GEFSv12 High- and Low-Skill Day-10 Tornado Forecasts

DOUGLAS E. MILLER[a] AND VITTORIO A. GENSINI[a]

[a] *Department of Earth, Atmosphere, and Environment, Northern Illinois University, DeKalb, Illinois*

(Manuscript received 29 June 2022, in final form 19 March 2023, accepted 2 May 2023)

ABSTRACT: On average, modern numerical weather prediction forecasts for daily tornado frequency exhibit no skill beyond day 10. However, in this extended-range lead window, there are particular model cycles that have exceptionally high forecast skill for tornadoes because of their ability to correctly simulate the future synoptic pattern. Here, model initial conditions that produced a more skillful forecast for tornadoes over the United States were exploited while also highlighting potential causes for low-skill cycles within the Global Ensemble Forecasting System, version 12 (GEFSv12). There were 88 high-skill and 91 low-skill forecasts in which the verifying day-10 synoptic pattern for tornado conditions revealed a western U.S. thermal trough and an eastern U.S. thermal ridge, a favorable configuration for tornadic storm occurrence. Initial conditions for high skill forecasts tended to exhibit warmer sea surface temperatures throughout the tropical Pacific Ocean and Gulf of Mexico, an active Madden–Julian oscillation, and significant modulation of Earth-relative atmospheric angular momentum. Low-skill forecasts were often initialized during La Niña and negative Pacific decadal oscillation conditions. Significant atmospheric blocking over eastern Russia—in which the GEFSv12 overforecast the duration and characteristics of the downstream flow—was a common physical process associated with low-skill forecasts. This work helps to increase our understanding of the common causes of high- or low-skill extended-range tornado forecasts and could serve as a helpful tool for operational forecasters.

SIGNIFICANCE STATEMENT: This research provides a framework for the anticipation of a more (or less) skillful 10-day tornado forecast in an operational numerical weather prediction system. High-skill forecasts were associated with substantial tropical convection and warm sea surface temperature throughout the Pacific Ocean and Gulf of Mexico, whereas the underlying cause of low-skill forecasts were typically associated with a blocking anticyclone over eastern Russia. These findings are important because they permit increased or decreased confidence in a long-range forecast of tornado occurrence based on a dynamical prediction system.

KEYWORDS: ENSO; Madden–Julian oscillation; Severe storms; Climate variability; Forecast verification/skill; Operational forecasting

1. Introduction and background

Operational severe convective storm (SCS; tornadoes, severe hail, and severe convective wind gusts) forecasts, represented as convective "outlooks," are issued by the National Oceanic and Atmospheric Administration's (NOAA) Storm Prediction Center (SPC) each day with a forecast lead time out to day 8 (https://www.spc.noaa.gov/misc/about.html). These outlooks provide individual areal hazard probabilities within 40 km of a point for days 1 and 2, whereas the combined SCS probability is provided for days 3–8. Day-1 SPC convective outlook skill increased over the period 1979–2014, and increases for days 2 and 3 have been noted since the early 2000s (Hitchens and Brooks 2012, 2014). Increases in skill are likely attributable to reductions in numerical weather prediction (NWP) horizontal grid spacing, increases in vertical resolution, better representation of microphysics (Dyer et al. 2016; Morrison et al. 2020), and improvements to our heuristic understanding of severe convective storms (Wurman et al. 2012). SPC outlooks are less likely to be issued in the day-4–8 period because of the decreasing confidence and skill in NWP solutions with lead time (Gensini and Tippett 2019). Deterministic and ensemble forecast products are typically useful for lead times of 1–3 days, where forecasters often utilize the quantification of environmental "ingredients" that are necessary for the formation and sustenance of severe convective storms (Maddox 1976; Johns and Doswell 1992; Brooks et al. 1994, 2003; Grams et al. 2012).

Several recent efforts have focused on evaluating skill and increasing our understanding of U.S. SCS occurrence at and beyond the day-4–8 period (e.g., Gensini and Marinaro 2016; Lepore et al. 2018; Gensini and Allen 2018; Baggett et al. 2018; Moore 2018; Gensini et al. 2019, 2020a; Miller et al. 2020; Moore and McGuire 2020; Miller et al. 2022). From a strictly dynamical perspective, daily tornado and hail predictions were found to be skillful relative to climatological no-skill reference forecasts out to—on average—days 9 and 11, respectively, using the Global Ensemble Forecast System (GEFS), version 10, during spring 2016 and 2017 (Gensini and Tippett 2019). Because the GEFS cannot explicitly simulate SCS hazards, a modified version (adding simulated convective precipitation) of the supercell composite parameter (SCP; Thompson et al. 2003) was used as a predictor, similar to an environmental ingredients-based approach (Gensini and Tippett 2019).

This study was motivated by Gensini and Tippett (2019), where a relatively large spread in GEFS individual SCS forecast skill was noted at the average lead time of zero skill (i.e.,

---

Corresponding author: Vittorio A. Gensini, vgensini@niu.edu

DOI: 10.1175/WAF-D-22-0122.1

GEFSv12 High- and Low-Skill Day-10 Tornado Forecasts

DOUGLAS E. MILLER[a] AND VITTORIO A. GENSINI[a]

[a] *Department of Earth, Atmosphere, and Environment, Northern Illinois University, DeKalb, Illinois*

(Manuscript received 29 June 2022, in final form 19 March 2023, accepted 2 May 2023)

ABSTRACT: On average, modern numerical weather prediction forecasts for daily tornado frequency exhibit no skill beyond day 10. However, in this extended-range lead window, there are particular model cycles that have exceptionally high forecast skill for tornadoes because of their ability to correctly simulate the future synoptic pattern. Here, model initial conditions that produced a more skillful forecast for tornadoes over the United States were exploited while also highlighting potential causes for low-skill cycles within the Global Ensemble Forecasting System, version 12 (GEFSv12). There were 88 high-skill and 91 low-skill forecasts in which the verifying day-10 synoptic pattern for tornado conditions revealed a western U.S. thermal trough and an eastern U.S. thermal ridge, a favorable configuration for tornadic storm occurrence. Initial conditions for high skill forecasts tended to exhibit warmer sea surface temperatures throughout the tropical Pacific Ocean and Gulf of Mexico, an active Madden–Julian oscillation, and significant modulation of Earth-relative atmospheric angular momentum. Low-skill forecasts were often initialized during La Niña and negative Pacific decadal oscillation conditions. Significant atmospheric blocking over eastern Russia—in which the GEFSv12 overforecast the duration and characteristics of the downstream flow—was a common physical process associated with low-skill forecasts. This work helps to increase our understanding of the common causes of high- or low-skill extended-range tornado forecasts and could serve as a helpful tool for operational forecasters.

SIGNIFICANCE STATEMENT: This research provides a framework for the anticipation of a more (or less) skillful 10-day tornado forecast in an operational numerical weather prediction system. High-skill forecasts were associated with substantial tropical convection and warm sea surface temperature throughout the Pacific Ocean and Gulf of Mexico, whereas the underlying cause of low-skill forecasts were typically associated with a blocking anticyclone over eastern Russia. These findings are important because they permit increased or decreased confidence in a long-range forecast of tornado occurrence based on a dynamical prediction system.

KEYWORDS: ENSO; Madden–Julian oscillation; Severe storms; Climate variability; Forecast verification/skill; Operational forecasting

1. Introduction and background

Operational severe convective storm (SCS; tornadoes, severe hail, and severe convective wind gusts) forecasts, represented as convective "outlooks," are issued by the National Oceanic and Atmospheric Administration's (NOAA) Storm Prediction Center (SPC) each day with a forecast lead time out to day 8 (https://www.spc.noaa.gov/misc/about.html). These outlooks provide individual areal hazard probabilities within 40 km of a point for days 1 and 2, whereas the combined SCS probability is provided for days 3–8. Day-1 SPC convective outlook skill increased over the period 1979–2014, and increases for days 2 and 3 have been noted since the early 2000s (Hitchens and Brooks 2012, 2014). Increases in skill are likely attributable to reductions in numerical weather prediction (NWP) horizontal grid spacing, increases in vertical resolution, better representation of microphysics (Dyer et al. 2016; Morrison et al. 2020), and improvements to our heuristic understanding of severe convective storms (Wurman et al. 2012). SPC outlooks are less likely to be issued in the day-4–8 period because of the decreasing confidence and skill in NWP solutions with lead time (Gensini and Tippett 2019). Deterministic and ensemble forecast products are typically useful for lead times of 1–3 days, where forecasters often utilize the quantification of environmental "ingredients" that are necessary for the formation and sustenance of severe convective storms (Maddox 1976; Johns and Doswell 1992; Brooks et al. 1994, 2003; Grams et al. 2012).

Several recent efforts have focused on evaluating skill and increasing our understanding of U.S. SCS occurrence at and beyond the day-4–8 period (e.g., Gensini and Marinaro 2016; Lepore et al. 2018; Gensini and Allen 2018; Baggett et al. 2018; Moore 2018; Gensini et al. 2019, 2020a; Miller et al. 2020; Moore and McGuire 2020; Miller et al. 2022). From a strictly dynamical perspective, daily tornado and hail predictions were found to be skillful relative to climatological no-skill reference forecasts out to—on average—days 9 and 11, respectively, using the Global Ensemble Forecast System (GEFS), version 10, during spring 2016 and 2017 (Gensini and Tippett 2019). Because the GEFS cannot explicitly simulate SCS hazards, a modified version (adding simulated convective precipitation) of the supercell composite parameter (SCP; Thompson et al. 2003) was used as a predictor, similar to an environmental ingredients-based approach (Gensini and Tippett 2019).

This study was motivated by Gensini and Tippett (2019), where a relatively large spread in GEFS individual SCS forecast skill was noted at the average lead time of zero skill (i.e.,

---

Corresponding author: Vittorio A. Gensini, vgensini@niu.edu

DOI: 10.1175/WAF-D-22-0122.1

day 10) for spring tornado frequency. Essentially, some GEFS forecasts were very skillful at day 10, whereas others had negative skill. This suggests that a model evaluation of high- and low-skill forecasts may lead to identifying underlying regimes that contribute to enhanced or reduced predictability, specifically for extreme events at extended lead times. Several recent studies have followed a similar approach. For example, the representation of upstream surface cyclones is important in correctly predicting atmospheric blocking events within the European Centre for Medium-Range Weather Forecast (ECMWF) Ensemble Prediction System (EPS; Maddison et al. 2019). An ECMWF day-6 forecast was also evaluated and showcased Europe forecast "bust" events to be characteristic of a Rex-type blocking event (Rodwell et al. 2013). The authors found initial conditions for the bust events contained significant troughing over the U.S. Rocky Mountains and large values of convective available potential energy throughout the Great Plains. In addition, statistically significant higher winter average geopotential height prediction skill for the Climate Forecast System, version 2 (CFSv2), was evident during the negative phase of the North Atlantic Oscillation (NAO; Hurrell et al. 2003) under El Niño conditions, whereas larger prediction skill was evident during positive NAO under La Niña conditions (Miller and Wang 2019a).

This type of model evaluation can provide operational forecasters with higher or lower confidence in model solutions if certain sources of predictability or global weather regimes are present at the time of model initialization. Moreover, it may reveal model biases or sources of error that could prove useful to developers. Societally, losses caused by tornadoes can be devastating, and anticipating these events at extended lead times allows for increased awareness and potential for mitigation. Thus, day-10 GEFS, version 12 (GEFSv12), 500-hPa geopotential height (Z500) forecasts over the United States are investigated in tandem with the occurrence of practically perfect tornado day hindcasts. Initial model conditions were then analyzed and composited for high- and low-skill day-10 forecasts. Results were then put into context related to potential sources of forcing and variability that contribute to high- and low-skill extended-range tornado forecasts.

2. Data and methods

a. Data

Daily-averaged (from the native hourly interval) data from ECMWF's ERA5 reanalysis (Hersbach et al. 2020) during boreal spring over the period 2000–19 were downloaded and used as the verifying field for the day-10 GEFSv12 Z500 forecast. Initial conditions, as well as the evolutions, were investigated from ERA5 by examining Z500, outgoing longwave radiation (OLR), and sea surface temperature (SST). Geopotential, zonal wind, and meridional wind at 200 hPa were used for the calculation of wave-activity flux (WaF). Zonal wind at each vertical level was used for the calculation of Earth-relative atmospheric angular momentum (AAM) following the methods outlined in Miller et al. (2022), which included constraining the zonal integral to the Northern Hemisphere.

Model evaluation was conducted using GEFS version 12 (GEFSv12; Guan et al. 2022) reforecasts over the period 2000–19 during boreal spring [March–May (MAM)]. GEFSv12 reforecasts, containing one control and four perturbation runs, were initialized once per day from 0000 UTC initial conditions and integrated to a forecast lead time of 16 days. An additional six-perturbation cycle integrated to day 35 is also provided once weekly on Wednesdays. GEFSv12 has recently been utilized to analyze the performance in precipitation forecasts in India (Saminathan et al. 2021; Nageswararao et al. 2022) and additional hydrometeorological applications (Guan et al. 2022). GEFSv12 reforecasts have also been shown to contain smaller central U.S. 2-m temperature biases throughout the year, and overall higher skill in predicting Z500 than its Subseasonal Experiment (SubX; Pegion et al. 2019) reforecast counterpart (Guan et al. 2022). Specifically, Z500 over the United States (63°–124°W, 24°–50°N) was the forecast verification field evaluated in section 3, and our focus was on the day-10 lead window for a 1200–1200 UTC convective day (forecast hours 252–276 for a 0000 UTC cycle).

Tornado practically perfect hindcasts (PPH; Hitchens et al. 2013; Gensini et al. 2020b) were used to identify tornado days that occurred on high- or low-skill day-10 Z500 forecasts. Tornado PPHs represent a Gaussian-smoothed statistical point process of event probabilities with the goal of resembling what a perfect SPC convective outlook may look like. Days on which the U.S. maximum tornado PPH probability exceeded the March–June 2000–19 75th percentile (17.27%) were examined. This corresponds to a tornado PPH probability in between an "enhanced" and "moderate" categorical threshold by the SPC, which are associated with 10% and 30% tornado PPH probabilities, respectively. Hereinafter, these 75th-percentile tornado PPH days are referred to as tornado days for brevity.

b. Teleconnection indices, atmospheric blocking, and wave activity flux

Several weather/climate teleconnection indices were explored to identify any reoccurring and potential significant signals that may be present during high- and low-skill tornado day forecasts. The Madden–Julian oscillation (MJO; Baggett et al. 2018; Gensini et al. 2019, 2020a) has recently been utilized as a source of subseasonal-to-seasonal (S2S) predictability for SCSs, and recent work discovered increased probabilities of SCSs 3–4 weeks following the strongest MJO convection signal (Miller et al. 2022). CPC pentad MJO indices (NWS/CPC 2022; Baxter et al. 2014) were used to represent the MJO (as in Miller et al. 2022), which represent the negative projection of 200-hPa velocity potential (χ200) anomalies onto the 10 time-lagged patterns of the first extended empirical orthogonal function of pentad χ200 anomalies.

Indices describing the modulation of the Arctic Oscillation (AO), the North Atlantic Oscillation (NAO), the El Niño–Southern Oscillation (ENSO), Pacific decadal oscillation (PDO), the Pacific–North American (PNA) pattern, and the quasi-biennial oscillation (QBO) were all investigated to examine if recurrent phases of teleconnections were present during model cycle initialization and day-10 tornado forecasts. All indices

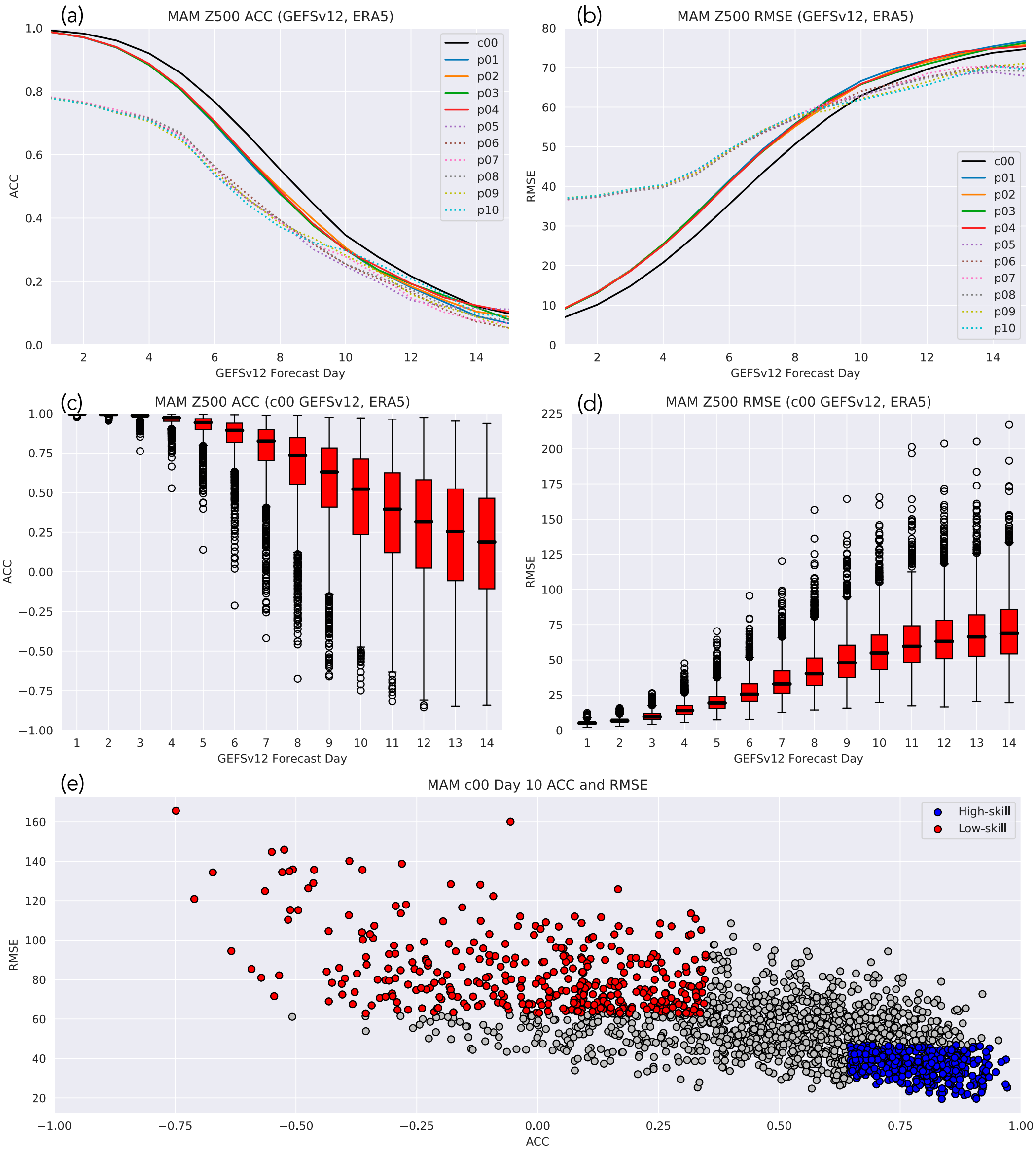

FIG. 1. The (a) anomaly correlation coefficient (ACC) and (b) root-mean-square error (RMSE) over the United States (63°–124°W, 24°–50°N) between the GEFSv12 and ERA5 Z500 fields from day 1 to day 15 for the control run and 10 perturbation members. Box-and-whisker plots for the control run (c) ACC and (d) RMSE for all MAM Z500 forecasts. (e) Scatterplot showing the ACC vs RMSE for all control member MAM Z500 forecasts. Red and blue circles represent the low- and high-skill forecasts, respectively.

were obtained and are available via the NOAA Physical Sciences Laboratory (PSL; https://psl.noaa.gov/data/climateindices/list/).

