Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL081724

Global Ensemble Forecast System (GEFS) Predictions of Days 1–15 U.S. Tornado and Hail Frequencies

Vittorio A. Gensini[1] **and Michael K. Tippett**[2]

[1]Department of Geographic and Atmospheric Sciences, Northern Illinois University, DeKalb, IL, USA, [2]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA

Key Points:
- Tornado and hail frequencies are skillfully predicted into the medium range by the Global Ensemble Forecast System
- Brier skill score approaches zero at day 9 for tornado and day 12 for hail
- Average forecast skill scores are higher for hail versus tornado activity

Correspondence to:
V. A. Gensini,
vgensini@niu.edu

Citation:
Gensini, V. A., & Tippett, M. K. (2019). Global Ensemble Forecast System (GEFS) predictions of days 1–15 U.S. tornado and hail frequencies. *Geophysical Research Letters*, *46*, 2922–2930. https://doi.org/10.1029/2018GL081724

Received 17 DEC 2018
Accepted 15 FEB 2019
Accepted article online 20 FEB 2019
Published online 11 MAR 2019

Abstract

U.S. severe convective storms have grown to represent a 10 billion dollar annual peril for the insurance industry, and their accurate prediction remains a challenging task. This study examines days 1–15 severe convective storm predictions from the Global Ensemble Forecast System (GEFS). GEFS forecasts are based on the Supercell Composite Parameter and verified against spatially smoothed tornado and hail reports over the periods 1 March to 31 May 2016–2017. Skill is analyzed for deterministic forecasts (ensemble mean Supercell Composite Parameter exceeding specified thresholds) and probabilistic forecasts (fraction of ensemble members exceeding specified thresholds). Deterministic forecasts of tornado and hail activities are statistically more skillful than a random no-skill reference to days 9 and 12, respectively. Probabilistic forecasts are skillful relative to climatological no-skill reference to day 9 for tornado and day 12 for hail activity. These results provide a useful baseline for further improvement of tornado and hail forecasts at these ranges.

Plain Language Summary

Tornadoes and hailstorms represent a 10 billion dollar annual peril for the insurance industry. Prediction of these extreme events at long lead times is a challenging task. This research shows that it is possible to use the Global Ensemble Forecast System to anticipate tornado and hail events in the United States at lead times beyond 1 week. This represents an important baseline for the improvement of tornado and hail forecasts at the subseasonal to seasonal time scales.

1. Introduction

In the United States, the National Oceanic and Atmospheric Administration's Storm Prediction Center (SPC) issues areal outlooks for severe convective storms at lead times spanning 1–8 days. At such lead times, an ingredients-based approach (Johns & Doswell, 1992) is often used in the forecast process. These ingredients include low static stability, high-surface water vapor mixing ratios, a lifting mechanism (e.g., a boundary or orographic feature) to get air parcels to the level of free convection, and adequate deep-layer vertical wind shear (Rasmussen, 2003; Rasmussen & Blanchard, 1998) to promote organized convection. More recently, storm-scale outputs from convection permitting models provide short-range (e.g., 12–36 hr) forecast guidance (Gallo et al., 2016, 2018). Accurately forecasting severe weather events remains a challenging task, but skill of SPC days 1–3 forecasts for severe convective storms has been increasing since the mid-1990s (Herman et al., 2017; Hitchens & Brooks, 2012; 2014; Hitchens et al., 2013).

Less attention has been given to forecast guidance at medium to extended lead times (e.g., 4 to 14 days), although there are indications that tornado and hail activities are predictable on subseasonal to seasonal time scales (Lepore et al., 2017, 2018). Guidance from current convection allowing models tends to be unavailable at these temporal ranges because of computational cost. However, at these longer leads, ingredients-based parameter evaluation remains a tractable approach given the horizontal grid spacing of available model guidance. Here we use a well-known severe weather composite index (the Supercell Composite Parameter; SCP) as an indication of conditions that are favorable for tornado or hail events. Forecasts of SCP values at lead times of 1–15 days are provided by the National Center for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS).

2. Data and Methods

2.1. Tornado and Hail Reports

Tornado and hail observations originate from the National Center for Environmental Information *Storm Data* (Schaefer & Edwards, 1999) for 2016 and 2017. These observations are known to contain spatial and nonmeteorological biases relating to spatial variations in population, inherent subjectivity in intensity rating, and other discontinuities in the record (Potvin et al., 2019; Verbout et al., 2006). Here we aggregate storm reports at daily (valid 1200–1200 UTC) resolution and calculate *practically perfect probabilities* on NCEP's ETA 212 Lambert conformal grid (approximately 40-km horizontal grid spacing) following equation (1) of Hitchens et al. (2013) with $\sigma = 0.75$. Skill calculations using $\sigma = 1.5$ on a 80-km grid gave qualitatively similar results. The basic premise of practically perfect probabilities is to construct probabilities based on storm reports that are consistent with SPC outlook probabilities by spatially smoothing the severe weather reports. The amount of smoothing reflects the level of detail and uncertainty that is typical in SPC outlooks. We define tornado activity as when the practically perfect probabilities of tornadoes exceed the 5% level and hail activity as when the practically perfect probabilities of hail exceed the 15% level. These probability thresholds represent the first rank of a categorical “SLIGHT” risk hazard used by the SPC in deterministic products. We then verify deterministic and probabilistic GEFS forecasts against these binary (yes/no) tornado and hail verification fields.

