Response to all reviewers for egusphere-2025-3145

We would like to thank the reviewers for their comments and constructive feedback. Please see our responses below, which will be provided in blue font.

OTHER COMMUNITY COMMENTS

Comment CC1 (Adam Scaife): At line 50 this manuscript discusses stochastic or partial physics approaches to event sets but this omits a whole new set of literature using full physics models to create large event sets using ensembles that are physically plausible. These include:
Extreme heatwave days: https://rmets.onlinelibrary.wiley.com/doi/10.1002/wea.7741
Monsoon rainfall: https://iopscience.iop.org/article/10.1088/1748-9326/ab7b98
and Extreme rainfall: https://www.nature.com/articles/s41612-020-00149-4.
Reply to Comment CC1 (Adam Scaife): We discuss the use of physics based models, e.g., large ensembles or reforecasts, to generate event sets and/or calculate extreme statistics starting at L71. We do cite the Kelder et al. (2020) paper. We will include the other suggested papers in this section.

Reviewer 1

This article describes a technique to generate synthetic tornado events from a reforecast dataset from the Global Ensemble Forecast System (GEFS). Synthetic events are created from short- and long-lead GEFS reforecasts by using simulated environments to predict the probability of a tornado outbreak occurrence, using previously establish formulas from Malloy and Tippett (2024). Precipitation, SRH, and CAPE are used empirically to derive a map of probabilities defining the outbreak index. The probabilities are further used to compute the expected number of outbreak tornadoes given the environment. Notably, these probabilities and indeces are computed across all GEFS ensemble members (5-11 members depending on the initialization) to build up a larger dataset of outbreak tornadoes than what is provided in the current observational record. These simulated environments are plausible realizations of tornado occurrences even if they didn't happen exactly in the past – they could in the future!

From this index the authors construct return intervals for outbreak tornadoes, comparing the GEFS-based intervals to those from the observation record, and further delineate how return intervals can describe seasonal differences, reporting trends (i.e., 2000-2009 vs. 2010-2019), climate teleconnections (ENSO), and local/regional tornado risk. A limited observational record cannot extrapolate beyond approximately 1 in 40 yr events while the simulated GEFS environments provide context for 1-in-100 and 1-in-1000 yr events at daily, yearly, and local levels.

I found the paper to be extremely easy to follow and carefully constructed – small technical edits provided below regarding the use of articles. I think this type of work will have broad implications across the insurance/reinsurance industry as the authors suggest, but also in the severe storms community as we grapple with how to improve our observational records (e.g., using radar observations to supplement the tornado record, hail record, etc.). My biggest comment on the manuscript is related to the physical constraints on the system that either allow, or don't allow as I suspect, the realization of 1-in-100 or 1-in-1000 yr outbreak tornado events. The authors propose that a 1-in-1000yr event in Nashville would amount to 17 tornadoes in a single day within a 1 degree by 1 degree box. If those tornadoes are equally spaced, they are separated by 33km in any E-W or N-S direction. That is an incredible density of distinct tornadoes/parent storms that doesn't seem physically realistic, both in part because of what we know about how severe storms interact with one another on storm-scales and how environments are impacted by storms. Shouldn't there be some physical limits to the number of tornadoes on any given day in an area? Extreme rainfall as a corollary does not have the same spatial constraints as a tornado – tornadoes only happen in one area of a storm whereas extreme rainfall is possible throughout (with preferential regions). So we can have side-by-side grid points/pixels with rainfall exceeding some return interval threshold but cannot assume that with tornadoes. So don't we have a physical limit to the number of realized tornadoes for a given time period? I don't think the authors can address this per se but I do think the authors should mention this in their concluding thoughts that this type of statistical technique may need to be adjusted to accommodate physical constraints that the physical imposes on tornado frequencies.

Reply: Thank you for the constructive feedback. We appreciate the comments about the physical constraints on the system/approach. We agree that the index approach does not represent storm-scale processes or features. To address those questions regarding physical constraints, high-resolution models, e.g. HRRR or WRF, could be implemented to incorporate storm-scale characteristics, such as tornado path/swath, or to parameterize physical constraints, such as limits to number of daily tornadoes generated in grid cell. However, the goal of this study was to capture tornado outbreak statistics at a broader spatial scale and therefore it is beyond the scope of this study to assess mesoscale physical constraints. For interest of reviewer, we calculated the total E/EF1+ tornadoes per grid cell per convective day in SPC report data. The histogram of tornadoes per grid cell per convective day in 1979–2022 SPC report data is shown in Fig. R1. We found that the greatest total E/EF1+ tornadoes per day per grid cell recorded was on 1992 June 15 with 17 tornadoes occurring in a 1°x1° grid cell (39°N, 98°W). In fact, if not strict with the convective day definition, the SPC report data has a few instances when >20 tornadoes occurred in a single grid cell for a 24-hour period.

