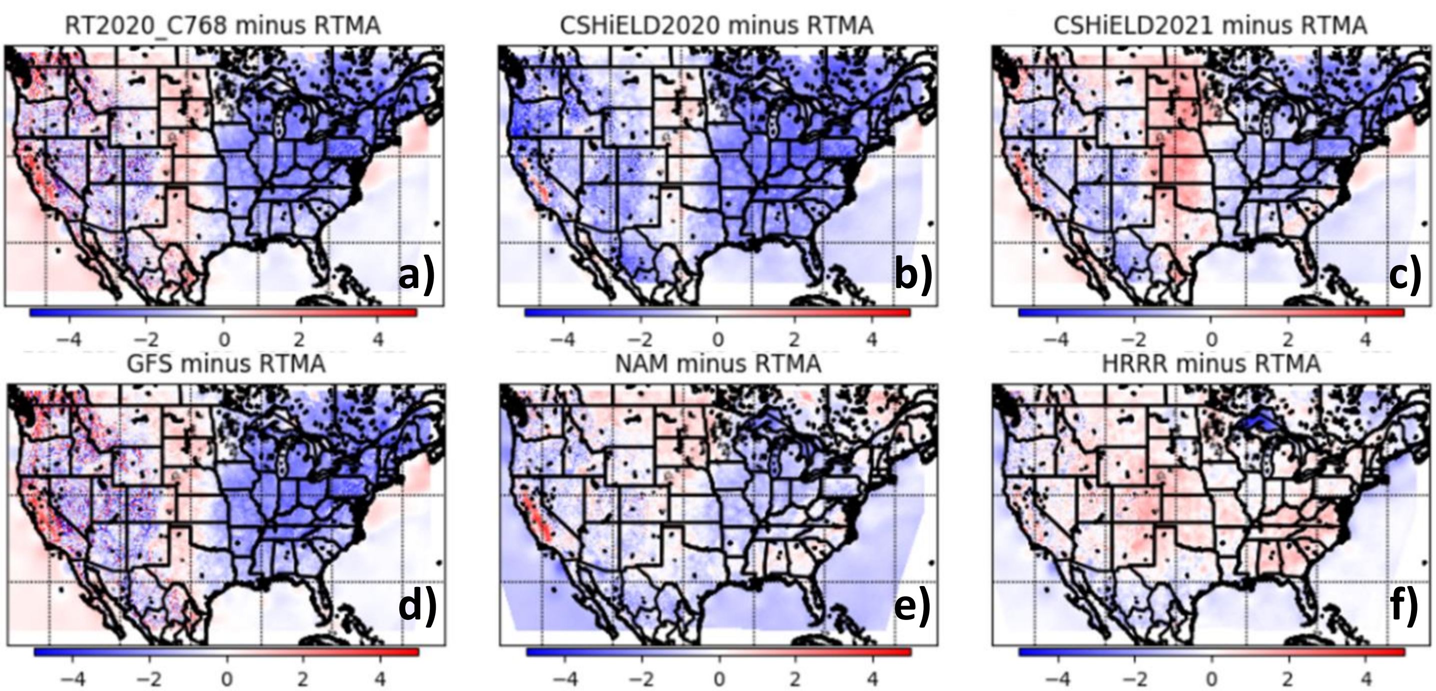


Figure 7: Maps of the April-August averaged 2-meter temperature BIAS (K) at 24 hours after the initialization time for SHiELD (a), C-SHiELD 2020 (b), C-SHiELD 2021 (c), GFS (d), NAM (e), and HRRR (f).

In addition to temperature, we validated the 2-meter dewpoint temperature (DPT2m) and specific humidity (SPFH2m) in order to examine the moisture biases in C-SHiELD. The patterns are similar in both models; we include SPFH2m since SHiELD does not make DPT2m output available. Figure 8 shows the April-August averaged DPT2m bias (K) and SPFH2m bias (g/kg) as a function of lead time.

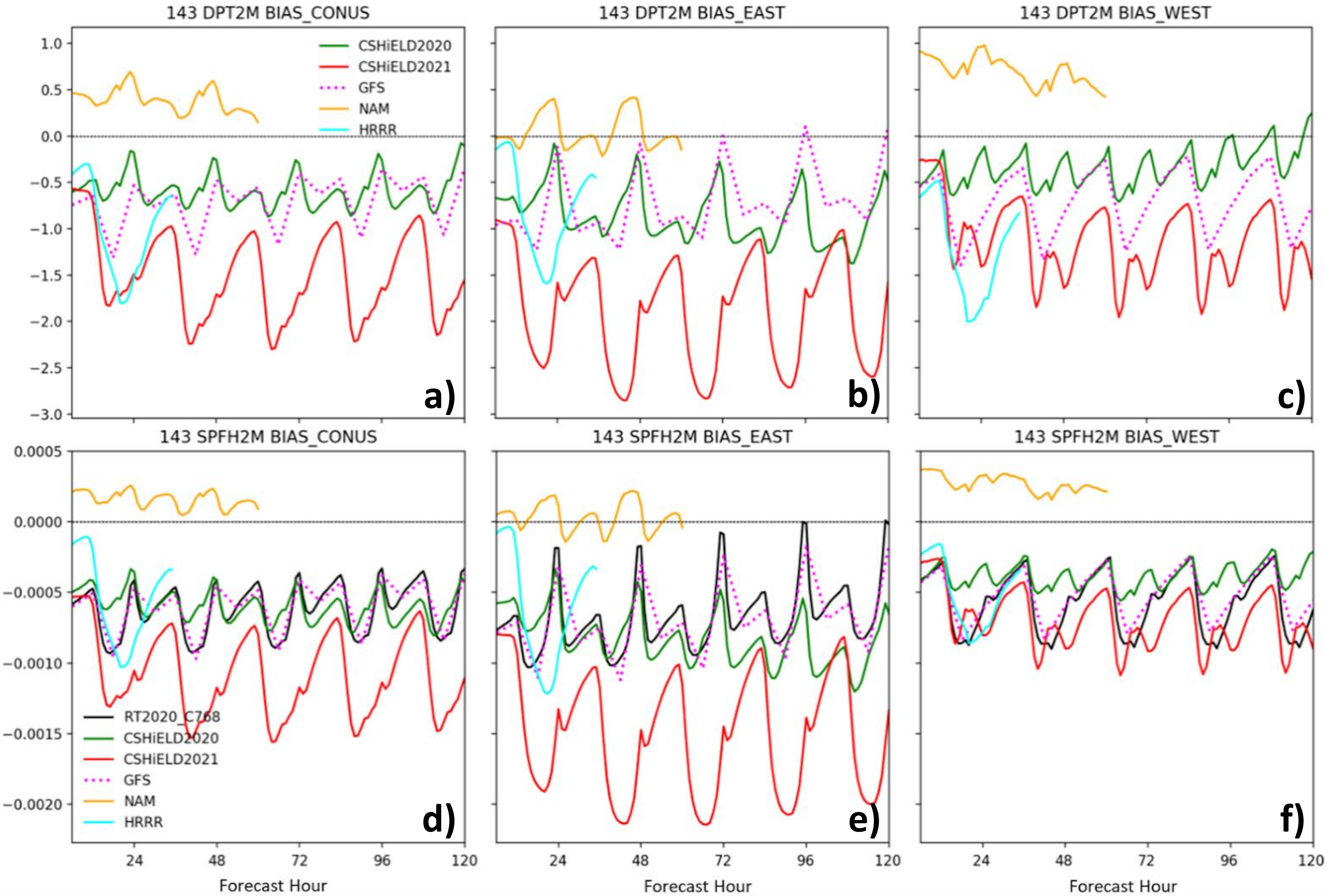


Figure 8: April-August averaged 2-meter dewpoint temperature BIAS (K; a-c) and 2-meter specific humidity BIAS (g/kg; d-f). RT2020_C768 refers to SHiELD. Column 1 is the Contiguous U.S. (CONUS), column 2 is the Eastern U.S. (EAST), and column 3 is the Western U.S. (WEST).

C-SHiELD 2021 has a significant dry bias, especially in the eastern U.S. The bias is greatest in the afternoon (inverted from C-SHiELD 2020). This is significantly drier in the eastern US than C-SHiELD 2020, which indicates that it produced significantly less moisture. This continues an ongoing problem in C-SHiELD (Harris et al. 2019). In the western US dry biases in C-SHiELD 2021 are less severe and more similar to the global models (GFS, SHiELD) but still degraded compared to the very good western U.S. moisture biases in C-SHiELD 2020. The NAM shows the smallest moisture bias (closest to zero) overall while the HRRR has a dry bias similar to C-SHiELD 2020 and the global models.

In order to investigate the biases in surface temperature and moisture, we examined the surface sensible and latent heating fluxes with the HRRR version 4 analysis used as a baseline.