In addition to teleconnections, two physical processes of interest on these time scales are atmospheric blocking and wave propagation. Atmospheric blocking is characterized as a quasi-stationary, high pressure system that acts to hamper the normal progression of the westerly jet and may induce extreme weather conditions (Sillmann and Croci-Maspoli 2009; Buehler et al. 2011; Dole et al. 2011; Miller et al. 2020, 2021; Miller and Wang 2022, 2019b). A two-dimensional blocking index was calculated following the methods of Miller et al. (2021), which identifies Z500 anomalies that exceed 1.5σ, have an area greater than 10° × 10°, and persist for at least 5 days. Blocking indices were calculated for day 0 (model initialization) and day 10.

Observed wave propagation for both high- and low-skill forecast groups were examined by calculating the WaF ($m^2\ s^{-2}$) at 200 hPa following Takaya and Nakamura (2001). WaF indicates a snapshot of the propagation direction of stationary or migrating waves and is nearly parallel to the group velocity. The divergence/convergence of WaF indicates regions where the wave packet is emitted/absorbed.

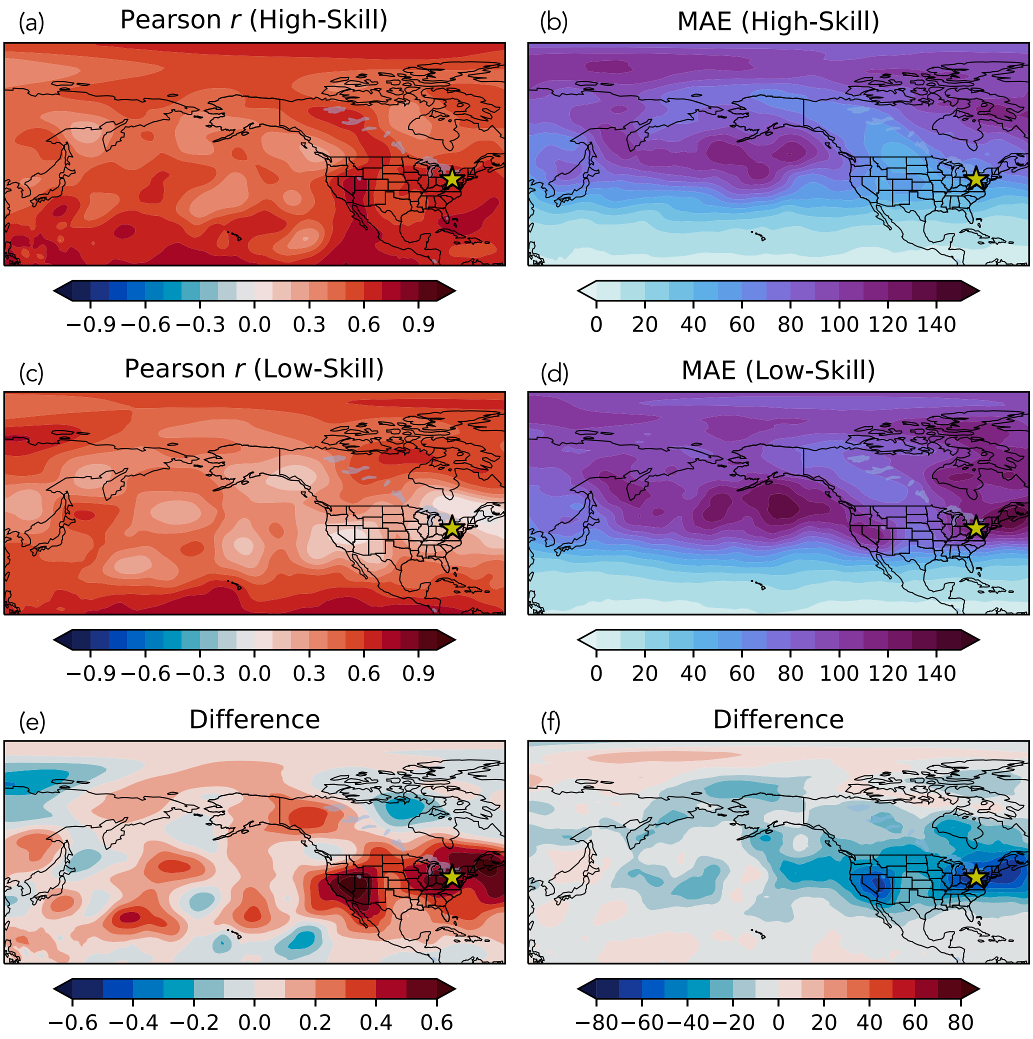

FIG. 2. Two-dimensional (a),(c) Pearson *r* and (b),(d) mean absolute error between the GEFSv12 day-10 forecasts and ERA5 for (top) high-skill and (middle) low-skill forecasts. Also shown is the difference between the high- and low-skill (e) ACC and (f) MAE. The yellow star indicates the location evaluated in Fig. 3, below.

c. *Defining windows of high and low forecast skill*

Anomaly correlation coefficient (ACC; Wilks 2011) and root-mean-square error (RMSE; Wilks 2011) were first calculated for the Z500 anomaly field over the United States between the ERA5 and GEFSv12 forecasts from days 1 to 15 for each ensemble member (Figs. 1a,b). Anomalies were created by removing the 2000–19 daily averages for both the ERA5 and GEFSv12. High- and low-skill days were chosen using the control member given initial exploratory findings that control forecasts had consistently higher skill than the perturbation members, which is similar to previous work (Rodwell et al. 2013). It is important to note that the results were similar when examining the ensemble mean, with composites being more smoothed out for model conditions. A day-10 control member forecast with an ACC ≥ 66th percentile and a RMSE ≤ 33rd percentile was labeled as a high-skill forecast. A low-skill forecast day contained an ACC ≤ 33rd percentile and a RMSE ≥ 66th percentile. Together, these ensure that a high-skill forecast is associated with relatively low errors related to magnitude (RMSE) and pattern location (ACC). A total of 88 high-skill and 91 low-skill forecasts were identified over the study period.

d. *Evaluation metrics and statistical significance*

Two-dimensional fields of Pearson *r* correlations and mean absolute error (MAE) for the day-10 forecasts were calculated for both forecast groups (Fig. 2). Calculations were made at each grid point between the ERA5 and GEFSv12 day-10 forecasts of Z500 anomalies. Both datasets were bilinearly interpolated to a common latitude/longitude grid (1° × 1°) for the calculations of ACC and RMSE.

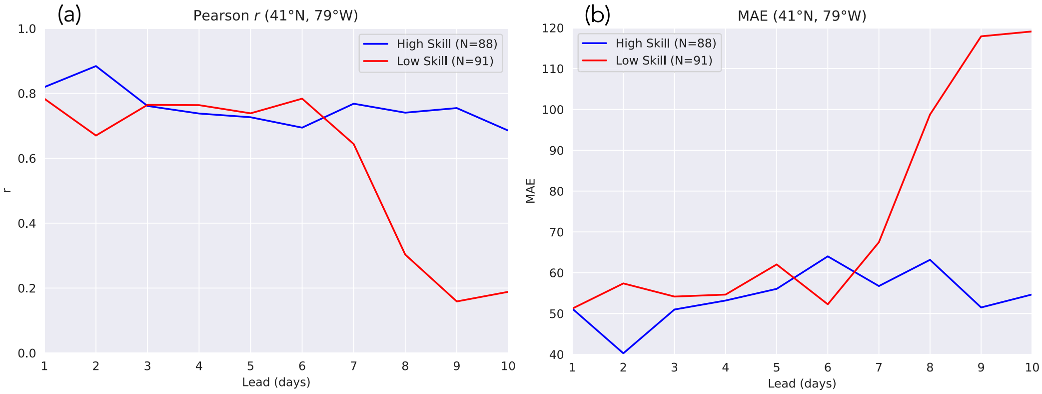

FIG. 3. The average (a) Pearson r and (b) MAE evolution from day 1 to day 10 at 41°N, 79°W for the high-skill (blue) and low-skill (red) groups. The sample size for each group is located in the legend.

Anomalies of Z500, OLR, SST, blocking frequency, and several teleconnection indices were constructed by removing the 2000–19 daily averages (i.e., seasonal cycle) to examine the initial state for the high- and low-skill groups. Statistical significance of the anomalies were tested using a two-tailed Student's t test, with the null hypothesis that the anomalies do not differ from zero. The statistical significance of blocking frequency anomalies was performed using a Mann–Whitney U test for the medians. Results were considered to be significant if the p value was less than 0.05 (95% confidence level). Field significance was performed using the false discovery rate (FDR; Wilks 2011) of $\alpha = 0.1$ for composite anomalies of several fields throughout the study.

3. Evaluation of GEFSv12 Z500 over the United States

The control run showed the highest average ACC through day 14 while also containing the lowest average RMSE through day 10 in comparison with the perturbation members (Figs. 1a,b). Given this, the rest of the focus herein is on the control member. Note that we are not examining the predictability of the ensemble system. Rather, we are trying to understand the initial conditions associated with individual forecasts, which by nature are individual members. The control member ACC and RMSE spread was small at shorter lead times, and increased the largest at day 6 (Figs. 1c,d). For example, the control run at day 10 contained spread of near perfect ACC/RMSE to outliers of ACC equal to -0.75 (opposite anomaly pattern from what was forecast) and RMSE > 150 m. The median ACC experienced the largest decreases from day 9 to 10. As stated previously, focus here is on the day-10 forecasts motivated by Gensini and Tippett (2019), where the average lead time of zero skill for all forecasts was day 10. The high-skill forecasts contain ACC and RMSE values greater than the 66th percentile (0.65) and less than the 33rd percentile values (46.81 m), respectively (Fig. 1e). In contrast, a low-skill forecast contains an ACC ≤ 0.35 and RMSE values ≥ 62.88 m. In total, 88 high-skill and 91 low-skill tornado events occurring at day 10 were found.

High-skill day-10 forecasts exhibited r values greater than 0.7 over the East and West Coast of the United States (Fig. 2a), which contrasted lower values (0–0.3) for low-skill forecasts (Fig. 2c). Major differences in ACC between the two groups were found over the United States, with values greater than 0.6 higher for high-skill forecasts (Fig. 2e). Other areas of higher Pearson r were located over the Pacific Ocean, northwest Canada, and Newfoundland. The largest MAE values for both groups were located over the Pacific Ocean (Figs. 2b,d), with the largest group differences over the U.S. East and West Coast, similar to the pattern found in Pearson r (Fig. 2f).

Lead-day averaged r and MAE values were also examined at a point over the eastern United States (41°N, 79°W; yellow dot in Fig. 2), which is collocated with the area of largest changes in metrics between the high- and low-skill groups. High-skill forecasts contained nearly constant r values between 0.7 and 0.8 from days 1 to 10 (Fig. 3a), while the MAE remained between 40 and 65 m (Fig. 3b). Low-skill forecasts had similar correlations and MAE values through day 6, followed by a significant decrease of Pearson r and increase of MAE through day 10. In essence, NWP evolution of the Z500 field was, on average, poor beyond day 6 for these forecasts.

4. Comparison between skillful and poorly predicted day-10 tornado days

a. *Observed characteristics of high- and low-skill forecasts*

Examining Z500 forecast evolution, locally significant positive Z500 anomalies were evident over the Alaska–Canadian border at day 1 for the high-skill groups (Fig. 4a). The observed Z500 for low-skill day-1 forecasts contained significant positive anomalies farther west relative to that of the high-skill forecasts (Fig. 4c), and it appears that these positive anomalies were the greatest differences among the two groups (Fig. 4e). Unsurprisingly, the GEFSv12 day-1 forecast (Figs. 4a,c, contours) for both groups closely resembles the ERA5. However, minor differences associated with the Pacific wave train were already occurring at day 1 for the low-skill forecasts (Fig. 4c). Observed day-10 patterns over the United States are relatively similar between both groups (Figs. 4b,d), showcasing a midlevel western U.S. thermal trough and eastern

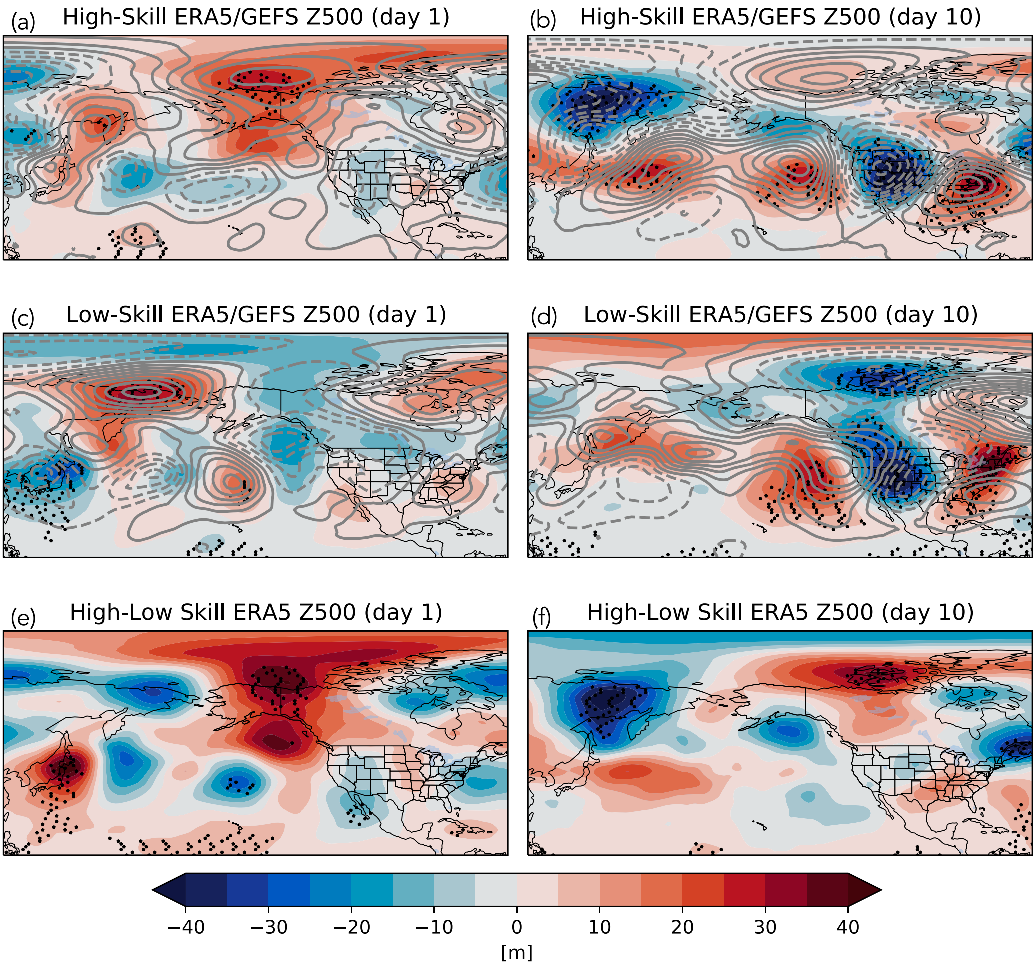

FIG. 4. ERA5 (shading) and GEFSv12 (contours) Z500 anomalies for (a),(c) day 1 and (b),(d) day 10 for the (top) high-skill group and (middle) low-skill group. Black dots indicate ERA5 anomalies statistically significant at the 95% confidence interval. Also shown are the difference between the high- and low-skill ERA5 Z500 anomalies for days (e) 1 and (f) 10. Black stippling indicates where anomalies are locally statistically significant at the 95% confidence level. No grid points were field significant.

U.S. thermal ridge synoptic pattern conducive for severe convective storms (Miller et al. 2020). The largest differences were associated with a significant north–south dipole in the day-10 Z500 pattern located over eastern Russia for the high-skill forecasts (Figs. 4b,f). Day-10 predicted Z500 field for the high-skill group closely resembles that of the ERA5. Not only is the U.S. trough west–east ridge pattern replicated, but the western Pacific north–south dipole is also evident (contours in Fig. 4b). GEFSv12 forecasts for the low-skill group produced a western U.S. midlevel thermal ridge, central U.S. thermal troughing, and southeast U.S. thermal ridging, which was out of phase relative to the ERA5 (Fig. 4d), leading to low Pearson r values on the U.S. West and East Coasts (Fig. 2c). A closer look at the evolution of the low-skill day-10 forecasts (Fig. 5) revealed that the GEFSv12 was too amplified with eastern Russia positive Z500 anomalies, especially at day 6 (Fig. 5c). By day 8 (Fig. 5d), the wave train was also out of phase, leading to rapidly decreasing and increasing ACC and RMSE, respectively. In fact, positive Z500 anomalies over eastern Russia at initialization were associated with a blocking anticyclone.