2.2. GEFS Data

GEFS forecasts initialized at 0000 UTC were downloaded for two recent boreal spring periods 1 March to 31 May 2016–2017 (184 forecast cycles). These global data consist of a 21-member ensemble (20 perturbed members and 1 control member) at 1° × 1° horizontal grid spacing spanning temporally from initialization to 384 hr (16 days). Native horizontal grid spacing is T574 (~34 km) for the first 8 days and T382 (~52 km) for the second 8 days (Zhou et al., 2017/07/19, 2017/07/19). SCP was calculated on the 6-hourly post processed GEFS output following a fixed-layer approach similar to equation (3) in Thompson et al. (2003)

$$\mathrm{SCP} = \frac{\mathrm{muCAPE}}{1,000\,\mathrm{J/kg}} \times \frac{0-3\,\mathrm{km\,SRH}}{100\,\mathrm{m^2/s^2}} \times \frac{0-6\,\mathrm{km\,BWS}}{20\,\mathrm{m/s}} \tag{1}$$

where muCAPE (J/kg) represents convective available potential energy (CAPE) associated with the most unstable parcel in the lowest 255 mb of the model. The 0- to 3-km SRH (m^2/s^2) refers to storm relative helicity integrated through the 0- to 3-km layer using the Bunkers storm motion technique (Bunkers et al., 2000), and 0- to 6-km BWS describes the bulk wind shear between 0- and 6-km above ground level. The 0- to 3-km SRH and muCAPE are available as post processed fields from GEFS. The 0- to 6-km BWS is the magnitude of the vector difference of winds at 10 m and at 6 km, calculated from GEFS data by vertically interpolating the isobaric u, v wind fields to AGL height coordinates. SCP values were set to zero if GEFS convective precipitation did not exceed 1 mm in the subsequent 6-hr period to limit our attention to severe weather environments where convection was forecast to initiate, similar to Trapp et al. (2009). We use the notation $\mathrm{SCP_{CP}}$ to refer to this new variable SCP conditional on convective precipitation greater than 1 mm.

For each forecast cycle, the first and last 12 hr were discarded leaving fifteen 1200–1200 UTC periods which we refer to as leads 1–15. For each period, the daily maximum $\mathrm{SCP_{CP}}$ value was extracted and bilinearly interpolated to NCEP's ETA 212 grid for direct comparison to the corresponding 1200–1200 UTC practically perfect tornado and severe hail probabilities. Forecast verification was conducted using deterministic (e.g., Heidke skill score, HSS) and probabilistic (e.g., Brier skill score, BSS) metrics.

2.3. Verification Methods

2.3.1. Deterministic

Binary (yes/no) forecasts for tornado and hail activities were constructed based on whether the ensemble mean $\mathrm{SCP_{CP}}$ value met or exceeded a value of 2 for hail and 4 for tornado. $\mathrm{SCP_{CP}}$ values of 2 and 4 were chosen after performing a percentile-percentile analysis on all ensemble mean $\mathrm{SCP_{CP}}$ and practically perfect values by lead day. We note that the results herein are not particularly sensitive to the choice of $\mathrm{SCP_{CP}}$ threshold, and other applications may benefit from differing thresholds by hazard or lead day. Deterministic forecast performance was measured using the HSS due to its preferred use for forecasting rare events (Doswell et al., 1990). The HSS measures the fractional improvement of the forecast over a random forecast. The range of the HSS is $-\infty$ to 1. Negative values indicate that the random forecast is better, 0 means no skill relative to the random forecast, and a perfect forecast obtains a HSS value of 1. Importantly, the random forecast reference in the HSS is one whose base rate is equal to the observed base rate. This random forecast is not