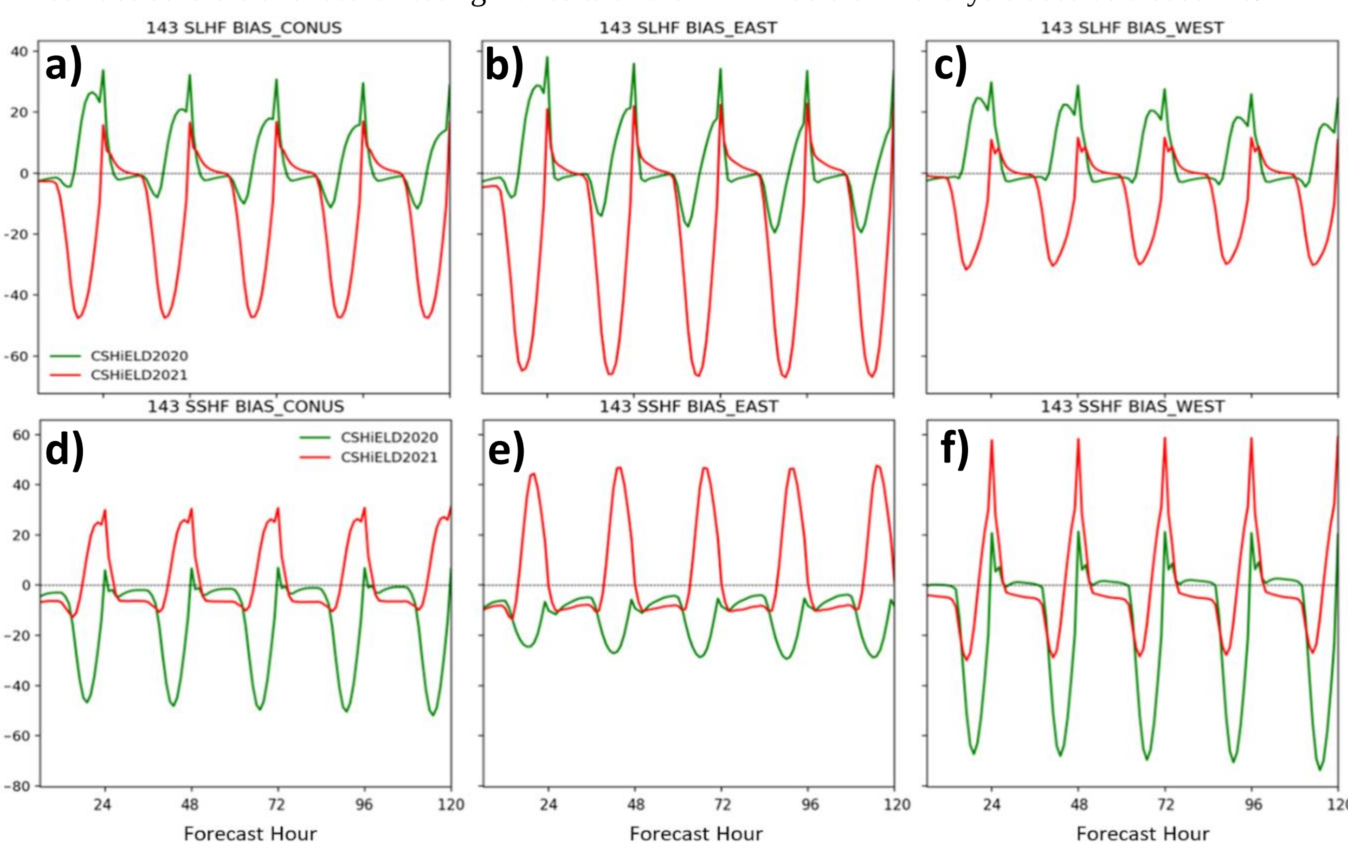


Figure 9: April-August averaged latent heating flux BIAS (W/m^2; a-c) and sensible heating flux BIAS (W/m^2; d-f) with the HRRR_v4 used as the reference. Column 1 is the Contiguous U.S. (CONUS), column 2 is the Eastern U.S. (EAST), and column 3 is the Western U.S. (WEST).

Figure 9 shows that C-SHiELD 2021 has a higher sensible heat flux and lower latent heat/moisture flux than C-SHiELD 2020, especially over the eastern U.S. This indicates that the changes between the two versions, perhaps by the change in the land surface model, altered the partitioning between the sensible and latent heat fluxes. The amplitude of the difference with HRRR is also larger in C-SHIELD 2021, especially in the eastern U.S. where we had noticed a severe dry bias. These differences in models' representation of the surface fluxes could explain why C-SHiELD 2021 is warmer and drier at the surface, especially in the eastern U.S., and this will certainly be addressed in future versions of C-SHiELD.

c. Convective Precipitation and Storm Analysis

Following the surface analyses, we validated the models' precipitation forecasts between April and August. Figure 10 shows the frequency biases and skill scores over the CONUS for 6-hour quantitative precipitation forecasts (QPF) compared to observed precipitation from the Stage IV multisensory estimates. As described in section 3, the precipitation skill was evaluated using two different metrics: the Fractional Skill Score (FSS; Roberts & Lean, 2008) and the Gilbert Skill Score (GSS; Hogan & Mason, 2012). The frequency bias (labeled as Fcst/Obs) is the ratio computed by dividing the number of forecast events by the number of observation events within a given precipitation threshold.

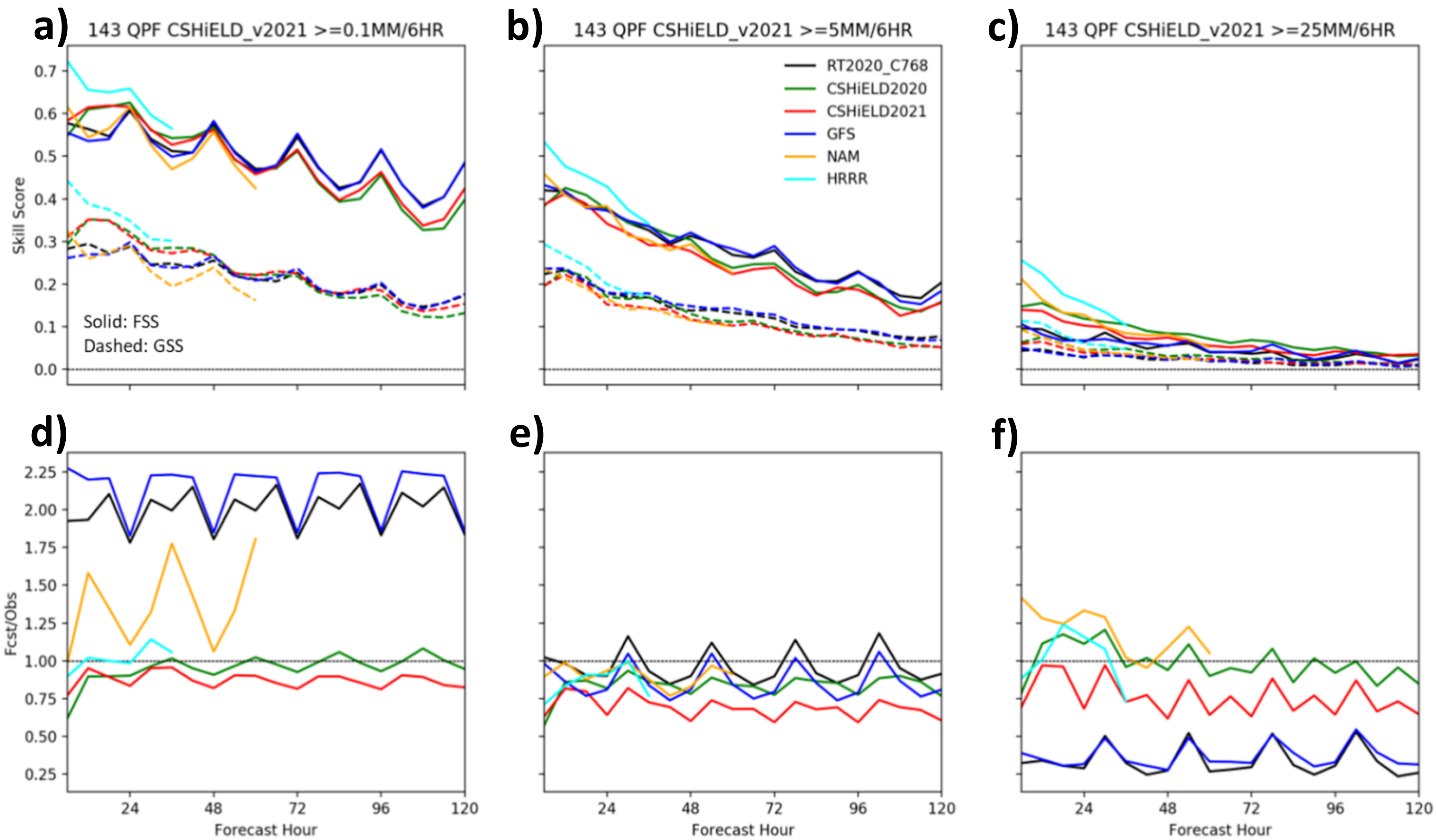


Figure 10: April-August averaged QPF FSS (solid line) and GSS (dotted line) skill scores (a-c) and frequency bias (d-f) over the CONUS.

Both versions of C-SHiELD along with the HRRR have biases that are close to zero at lighter precipitation rates (0.1mm/6hr) and heavier precipitation rates (>=25mm/6hr). However, C-SHiELD 2021 has a slight negative bias when compared to C-SHiELD 2020. The 3km NAM (along with the 13-km models) shows a positive bias at lighter precipitation rates, but it has biases close to those of C-SHiELD 2020 and 2021 at moderate and heavy precipitation rates. SHiELD and the GFS have biases that are close to zero at moderate precipitation rates, but they have strong positive and negative biases at lighter and heavier rates, respectively. This is expected as the global models have a tendency to produce excess drizzle as well as difficulty representing the intense storms that produce the heaviest rainfall rates.

In terms of skill, both versions of C-SHiELD have skill scores that are comparable to those of the NAM. However, the HRRR has a slightly higher skill than C-SHiELD, possibly due to the radar data assimilation that improves the precipitation forecast in the first few hours of the forecast. By the end of the HRRR's 36-hour forecast time, its skill is comparable to the other models, especially for heavier precipitation rates. The 13km SHiELD and GFS also have slightly higher skill than C-SHiELD for light and moderate precipitation rates, but they are noticeably worse with heavier precipitation rates. In addition, C-SHiELD 2021 and C-SHiELD 2020 have comparable skill scores across all thresholds.