Atmospheric blocking is a significant producer of extreme weather (Sillmann and Croci-Maspoli 2009; Buehler et al. 2011; Dole et al. 2011; Miller and Wang 2019b; Miller et al. 2020; Miller and Wang 2022), and operational models have difficulty forecasting the initiation, duration, and cessation of blocking events (Tibaldi and Molteni 1990; Matsueda 2011; Reynolds et al. 2017). A two-dimensional blocking index was calculated and examined for both the high- and low-skill groups. Field significant positive anomalies of blocking frequency were located over eastern Russia at day 0 for the low-skill forecasts, hinting at a potential problem for the evolution of the flow (Fig. 6c). Although blocking onset mechanisms have been extensively studied—showing to be complex and varying

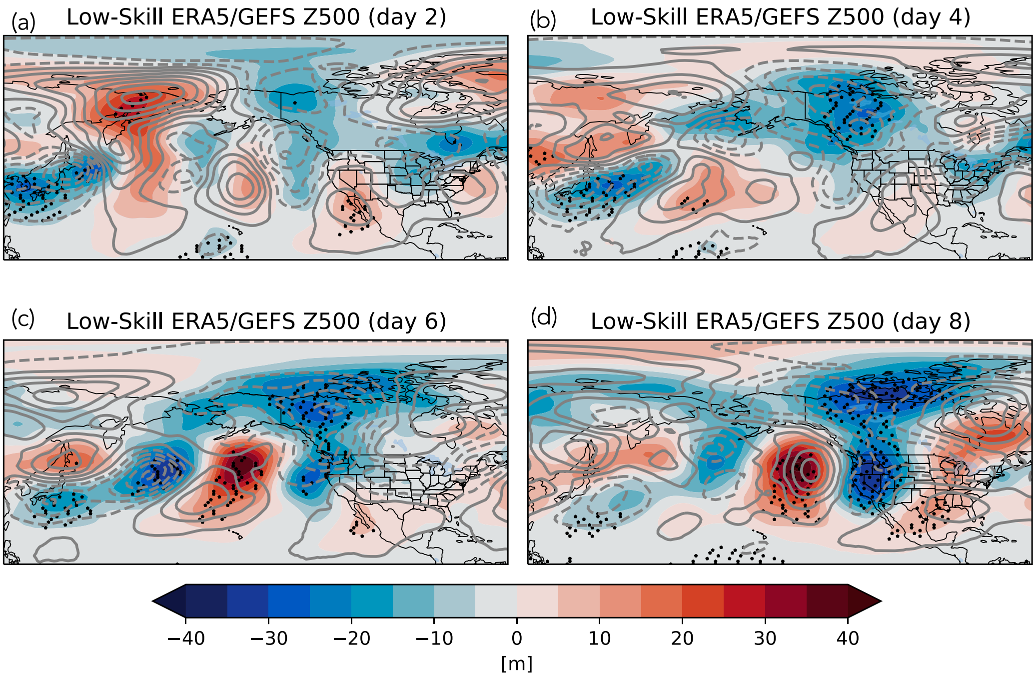

FIG. 5. As in Figs. 4a–d, but for the low-skill group at (a) day 2, (b) day 4, (c) day 6, and (d) day 8. Black stippling indicates where anomalies are locally statistically significant at the 95% confidence level.

by region (Miller and Wang 2022)—dynamics involved in blocking maintenance and cessation remain challenging. Owing to the unsettled theories of blocking decay, operational models may fail to predict the duration of the blocking life cycle (Reynolds et al. 2017). Since blocking can significantly influence the downstream evolution of the synoptic flow (Rex 1950; Miller and Wang 2022), an overforecast blocking duration, as we see here, can cause significant error growth in the model solution.

Differences in the WaF propagation was also found to be a potential cause for poor model performance of the low-skill groups. The wave propagation pathway from the ERA5 was different between the high- and low-skill groups, specifically at days 6 and 8 (Fig. 7). WaF vectors for the high-skill forecasts evolved from the central Pacific to the northwest United States, showcasing a significant meridional component at day 6 (Fig. 7a). The convergence intensified over the U.S. West Coast,

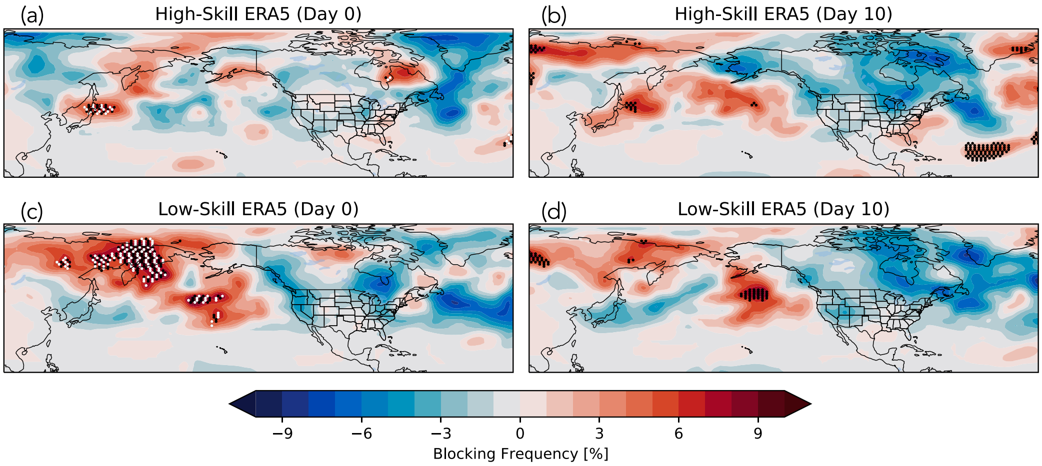

FIG. 6. ERA5 blocking frequency anomalies (% of days) for (a),(c) day-0 and (b),(d) day-10 (top) high-skill forecasts and (bottom) low-skill forecasts. Black stippling indicates where anomalies are locally statistically significant at the 95% confidence level, whereas white stippling indicates anomalies that are field significant at the FDR level of $\alpha = 0.1$.

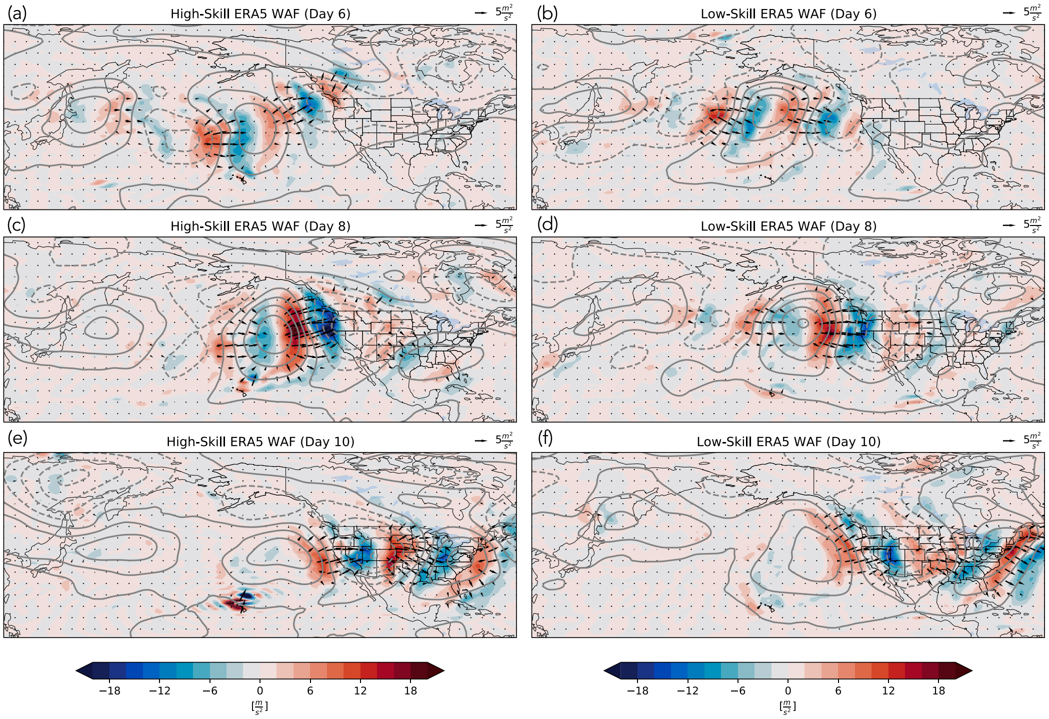

FIG. 7. ERA5 wave activity flux (vectors; $m^2 s^{-2}$) and convergence (shading; m s^{-2}) for days (a),(b) 6; (c),(d) 8; and (e),(f) 10 of the (left) high-skill and (right) low-skill forecasts. The scale for WaF vectors is shown in the top-right corners of the panels.

working to amplify the thermal trough over the western United States at day 8 (Fig. 7c). The GEFSv12 reproduced the meridional component of the WaF vectors with similar convergence off the U.S. West Coast in the high-skill forecasts (Figs. 8a,c,e). In contrast with the high-skill forecasts, the low-skill evolution showed a zonal propagation with almost no meridional influence at days 6 and 8 (Figs. 7b,d), and the WaF convergence over the western United States was weaker than the observed WaF during high-skill forecasts. GEFSv12 WaF at day 6 showed a waveguide extending from Japan to the west of Hawaii, with no zonal propagation to the United States (Fig. 8b), contrary to what was observed in the ERA5 (Fig. 7b). We suspect the overpredicted blocking anticyclone was preventing the zonal waveguide from occurring and evolving in GEFSv12.

Several sources of subseasonal-to-seasonal variability (Merryfield et al. 2020) have been shown to modulate severe convective storm occurrence, including the MJO (Thompson and Roundy 2013; Barrett and Gensini 2013; Baggett et al. 2018; Gensini et al. 2020a; Moore and McGuire 2020; Miller et al. 2022), the ENSO (Allen et al. 2015), the AAM budget (Gensini and Marinaro 2016; Moore 2018; Gensini et al. 2020a), land surface (Frye and Mote 2010), and sea ice (Trapp and Hoogewind 2018). To examine this variability, OLR and SST states at initialization were analyzed for both skill groups. Near opposite OLR patterns were present, especially over the western tropical Pacific Ocean (Figs. 9a,c,e). Locally statistically significant negative OLR anomalies were present over the "Maritime Continent" along with positive anomalies over the central and eastern tropical Pacific for the high-skill forecasts, whereas the low-skill group exhibited positive anomalies over the Maritime Continent (negative OLR anomalies represent increased cloudiness/precipitation). The negative OLR anomalies were associated with the MJO (section 4b).

As mentioned, a stark contrast in SST was apparent between the high- and low-skill initialization states. Skillful forecasts contained field significant warmer temperatures in the central tropical Pacific Ocean, off the west coast of the United States, and the Gulf of Mexico (Fig. 9b). In contrast, SSTs during initialization for the low-skill forecasts contained La Niña–like conditions with field significant (FDR value < 0.1) colder SSTs than the seasonal cycle throughout the tropical Pacific Ocean and off the west coast of the United States and Canada (Fig. 9d). These spring conditions are known to be associated with increased probabilities of severe weather (Lee et al. 2016), yet the GEFSv12 fails to predict the pattern evolution accurately. Differences between the two skill groups (Fig. 9f) highlight the opposite patterns with field significant differences occurring throughout the domain.

b. *The MJO and climate indices*

The MJO is a primary source of subseasonal predictability and has recently been exploited for its relation to U.S. severe convective weather (Gensini et al. 2019; Baggett et al. 2018; Miller et al. 2022). The CPC pentad MJO indices and AAM for 20 days before/after the day of GEFSv12 initialization for

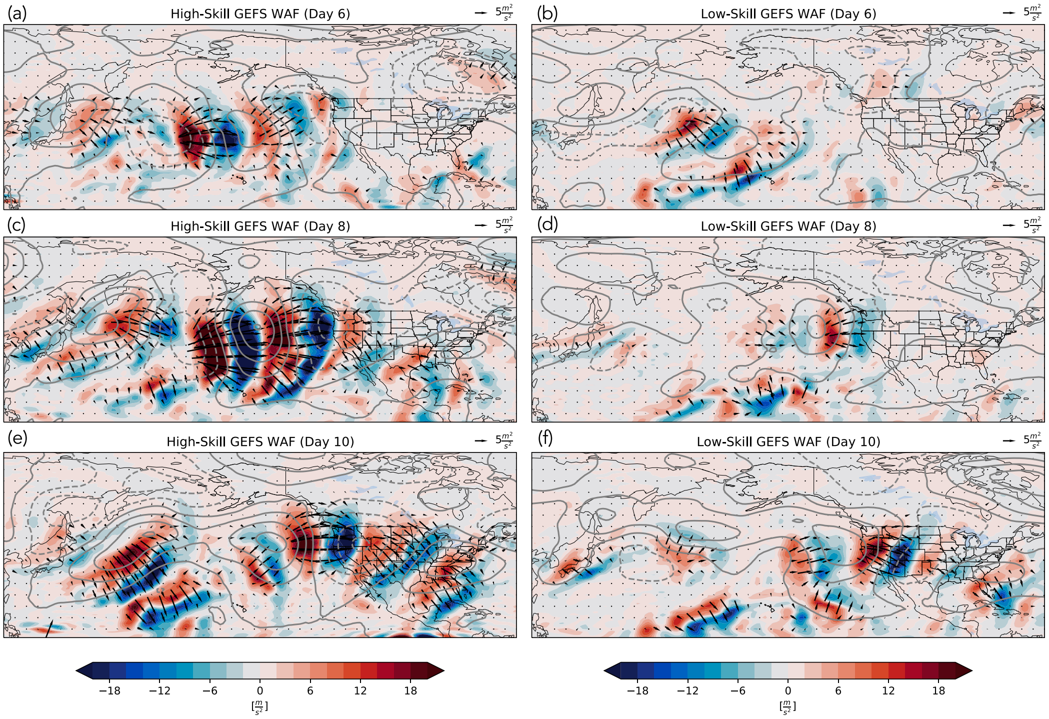

FIG. 8. As in Fig. 7, but for the GEFSv12.

both forecast groups were examined (Fig. 10). Major differences were evident, with the skillful forecasts being initialized ~20 days following a strong active MJO which appears to propagate, on average, rather quickly (Fig. 10a). A significant increase in AAM is apparent between days −15 and −5 (Fig. 10b), followed by a minimum around the time of peak severe weather frequency (day 10). The noted increases in AAM following the MJO were reminiscent of the modulation

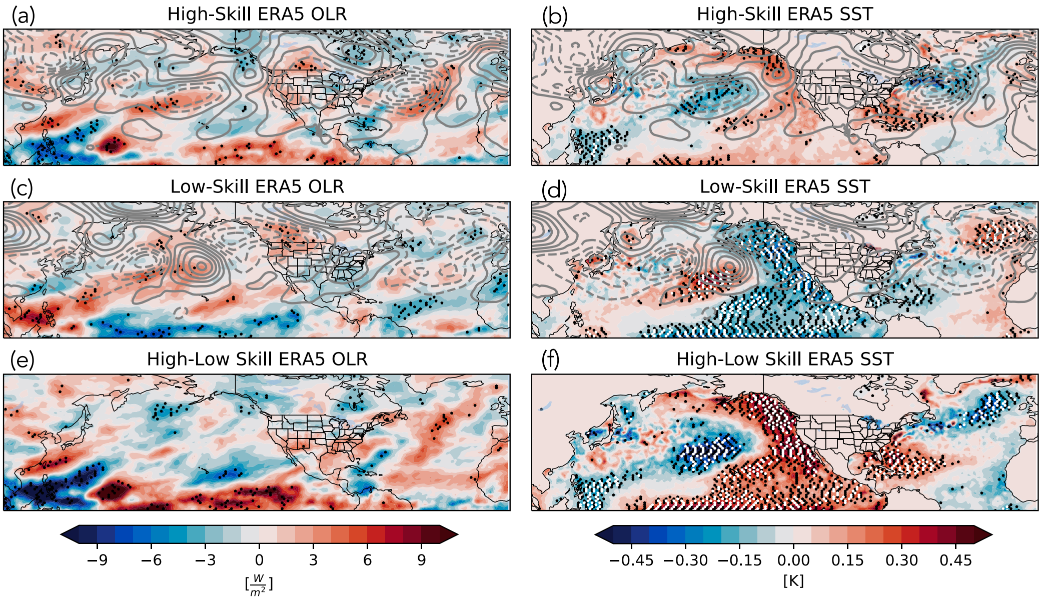

FIG. 9. ERA5 (a),(c) outgoing longwave radiation (OLR; W m^{-2}) anomalies and (b),(d) sea surface temperature (SST; K) anomalies for the (top) high-skill group and (bottom) low-skill group. Gray contours represent Z500 anomalies (m). Also shown is the difference of (e) OLR and (f) SST between the two skill groups. Black stippling indicates where anomalies are locally statistically significant at the 95% confidence level, whereas white stippling indicates anomalies that are field significant at the FDR level of $\alpha = 0.1$.