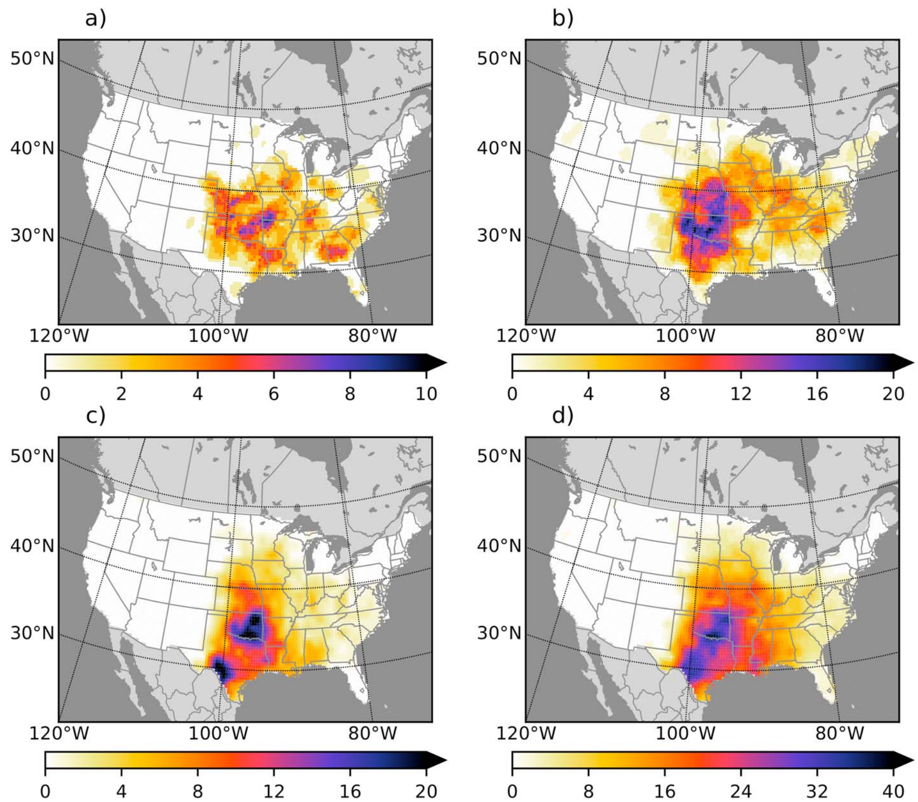

Figure 1. March–May 2016–2017 frequency of (a) practically perfect tornado probabilities ≥ 5%, (b) practically perfect hail probabilities ≥15%, (c) Global Ensemble Forecast System ensemble mean day 1 SCP_{CP} ≥4, and (d) Global Ensemble Forecast System member mean day 1 SCP_{CP} ≥2. SCP = Supercell Composite Parameter.

an especially realistic reference for a weather or climate application where seasonality is strong. In fact, a no-skill climatological forecast might have positive HSS. Therefore, the HSS was also calculated by lead day for a "seasonally informed" or climatological no-skill forecast given by a randomly selected March–May practically perfect verification set from the period 1979 to 2015. To determine statistical significance of the HSS departures, all practically perfect verification sets were generated and ranked. We describe statistically significant ($p \leq 0.05$ significance level) skill when the GEFS forecast meets or exceeds the 95th percentile HSS of the practically perfect "no-skill" forecasts.

2.3.2. Probabilistic

GEFS probabilities of SCP_{CP} meeting/exceeding 2 for hail and 4 for tornado were generated at all CONUS grid points for each forecast cycle and lead day. For example, if 10 of the 21 GEFS members recorded SCP_{CP} values ≥2, then the probability assigned is 47.6%. Probabilistic forecast skill was measured using the BSS (Murphy, 1973) by using the daily climatological practically perfect probabilities as a reference forecast from the base period 1979–2015. Positive BSS values indicate more skill than a climatological forecast, and negative values indicate less skill than a climatological forecast. A score of 0 indicates no skill versus a forecast of climatology.

Receiver operating characteristic (ROC) curves provide a measure of probabilistic discrimination, namely, whether higher forecast probabilities are associated with more hits (correctly forecasting occurrence) and fewer false alarms (incorrectly forecasting occurrence). ROC curves plot the false alarm rate (x axis) against the probability of detection (y axis) as the forecast probability threshold used to classify decreases. Ideally, the hit rate increases faster than the false alarm rate as the forecast probability threshold decreases. The no-skill line on a ROC curve is the one-to-one line where the hit rate and false alarm rate are equal. ROC curves do not measure reliability, but they do indicate the potential usefulness of the forecast if properly calibrated. Area under the ROC curve (AUC) is a common measure for forecast skill and can be measured relative to the no-skill line or other forecast methods (e.g., climatology).

3. Results

3.1. Spatial Distribution

We first examine the spatial patterns of day 1 forecast SCP_{CP} and corresponding practically perfect probability exceedances for tornado and hail reports. Frequency of 2016–2017 March–May practically perfect tornado events ≥ 5% (Figure 1a) and hail events ≥ 15% (Figure 1b) were compared to frequency of day 1 forecast SCP_{CP} values exceeding 4 (Figure 1c) and 2 (Figure 1d) for tornado and hail, respectively. These SCP_{CP} thresholds for day 1 were chosen based on a comparison of the percentiles of SCP and practically perfect probabilities so that the frequency of SCP_{CP} exceedances approximately match those of the practically perfect probabilities. Subjective examination of these plots indicates an area of overforecasting (i.e., false alarms) in the vicinity of Del Rio, Texas, which has been noted in previous research as an area where convective inhibition often suppresses thunderstorm development despite high values of moisture and instability (Gensini & Ashley, 2011; Gensini et al., 2014). In addition, some underforecasting is noted across portions of the southern High Plains when compared to observed tornado and hail reports. These biases could be related to the choice of SCP_{CP} threshold, GEFS model biases, and/or geographic dependence of tornado and hail reports (e.g., population density). A large area of spatial agreement is noted across portions of the southern Great Plains and Midwest. Overall, this analysis suggests that there is sufficient subjective agreement between SCP_{CP} exceedances and tornado/hail events to warrant more quantitative deterministic and probabilistic verification.