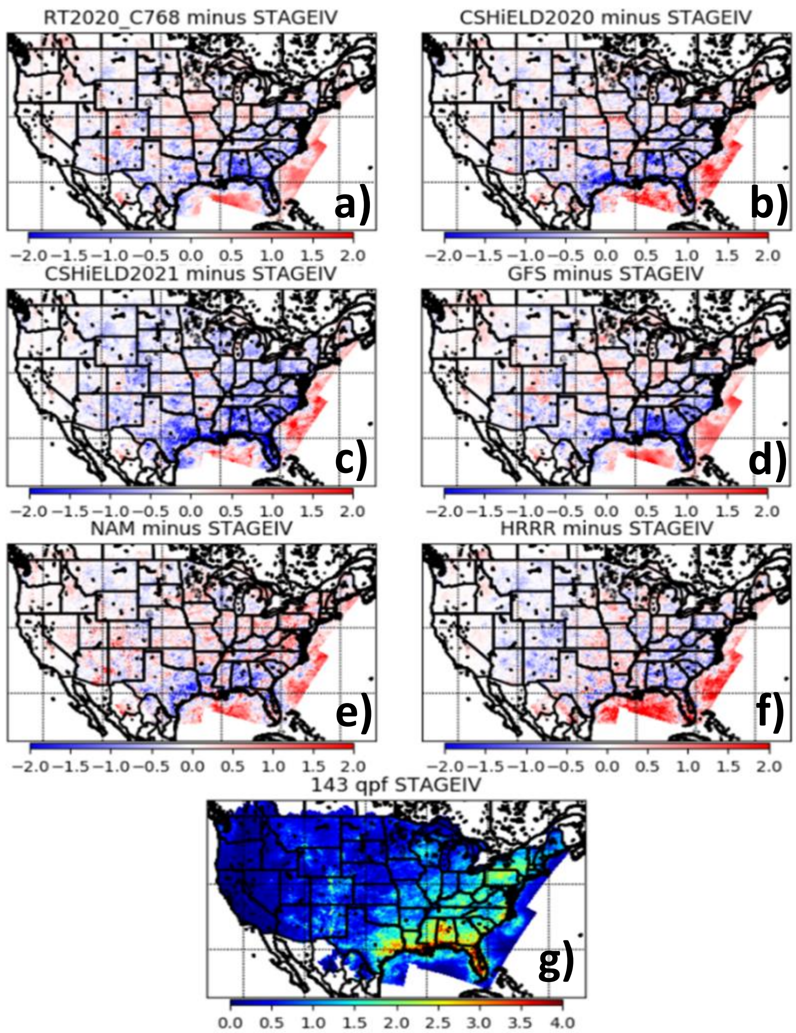


Figure 11: Maps of the April-August averaged QPF BIAS (mm/day) at 24 hours after the initialization time for SHiELD (a), C-SHiELD 2020 (b), C-SHiELD 2021 (c), GFS (d), NAM (e), and HRRR (f). The observed precipitation from STAGEIV (mm/day) is shown in (g).

Figure 11 shows maps of the April-August averaged precipitation biases relative to the STAGEIV observations at 24 hours after the initialization times. From this, it is evident that C-SHiELD 2021 has the strongest negative precipitation bias out of all models over the Southeastern U.S during this time period. Interestingly, this negative bias seems to largely disappear when we examine only the April-May biases (Figure 12), which suggests that there is some seasonal variability in the precipitation biases.

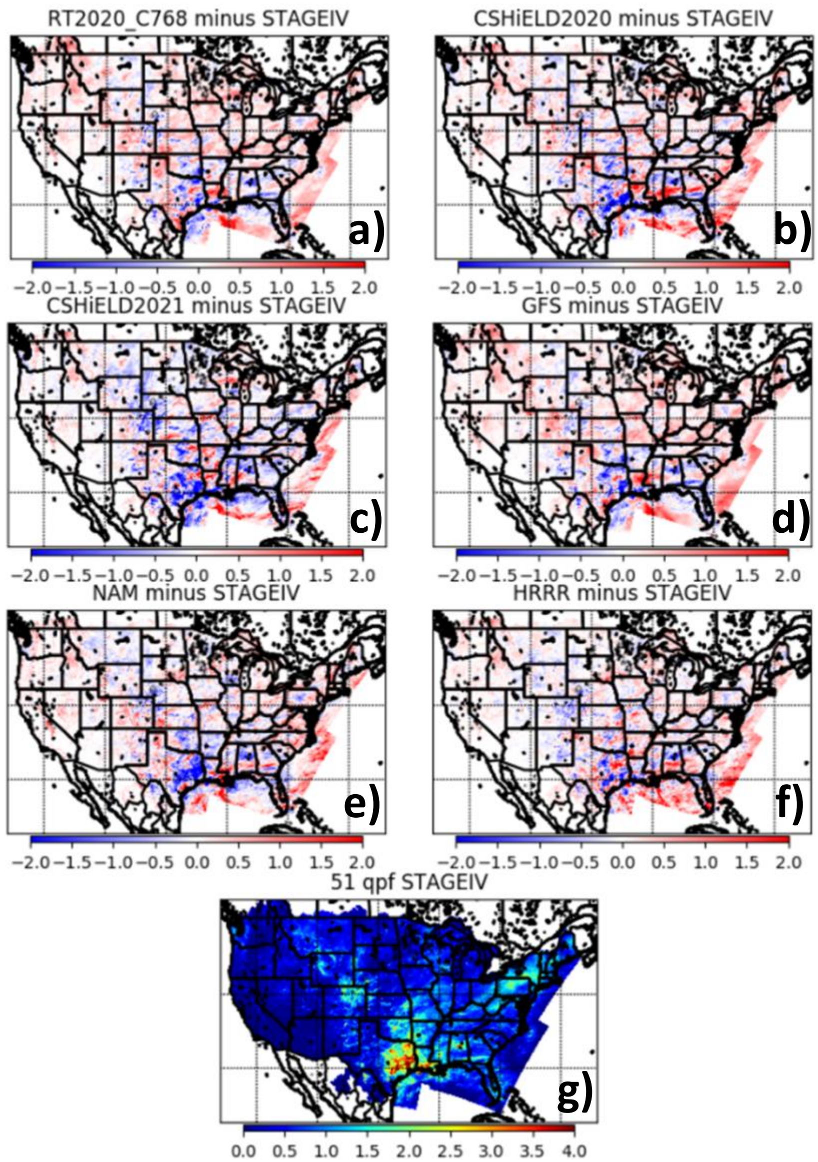


Figure 12: Maps of the April-May averaged QPF BIAS (mm/day) at 24 hours after the initialization time for SHiELD (a), C-SHiELD 2020 (b), C-SHiELD 2021 (c), GFS (d), NAM (e), and HRRR (f). The observed precipitation from STAGEIV (mm/day) is shown in (g).

The negative bias of C-SHiELD 2021 during the summer months (June-August) may be attributed to the model's poor prediction of the heavy precipitation produced by tropical cyclones, but this warrants further investigation. Overall, both versions of C-SHiELD show respectable skill in forecasting precipitation when compared to the other regional models.

In order to investigate C-SHiELD's severe weather forecasting capabilities, we validated the 2-5 km updraft helicity (UH) fields against storm reports from NOAA's Storm Prediction Center (SPC) from April through August (Figure 13) using the surrogate severe technique described in greater detail in Sobash et al. (2011). The UH fields are computed by the model, and they are useful in identifying extreme updraft rotation and supercellular storms. Thus, the UH is often used to predict the probability of severe storms in model simulations.

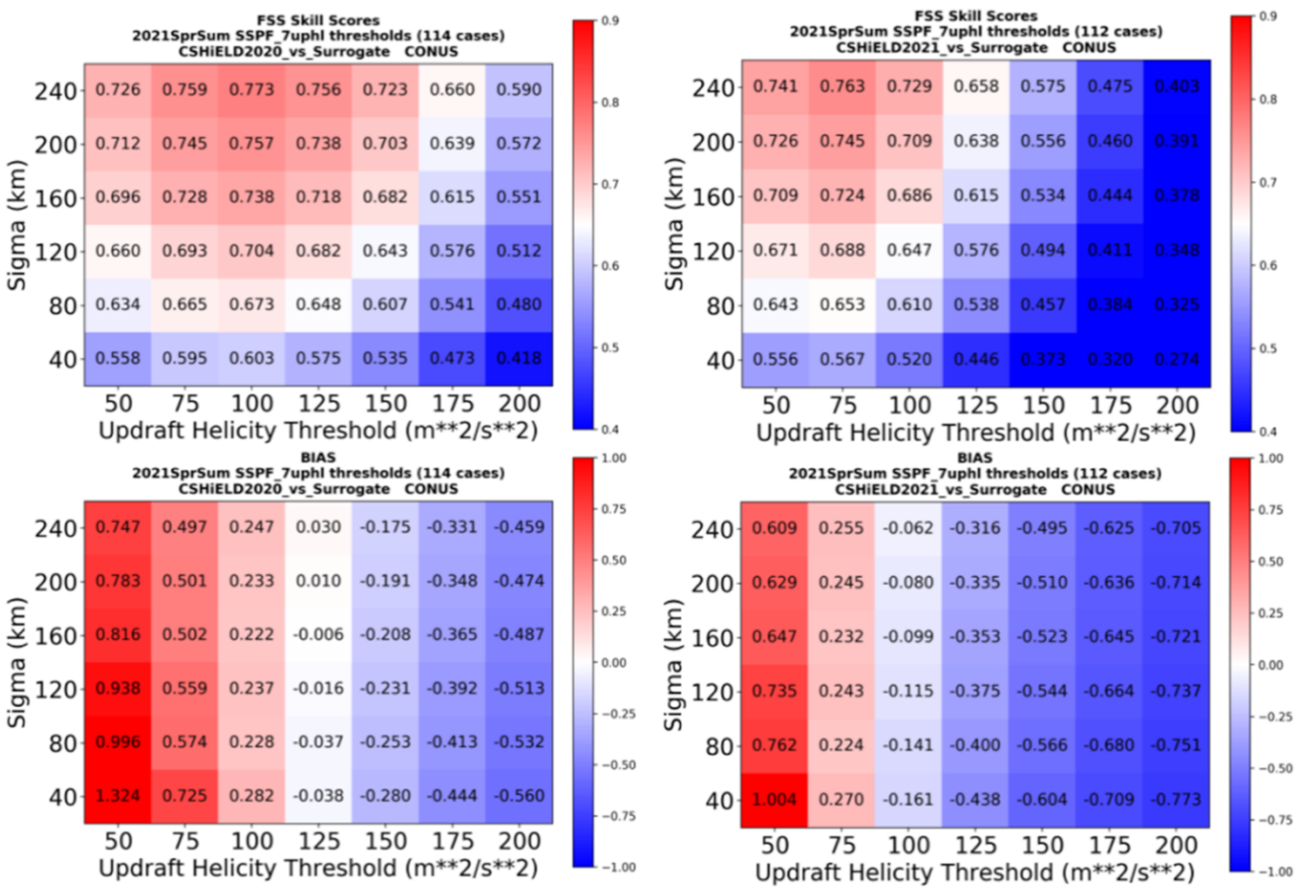


Figure 13: April-August surrogate severe forecast skill scores (FSS; top) and BIAS scores (bottom). C-SHiELD 2020 is on the left, and C-SHiELD 2021 is on the right. All forecasts were initialized at 00Z. Higher numbers are better for FSS, and closer to 0 is better for BIAS score.