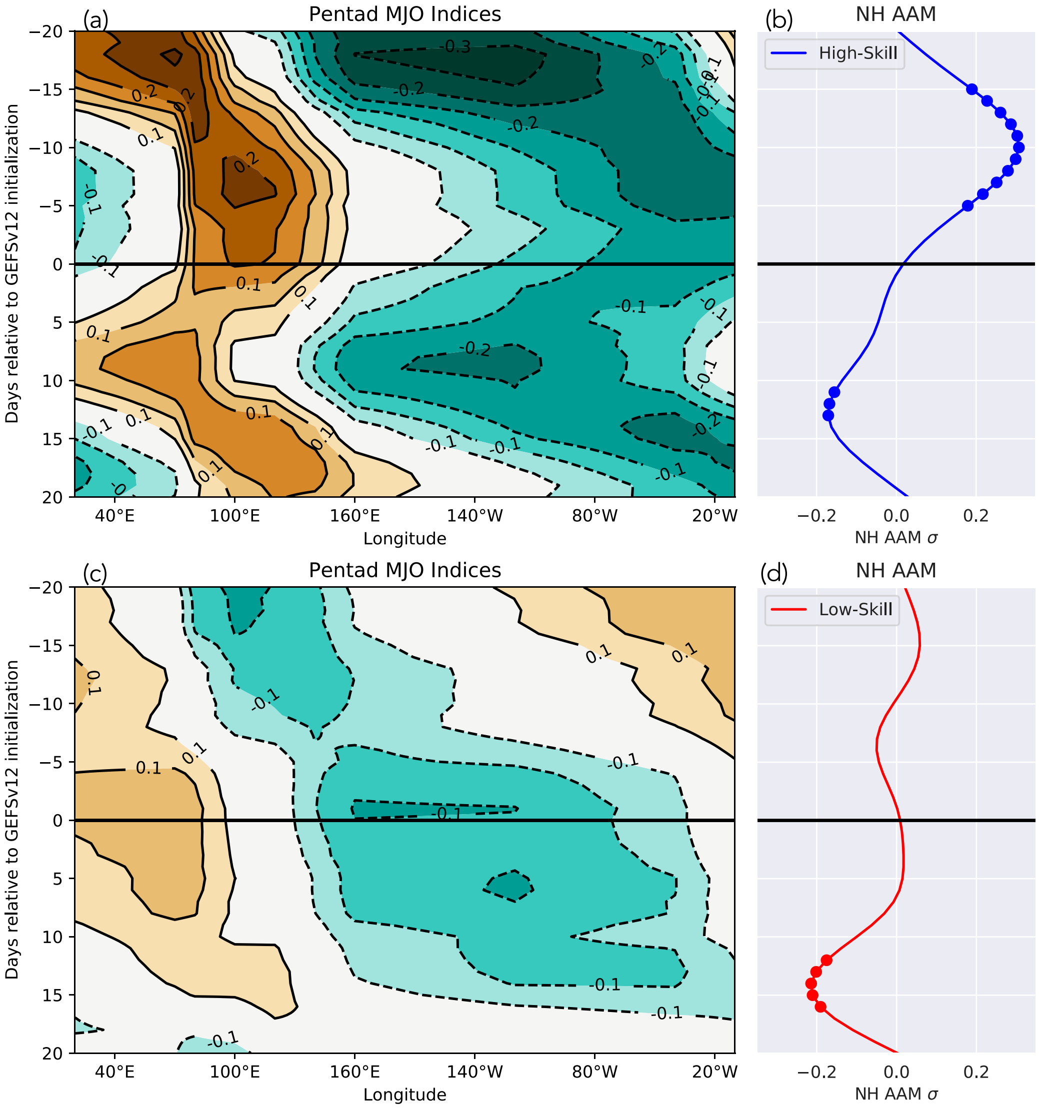

FIG. 10. (a),(c) Observed CPC pentad MJO indices and (b),(d) Northern Hemisphere AAM for 20 days prior to 20 days after the date of initialization for the (top) high-skill forecast group and (bottom) low-skill group. Green shading represents "active convection." Colored dots in (b) and (d) indicate statistically significant AAM anomalies at the 95% confidence level.

following the minimum pentad index discussed in Miller et al. (2022). Low-skill forecasts had a slower propagating, weaker MJO signal, and despite the significant decrease ~12 days following model cycle initialization, there was no apparent preceding peak in AAM (Figs. 10c,d). These results were consistent with the background state of SST (Fig. 9), where the MJO was found to propagate faster during warm Pacific SST conditions (El Niño), whereas slower propagation occurred during relatively cold Pacific SST conditions (Pohl and Matthews 2007). In summary, MJO and AAM characteristics *prior* to model initialization indicate distinguishing signals to identify "forecasts of opportunity" and periods of lower forecast skill, potentially allowing for a "forecast of forecast skill."

Various weather/climate teleconnection indices were examined, including the AO, NAO, Niño-3.4 index, PDO, PNA, and QBO for the initialization date and day 10 (Fig. 11). No teleconnection indices were significant precursor signals for the skillful forecasts. Statistically significant negative indices for low-skill forecasts existed for the Niño-3.4 and PDO at day 0, confirming the La Niña conditions evident in Fig. 9d and the discussion surrounding the observed MJO. The La Niña and negative PDO conditions were still significant at day 10 for the low-skill forecasts. Additionally, the negative PNA pattern occurred for both forecast skill groups, congruent with the "high-low-high" Z500 pattern over the eastern Pacific and the United States (Figs. 4b,d). In general, caution should

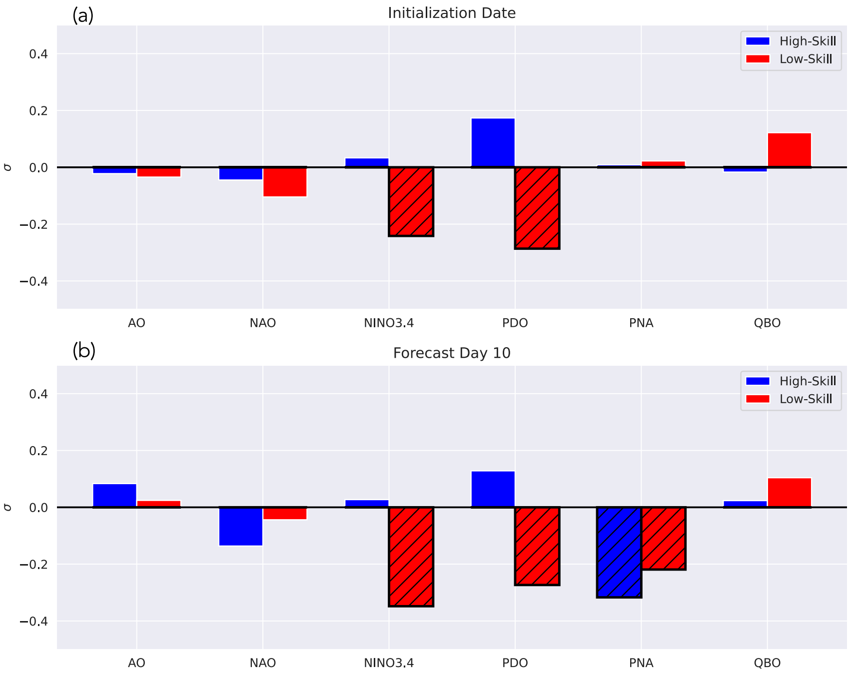

FIG. 11. Mean values of several teleconnection indices for high- (blue) and low-skill (red) forecasts for the (a) initialization date and (b) forecast day 10. Hatched bars indicates statistically significant values at the 95% confidence level.

be practiced when examining such teleconnection indices for this purpose, as they represent a particular state of the atmosphere in specific geographic regions that individually may or may not be related to the evolution of the physical synoptic weather pattern over the CONUS.

5. Discussion and conclusions

Despite the need for extended-range prediction of severe convective storms, the current predictability limit for daily forecasts is, on average, ~10 days. As a result of chaos, skill in NWP solutions decrease as model integration through time increases. However, instances exist in which forecast solutions are more or less skillful at day 10 (and beyond), with the former providing "forecasts of opportunity." To examine these instances, atmospheric and oceanic states (and their time evolution) that supported a high- or low-skill day-10 forecast were examined for 75th percentile tornado day hindcasts. The day-10 GEFSv12 Z500 field over the United States during these active tornado days was evaluated and split into high- and low-skill groups based on Pearson *r* correlation and mean absolute error. The observed Z500 pattern supportive of tornado days over the United States at day 10 were similar for both groups, characteristic of a midlevel western U.S. thermal trough and eastern U.S. thermal ridge.

High-skill GEFSv12 forecasts reproduced the geopotential height pattern at 500 hPa well, whereas the pattern for the low-skill forecasts was out of phase and propagated too quickly. High-skill forecasts were also associated with northward wave activity flux propagation and strong convergence off the west coast of the United States. These forecasts also were characteristic of field significant negative OLR anomalies over the Maritime Continent and warmer-than-average SSTs throughout the tropical Pacific and Gulf of Mexico. Additionally, high-skill forecasts exhibited a strong MJO 15–20 days prior to model initialization, which lead to a significant modulation of AAM, which is consistent with previous research (Gensini et al. 2019, 2020a; Miller et al. 2022).

In contrast to the high-skill forecasts, WaF propagation for the low-skill forecasts contained nearly zonal WaF vectors. Colder than normal SSTs were evident throughout the tropical and eastern Pacific, reminiscent of La Niña conditions. A weaker and slower MJO was evident relative to the high-skill group, and no significant peak in AAM occurred prior to forecast initialization. The most prominent feature with low-skill forecasts was a blocking anticyclone over eastern Russia during model initialization, in which the GEFSv12 overforecast the duration of the block. This is perhaps not too surprising, as theories of atmospheric blocking onset, maintenance, and decay are still unsettled, and research has shown that models often fail to predict the duration of atmospheric blocking events (Reynolds et al. 2017).

This research is the first step in understanding under what initialization conditions an extended-range GEFSv12 forecast for tornado conditions may promote higher or lower forecast skill. Inherently, this study exposes forecasts of opportunity, which could be leveraged as windows of enhanced dynamical model forecast skill for operational forecasters. These results promote the further possibility of aiding in the creation of medium-range tornado outlooks—but most importantly—permit the assessment of forecast confidence and allow for the potential to "forecast the forecast skill." For example, if a blocking anticyclone is present over the Northwest Pacific, GEFSv12 forecasts may not evolve properly due to the overprediction of blocking duration and subsequent impacts on the downstream flow. Additional model diagnostic studies for other variables may further aid in our understanding of potential biases and sources of errors for extended-range prediction of extreme weather events like tornadoes.

Acknowledgments. This work was supported by the National Science Foundation (award #2048770). We acknowledge the Northern Illinois University Center for Research Computing and Data for providing computing resources. This research also used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

Data availability statement. ERA5 data were downloaded from the ECMWF Copernicus Climate Change Service (C3S; https://cds.climate.copernicus.eu/). The GEFSv12 data are available via the Amazon web server (https://noaa-gefs-retrospective.s3.amazonaws.com/index.html). The CPC Pentad MJO indices are available online (https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_mjo_index/pentad.html). PPH data are available online (https://atlas.niu.edu/pperfect/BAMS/). Teleconnection indices were obtained from NOAA's PSL (https://psl.noaa.gov/data/climateindices/list/).

REFERENCES

Allen, J. T., M. K. Tippett, and A. H. Sobel, 2015: Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nat. Geosci.*, **8**, 278–283, https://doi.org/10.1038/ngeo2385.

Baggett, C. F., K. M. Nardi, S. J. Childs, S. N. Zito, E. A. Barnes, and E. D. Maloney, 2018: Skillful subseasonal forecasts of weekly tornado and hail activity using the Madden–Julian Oscillation. *J. Geophys. Res. Atmos.*, **123**, 12 661–12 675, https://doi.org/10.1029/2018JD029059.

Barrett, B. S., and V. A. Gensini, 2013: Variability of central United States April–May tornado day likelihood by phase of the Madden–Julian Oscillation. *Geophys. Res. Lett.*, **40**, 2790–2795, https://doi.org/10.1002/grl.50522.

Baxter, S., S. Weaver, J. Gottschalck, and Y. Xue, 2014: Pentad evolution of wintertime impacts of the Madden–Julian oscillation over the contiguous United States. *J. Climate*, **27**, 7356–7367, https://doi.org/10.1175/JCLI-D-14-00105.1.

Brooks, H. E., C. A. Doswell III, and J. Cooper, 1994: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting*, **9**, 606–618, https://doi.org/10.1175/1520-0434(1994)009<0606:OTEOTA>2.0.CO;2.

——, J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67**, 73–94, https://doi.org/10.1016/S0169-8095(03)00045-0.

Buehler, T., C. C. Raible, and T. F. Stocker, 2011: The relationship of winter season North Atlantic blocking frequencies to extreme cold or dry spells in the ERA-40. *Tellus*, **63A**, 212–222, https://doi.org/10.1111/j.1600-0870.2010.00492.x.

Dole, R., and Coauthors, 2011: Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.*, **38**, L06702, https://doi.org/10.1029/2010GL046582.

Dyer, J., C. Zarzar, P. Amburn, R. Dumais, J. Raby, and J. A. Smith, 2016: Defining the influence of horizontal grid spacing on ensemble uncertainty within a regional modeling framework. *Wea. Forecasting*, **31**, 1997–2017, https://doi.org/10.1175/WAF-D-16-0030.1.

Frye, J. D., and T. L. Mote, 2010: Convection initiation along soil moisture boundaries in the Southern Great Plains. *Mon. Wea. Rev.*, **138**, 1140–1151, https://doi.org/10.1175/2009MWR2865.1.

Gensini, V. A., and A. Marinaro, 2016: Tornado frequency in the United States related to global relative angular momentum. *Mon. Wea. Rev.*, **144**, 801–810, https://doi.org/10.1175/MWR-D-15-0289.1.

——, and J. T. Allen, 2018: U.S. hail frequency and the global wind oscillation. *Geophys. Res. Lett.*, **45**, 1611–1620, https://doi.org/10.1002/2017GL076822.

——, and M. K. Tippett, 2019: Global Ensemble Forecast System (GEFS) predictions of days 1–15 U.S. tornado and hail frequencies. *Geophys. Res. Lett.*, **46**, 2922–2930, https://doi.org/10.1029/2018GL081724.

——, D. Gold, J. T. Allen, and B. S. Barrett, 2019: Extended U.S. tornado outbreak during late May 2019: A forecast of opportunity. *Geophys. Res. Lett.*, **46**, 10 150–10 158, https://doi.org/10.1029/2019GL084470.

——, B. S. Barrett, J. T. Allen, D. Gold, and P. Sirvatka, 2020a: The Extended-Range Tornado Activity Forecast (ERTAF) project. *Bull. Amer. Meteor. Soc.*, **101**, E700–E709, https://doi.org/10.1175/BAMS-D-19-0188.1.

——, A. M. Haberlie, and P. T. Marsh, 2020b: Practically perfect hindcasts of severe convective storms. *Bull. Amer. Meteor. Soc.*, **101**, E1259–E1278, https://doi.org/10.1175/BAMS-D-19-0321.1.

Grams, J. S., R. L. Thompson, D. V. Snively, J. A. Prentice, G. M. Hodges, and L. J. Reames, 2012: A climatology and comparison of parameters for significant tornado events in the United States. *Wea. Forecasting*, **27**, 106–123, https://doi.org/10.1175/WAF-D-11-00008.1.

Guan, H., and Coauthors, 2022: GEFSV12 reforecast dataset for supporting subseasonal and hydrometeorological applications. *Mon. Wea. Rev.*, **150**, 647–665, https://doi.org/10.1175/MWR-D-21-0245.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hitchens, N. M., and H. E. Brooks, 2012: Evaluation of the Storm Prediction Center's day 1 convective outlooks. *Wea. Forecasting*, **27**, 1580–1585. https://doi.org/10.1175/WAF-D-12-00061.1

——, and ——, 2014: Evaluation of the Storm Prediction Center's convective outlooks from day 3 through day 1. *Wea. Forecasting*, **29**, 1134–1142, https://doi.org/10.1175/WAF-D-13-00132.1.

——, ——, and M. P. Kay, 2013: Objective limits on forecasting skill of rare events. *Wea. Forecasting*, **28**, 525–534, https://doi.org/10.1175/WAF-D-12-00113.1.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, *Geophys. Monogr.*, Vol. 134, Amer. Geophys. Union, 1–35, https://doi.org/10.1029/134GM01.

Johns, R. H., and C. A. Doswell, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612, https://doi.org/10.1175/1520-0434(1992)007<0588:SLSF>2.0.CO;2.

Lee, S.-K., A. T. Wittenberg, D. B. Enfield, S. J. Weaver, C. Wang, and R. Atlas, 2016: U.S. regional tornado outbreaks and their links to spring ENSO phases and North Atlantic SST variability. *Environ. Res. Lett.*, **11**, 044008, https://doi.org/10.1088/1748-9326/11/4/044008.

Lepore, C., M. K. Tippett, and J. T. Allen, 2018: CFSV2 monthly forecasts of tornado and hail activity. *Wea. Forecasting*, **33**, 1283–1297, https://doi.org/10.1175/WAF-D-18-0054.1.

Maddison, J., S. Gray, O. Martínez-Alvarado, and K. Williams, 2019: Upstream cyclone influence on the predictability of block onsets over the Euro-Atlantic region. *Mon. Wea. Rev.*, **147**, 1277–1296, https://doi.org/10.1175/MWR-D-18-0226.1.

Maddox, R. A., 1976: An evaluation of tornado proximity wind and stability data. *Mon. Wea. Rev.*, **104**, 133–142, https://doi.org/10.1175/1520-0493(1976)104<0133:AEOTPW>2.0.CO;2.

Matsueda, M., 2011: Predictability of Euro-Russian blocking in summer of 2010. *Geophys. Res. Lett.*, **38**, L06801, https://doi.org/10.1029/2010GL046557.

Merryfield, W. J., and Coauthors, 2020: Current and emerging developments in subseasonal to decadal prediction. *Bull. Amer. Meteor. Soc.*, **101**, E869–E896, https://doi.org/10.1175/BAMS-D-19-0037.1.

Miller, D. E., and Z. Wang, 2019a: Assessing seasonal predictability sources and windows of high predictability in the Climate Forecast System, version 2. *J. Climate*, **32**, 1307–1326, https://doi.org/10.1175/JCLI-D-18-0389.1.

——, and ——, 2019b: Skillful seasonal prediction of Eurasian winter blocking and extreme temperature frequency. *Geophys. Res. Lett.*, **46**, 11 530–11 538, https://doi.org/10.1029/2019GL085035.

——, and ——, 2022: Northern Hemisphere winter blocking: Differing onset mechanisms across regions. *J. Atmos. Sci.*, **79**, 1291–1309, https://doi.org/10.1175/JAS-D-21-0104.1.