3.2. Deterministic Forecasts

Qualitatively, it can be useful to visualize severe weather (or other phenomena) forecast guidance in the medium range by use of a so-called “chiclet” plot, where forecasts run diagonally toward the top right of the plot and vertical columns share a constant verification date (Carbin et al., 2016). Although such displays are limited to scalar quantities (e.g., map or box averages), they permit the display of many forecasts with the same target and varying lead times. Chiclet plots for boreal spring 2016 (Figure 2a) and 2017 (Figure 2b) give an overview of all forecasts examined in this study. Color fill chiclets (top panel in both figures) indicate the CONUS Lambert area (km^2) covered by ensemble mean SCP_{CP} values ≥ 2. Bar chart plots (bottom panels in both figures) show the verifying practically perfect area for tornado (violet) and hail (green) at the previously discussed “SLIGHT” risk categorical thresholds. There are examples of good (e.g., 26 April 2016; Figure 2a) and poor (e.g., 28 April 2017; Figure 2b) visual agreement between CONUS mean SCP_{CP} area ≥ 2 and hail/tornado CONUS SLIGHT risk practically perfect area. Overall, the picture is of forecasts whose timing of active and inactive periods is fairly good up to a week in advance. Interestingly, there appear to be several examples of increased mean SCP_{CP} values ≥ 2 emerging in the 8–10 lead day window with other events showing a potential signal beyond day 12.

Quantitative analysis of the areal coverage information in the chiclet forecasts was performed through examination of the Spearman rank correlation (Figure 3a). For each lead day, CONUS Lambert areas were calculated for SCP_{CP} ≥ 2 (hail) and ≥ 4 (tornado) thresholds and compared to CONUS Lambert areas of the respective practically perfect thresholds for categorical SLIGHT risk (i.e., 15% for hail and 5% for tornado). Results indicate that Spearman rank correlation decreases in a linear fashion from ~0.7 to ~0.1 for hail and ~0.5 to ~0.0 for tornado as the lead day goes from 1 to 15. Both tornado and hail are significantly correlated using a p value of 0.01 until day 13 for hail and day 9 for tornado (days with dot markers plotted in Figure 3a). GEFS mean SCP_{CP} CONUS Lambert area is slightly more correlated (average of +0.08 over all lead days) with hail practically perfect threshold area versus tornado. These findings indicate that the activity forecasts are skillfully predicting the size of the area impacted by severe convective storms but does not measure the spatial correspondence between forecasts and reports.

The HSS was also calculated by lead day and provides a measure of how well forecasts match the spatial features of the practically perfect probability exceedances (Figure 3b). Like Spearman rank correlation, HSS again decreases in a linear fashion by lead day. Higher values of HSS are found for hail, suggesting that GEFS mean SCP_{CP} is more skillfull at predicting practically perfect hail thresholds versus practically perfect tornado thresholds, presumably because of the inherently greater uncertainty of tornado occurrence. HSS values greater than 0 indicate skill over that of random chance. However, it is probably more informative to measure HSS against a randomly substituted seasonally informed value. Thus, the dashed lines in Figure 3b indicate a randomly substituted no-skill forecast as described in section 2.3.1. Forecasts for hail practically perfect thresholds using GEFS mean SCP_{CP} ≥2 show positive skill for all lead times when compared to random chance and statistically significant positive skill through day 11 when compared to a seasonally