We computed the FSS skill and Bias scores as functions of the UH thresholds and the smoothing radii (sigma) for the first 12-36 hours of the model runs, as shown above in Figure 13. Both models have the highest FSS at UH thresholds between 50 and 100 m^2/s^2 at the highest smoothing radius of 240 km. C-SHiELD 2021 has slightly improved skill and smaller positive biases (by about 0.14 m^2/s^2) than C-SHiELD 2020 at these same UH thresholds, which indicates that C-SHiELD 2021 does not over-forecast strong storms at lower UH thresholds as much as C-SHiELD 2020. Conversely, C-SHiELD 2021 shows lower skill and a stronger negative bias at higher UH thresholds (at and above 150 m^2/s^2), which indicates that it may be under forecasting stronger storms at higher UH thresholds. Overall C-SHiELD 2021 shows a shift towards lower UH values. As we will see, the surrogate severe forecast skill and bias can change significantly when we examine individual severe weather events instead of analyzing over longer time periods.

i. Case 1: May 2-5, 2021 Tornado Outbreak

We have selected a couple of severe weather outbreaks during the spring and summer of 2021 in order to examine how C-SHiELD performed with individual events. The first outbreak occurred during a three-day period from May 2-5, 2021 and was driven by a surface low-pressure system out ahead of a large 500-mb trough over the Central Plains. Southerly flow on the eastern side of the low-pressure system advected warm, moist air into the Southeastern US from the Gulf of Mexico while the upper-level winds provided moderately strong vertical wind shear over the region. As the surface low moved toward the Northeast, convective activity occurred from Florida up to Virginia, resulting in heavy precipitation in many areas. In addition, 128 tornadoes were reported during this three-day period (SPC, 2022). Figure 14 shows the accumulated precipitation biases relative to STAGEIV for all models from 00Z on May 3 to 00Z on May 5. All six models were initialized at 00Z on May 3.

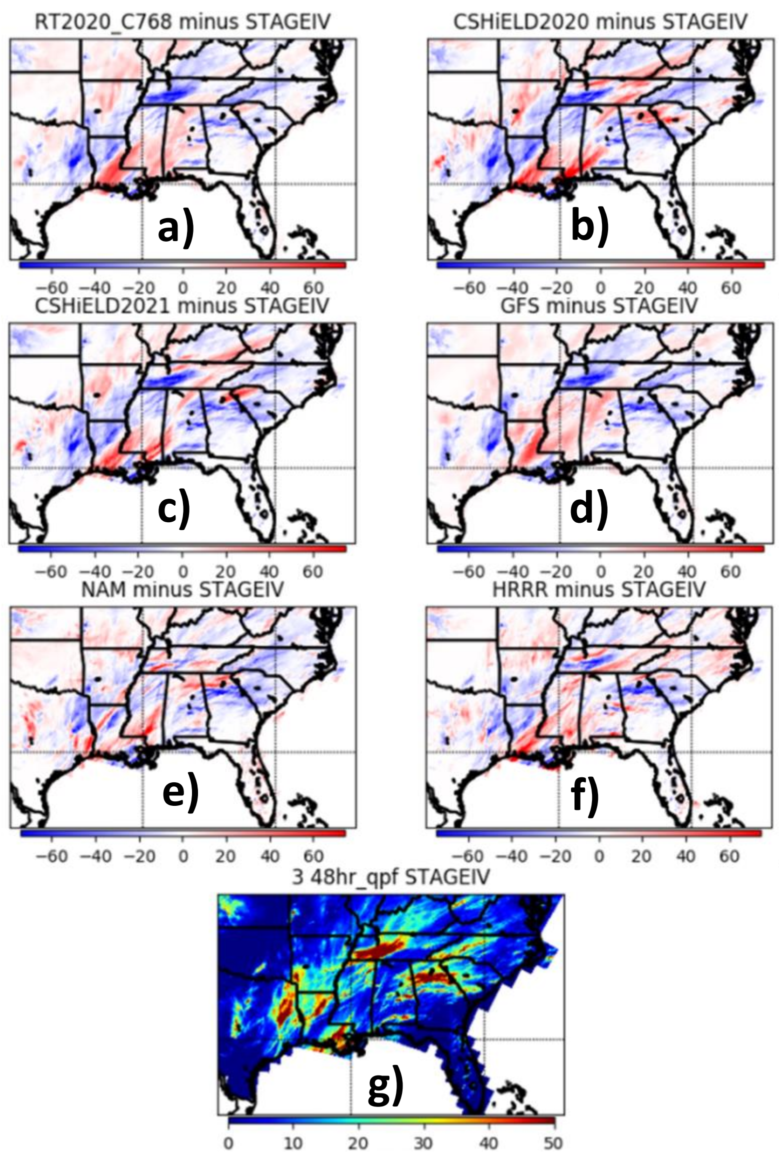


Figure 14: Maps of the 48-hour accumulated precipitation bias (mm/day) from 00Z on May 3 to 00Z on May 5 for SHiELD (a), C-SHiELD 2020 (b), C-SHiELD 2021 (c), GFS (d), NAM (e), and HRRR (f). The observed precipitation from STAGEIV (mm/day) is shown in (g). The forecasts were initialized at 00Z on May 3.

The 3km NAM and HRRR show the smallest precipitation biases (lightest colors) overall, but both C-SHiELD versions show similar bias patterns to the NAM and especially the HRRR. The 13-km SHiELD and GFS performed similarly to C-SHiELD with all four models showing the largest biases over western Tennessee and central Louisiana. However, C-SHiELD

2021 shows some improvement from C-SHiELD 2020 over Kentucky, Eastern Tennessee, and North Carolina. Indeed, all models show similar bias patterns: a difficulty predicting the heavy precipitation from Texas into Tennessee, and in northern Georgia, and too much precipitation in between these two bands, possibly pointing to challenges in predicting the precise evolution of this storm system. The skill scores and biases of the 6-hour QPF during the event (relative to STAGEIV) shown in Figure 15 reveal that C-SHiELD 2021 performed roughly on par with C-SHiELD 2020 and the HRRR in terms of bias, but the HRRR had higher skill during this event. The 3-km NAM along with the 13-km models had a positive bias at the lighter precipitation rates.

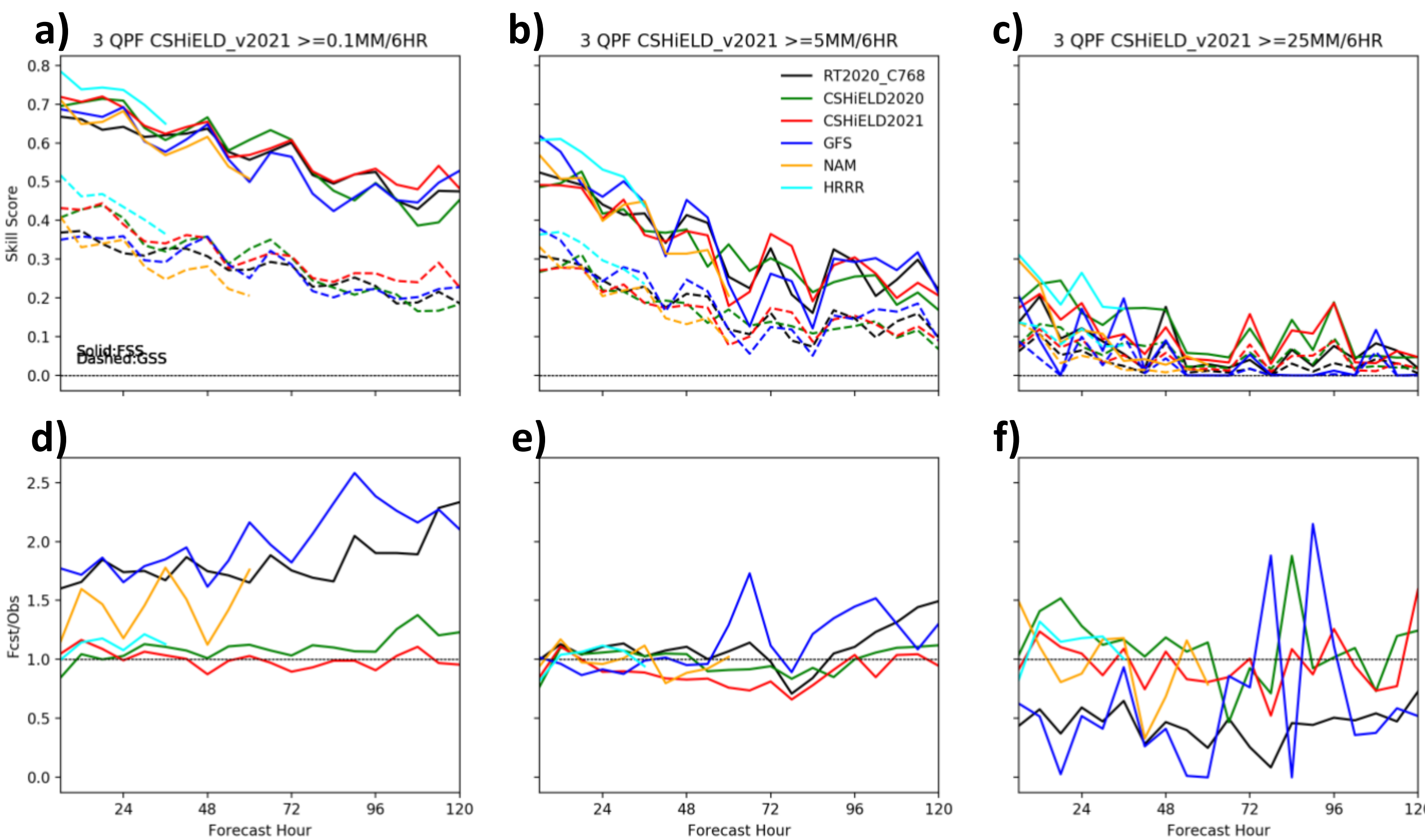


Figure 15: May 2-5, 2021 averaged QPF FSS (solid line) and GSS (dotted line) skill scores (a-c) and frequency bias (d-f) over the CONUS.