——, ——, R. J. Trapp, and D. S. Harnos, 2020: Hybrid prediction of weekly tornado activity out to week 3: Utilizing weather regimes. *Geophys. Res. Lett.*, **47**, e2020GL087253, https://doi.org/10.1029/2020GL087253.

——, ——, B. Li, D. S. Harnos, and T. Ford, 2021: Skillful subseasonal prediction of U.S. extreme warm days and standardized precipitation index in boreal summer. *J. Climate*, **34**, 5887–5898, https://doi.org/10.1175/JCLI-D-20-0878.1.

——, V. A. Gensini, and B. S. Barrett, 2022: Madden–Julian Oscillation influences United States springtime tornado and hail frequency. *npj Climate Atmos. Sci.*, **5**, 37, https://doi.org/10.1038/s41612-022-00263-5.

Moore, T. W., 2018: Annual and seasonal tornado activity in the United States and the global wind oscillation. *Climate Dyn.*, **50**, 4323–4334, https://doi.org/10.1007/s00382-017-3877-5.

——, and M. P. McGuire, 2020: Tornado-days in the United States by phase of the Madden–Julian Oscillation and global wind oscillation. *Climate Dyn.*, **54**, 17–36, https://doi.org/10.1007/s00382-019-04983-y.

Morrison, H., and Coauthors, 2020: Confronting the challenge of modeling cloud and precipitation microphysics. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001689, https://doi.org/10.1029/2019MS001689.

Nageswararao, M., Y. Zhu, and V. Tallapragada, 2022: Prediction skill of GEFSV12 for southwest summer monsoon rainfall and associated extreme rainfall events on extended range scale over India. *Wea. Forecasting*, **37**, 1135–1156, https://doi.org/10.1175/WAF-D-21-0184.1.

NWS/CPC, 2022: Pentad MJO indices. Accessed 1 January 2022, https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_mjo_index/pentad.html.

Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100**, 2043–2060, https://doi.org/10.1175/BAMS-D-18-0270.1.

Pohl, B., and A. J. Matthews, 2007: Observed changes in the lifetime and amplitude of the Madden–Julian Oscillation associated with interannual ENSO sea surface temperature anomalies. *J. Climate*, **20**, 2659–2674, https://doi.org/10.1175/JCLI4230.1.

Rex, D. F., 1950: Blocking action in the middle troposphere and its effect upon regional climate. *Tellus*, **2**, 275–301, https://doi.org/10.3402/tellusa.v2i4.8603.

Reynolds, D. D., A. R. Lupo, A. D. Jensen, and P. S. Market, 2017: The predictability of Northern Hemispheric blocking using an ensemble mean forecast system. *Proceedings*, **1**, 87, https://doi.org/10.3390/ecas2017-04128.

Rodwell, M. J., and Coauthors, 2013: Characteristics of occasional poor medium-range weather forecasts for Europe. *Bull. Amer. Meteor. Soc.*, **94**, 1393–1405, https://doi.org/10.1175/BAMS-D-12-00099.1.

Saminathan, S., H. Medina, S. Mitra, and D. Tian, 2021: Improving short to medium range GEFS precipitation forecast in India. *J. Hydrol.*, **598**, 126431, https://doi.org/10.1016/j.jhydrol.2021.126431.

Sillmann, J., and M. Croci-Maspoli, 2009: Present and future atmospheric blocking and its impact on European mean and extreme climate. *Geophys. Res. Lett.*, **36**, L10702, https://doi.org/10.1029/2009GL038259.

Takaya, K., and H. Nakamura, 2001: A formulation of a phase-independent wave-activity flux for stationary and migratory quasigeostrophic eddies on a zonally varying basic flow. *J. Atmos. Sci.*, **58**, 608–627, https://doi.org/10.1175/1520-0469(2001)058<0608:AFOAPI>2.0.CO;2.

Thompson, D. B., and P. E. Roundy, 2013: The relationship between the Madden–Julian Oscillation and U.S. violent tornado outbreaks in the spring. *Mon. Wea. Rev.*, **141**, 2087–2095, https://doi.org/10.1175/MWR-D-12-00173.1.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261, https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

Tibaldi, S., and F. Molteni, 1990: On the operational predictability of blocking. *Tellus*, **42**, 343–365, https://doi.org/10.3402/tellusa.v42i3.11882.

Trapp, R. J., and K. A. Hoogewind, 2018: Exploring a possible connection between U.S. tornado activity and arctic sea ice. *npj Climate Atmos. Sci.*, **1**, 14, https://doi.org/10.1038/s41612-018-0025-9.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Vol. 100, Academic Press, 704 pp.

Wurman, J., D. Dowell, Y. Richardson, P. Markowski, E. Rasmussen, D. Burgess, L. Wicker, and H. B. Bluestein, 2012: The second Verification of the Origins of Rotation in Tornadoes Experiment: VORTEX2. *Bull. Amer. Meteor. Soc.*, **93**, 1147–1170, https://doi.org/10.1175/BAMS-D-11-00010.1.

day 10) for spring tornado frequency. Essentially, some GEFS forecasts were very skillful at day 10, whereas others had negative skill. This suggests that a model evaluation of high- and low-skill forecasts may lead to identifying underlying regimes that contribute to enhanced or reduced predictability, specifically for extreme events at extended lead times. Several recent studies have followed a similar approach. For example, the representation of upstream surface cyclones is important in correctly predicting atmospheric blocking events within the European Centre for Medium-Range Weather Forecast (ECMWF) Ensemble Prediction System (EPS; Maddison et al. 2019). An ECMWF day-6 forecast was also evaluated and showcased Europe forecast "bust" events to be characteristic of a Rex-type blocking event (Rodwell et al. 2013). The authors found initial conditions for the bust events contained significant troughing over the U.S. Rocky Mountains and large values of convective available potential energy throughout the Great Plains. In addition, statistically significant higher winter average geopotential height prediction skill for the Climate Forecast System, version 2 (CFSv2), was evident during the negative phase of the North Atlantic Oscillation (NAO; Hurrell et al. 2003) under El Niño conditions, whereas larger prediction skill was evident during positive NAO under La Niña conditions (Miller and Wang 2019a).

This type of model evaluation can provide operational forecasters with higher or lower confidence in model solutions if certain sources of predictability or global weather regimes are present at the time of model initialization. Moreover, it may reveal model biases or sources of error that could prove useful to developers. Societally, losses caused by tornadoes can be devastating, and anticipating these events at extended lead times allows for increased awareness and potential for mitigation. Thus, day-10 GEFS, version 12 (GEFSv12), 500-hPa geopotential height (Z500) forecasts over the United States are investigated in tandem with the occurrence of practically perfect tornado day hindcasts. Initial model conditions were then analyzed and composited for high- and low-skill day-10 forecasts. Results were then put into context related to potential sources of forcing and variability that contribute to high- and low-skill extended-range tornado forecasts.

2. Data and methods

a. Data

Daily-averaged (from the native hourly interval) data from ECMWF's ERA5 reanalysis (Hersbach et al. 2020) during boreal spring over the period 2000–19 were downloaded and used as the verifying field for the day-10 GEFSv12 Z500 forecast. Initial conditions, as well as the evolutions, were investigated from ERA5 by examining Z500, outgoing longwave radiation (OLR), and sea surface temperature (SST). Geopotential, zonal wind, and meridional wind at 200 hPa were used for the calculation of wave-activity flux (WaF). Zonal wind at each vertical level was used for the calculation of Earth-relative atmospheric angular momentum (AAM) following the methods outlined in Miller et al. (2022), which included constraining the zonal integral to the Northern Hemisphere.

Model evaluation was conducted using GEFS version 12 (GEFSv12; Guan et al. 2022) reforecasts over the period 2000–19 during boreal spring [March–May (MAM)]. GEFSv12 reforecasts, containing one control and four perturbation runs, were initialized once per day from 0000 UTC initial conditions and integrated to a forecast lead time of 16 days. An additional six-perturbation cycle integrated to day 35 is also provided once weekly on Wednesdays. GEFSv12 has recently been utilized to analyze the performance in precipitation forecasts in India (Saminathan et al. 2021; Nageswararao et al. 2022) and additional hydrometeorological applications (Guan et al. 2022). GEFSv12 reforecasts have also been shown to contain smaller central U.S. 2-m temperature biases throughout the year, and overall higher skill in predicting Z500 than its Subseasonal Experiment (SubX; Pegion et al. 2019) reforecast counterpart (Guan et al. 2022). Specifically, Z500 over the United States (63°–124°W, 24°–50°N) was the forecast verification field evaluated in section 3, and our focus was on the day-10 lead window for a 1200–1200 UTC convective day (forecast hours 252–276 for a 0000 UTC cycle).

Tornado practically perfect hindcasts (PPH; Hitchens et al. 2013; Gensini et al. 2020b) were used to identify tornado days that occurred on high- or low-skill day-10 Z500 forecasts. Tornado PPHs represent a Gaussian-smoothed statistical point process of event probabilities with the goal of resembling what a perfect SPC convective outlook may look like. Days on which the U.S. maximum tornado PPH probability exceeded the March–June 2000–19 75th percentile (17.27%) were examined. This corresponds to a tornado PPH probability in between an "enhanced" and "moderate" categorical threshold by the SPC, which are associated with 10% and 30% tornado PPH probabilities, respectively. Hereinafter, these 75th-percentile tornado PPH days are referred to as tornado days for brevity.

b. Teleconnection indices, atmospheric blocking, and wave activity flux

Several weather/climate teleconnection indices were explored to identify any reoccurring and potential significant signals that may be present during high- and low-skill tornado day forecasts. The Madden–Julian oscillation (MJO; Baggett et al. 2018; Gensini et al. 2019, 2020a) has recently been utilized as a source of subseasonal-to-seasonal (S2S) predictability for SCSs, and recent work discovered increased probabilities of SCSs 3–4 weeks following the strongest MJO convection signal (Miller et al. 2022). CPC pentad MJO indices (NWS/CPC 2022; Baxter et al. 2014) were used to represent the MJO (as in Miller et al. 2022), which represent the negative projection of 200-hPa velocity potential (χ200) anomalies onto the 10 time-lagged patterns of the first extended empirical orthogonal function of pentad χ200 anomalies.

Indices describing the modulation of the Arctic Oscillation (AO), the North Atlantic Oscillation (NAO), the El Niño–Southern Oscillation (ENSO), Pacific decadal oscillation (PDO), the Pacific–North American (PNA) pattern, and the quasi-biennial oscillation (QBO) were all investigated to examine if recurrent phases of teleconnections were present during model cycle initialization and day-10 tornado forecasts. All indices

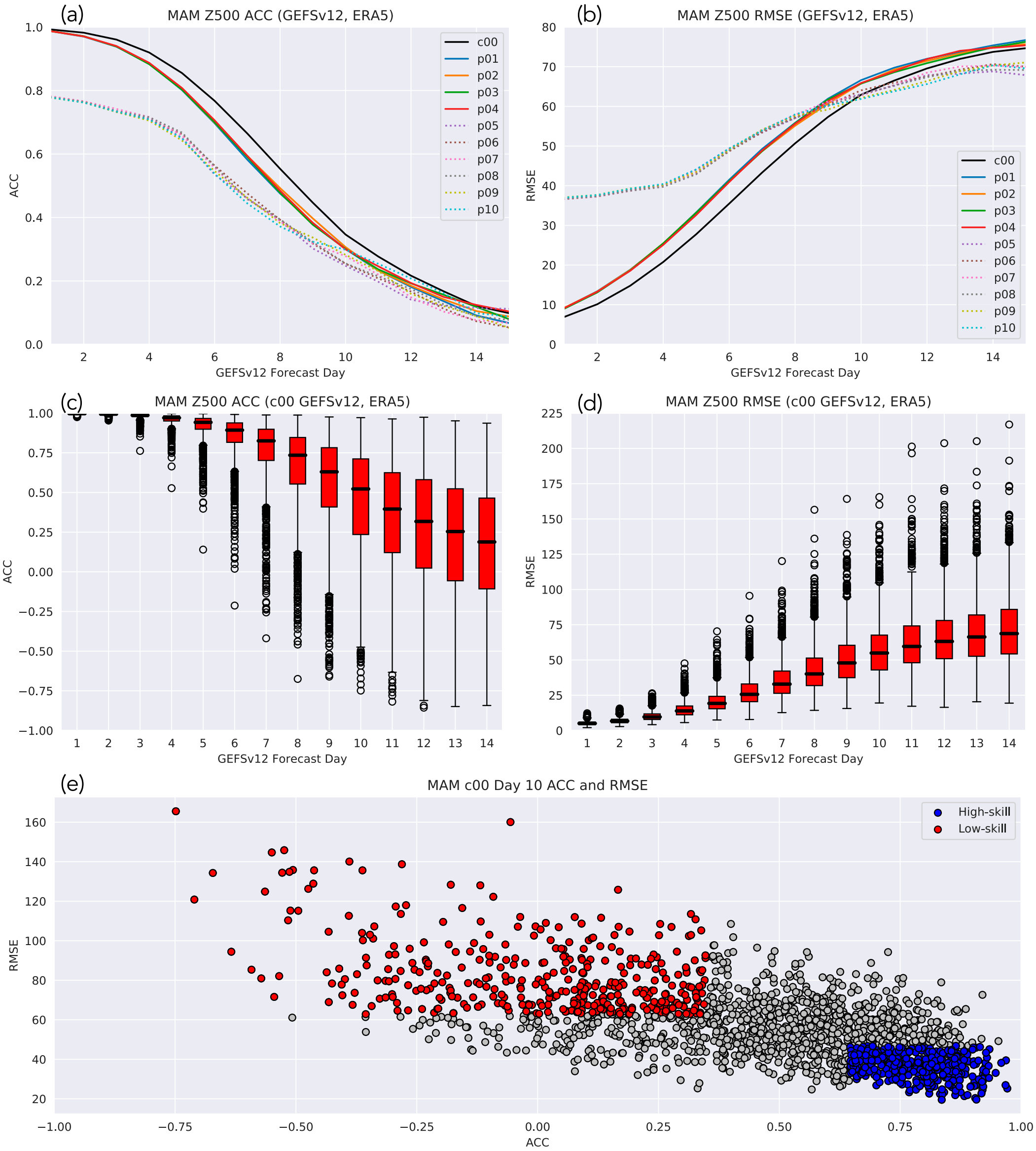

FIG. 1. The (a) anomaly correlation coefficient (ACC) and (b) root-mean-square error (RMSE) over the United States (63°–124°W, 24°–50°N) between the GEFSv12 and ERA5 Z500 fields from day 1 to day 15 for the control run and 10 perturbation members. Box-and-whisker plots for the control run (c) ACC and (d) RMSE for all MAM Z500 forecasts. (e) Scatterplot showing the ACC vs RMSE for all control member MAM Z500 forecasts. Red and blue circles represent the low- and high-skill forecasts, respectively.

were obtained and are available via the NOAA Physical Sciences Laboratory (PSL; https://psl.noaa.gov/data/climateindices/list/).

In addition to teleconnections, two physical processes of interest on these time scales are atmospheric blocking and wave propagation. Atmospheric blocking is characterized as a quasi-stationary, high pressure system that acts to hamper the normal progression of the westerly jet and may induce extreme weather conditions (Sillmann and Croci-Maspoli 2009; Buehler et al. 2011; Dole et al. 2011; Miller et al. 2020, 2021; Miller and Wang 2022, 2019b). A two-dimensional blocking index was calculated following the methods of Miller et al. (2021), which identifies Z500 anomalies that exceed 1.5σ, have an area greater than 10° × 10°, and persist for at least 5 days. Blocking indices were calculated for day 0 (model initialization) and day 10.

Observed wave propagation for both high- and low-skill forecast groups were examined by calculating the WaF ($m^2\ s^{-2}$) at 200 hPa following Takaya and Nakamura (2001). WaF indicates a snapshot of the propagation direction of stationary or migrating waves and is nearly parallel to the group velocity. The divergence/convergence of WaF indicates regions where the wave packet is emitted/absorbed.