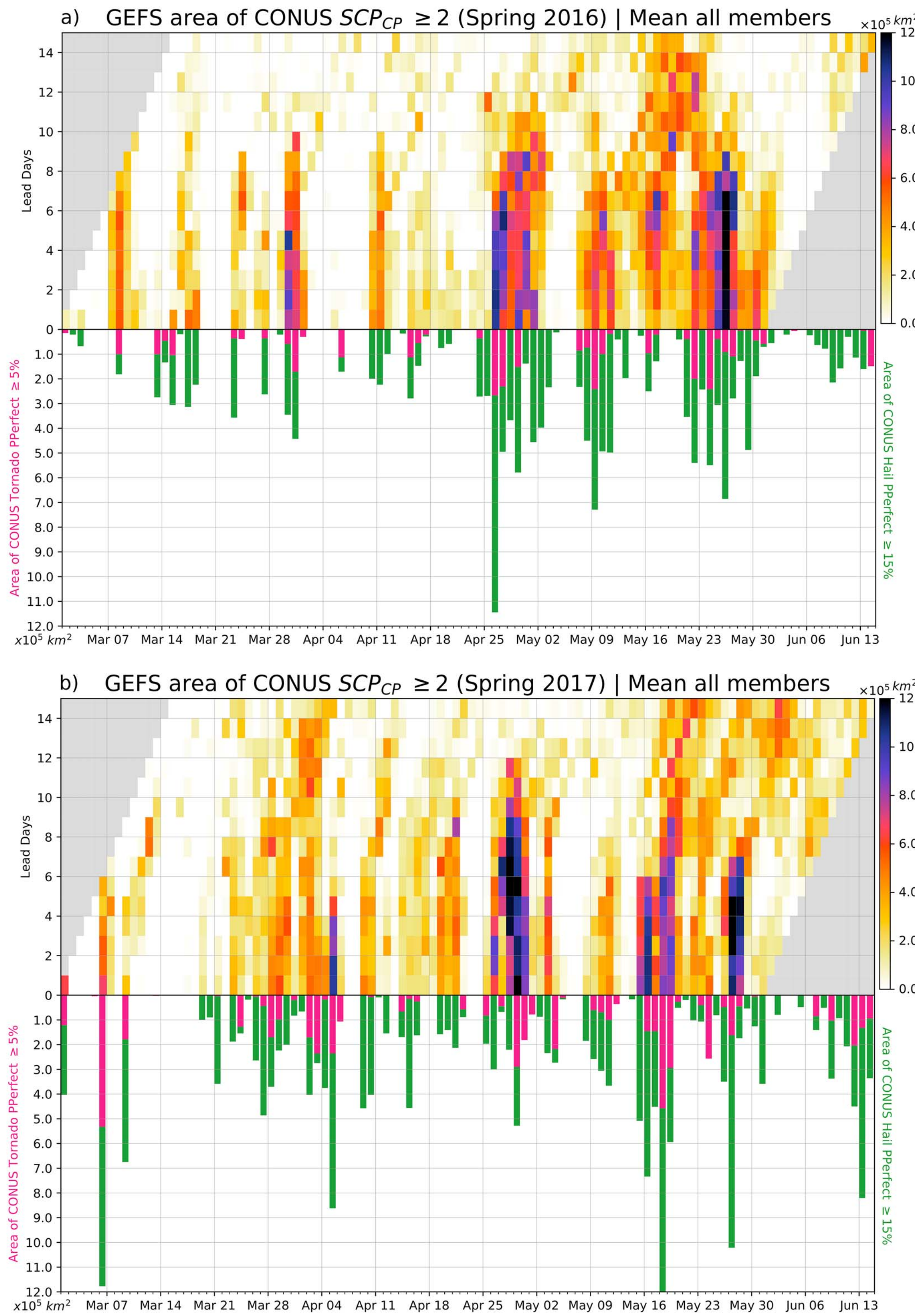

Figure 2. GEFS chiclet plots for 1 March to 31 May (a) 2016 and (b) 2017. Color fill corresponds to the GEFS member mean CONUS area of SCP_{CP} ≥ 2. Colored bars represent the area of practically perfect tornado probabilities ≥ 5% (violet) and area of practically perfect hail probabilities ≥ 15% (green). GEFS = Global Ensemble Forecast System; SCP = Supercell Composite Parameter.

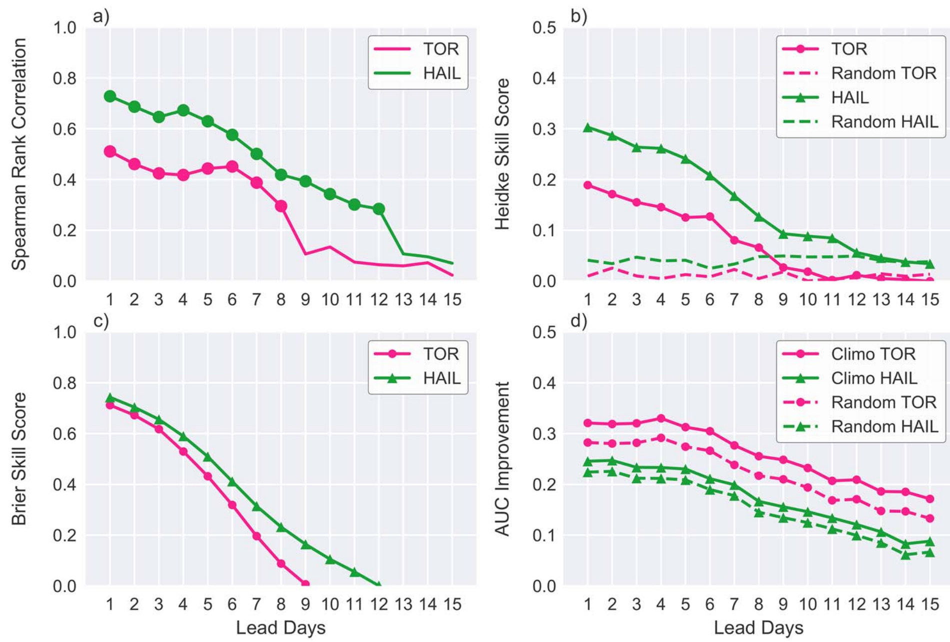

Figure 3. (a) Spearman rank correlation of forecast area, (b) Heidke skill score of deterministic forecasts, (c) Brier skill score, and (d) area under receiver operating characteristic curve improvement by lead day for Global Ensemble Forecast System SCP_{CP} forecasts during 1 March to 31 May 2016–2017. SCP = Supercell Composite Parameter; TOR = tornado.

informed random forecast. Tornado forecasts using GEFS mean SCP_{CP} ≥4 show statistically significant positive skill through day 8 against a seasonally informed random forecast. This suggests, in an average sense, that forecasters can make skillful forecasts for tornado and hail SLIGHT risk areas up to days 9 and 12, respectively, using ensemble mean SCP_{CP}.