As the system's cold front approached the region on the evening of May 4, 2021, the convection became more organized into a squall line. Figure 16 shows the composite reflectivity imagery from 00Z on May 5, 2021 as the squall line became organized over the Southeastern U.S.

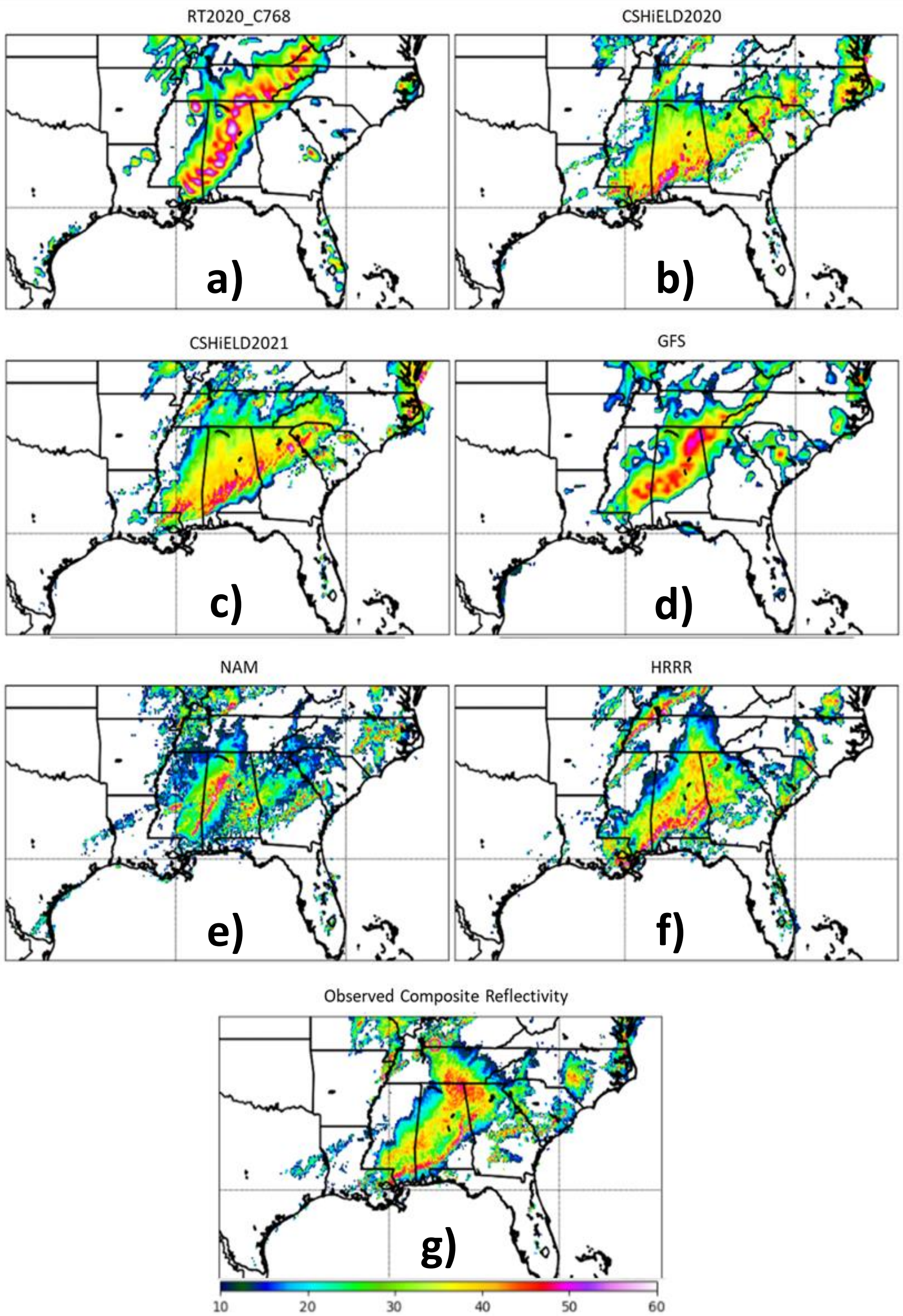


Figure 16: Maps of the simulated composite reflectivity (dBZ) at 00Z on May 5 for SHiELD (a), C-SHiELD 2020 (b), C-SHiELD 2021 (c), GFS (d), NAM (e), and HRRR (f). The observed radar (dBZ) is shown in (g). The forecasts were initialized at 00Z on May 4.

C-SHiELD 2021 did a decent job of simulating the line of heaviest precipitation, but it overestimated the light and moderate reflectivity values around the main squall line (as did C-SHiELD 2020). The 3km NAM did not seem to properly simulate the main squall line while the HRRR did the best job overall of predicting this particular line of storms; however, the placement of the most intense band of storms was not as good as in either version of C-SHiELD. It is also clear from Figure 16 that the 3-km model shows far more structure in the squall line, and much better storm placement, than do the 13-km global models with parameterized convection. The surrogate severe validations for these three days show that C-SHiELD 2020 and 2021 predicted the severe weather probability fields almost perfectly 12-36 hours in advance (Figure 17 and Figure 18).

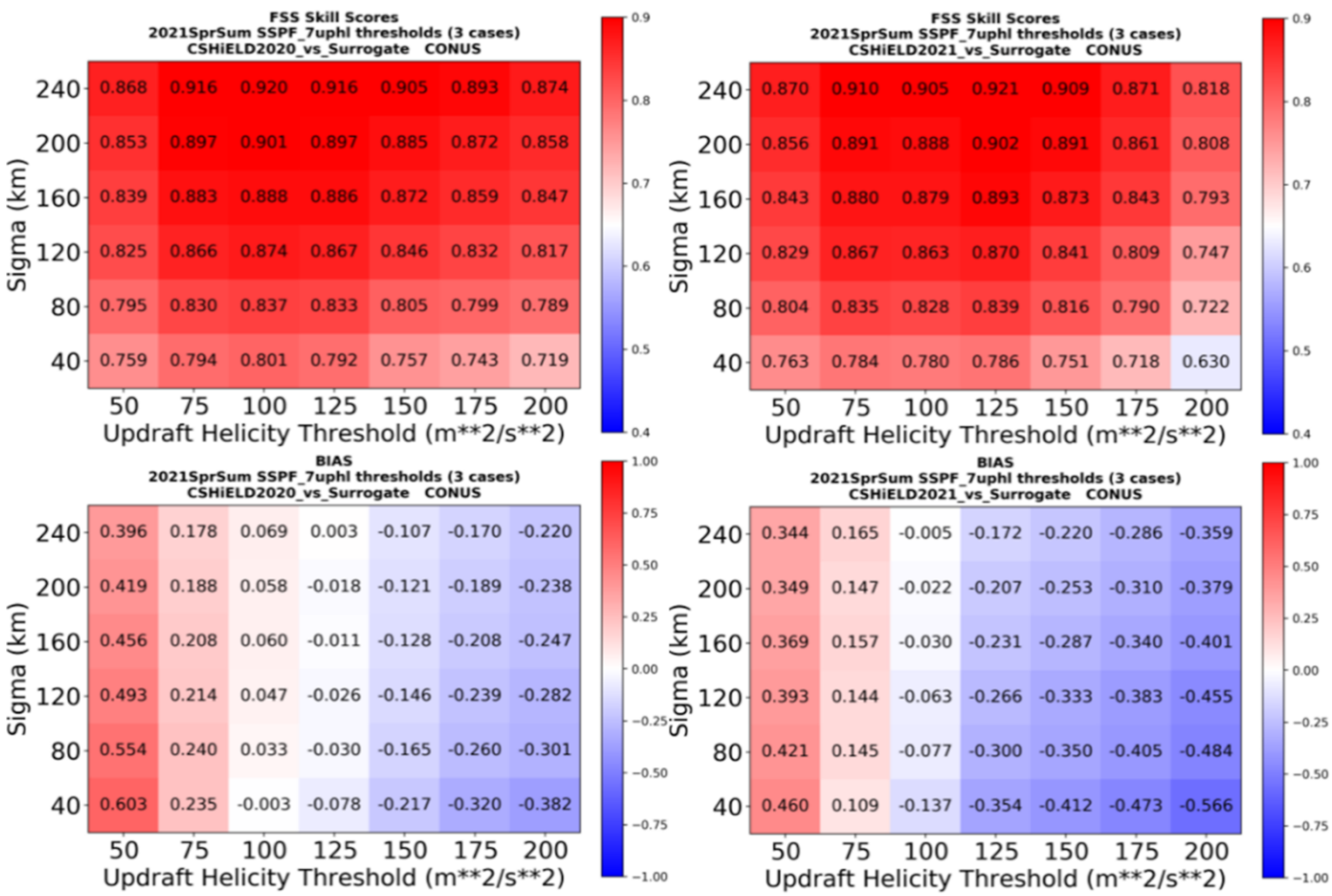


Figure 17: May 2-5, 2021 surrogate severe forecast skill scores (FSS; top) and BIAS scores (bottom). C-SHiELD 2020 is on the left, and C-SHiELD 2021 is on the right. All forecasts were initialized at 00Z. Higher numbers are better for FSS, and closer to 0 is better for BIAS score.