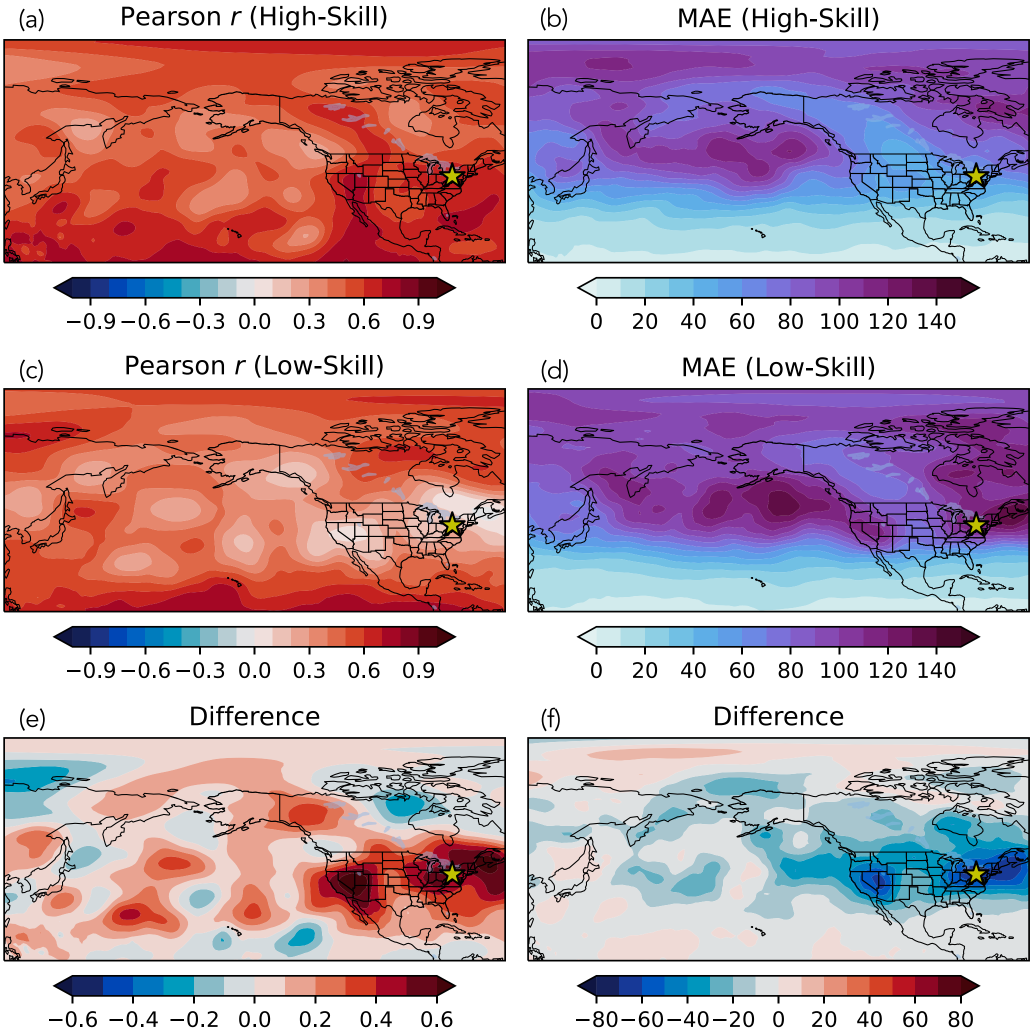

FIG. 2. Two-dimensional (a),(c) Pearson *r* and (b),(d) mean absolute error between the GEFSv12 day-10 forecasts and ERA5 for (top) high-skill and (middle) low-skill forecasts. Also shown is the difference between the high- and low-skill (e) ACC and (f) MAE. The yellow star indicates the location evaluated in Fig. 3, below.

c. *Defining windows of high and low forecast skill*

Anomaly correlation coefficient (ACC; Wilks 2011) and root-mean-square error (RMSE; Wilks 2011) were first calculated for the Z500 anomaly field over the United States between the ERA5 and GEFSv12 forecasts from days 1 to 15 for each ensemble member (Figs. 1a,b). Anomalies were created by removing the 2000–19 daily averages for both the ERA5 and GEFSv12. High- and low-skill days were chosen using the control member given initial exploratory findings that control forecasts had consistently higher skill than the perturbation members, which is similar to previous work (Rodwell et al. 2013). It is important to note that the results were similar when examining the ensemble mean, with composites being more smoothed out for model conditions. A day-10 control member forecast with an ACC ≥ 66th percentile and a RMSE ≤ 33rd percentile was labeled as a high-skill forecast. A low-skill forecast day contained an ACC ≤ 33rd percentile and a RMSE ≥ 66th percentile. Together, these ensure that a high-skill forecast is associated with relatively low errors related to magnitude (RMSE) and pattern location (ACC). A total of 88 high-skill and 91 low-skill forecasts were identified over the study period.

d. *Evaluation metrics and statistical significance*

Two-dimensional fields of Pearson *r* correlations and mean absolute error (MAE) for the day-10 forecasts were calculated for both forecast groups (Fig. 2). Calculations were made at each grid point between the ERA5 and GEFSv12 day-10 forecasts of Z500 anomalies. Both datasets were bilinearly interpolated to a common latitude/longitude grid (1° × 1°) for the calculations of ACC and RMSE.

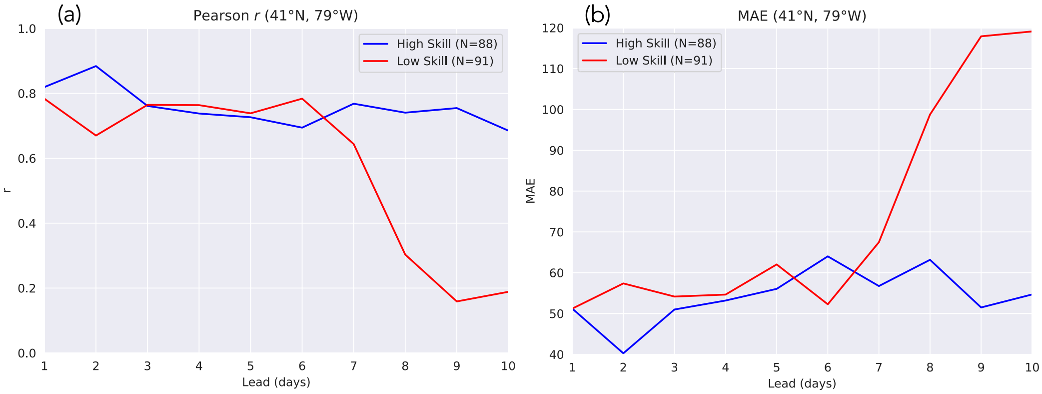

FIG. 3. The average (a) Pearson r and (b) MAE evolution from day 1 to day 10 at 41°N, 79°W for the high-skill (blue) and low-skill (red) groups. The sample size for each group is located in the legend.

Anomalies of Z500, OLR, SST, blocking frequency, and several teleconnection indices were constructed by removing the 2000–19 daily averages (i.e., seasonal cycle) to examine the initial state for the high- and low-skill groups. Statistical significance of the anomalies were tested using a two-tailed Student's t test, with the null hypothesis that the anomalies do not differ from zero. The statistical significance of blocking frequency anomalies was performed using a Mann–Whitney U test for the medians. Results were considered to be significant if the p value was less than 0.05 (95% confidence level). Field significance was performed using the false discovery rate (FDR; Wilks 2011) of $\alpha = 0.1$ for composite anomalies of several fields throughout the study.

3. Evaluation of GEFSv12 Z500 over the United States

The control run showed the highest average ACC through day 14 while also containing the lowest average RMSE through day 10 in comparison with the perturbation members (Figs. 1a,b). Given this, the rest of the focus herein is on the control member. Note that we are not examining the predictability of the ensemble system. Rather, we are trying to understand the initial conditions associated with individual forecasts, which by nature are individual members. The control member ACC and RMSE spread was small at shorter lead times, and increased the largest at day 6 (Figs. 1c,d). For example, the control run at day 10 contained spread of near perfect ACC/RMSE to outliers of ACC equal to -0.75 (opposite anomaly pattern from what was forecast) and RMSE > 150 m. The median ACC experienced the largest decreases from day 9 to 10. As stated previously, focus here is on the day-10 forecasts motivated by Gensini and Tippett (2019), where the average lead time of zero skill for all forecasts was day 10. The high-skill forecasts contain ACC and RMSE values greater than the 66th percentile (0.65) and less than the 33rd percentile values (46.81 m), respectively (Fig. 1e). In contrast, a low-skill forecast contains an ACC ≤ 0.35 and RMSE values ≥ 62.88 m. In total, 88 high-skill and 91 low-skill tornado events occurring at day 10 were found.

High-skill day-10 forecasts exhibited r values greater than 0.7 over the East and West Coast of the United States (Fig. 2a), which contrasted lower values (0–0.3) for low-skill forecasts (Fig. 2c). Major differences in ACC between the two groups were found over the United States, with values greater than 0.6 higher for high-skill forecasts (Fig. 2e). Other areas of higher Pearson r were located over the Pacific Ocean, northwest Canada, and Newfoundland. The largest MAE values for both groups were located over the Pacific Ocean (Figs. 2b,d), with the largest group differences over the U.S. East and West Coast, similar to the pattern found in Pearson r (Fig. 2f).

Lead-day averaged r and MAE values were also examined at a point over the eastern United States (41°N, 79°W; yellow dot in Fig. 2), which is collocated with the area of largest changes in metrics between the high- and low-skill groups. High-skill forecasts contained nearly constant r values between 0.7 and 0.8 from days 1 to 10 (Fig. 3a), while the MAE remained between 40 and 65 m (Fig. 3b). Low-skill forecasts had similar correlations and MAE values through day 6, followed by a significant decrease of Pearson r and increase of MAE through day 10. In essence, NWP evolution of the Z500 field was, on average, poor beyond day 6 for these forecasts.

4. Comparison between skillful and poorly predicted day-10 tornado days

a. *Observed characteristics of high- and low-skill forecasts*

Examining Z500 forecast evolution, locally significant positive Z500 anomalies were evident over the Alaska–Canadian border at day 1 for the high-skill groups (Fig. 4a). The observed Z500 for low-skill day-1 forecasts contained significant positive anomalies farther west relative to that of the high-skill forecasts (Fig. 4c), and it appears that these positive anomalies were the greatest differences among the two groups (Fig. 4e). Unsurprisingly, the GEFSv12 day-1 forecast (Figs. 4a,c, contours) for both groups closely resembles the ERA5. However, minor differences associated with the Pacific wave train were already occurring at day 1 for the low-skill forecasts (Fig. 4c). Observed day-10 patterns over the United States are relatively similar between both groups (Figs. 4b,d), showcasing a midlevel western U.S. thermal trough and eastern

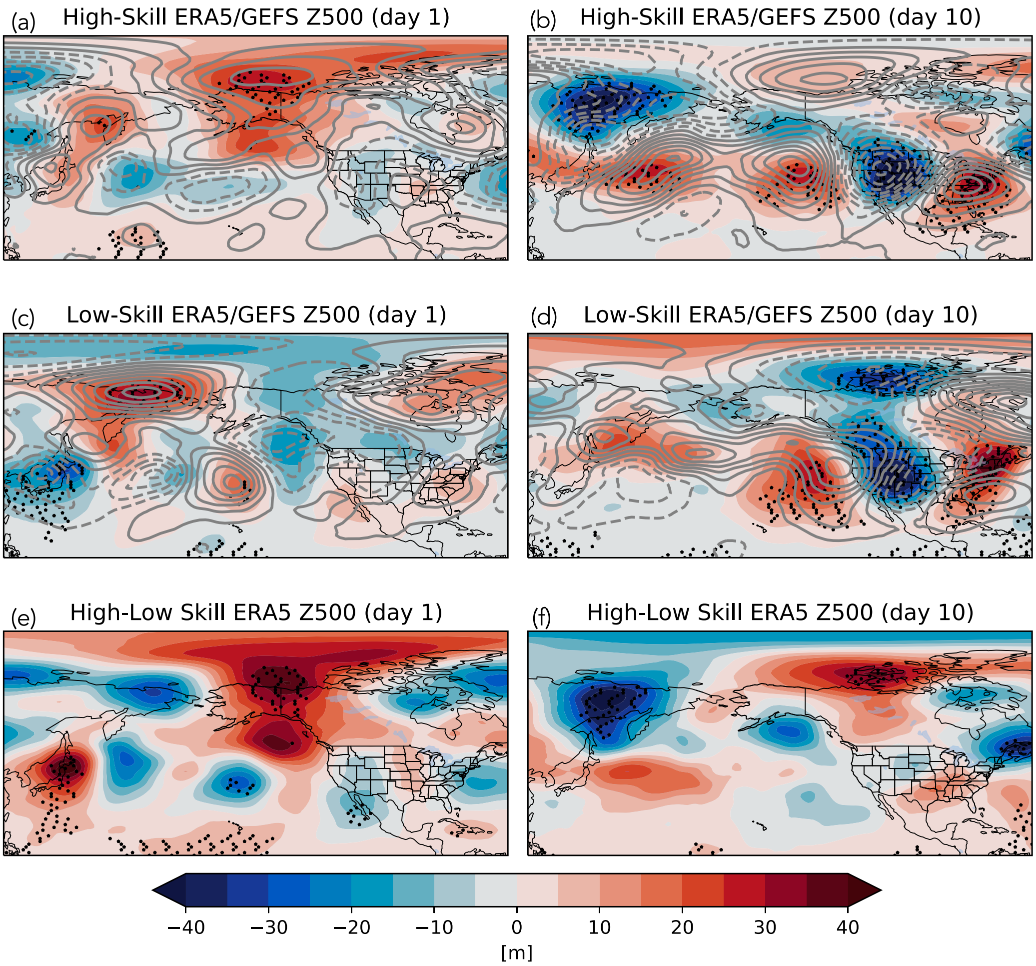

FIG. 4. ERA5 (shading) and GEFSv12 (contours) Z500 anomalies for (a),(c) day 1 and (b),(d) day 10 for the (top) high-skill group and (middle) low-skill group. Black dots indicate ERA5 anomalies statistically significant at the 95% confidence interval. Also shown are the difference between the high- and low-skill ERA5 Z500 anomalies for days (e) 1 and (f) 10. Black stippling indicates where anomalies are locally statistically significant at the 95% confidence level. No grid points were field significant.

U.S. thermal ridge synoptic pattern conducive for severe convective storms (Miller et al. 2020). The largest differences were associated with a significant north–south dipole in the day-10 Z500 pattern located over eastern Russia for the high-skill forecasts (Figs. 4b,f). Day-10 predicted Z500 field for the high-skill group closely resembles that of the ERA5. Not only is the U.S. trough west–east ridge pattern replicated, but the western Pacific north–south dipole is also evident (contours in Fig. 4b). GEFSv12 forecasts for the low-skill group produced a western U.S. midlevel thermal ridge, central U.S. thermal troughing, and southeast U.S. thermal ridging, which was out of phase relative to the ERA5 (Fig. 4d), leading to low Pearson r values on the U.S. West and East Coasts (Fig. 2c). A closer look at the evolution of the low-skill day-10 forecasts (Fig. 5) revealed that the GEFSv12 was too amplified with eastern Russia positive Z500 anomalies, especially at day 6 (Fig. 5c). By day 8 (Fig. 5d), the wave train was also out of phase, leading to rapidly decreasing and increasing ACC and RMSE, respectively. In fact, positive Z500 anomalies over eastern Russia at initialization were associated with a blocking anticyclone.

Atmospheric blocking is a significant producer of extreme weather (Sillmann and Croci-Maspoli 2009; Buehler et al. 2011; Dole et al. 2011; Miller and Wang 2019b; Miller et al. 2020; Miller and Wang 2022), and operational models have difficulty forecasting the initiation, duration, and cessation of blocking events (Tibaldi and Molteni 1990; Matsueda 2011; Reynolds et al. 2017). A two-dimensional blocking index was calculated and examined for both the high- and low-skill groups. Field significant positive anomalies of blocking frequency were located over eastern Russia at day 0 for the low-skill forecasts, hinting at a potential problem for the evolution of the flow (Fig. 6c). Although blocking onset mechanisms have been extensively studied—showing to be complex and varying

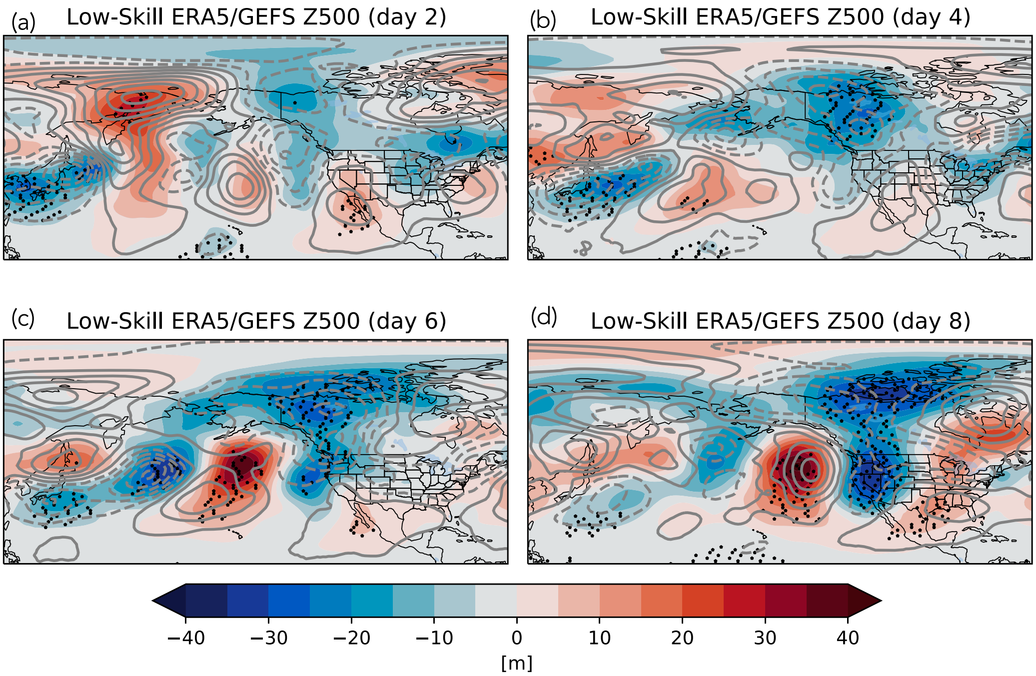

FIG. 5. As in Figs. 4a–d, but for the low-skill group at (a) day 2, (b) day 4, (c) day 6, and (d) day 8. Black stippling indicates where anomalies are locally statistically significant at the 95% confidence level.

by region (Miller and Wang 2022)—dynamics involved in blocking maintenance and cessation remain challenging. Owing to the unsettled theories of blocking decay, operational models may fail to predict the duration of the blocking life cycle (Reynolds et al. 2017). Since blocking can significantly influence the downstream evolution of the synoptic flow (Rex 1950; Miller and Wang 2022), an overforecast blocking duration, as we see here, can cause significant error growth in the model solution.