3.3. Probabilistic Forecasts

Probability forecasts can convey more information about the likelihood of severe weather occurrence, and here they are based on individual GEFS member SCP_{CP} values to examine probabilities of threshold exceedance. The BSS was calculated for this purpose to understand the relative skill of GEFS as an ensemble system to forecast categorical SLIGHT risk events for tornado and hail versus climatology. That is, when the BSS $<$ 0, a forecast of climatology would be more skillful than probabilities derived from GEFS member SCP_{CP} values. BSS values first become $<$ 0 at days 9 and 12 for tornado and hail activities, respectively (Figure 3c).

ROC curves for hail (Figure 4a) and tornado (Figure 4b) show improvement at all leads over the climatological (black solid line) and randomly substituted seasonally informed no-skill (black dashed line) forecasts.

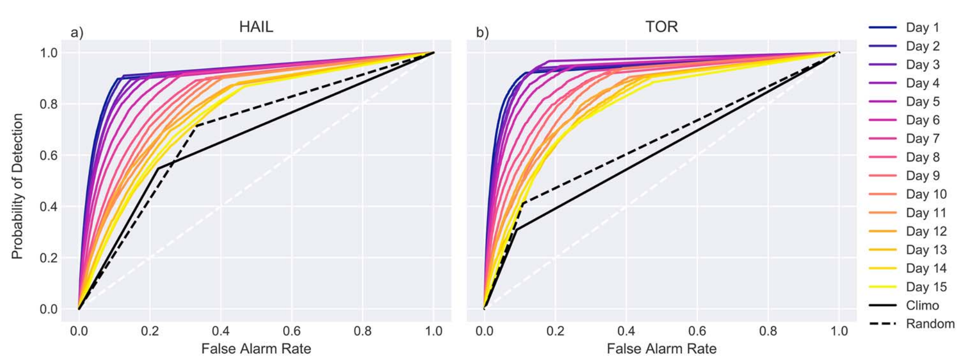

Figure 4. Receiver operating characteristic curves for occurrence of (a) hail practically perfect ≥ 15% and (b) tornado practically perfect ≥ 5%. Predictor is the Global Ensemble Forecast System probability of SCP_{CP} ≥ 2 for hail and ≥ 4 for tornado. Climatological receiver operating characteristic curve is calculated from the 1979–2015 respective daily practically perfect averages. SCP = Supercell Composite Parameter; TOR = tornado.