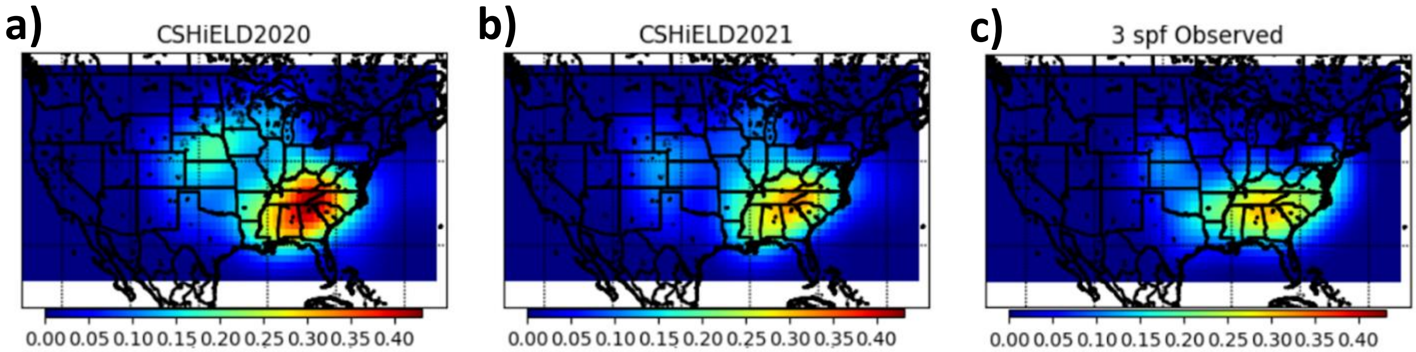


Figure 18: Maps of the May 2-5, 2021 surrogate severe probability fields (SSPF) at an updraft helicity threshold of 75 m**2/s**2 and smoothing radius of 240 km for C-SHiELD 2020 (a), C-SHiELD 2021 (b), and SPC observations (c).

Notably C-SHiELD 2021 has a very small bias at this threshold. The FSS is greater than 0.8 for all UH thresholds in both models, and the bias is slightly lower for C-SHiELD 2021. For UH thresholds between 75 and 150 m^2/s^2, both versions of C-SHiELD have similar FSS of greater than 0.9, which indicates nearly perfect forecasts of severe storms within these thresholds. Again, it's worth noting that these high scores are present at the smoothing radius of 240 km, but the FSS values are still close to 0.9 for lower smoothing radii.

ii. Case 2: July 28-30, 2021 Severe Outbreak

The second severe weather outbreak occurred in late July 2021 as a low-level wave formed just downstream of an upper-level trough in the Mid-Atlantic region, which already had a warm, unstable air mass in place. The outbreak began over the Ohio Valley, but the convection eventually re-developed over Pennsylvania and New Jersey as the shortwave moved to the East. Precipitation for this event was generally lighter than the May event, with the heaviest precipitation concentrated in the regions that experienced the strongest convective bands. The Mid-Atlantic region reported more than 40 tornadoes and multiple instances of damaging winds during this event (SPC, 2022). Figure 19 shows the accumulated precipitation bias relative to STAGEIV from 00Z on July 28 to 00Z on July 30.

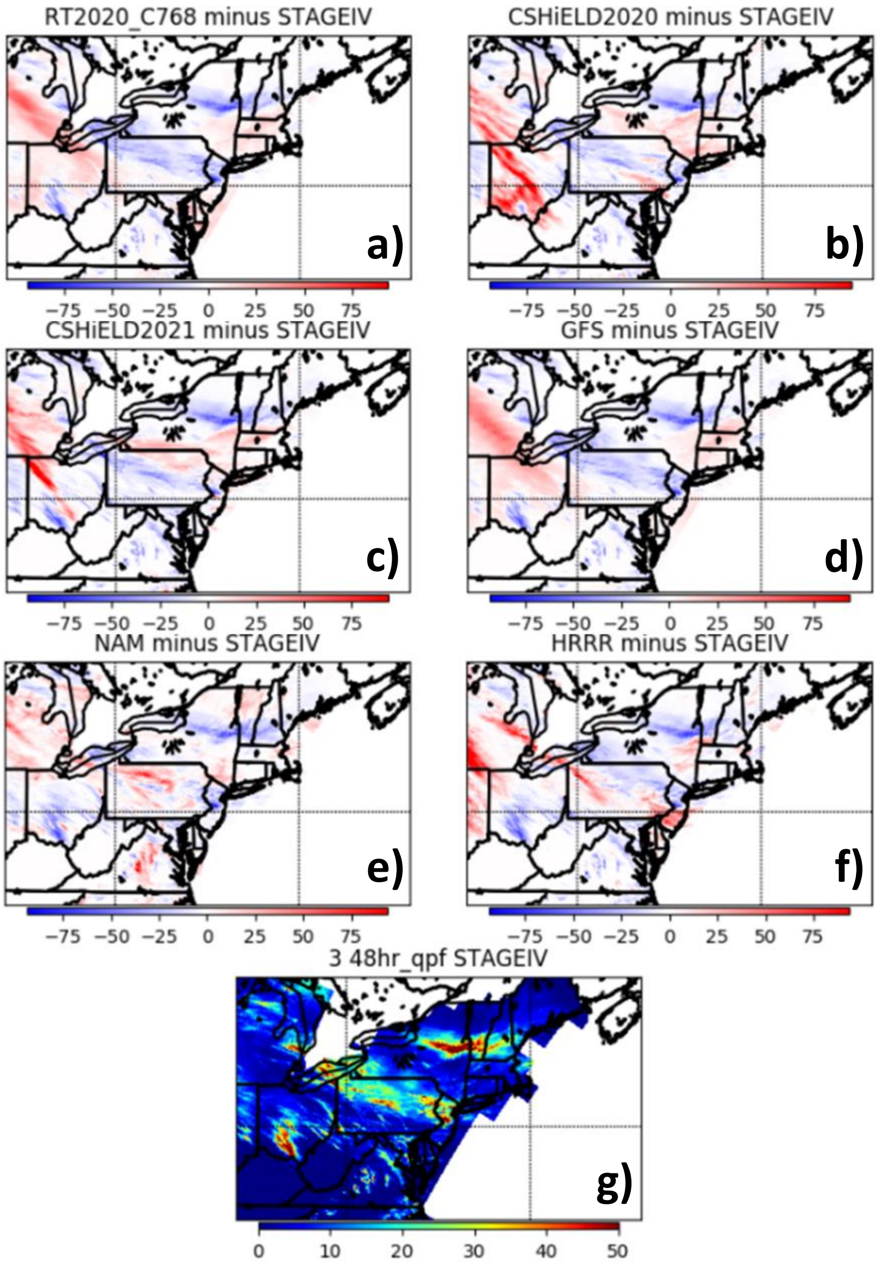


Figure 19: Maps of the 48-hour accumulated precipitation bias (mm/day) from 00Z on July 28 to 00Z on July 30 for SHiELD (a), C-SHiELD 2020 (b), C-SHiELD 2021 (c), GFS (d), NAM (e), and HRRR (f). The observed precipitation from STAGEIV (mm/day) is shown in (g). The forecasts were initialized at 00Z on July 28.

C-SHiELD 2021 has smaller biases than C-SHiELD 2020 over Ohio, eastern Pennsylvania, and New Jersey. The 3km NAM and HRRR overpredicted the precipitation in Pennsylvania and southern New Jersey, and C-SHiELD 2020 generally has better skill scores and bias than C-SHiELD 2021 during this event (Figure 20).