Differences in the WaF propagation was also found to be a potential cause for poor model performance of the low-skill groups. The wave propagation pathway from the ERA5 was different between the high- and low-skill groups, specifically at days 6 and 8 (Fig. 7). WaF vectors for the high-skill forecasts evolved from the central Pacific to the northwest United States, showcasing a significant meridional component at day 6 (Fig. 7a). The convergence intensified over the U.S. West Coast,

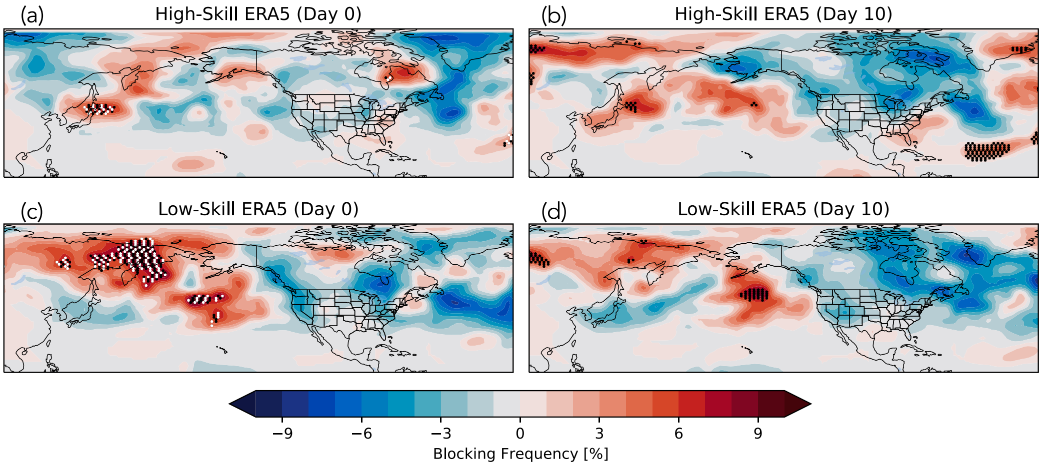

FIG. 6. ERA5 blocking frequency anomalies (% of days) for (a),(c) day-0 and (b),(d) day-10 (top) high-skill forecasts and (bottom) low-skill forecasts. Black stippling indicates where anomalies are locally statistically significant at the 95% confidence level, whereas white stippling indicates anomalies that are field significant at the FDR level of $\alpha = 0.1$.

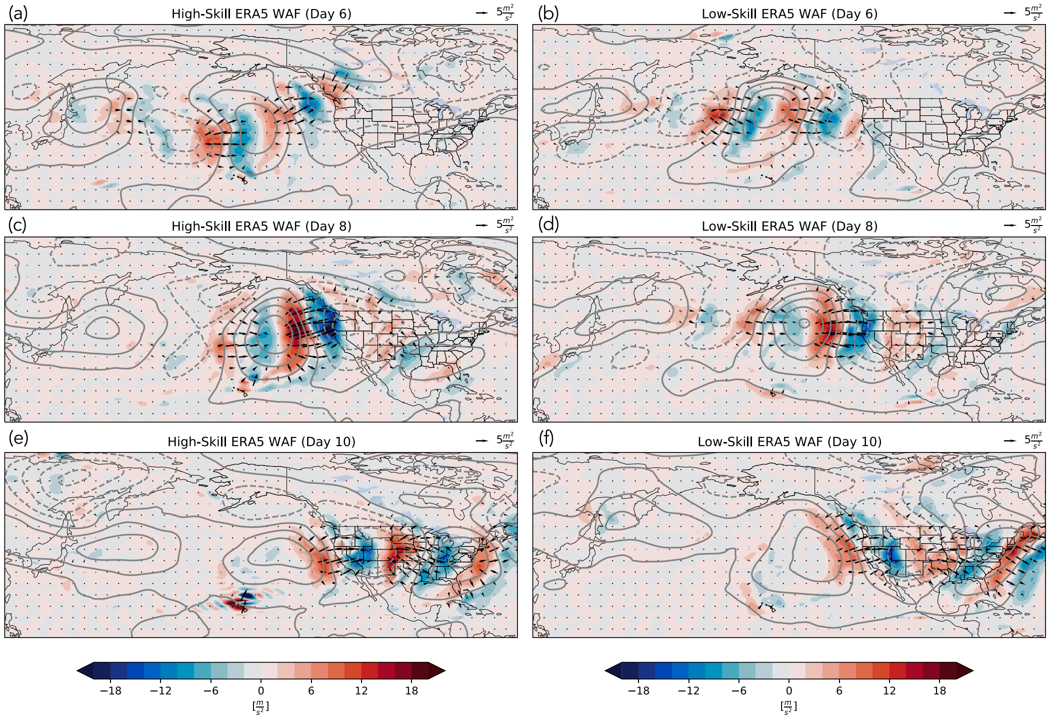

FIG. 7. ERA5 wave activity flux (vectors; $m^2 s^{-2}$) and convergence (shading; $m s^{-2}$) for days (a),(b) 6; (c),(d) 8; and (e),(f) 10 of the (left) high-skill and (right) low-skill forecasts. The scale for WaF vectors is shown in the top-right corners of the panels.

working to amplify the thermal trough over the western United States at day 8 (Fig. 7c). The GEFSv12 reproduced the meridional component of the WaF vectors with similar convergence off the U.S. West Coast in the high-skill forecasts (Figs. 8a,c,e). In contrast with the high-skill forecasts, the low-skill evolution showed a zonal propagation with almost no meridional influence at days 6 and 8 (Figs. 7b,d), and the WaF convergence over the western United States was weaker than the observed WaF during high-skill forecasts. GEFSv12 WaF at day 6 showed a waveguide extending from Japan to the west of Hawaii, with no zonal propagation to the United States (Fig. 8b), contrary to what was observed in the ERA5 (Fig. 7b). We suspect the overpredicted blocking anticyclone was preventing the zonal waveguide from occurring and evolving in GEFSv12.

Several sources of subseasonal-to-seasonal variability (Merryfield et al. 2020) have been shown to modulate severe convective storm occurrence, including the MJO (Thompson and Roundy 2013; Barrett and Gensini 2013; Baggett et al. 2018; Gensini et al. 2020a; Moore and McGuire 2020; Miller et al. 2022), the ENSO (Allen et al. 2015), the AAM budget (Gensini and Marinaro 2016; Moore 2018; Gensini et al. 2020a), land surface (Frye and Mote 2010), and sea ice (Trapp and Hoogewind 2018). To examine this variability, OLR and SST states at initialization were analyzed for both skill groups. Near opposite OLR patterns were present, especially over the western tropical Pacific Ocean (Figs. 9a,c,e). Locally statistically significant negative OLR anomalies were present over the "Maritime Continent" along with positive anomalies over the central and eastern tropical Pacific for the high-skill forecasts, whereas the low-skill group exhibited positive anomalies over the Maritime Continent (negative OLR anomalies represent increased cloudiness/precipitation). The negative OLR anomalies were associated with the MJO (section 4b).

As mentioned, a stark contrast in SST was apparent between the high- and low-skill initialization states. Skillful forecasts contained field significant warmer temperatures in the central tropical Pacific Ocean, off the west coast of the United States, and the Gulf of Mexico (Fig. 9b). In contrast, SSTs during initialization for the low-skill forecasts contained La Niña–like conditions with field significant (FDR value < 0.1) colder SSTs than the seasonal cycle throughout the tropical Pacific Ocean and off the west coast of the United States and Canada (Fig. 9d). These spring conditions are known to be associated with increased probabilities of severe weather (Lee et al. 2016), yet the GEFSv12 fails to predict the pattern evolution accurately. Differences between the two skill groups (Fig. 9f) highlight the opposite patterns with field significant differences occurring throughout the domain.

b. *The MJO and climate indices*

The MJO is a primary source of subseasonal predictability and has recently been exploited for its relation to U.S. severe convective weather (Gensini et al. 2019; Baggett et al. 2018; Miller et al. 2022). The CPC pentad MJO indices and AAM for 20 days before/after the day of GEFSv12 initialization for

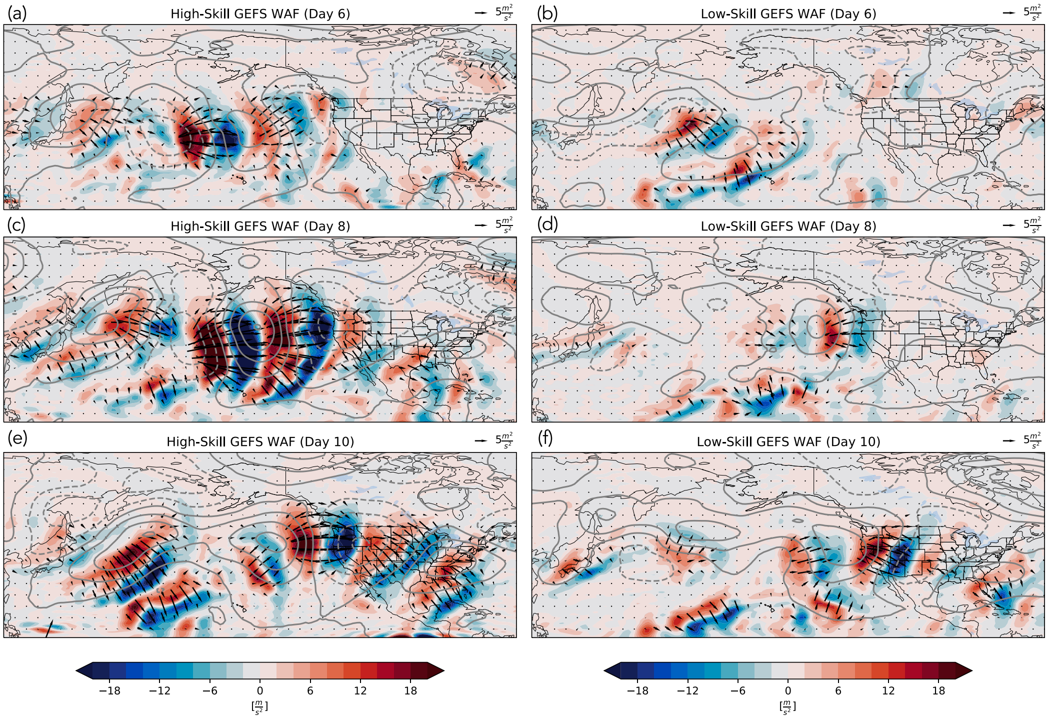

FIG. 8. As in Fig. 7, but for the GEFSv12.

both forecast groups were examined (Fig. 10). Major differences were evident, with the skillful forecasts being initialized ~20 days following a strong active MJO which appears to propagate, on average, rather quickly (Fig. 10a). A significant increase in AAM is apparent between days −15 and −5 (Fig. 10b), followed by a minimum around the time of peak severe weather frequency (day 10). The noted increases in AAM following the MJO were reminiscent of the modulation

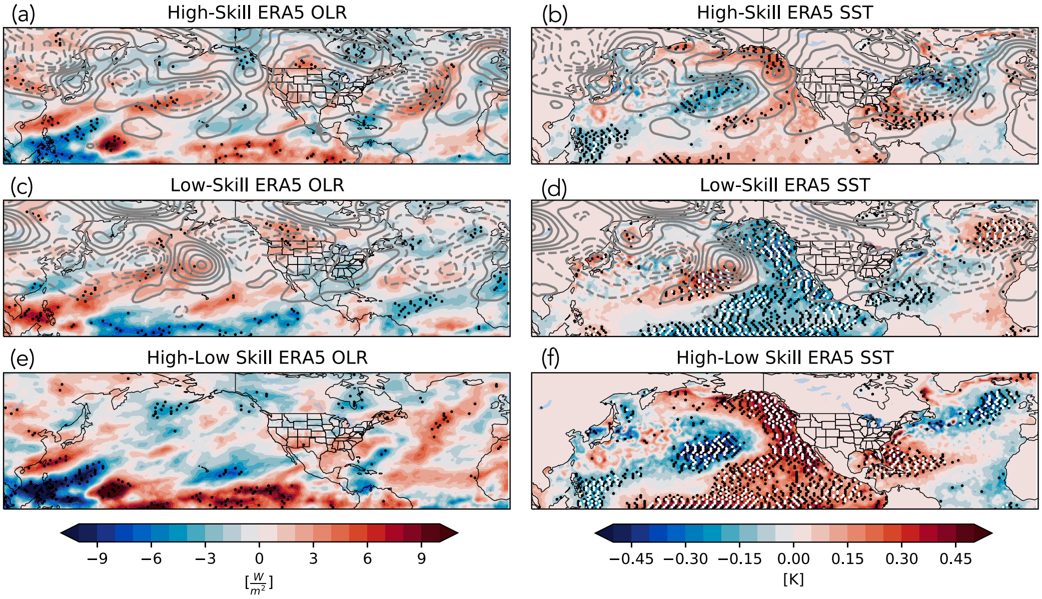

FIG. 9. ERA5 (a),(c) outgoing longwave radiation (OLR; W m^{-2}) anomalies and (b),(d) sea surface temperature (SST; K) anomalies for the (top) high-skill group and (bottom) low-skill group. Gray contours represent Z500 anomalies (m). Also shown is the difference of (e) OLR and (f) SST between the two skill groups. Black stippling indicates where anomalies are locally statistically significant at the 95% confidence level, whereas white stippling indicates anomalies that are field significant at the FDR level of $\alpha = 0.1$.

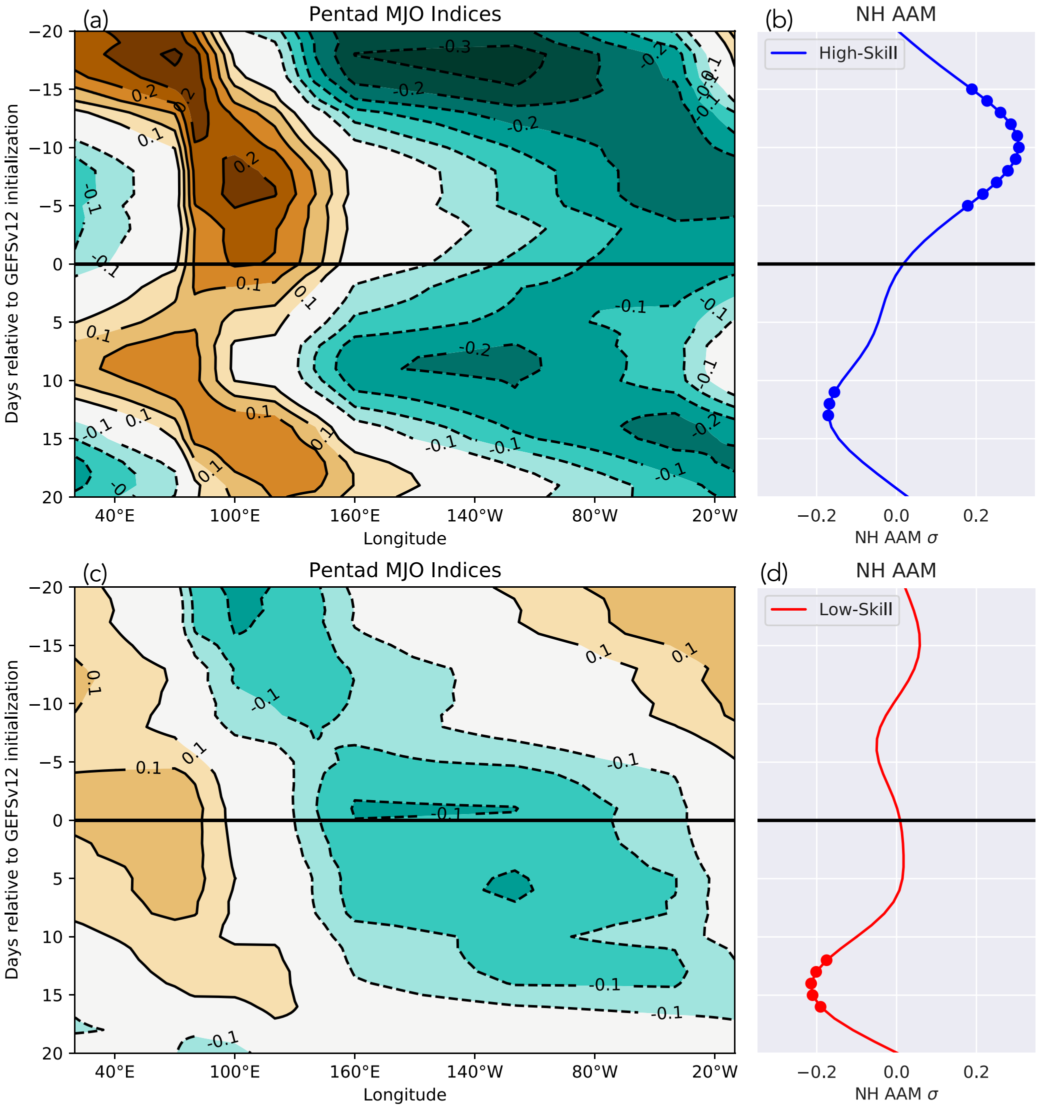

FIG. 10. (a),(c) Observed CPC pentad MJO indices and (b),(d) Northern Hemisphere AAM for 20 days prior to 20 days after the date of initialization for the (top) high-skill forecast group and (bottom) low-skill group. Green shading represents "active convection." Colored dots in (b) and (d) indicate statistically significant AAM anomalies at the 95% confidence level.

following the minimum pentad index discussed in Miller et al. (2022). Low-skill forecasts had a slower propagating, weaker MJO signal, and despite the significant decrease ~12 days following model cycle initialization, there was no apparent preceding peak in AAM (Figs. 10c,d). These results were consistent with the background state of SST (Fig. 9), where the MJO was found to propagate faster during warm Pacific SST conditions (El Niño), whereas slower propagation occurred during relatively cold Pacific SST conditions (Pohl and Matthews 2007). In summary, MJO and AAM characteristics *prior* to model initialization indicate distinguishing signals to identify "forecasts of opportunity" and periods of lower forecast skill, potentially allowing for a "forecast of forecast skill."