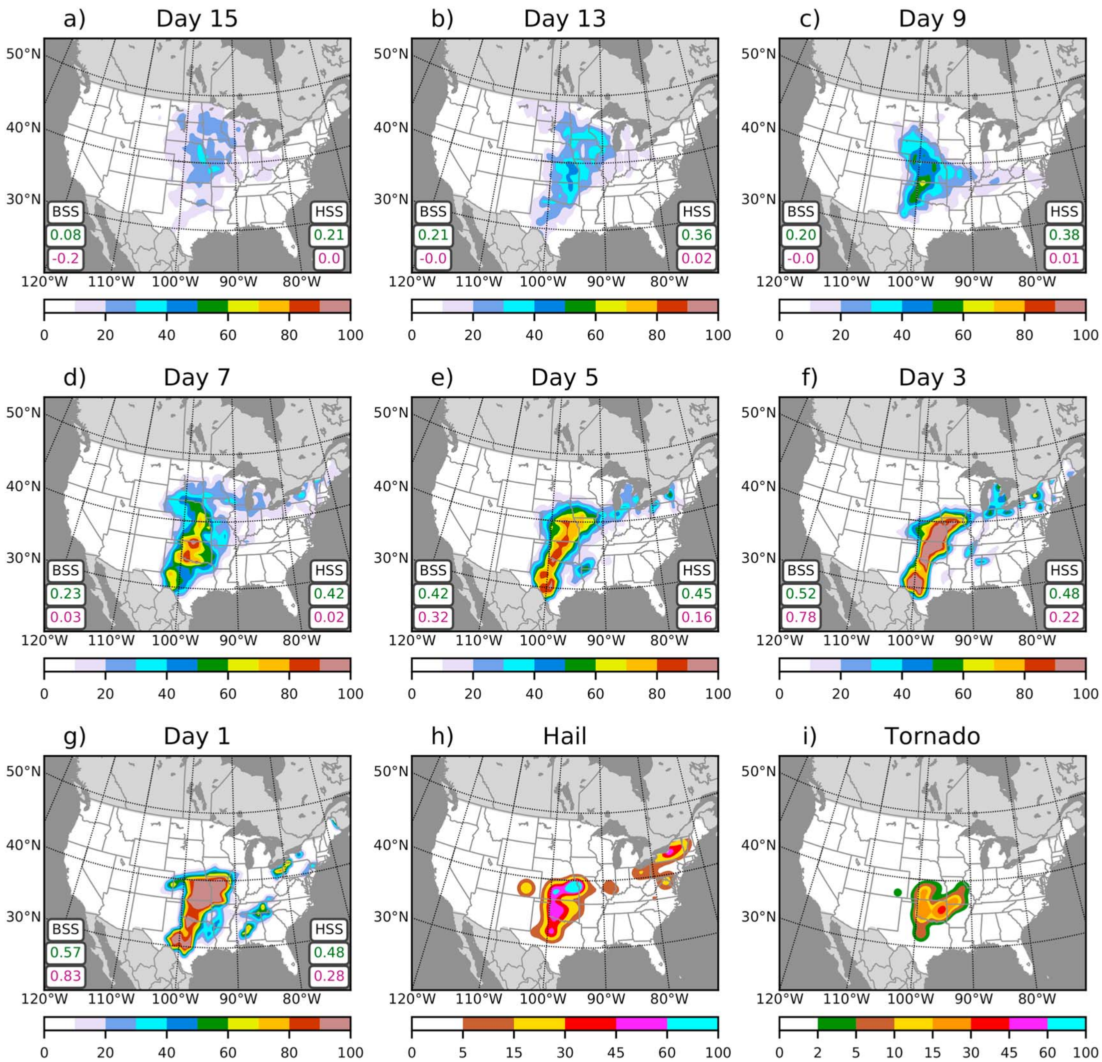

Figure 5. Lead day progression (a–g) of the Global Ensemble Forecast System probabilistic forecast for SCP_{CP} values ≥ 2 valid 1200 UTC 18 May to 1200 UTC 19 May 2017. Practically perfect verification maps are also shown for (h) hail and (i) tornado reports following probability thresholds used by the SPC. Brier skill score (BSS) and Heidke skill score (HSS) values for hail (green) and tornado (violet) are shown in the lower corners of each lead panel. SCP = Supercell Composite Parameter.

AUC improvements over the climatological and seasonally informed no-skill "random" forecast are shown in Figure 3d. Both hazards exhibited a greater improvement over climatology in comparison to a random forecast (i.e., random seasonally informed GEFS forecasts are more skillful when compared to a forecast of climatology). Additionally, improvement over the random no-skill and climatological forecasts is greater for tornado versus hail activity. This makes sense given the lower climatological frequency of tornado events relative to hail events. For both tornado and hail, the greatest contribution to AUC improvement originates from lower false alarm rates at the highest ($\approx \geq$ 80%) probabilities of detection against a climatological forecast. Positive AUC improvement at all leads suggests the potential for forecast improvement given proper calibration.

3.4. Case Study

A case study was examined for the period 1200 UTC 18 May to 1200 UTC 19 May 2017 to illustrate the spatio temporal forecast evolution of a high-impact event (the day that exhibited the largest sum of verified tornado and hail SLIGHT risk areas) by using the associated GEFS exceedance probabilities of an SCP_{CP} value ≥ 2 (Figure 5). For this event, a clear signal seems to emerge by ~ lead day 9, where over half of the members that predict SCP_{CP} would exceed a value of 2 in portions of south central Kansas and north central Oklahoma. SCP_{CP} probabilities of meeting/exceeding a value of 2 generally increased and expanded as lead day decreased. For this event, one can see a large spatial overlap between the highest probabilities of exceedance and the highest practically perfect tornado and hail probabilities. This case also shows the aforementioned overforecasting in the Del Rio, Texas region. GEFS SCP_{CP} probabilities were over 90% in this region beginning around lead day 3, yet no tornadoes or large hail reports were received from this area for this event. Further analysis of this case indicates that nearly all GEFS members predicted convective precipitation ≥ 1 mm in this region. Composite RADAR reflectivity summaries and upper-level radiosonde data from Del Rio, Texas, for this day show a strongly "capped" atmosphere. Thus, this appears to be a case where the GEFS convective parameterization scheme was too aggressive in the initiation of precipitation in this region. Future studies could examine using different convective precipitation thresholds in the noted regions of the greatest false alarm.

4. Summary and Discussion

This study explored the skill of the GEFS in forecasting smoothed probabilities of tornado and hail frequencies for two recent severe weather seasons. Specifically, the SCP was derived from GEFS post processed isobaric output and used to predict tornado and hail report activities based on gridded practically perfect verification. Using GEFS ensemble mean values of SCP, HSS values indicated that skillful deterministic forecasts of activity area could be made over a no-skill forecast to lead day 9 for tornado and day 12 for hail. Forecast skill was found to be greater for hail events versus tornado events for all lead days. However, forecasters typically have climatological information at various lead times that can also serve as a probabilistic forecast, and this is usually a preferable metric to assess forecast skill in weather and climate applications. Thus, using the BSS, with climatology as a reference forecast, it was concluded that practically perfect thresholds of tornado and hail occurrence can be skillfully predicted over climatology to lead days 9 and 12, respectively. In regard to probabilistic verification, 2016–2017 GEFS probabilistic SCP_{CP} forecasts on random practically perfect years are more skillfull than climatology. ROC curves indicate that this increased skill is due to improvement over climatology at higher probability of SCP_{CP} exceedance values.

Additional analysis showed the largest number of forecast false alarms in southern Texas, whereas the greatest number of forecast misses were found across regions of the U.S. High Plains. The greatest number of forecast hits were found across central portions of Oklahoma and Kansas. Future studies may want to examine how to improve forecast skill in the geographic locations featuring many misses and false alarms. In this study, we were only able to examine two recent severe weather seasons, and it is likely that further information could be gleaned from additional analysis of other years and seasons. Furthermore, other parameters may be more skillful in forecasting U.S. tornado and hail frequencies. We chose the SCP in this study due to its common use in the operational weather forecasting community and attempts at diagnostic verification by previous research. Most importantly, these results provide an important baseline for the improvement of tornado and hail forecasts by ensemble-based dynamical prediction systems in the short to medium forecast range.