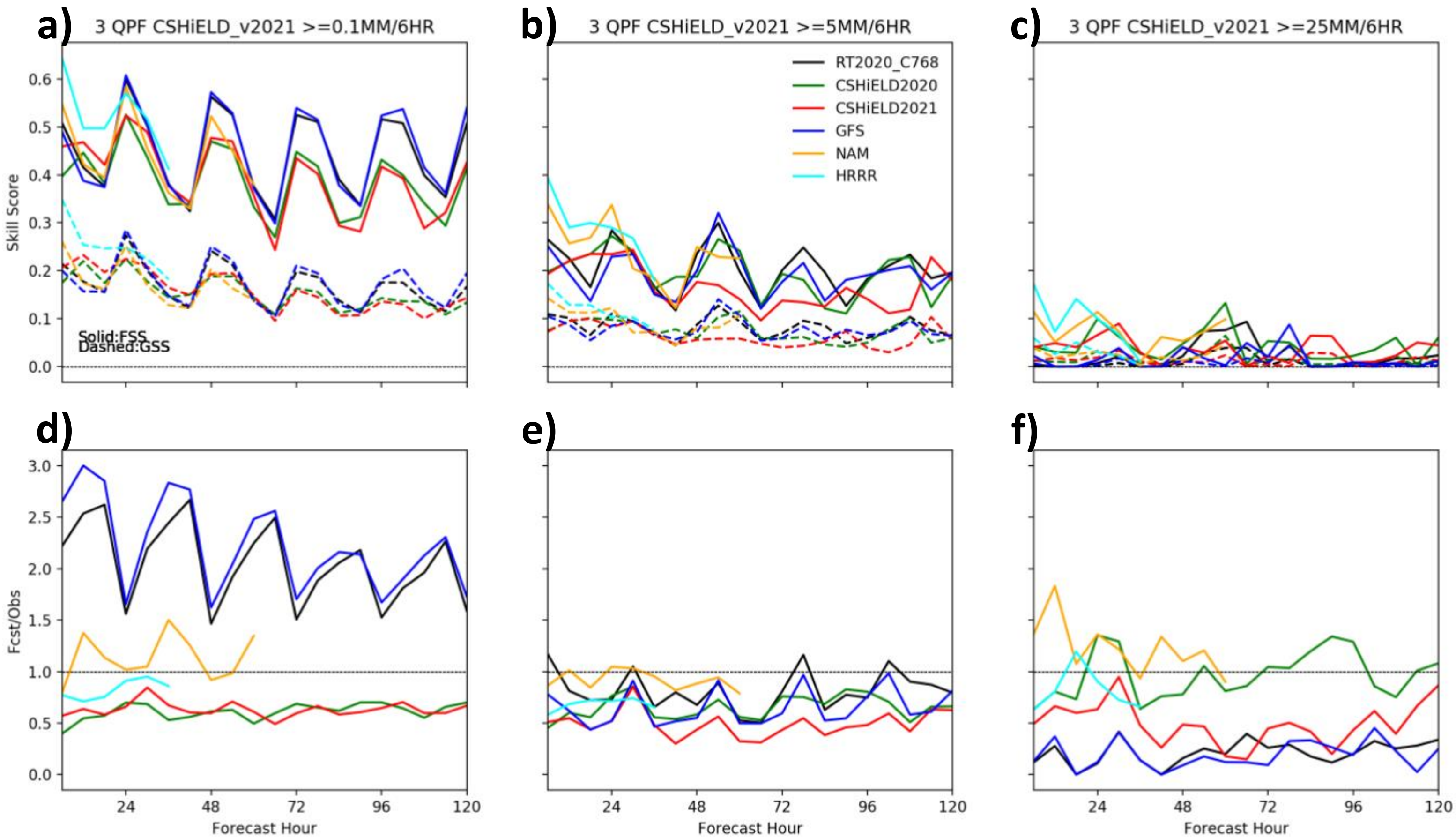


Figure 20: July 28-30, 2021 averaged QPF FSS (solid line) and GSS (dotted line) skill scores (a-c) and frequency bias (d-f) over the CONUS.

C-SHiELD 2021 also did a decent job of simulating the main areas of convection moving across Pennsylvania and New York on the evening of July 29 (Figure 21). Again, all models under-predicted the precipitation in the heaviest precipitation areas, especially from central New York state into New Hampshire and in Ohio.

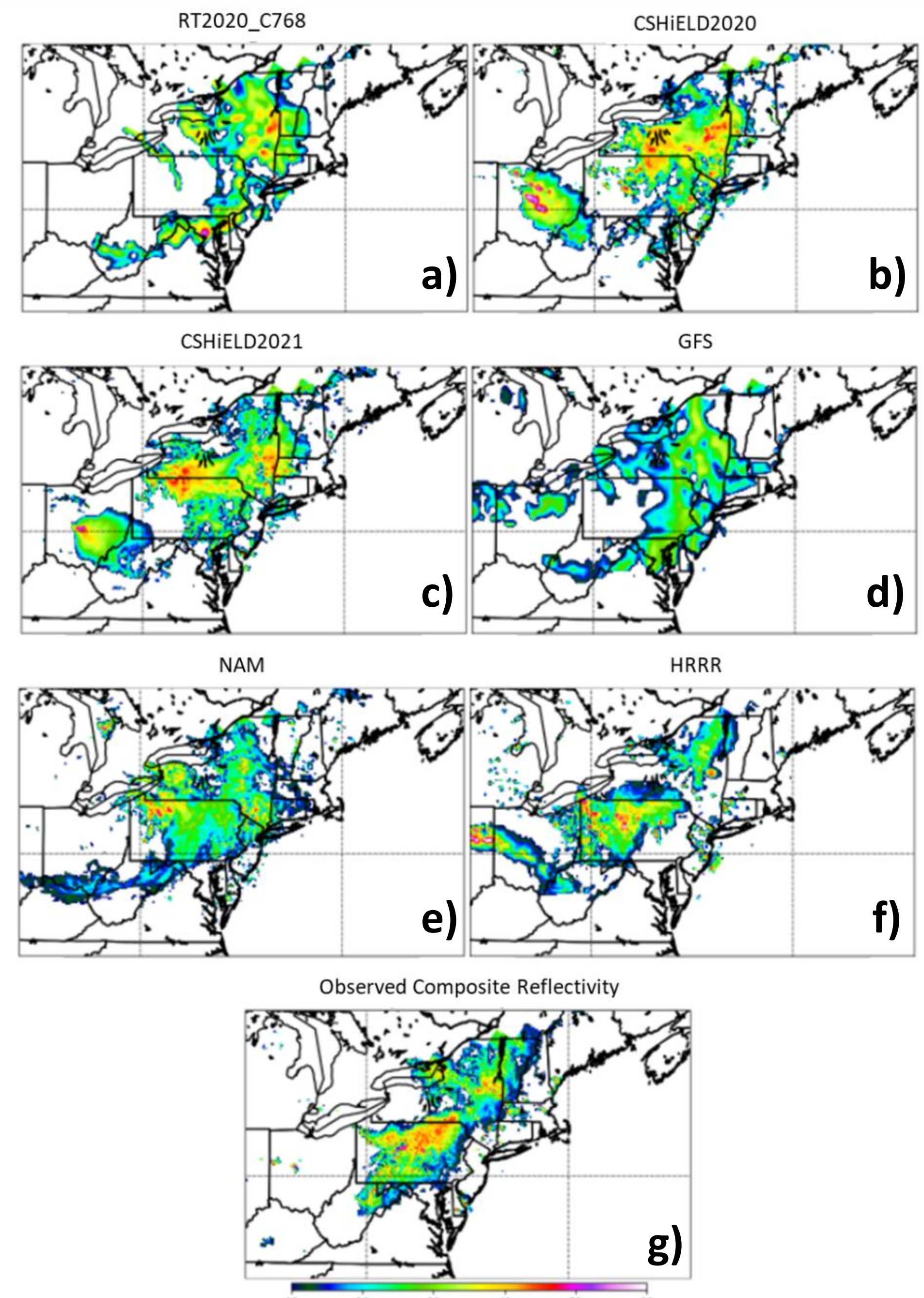


Figure 21: Maps of the simulated composite reflectivity (dBZ) at 00Z on July 30 for SHiELD (a), C-SHiELD 2020 (b), C-SHiELD 2021 (c), GFS (d), NAM (e), and HRRR (f). The observed radar (dBZ) is shown in (g). The forecasts were initialized at 00Z on July 29.

When we validated the surrogate severe forecasts for this event against the SPC observations, we found FSS values at or above 0.75 with both versions of C-SHiELD for UH thresholds between 75 and 150 m^2/s^2 (Figure 22). Though these skill scores are respectable, they indicate that both models did a better job of forecasting the severe probability fields during the May outbreak. However, the bias is substantially reduced in C-SHiELD 2021 (by as much as 0.25 m^2/s^2), which indicates that C-SHiELD 2020 over-predicted the storm activity at lower UH thresholds. This finding is illustrated in Figure 23.

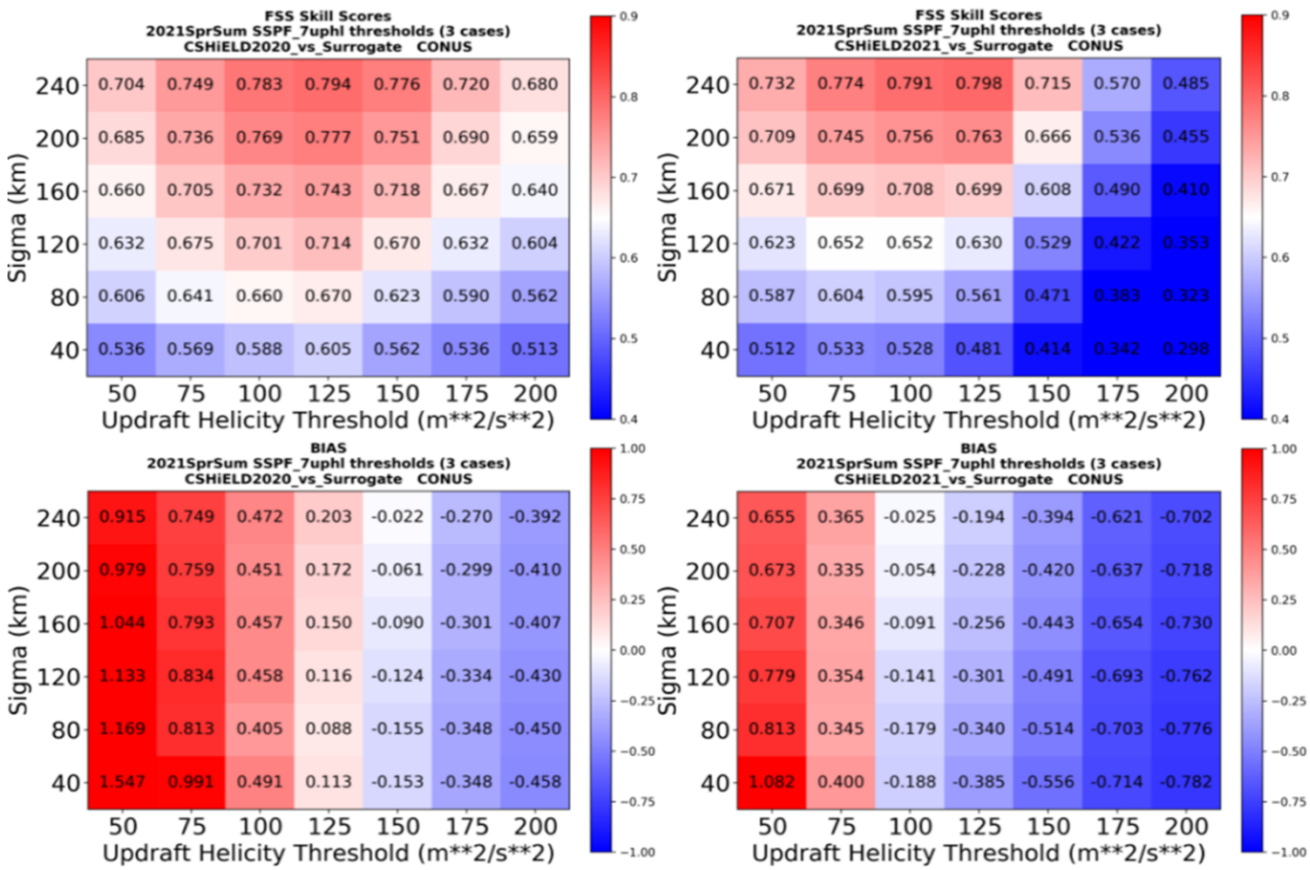


Figure 22: July 28-30, 2021 surrogate severe forecast skill scores (FSS; top) and BIAS scores (bottom). C-SHiELD 2020 is on the left, and C-SHiELD 2021 is on the right. All forecasts were initialized at 00Z. Higher numbers are better for FSS, and closer to 0 is better for BIAS score.