Various weather/climate teleconnection indices were examined, including the AO, NAO, Niño-3.4 index, PDO, PNA, and QBO for the initialization date and day 10 (Fig. 11). No teleconnection indices were significant precursor signals for the skillful forecasts. Statistically significant negative indices for low-skill forecasts existed for the Niño-3.4 and PDO at day 0, confirming the La Niña conditions evident in Fig. 9d and the discussion surrounding the observed MJO. The La Niña and negative PDO conditions were still significant at day 10 for the low-skill forecasts. Additionally, the negative PNA pattern occurred for both forecast skill groups, congruent with the "high-low-high" Z500 pattern over the eastern Pacific and the United States (Figs. 4b,d). In general, caution should

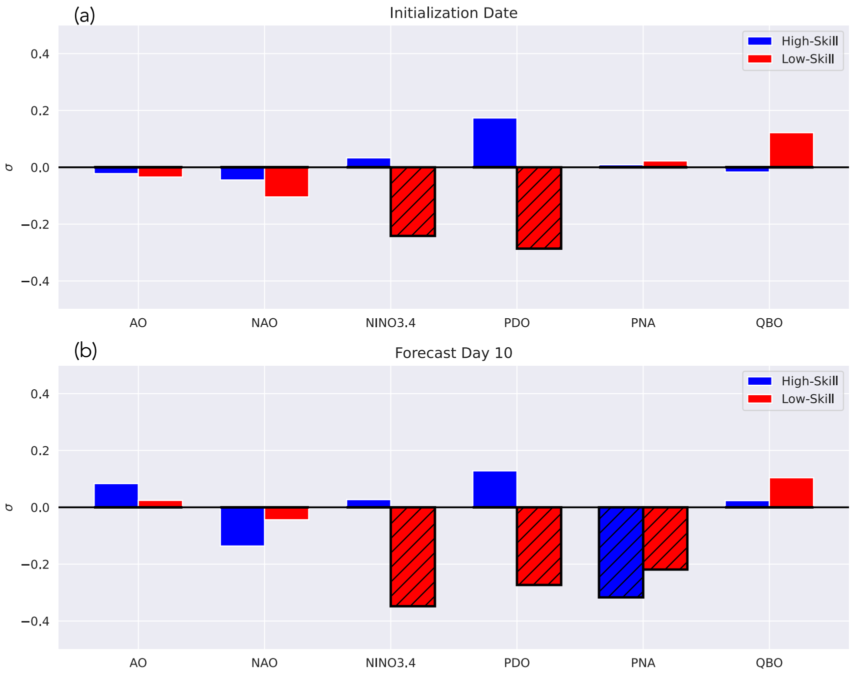

FIG. 11. Mean values of several teleconnection indices for high- (blue) and low-skill (red) forecasts for the (a) initialization date and (b) forecast day 10. Hatched bars indicates statistically significant values at the 95% confidence level.

be practiced when examining such teleconnection indices for this purpose, as they represent a particular state of the atmosphere in specific geographic regions that individually may or may not be related to the evolution of the physical synoptic weather pattern over the CONUS.

5. Discussion and conclusions

Despite the need for extended-range prediction of severe convective storms, the current predictability limit for daily forecasts is, on average, ~10 days. As a result of chaos, skill in NWP solutions decrease as model integration through time increases. However, instances exist in which forecast solutions are more or less skillful at day 10 (and beyond), with the former providing "forecasts of opportunity." To examine these instances, atmospheric and oceanic states (and their time evolution) that supported a high- or low-skill day-10 forecast were examined for 75th percentile tornado day hindcasts. The day-10 GEFSv12 Z500 field over the United States during these active tornado days was evaluated and split into high- and low-skill groups based on Pearson *r* correlation and mean absolute error. The observed Z500 pattern supportive of tornado days over the United States at day 10 were similar for both groups, characteristic of a midlevel western U.S. thermal trough and eastern U.S. thermal ridge.

High-skill GEFSv12 forecasts reproduced the geopotential height pattern at 500 hPa well, whereas the pattern for the low-skill forecasts was out of phase and propagated too quickly. High-skill forecasts were also associated with northward wave activity flux propagation and strong convergence off the west coast of the United States. These forecasts also were characteristic of field significant negative OLR anomalies over the Maritime Continent and warmer-than-average SSTs throughout the tropical Pacific and Gulf of Mexico. Additionally, high-skill forecasts exhibited a strong MJO 15–20 days prior to model initialization, which lead to a significant modulation of AAM, which is consistent with previous research (Gensini et al. 2019, 2020a; Miller et al. 2022).

In contrast to the high-skill forecasts, WaF propagation for the low-skill forecasts contained nearly zonal WaF vectors. Colder than normal SSTs were evident throughout the tropical and eastern Pacific, reminiscent of La Niña conditions. A weaker and slower MJO was evident relative to the high-skill group, and no significant peak in AAM occurred prior to forecast initialization. The most prominent feature with low-skill forecasts was a blocking anticyclone over eastern Russia during model initialization, in which the GEFSv12 overforecast the duration of the block. This is perhaps not too surprising, as theories of atmospheric blocking onset, maintenance, and decay are still unsettled, and research has shown that models often fail to predict the duration of atmospheric blocking events (Reynolds et al. 2017).

This research is the first step in understanding under what initialization conditions an extended-range GEFSv12 forecast for tornado conditions may promote higher or lower forecast skill. Inherently, this study exposes forecasts of opportunity, which could be leveraged as windows of enhanced dynamical model forecast skill for operational forecasters. These results promote the further possibility of aiding in the creation of medium-range tornado outlooks—but most importantly—permit the assessment of forecast confidence and allow for the potential to "forecast the forecast skill." For example, if a blocking anticyclone is present over the Northwest Pacific, GEFSv12 forecasts may not evolve properly due to the overprediction of blocking duration and subsequent impacts on the downstream flow. Additional model diagnostic studies for other variables may further aid in our understanding of potential biases and sources of errors for extended-range prediction of extreme weather events like tornadoes.

Acknowledgments. This work was supported by the National Science Foundation (award #2048770). We acknowledge the Northern Illinois University Center for Research Computing and Data for providing computing resources. This research also used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

Data availability statement. ERA5 data were downloaded from the ECMWF Copernicus Climate Change Service (C3S; https://cds.climate.copernicus.eu/). The GEFSv12 data are available via the Amazon web server (https://noaa-gefs-retrospective.s3.amazonaws.com/index.html). The CPC Pentad MJO indices are available online (https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_mjo_index/pentad.html). PPH data are available online (https://atlas.niu.edu/pperfect/BAMS/). Teleconnection indices were obtained from NOAA's PSL (https://psl.noaa.gov/data/climateindices/list/).

REFERENCES

Allen, J. T., M. K. Tippett, and A. H. Sobel, 2015: Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nat. Geosci.*, **8**, 278–283, https://doi.org/10.1038/ngeo2385.

Baggett, C. F., K. M. Nardi, S. J. Childs, S. N. Zito, E. A. Barnes, and E. D. Maloney, 2018: Skillful subseasonal forecasts of weekly tornado and hail activity using the Madden–Julian Oscillation. *J. Geophys. Res. Atmos.*, **123**, 12 661–12 675, https://doi.org/10.1029/2018JD029059.

Barrett, B. S., and V. A. Gensini, 2013: Variability of central United States April–May tornado day likelihood by phase of the Madden–Julian Oscillation. *Geophys. Res. Lett.*, **40**, 2790–2795, https://doi.org/10.1002/grl.50522.

Baxter, S., S. Weaver, J. Gottschalck, and Y. Xue, 2014: Pentad evolution of wintertime impacts of the Madden–Julian oscillation over the contiguous United States. *J. Climate*, **27**, 7356–7367, https://doi.org/10.1175/JCLI-D-14-00105.1.

Brooks, H. E., C. A. Doswell III, and J. Cooper, 1994: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting*, **9**, 606–618, https://doi.org/10.1175/1520-0434(1994)009<0606:OTEOTA>2.0.CO;2.

——, J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67**, 73–94, https://doi.org/10.1016/S0169-8095(03)00045-0.

Buehler, T., C. C. Raible, and T. F. Stocker, 2011: The relationship of winter season North Atlantic blocking frequencies to extreme cold or dry spells in the ERA-40. *Tellus*, **63A**, 212–222, https://doi.org/10.1111/j.1600-0870.2010.00492.x.

Dole, R., and Coauthors, 2011: Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.*, **38**, L06702, https://doi.org/10.1029/2010GL046582.

Dyer, J., C. Zarzar, P. Amburn, R. Dumais, J. Raby, and J. A. Smith, 2016: Defining the influence of horizontal grid spacing on ensemble uncertainty within a regional modeling framework. *Wea. Forecasting*, **31**, 1997–2017, https://doi.org/10.1175/WAF-D-16-0030.1.

Frye, J. D., and T. L. Mote, 2010: Convection initiation along soil moisture boundaries in the Southern Great Plains. *Mon. Wea. Rev.*, **138**, 1140–1151, https://doi.org/10.1175/2009MWR2865.1.

Gensini, V. A., and A. Marinaro, 2016: Tornado frequency in the United States related to global relative angular momentum. *Mon. Wea. Rev.*, **144**, 801–810, https://doi.org/10.1175/MWR-D-15-0289.1.

——, and J. T. Allen, 2018: U.S. hail frequency and the global wind oscillation. *Geophys. Res. Lett.*, **45**, 1611–1620, https://doi.org/10.1002/2017GL076822.

——, and M. K. Tippett, 2019: Global Ensemble Forecast System (GEFS) predictions of days 1–15 U.S. tornado and hail frequencies. *Geophys. Res. Lett.*, **46**, 2922–2930, https://doi.org/10.1029/2018GL081724.

——, D. Gold, J. T. Allen, and B. S. Barrett, 2019: Extended U.S. tornado outbreak during late May 2019: A forecast of opportunity. *Geophys. Res. Lett.*, **46**, 10 150–10 158, https://doi.org/10.1029/2019GL084470.

——, B. S. Barrett, J. T. Allen, D. Gold, and P. Sirvatka, 2020a: The Extended-Range Tornado Activity Forecast (ERTAF) project. *Bull. Amer. Meteor. Soc.*, **101**, E700–E709, https://doi.org/10.1175/BAMS-D-19-0188.1.

——, A. M. Haberlie, and P. T. Marsh, 2020b: Practically perfect hindcasts of severe convective storms. *Bull. Amer. Meteor. Soc.*, **101**, E1259–E1278, https://doi.org/10.1175/BAMS-D-19-0321.1.

Grams, J. S., R. L. Thompson, D. V. Snively, J. A. Prentice, G. M. Hodges, and L. J. Reames, 2012: A climatology and comparison of parameters for significant tornado events in the United States. *Wea. Forecasting*, **27**, 106–123, https://doi.org/10.1175/WAF-D-11-00008.1.

Guan, H., and Coauthors, 2022: GEFSV12 reforecast dataset for supporting subseasonal and hydrometeorological applications. *Mon. Wea. Rev.*, **150**, 647–665, https://doi.org/10.1175/MWR-D-21-0245.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hitchens, N. M., and H. E. Brooks, 2012: Evaluation of the Storm Prediction Center's day 1 convective outlooks. *Wea. Forecasting*, **27**, 1580–1585. https://doi.org/10.1175/WAF-D-12-00061.1

——, and ——, 2014: Evaluation of the Storm Prediction Center's convective outlooks from day 3 through day 1. *Wea. Forecasting*, **29**, 1134–1142, https://doi.org/10.1175/WAF-D-13-00132.1.

——, ——, and M. P. Kay, 2013: Objective limits on forecasting skill of rare events. *Wea. Forecasting*, **28**, 525–534, https://doi.org/10.1175/WAF-D-12-00113.1.

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, *Geophys. Monogr.*, Vol. 134, Amer. Geophys. Union, 1–35, https://doi.org/10.1029/134GM01.

Johns, R. H., and C. A. Doswell, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612, https://doi.org/10.1175/1520-0434(1992)007<0588:SLSF>2.0.CO;2.

Lee, S.-K., A. T. Wittenberg, D. B. Enfield, S. J. Weaver, C. Wang, and R. Atlas, 2016: U.S. regional tornado outbreaks and their links to spring ENSO phases and North Atlantic SST variability. *Environ. Res. Lett.*, **11**, 044008, https://doi.org/10.1088/1748-9326/11/4/044008.

Lepore, C., M. K. Tippett, and J. T. Allen, 2018: CFSV2 monthly forecasts of tornado and hail activity. *Wea. Forecasting*, **33**, 1283–1297, https://doi.org/10.1175/WAF-D-18-0054.1.

Maddison, J., S. Gray, O. Martínez-Alvarado, and K. Williams, 2019: Upstream cyclone influence on the predictability of block onsets over the Euro-Atlantic region. *Mon. Wea. Rev.*, **147**, 1277–1296, https://doi.org/10.1175/MWR-D-18-0226.1.

Maddox, R. A., 1976: An evaluation of tornado proximity wind and stability data. *Mon. Wea. Rev.*, **104**, 133–142, https://doi.org/10.1175/1520-0493(1976)104<0133:AEOTPW>2.0.CO;2.

Matsueda, M., 2011: Predictability of Euro-Russian blocking in summer of 2010. *Geophys. Res. Lett.*, **38**, L06801, https://doi.org/10.1029/2010GL046557.

Merryfield, W. J., and Coauthors, 2020: Current and emerging developments in subseasonal to decadal prediction. *Bull. Amer. Meteor. Soc.*, **101**, E869–E896, https://doi.org/10.1175/BAMS-D-19-0037.1.

Miller, D. E., and Z. Wang, 2019a: Assessing seasonal predictability sources and windows of high predictability in the Climate Forecast System, version 2. *J. Climate*, **32**, 1307–1326, https://doi.org/10.1175/JCLI-D-18-0389.1.

——, and ——, 2019b: Skillful seasonal prediction of Eurasian winter blocking and extreme temperature frequency. *Geophys. Res. Lett.*, **46**, 11 530–11 538, https://doi.org/10.1029/2019GL085035.

——, and ——, 2022: Northern Hemisphere winter blocking: Differing onset mechanisms across regions. *J. Atmos. Sci.*, **79**, 1291–1309, https://doi.org/10.1175/JAS-D-21-0104.1.

——, ——, R. J. Trapp, and D. S. Harnos, 2020: Hybrid prediction of weekly tornado activity out to week 3: Utilizing weather regimes. *Geophys. Res. Lett.*, **47**, e2020GL087253, https://doi.org/10.1029/2020GL087253.

——, ——, B. Li, D. S. Harnos, and T. Ford, 2021: Skillful subseasonal prediction of U.S. extreme warm days and standardized precipitation index in boreal summer. *J. Climate*, **34**, 5887–5898, https://doi.org/10.1175/JCLI-D-20-0878.1.

——, V. A. Gensini, and B. S. Barrett, 2022: Madden–Julian Oscillation influences United States springtime tornado and hail frequency. *npj Climate Atmos. Sci.*, **5**, 37, https://doi.org/10.1038/s41612-022-00263-5.

Moore, T. W., 2018: Annual and seasonal tornado activity in the United States and the global wind oscillation. *Climate Dyn.*, **50**, 4323–4334, https://doi.org/10.1007/s00382-017-3877-5.

——, and M. P. McGuire, 2020: Tornado-days in the United States by phase of the Madden–Julian Oscillation and global wind oscillation. *Climate Dyn.*, **54**, 17–36, https://doi.org/10.1007/s00382-019-04983-y.

Morrison, H., and Coauthors, 2020: Confronting the challenge of modeling cloud and precipitation microphysics. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001689, https://doi.org/10.1029/2019MS001689.

Nageswararao, M., Y. Zhu, and V. Tallapragada, 2022: Prediction skill of GEFSV12 for southwest summer monsoon rainfall and associated extreme rainfall events on extended range scale over India. *Wea. Forecasting*, **37**, 1135–1156, https://doi.org/10.1175/WAF-D-21-0184.1.

NWS/CPC, 2022: Pentad MJO indices. Accessed 1 January 2022, https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_mjo_index/pentad.html.

Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100**, 2043–2060, https://doi.org/10.1175/BAMS-D-18-0270.1.

Pohl, B., and A. J. Matthews, 2007: Observed changes in the lifetime and amplitude of the Madden–Julian Oscillation associated with interannual ENSO sea surface temperature anomalies. *J. Climate*, **20**, 2659–2674, https://doi.org/10.1175/JCLI4230.1.

Rex, D. F., 1950: Blocking action in the middle troposphere and its effect upon regional climate. *Tellus*, **2**, 275–301, https://doi.org/10.3402/tellusa.v2i4.8603.

Reynolds, D. D., A. R. Lupo, A. D. Jensen, and P. S. Market, 2017: The predictability of Northern Hemispheric blocking using an ensemble mean forecast system. *Proceedings*, **1**, 87, https://doi.org/10.3390/ecas2017-04128.

Rodwell, M. J., and Coauthors, 2013: Characteristics of occasional poor medium-range weather forecasts for Europe. *Bull. Amer. Meteor. Soc.*, **94**, 1393–1405, https://doi.org/10.1175/BAMS-D-12-00099.1.

Saminathan, S., H. Medina, S. Mitra, and D. Tian, 2021: Improving short to medium range GEFS precipitation forecast in India. *J. Hydrol.*, **598**, 126431, https://doi.org/10.1016/j.jhydrol.2021.126431.

Sillmann, J., and M. Croci-Maspoli, 2009: Present and future atmospheric blocking and its impact on European mean and extreme climate. *Geophys. Res. Lett.*, **36**, L10702, https://doi.org/10.1029/2009GL038259.

Takaya, K., and H. Nakamura, 2001: A formulation of a phase-independent wave-activity flux for stationary and migratory quasigeostrophic eddies on a zonally varying basic flow. *J. Atmos. Sci.*, **58**, 608–627, https://doi.org/10.1175/1520-0469(2001)058<0608:AFOAPI>2.0.CO;2.

Thompson, D. B., and P. E. Roundy, 2013: The relationship between the Madden–Julian Oscillation and U.S. violent tornado outbreaks in the spring. *Mon. Wea. Rev.*, **141**, 2087–2095, https://doi.org/10.1175/MWR-D-12-00173.1.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261, https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

Tibaldi, S., and F. Molteni, 1990: On the operational predictability of blocking. *Tellus*, **42**, 343–365, https://doi.org/10.3402/tellusa.v42i3.11882.

Trapp, R. J., and K. A. Hoogewind, 2018: Exploring a possible connection between U.S. tornado activity and arctic sea ice. *npj Climate Atmos. Sci.*, **1**, 14, https://doi.org/10.1038/s41612-018-0025-9.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Vol. 100, Academic Press, 704 pp.

Wurman, J., D. Dowell, Y. Richardson, P. Markowski, E. Rasmussen, D. Burgess, L. Wicker, and H. B. Bluestein, 2012: The second Verification of the Origins of Rotation in Tornadoes Experiment: VORTEX2. *Bull. Amer. Meteor. Soc.*, **93**, 1147–1170, https://doi.org/10.1175/BAMS-D-11-00010.1.