Acknowledgments
GEFS data are available from the NCEI HAS data archive at https://www.ncdc.noaa.gov/has/HAS.DsSelect.

References

Bunkers, M. J., Klimowski, B. A., Zeitler, J. W., Thompson, R. L., & Weisman, M. L. (2000). Predicting supercell motion using a new hodograph technique. *Weather Forecasting*, *15*(1), 61–79.

Carbin, G. W., Tippett, M. K., Lillo, S. P., & Brooks, H. E. (2016). Visualizing long-range severe thunderstorm environment guidance from CFSv2. *Bulletin of the American Meteorological Society*, *97*(6), 1021–1031.

Doswell, C. A. III, Davies-Jones, R., & Keller, D. L. (1990). On summary measures of skill in rare event forecasting based on contingency tables. *Weather Forecasting*, *5*(4), 576–585.

Gallo, B. T., Clark, A. J., & Dembek, S. R. (2016). Forecasting tornadoes using convection-permitting ensembles. *Weather Forecasting*, *31*, 273–295.

Gallo, B. T., Clark, A. J., Smith, B. T., Thompson, R. L., Jirak, I., & Dembek, S. R. (2018). Blended probabilistic tornado forecasts: Combining climatological frequencies with NSSL–WRF ensemble forecasts. *Weather Forecasting*, *33*(2), 443–460.

Gensini, V. A., & Ashley, W. L. (2011). Climatology of potentially severe convective environments from the North American regional reanalysis. *Electronic Journal of Severe Storms Meteorology*, *6*, 1–40.

Gensini, V. A., Mote, T. L., & Brooks, H. E. (2014). Severe-thunderstorm reanalysis environments and collocated radiosonde observations. *Journal of Applied Meteorology and Climatology*, *53*, 742–751. https://doi.org/10.1175/JAMC-D-13-0263.1

Herman, G. R., Nielsen, E. R., & Schumacher, R. S. (2017). Probabilistic verification of Storm Prediction Center convective outlooks. *Weather Forecasting*, *33*(1), 161–184. https://doi.org/10.1175/WAF-D-17-0104.1

Hitchens, N. M., & Brooks, H. E. (2012). Evaluation of the Storm Prediction Center's day 1 convective outlooks. *Weather Forecasting*, *27*(6), 1580–1585.

Hitchens, N. M., & Brooks, H. E. (2014). Evaluation of the Storm Prediction Center's convective outlooks from day 3 through day 1. *Weather Forecasting*, *29*(5), 1134–1142.

Hitchens, N. M., Brooks, H. E., & Kay, M. P. (2013). Objective limits on forecasting skill of rare events. *Weather Forecasting*, *28*(2), 525–534.

Johns, R. H., & Doswell, C. A. III (1992). Severe local storms forecasting. *Weather Forecasting*, *7*(4), 588–612.

Lepore, C., Tippett, M. K., & Allen, J. T. (2017). ENSO-based probabilistic forecasts of March–May US tornado and hail activity. *Geophysical Research Letters*, *44*, 9093–9101. https://doi.org/10.1002/2017GL074781

Lepore, C., Tippett, M. K., & Allen, J. T. (2018). CFSv2 monthly forecasts of tornado and hail activity. *Weather Forecasting*, *33*, 1283–1297.

Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology and Climatology*, *12*(4), 595–600.

Potvin, C. K., Broyles, C., Skinner, P. S., Brooks, H. E., & Rasmussen, E. (2019). A Bayesian hierarchical modeling framework for correcting reporting bias in the U.S. tornado database. *Weather and Forecasting*, *34*(1), 15–30.

Rasmussen, E. N. (2003). Refined supercell and tornado forecast parameters. *Weather Forecasting*, *18*(3), 530–535.

Rasmussen, E. N., & Blanchard, D. O. (1998). A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Weather Forecasting*, *13*(4), 1148–1164.

Schaefer, J. T., & Edwards, R. (1999). The SPC tornado/severe thunderstorm database. In *Preprints, 11th conf. applied climatology*, Amer. Meteor. Soc., Dallas, TX, pp. 215–220.

Thompson, R. L., Edwards, R., Hart, J. A., Elmore, K. L., & Markowski, P. (2003). Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Weather Forecasting*, *18*, 1243–1261.

Trapp, R. J., Diffenbaugh, N. S., & Gluhovsky, A. (2009). Transient response of severe thunderstorm forcing to elevated greenhouse gas concentrations. *Geophysical Research Letters*, *36*, L01703. https://doi.org/10.1029/2008GL036203

Verbout, S. M., Brooks, H. E., Leslie, L. M., & Schultz, D. M. (2006). Evolution of the US Tornado database: 1954-2003. *Weather Forecasting*, *21*, 86–93.

Zhou, X., Zhu, Y., Hou, D., Luo, Y., Peng, J., & Wobus, R. (2017/07/19). Performance of the new NCEP Global Ensemble Forecast System in a parallel experiment. *Weather Forecasting*, *32*, 1989–2004.