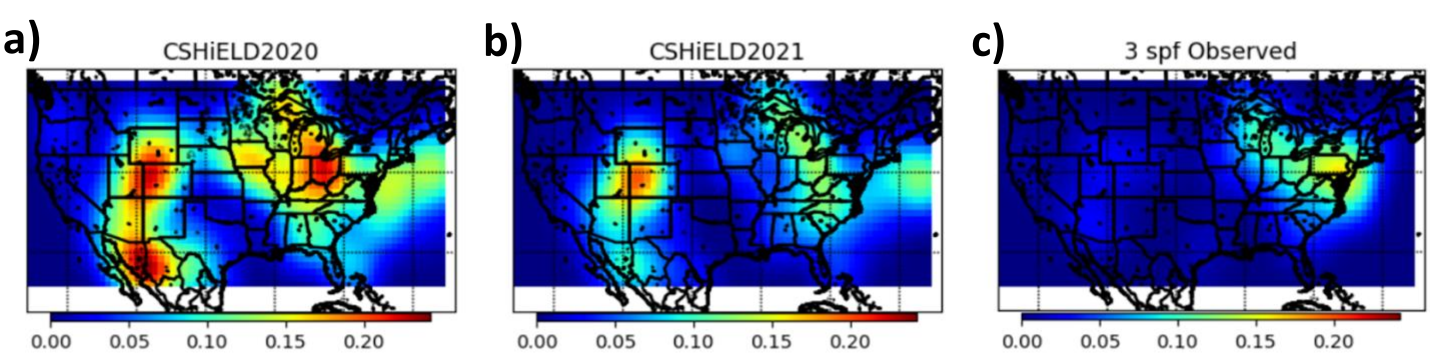


Figure 23: Maps of the July 28-30, 2021 surrogate severe probability fields (SSPF) at an updraft helicity threshold of 75 m**2/s**2 and smoothing radius of 240 km for C-SHiELD 2020 (a), C-SHiELD 2021 (b), and SPC observations (c).

5. Conclusion

The FV3-based GFDL SHiELD is a modeling system with a wide range of forecasting capabilities and applications, including the C-SHiELD configuration for the explicit prediction of convective storms. We analyzed the 3-km nests of C-SHiELD 2021 and C-SHiELD 2020 in forecasts for the spring and summer of 2021 in order to evaluate their forecast skill and observe any year-over-year improvements. During this time period, C-SHiELD 2021 showed significant improvements in large-scale 500-mb prediction over C-SHiELD 2020 and performed similarly to the 13-km SHiELD, which uses a global domain with an identical resolution and configuration. C-SHiELD 2021 also showed improvements in the 2-meter temperature bias and RMSE, but it was warmer and drier at the surface than C-SHiELD 2020. This could be attributed to its over-representation of the surface sensible heating flux and under-representation of the moisture/latent heat flux compared to C-SHiELD 2020. This may have been due to the change in land surface model from the NOAH LSM to Noah-MP, or also due to re-configuration of the TKE-EDMF PBL scheme. For precipitation forecasts, C-SHiELD 2021 performs similarly to other mesoscale models (C-SHiELD 2020, 3-km NAM, and HRRR) in terms of forecast skill, but it had a noticeable dry bias during the summer months. Conversely, C-SHiELD 2021 shows a smaller precipitation bias than the other mesoscale models during the spring, which suggests that there may be strong seasonal variability in the model's performance with the quantitative precipitation forecasts. Both versions of C-SHiELD show respectable skill with the 12-36 hour surrogate severe weather prediction at multiple UH thresholds during the spring and summer months, including extremely accurate forecasts of the severe weather probability fields during the May 2-5 tornado outbreak. Overall, C-SHiELD 2021 showed encouraging results and interesting improvements over its predecessor during this analysis. All surface flux and precipitation biases will be examined and addressed in future versions of C-SHiELD.

Acknowledgments

We would like to thank everyone in the FV3 team here at GFDL for continuing the important work on the development of SHiELD. Lucas Harris, Kai-Yuan Cheng, and Linjiong Zhou all provided valuable feedback and suggestions for this project, and Matt Morin provided all of the real-time model data used in the analyses. The Python scripts and data files used for the surrogate severe storm validations were provided by Bill Stern.

References

ECMWF (2021). ERA5: Data Documentation. Available at https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation. Last accessed 2 February 2022.

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

GFDL (2021). SHiELD: System for High-resolution prediction on Earth-to-Local Domains. Available at https://www.gfdl.noaa.gov/shield/#C-SHiELD. Last accessed 2 February 2022.

GSL (2020). The High-Resolution Rapid Refresh (HRRR). Available at https://rapidrefresh.noaa.gov/hrrr/. Last accessed 2 February 2022.

Han, J., & Bretherton, C. S. (2019). TKE-based moist eddy-diffusivity mass-flux (EDMF) parameterization for vertical turbulent mixing. Weather and Forecasting, *34*(4), 869-886.

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.

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

Harris, L. M., & Lin, S. J. (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., Rees, S. L., Morin, M., Zhou, L., & Stern, W. F. (2019). Explicit prediction of continental convection in a skillful variable-resolution global model. Journal of Advances in Modeling Earth Systems, 11, 1847– 1869. https://doi.org/10.1029/2018MS001542

Harris, L., Zhou, L., Lin, S., Chen, J., Chen, X., Gao, K., Morin, M., Rees, S., Sun, Y., Tong, M., Xiang, B., Bender, M., Benson, R., Cheng, K., Clark, S., Elbert, O.D., Hazelton, A., Huff, J.J., Kaltenbaugh, A., Liang, Z., Marchok, T., Shin, H.H., Stern, W. (2020). GFDL SHiELD: A Unified System for Weather-to-Seasonal Prediction. Journal of Advances in Modeling Earth Systems 12.. doi:10.1029/2020ms002223

Harris, L., Chen, X., Putman, W., Zhou, L., Chen, J. (2021). A Scientific Description of the GFDL Finite-Volume Cubed-Sphere Dynamical Core. NOAA technical memorandum OAR GFDL ; 2021-001. https://doi.org/10.25923/6nhs-5897

Hogan, R. J., & Mason, I. J. (2012). Deterministic forecasts of binary events. In I. T. Jolliffe & D. B. Stephenson (Eds.), *Forecast verification: A practitioner's guide in atmospheric science* (2nd ed., pp. 31– 59). Wiley. Retrieved from https://onlinelibrary.wiley.com/doi/book/10.1002/9781119960003

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

MET (2020). Overview of MET. Available at https://met.readthedocs.io/en/v9.1.2/Users_Guide/overview.html. Last accessed 2 February 2022.

NCAR (2022). NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Available at https://data.eol.ucar.edu/dataset/21.093. Last accessed 2 February 2022.

NCEP (2021). NCEP Products Inventory: Real-Time Mesoscale Analysis (RTMA) Products. Available at https://www.nco.ncep.noaa.gov/pmb/products/rtma/#RTMA2p5. Last accessed 2 February 2022.

NCEP (2017). North American Mesoscale. Available at https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/nam.php#. Last accessed 2 February 2022.

Potvin, C. K., Carley, J. R., Clark, A. J., Wicker, L. J., Skinner, P. S., Reinhart, A. E., Gallo, B. T., Kain, J. S., Romine, G. S., Aligo, E. A., Brewster, K. A., Dowell, D. C., Harris, L. M., Jirak, I. L., Kong, F., Supinie, T. A., Thomas, K. W., Wang, X., Wang, Y., & Xue, M. (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

Snook, N., Kong, F., Brewster, K. A., Xue, M., Thomas, K. W., Supinie, T. A., Perfater, S., & Albright, B. (2019). Evaluation of convection-permitting precipitation forecast products using WRF, NMMB, and FV3 for the 2016–17 NOAA hydrometeorology testbed flash flood and intense rainfall experiments. Weather and Forecasting, 34(3), 781– 804. https://doi.org/10.1175/WAF-D-18-0155.1

Sobash, R. A., Kain, J. S., Bright, D. R., Dean, A. R., Coniglio, M. C., & Weiss, S. J. (2011). Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. Weather and Forecasting, 26(5), 714– 728. https://doi.org/10.1175/WAF-D-10-05046.1

SPC (2022). SPC Severe Weather Events Archive. Available at https://www.spc.noaa.gov/exper/archive/events/. Last accessed 2 February 2022.

Zhang, C., Xue, M., Supinie, T. A., Kong, F., Snook, N., Thomas, K. W., Brewster, K., Jung, Y., Harris, L. M., & Lin, S. J. (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