We will address/discuss this limitation and possible future work and in the Conclusions section around L380: "Future work could address the physical constraints on the approach. For instance, the 100+ year return levels of tornadoes per day in Dallas, Nashville, and Chicago in Fig. 9 might not be physically realistic considering storm-scale processes. High-resolution models, e.g., High-Resolution Rapid Refresh (HRRR) model, could be used to determine the physical constraints of how many tornadoes could occur in a grid cell per day, and/or storm-scale quantities used in short-term forecasting, such as updraft helicity, could provide insight into subgrid processes. The main drawback of using high-resolution forecast models is the relatively smaller ensemble sizes and forecast lengths, reducing the sample size and hence the spatial smoothness of risk maps. Overall, our approach is consistent with the use of large-scale environments, and while does not capture storm-scale processes, it has a major advantage of generating large sample sizes for estimating extremes."

Specific comments

C 1.1 — Line 107: Are the 6-hourly periods the standard 00-06, 06-12 UTC, etc.? Or arbitrary?

Reply: The 6-hourly periods are the standard 00-06, 06-12, 12-18, and 18-00 UTC.

We will add that clarification to sentence at L108: "A one in the gridded SPC report dataset means that an outbreak tornado occurred in a given grid cell and 6-hourly period (00-06, 06-12, 12-18, or

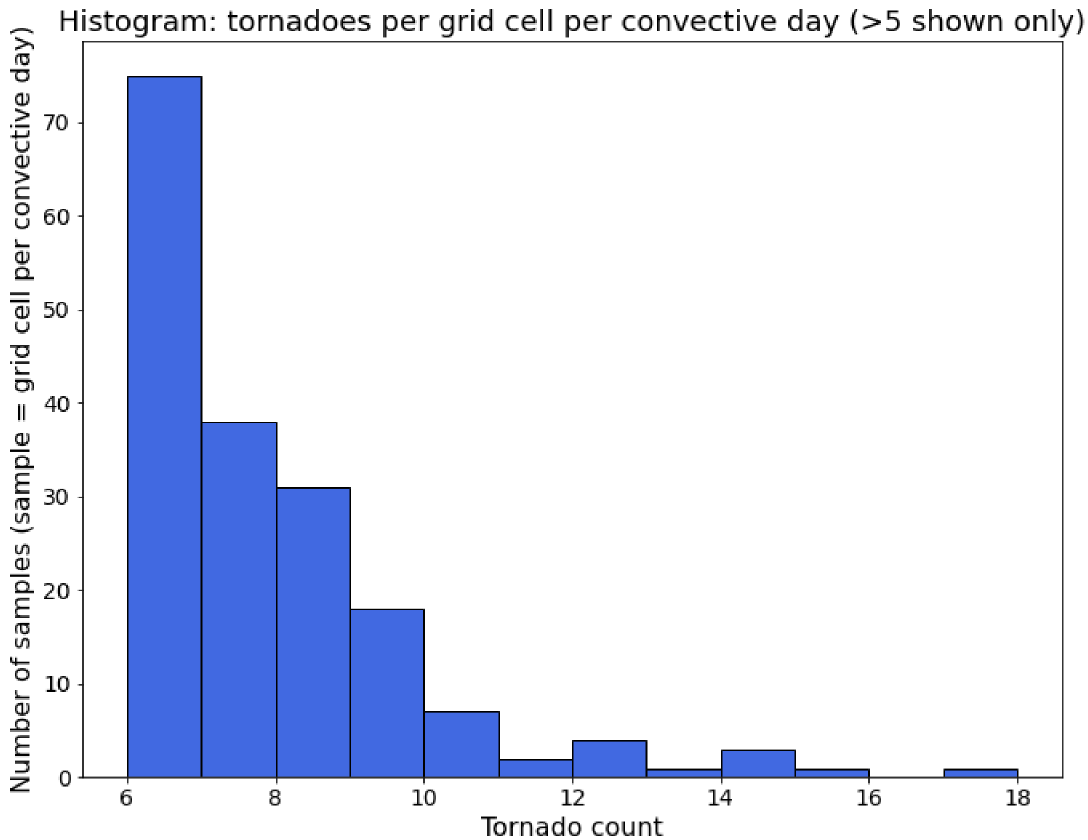

Figure R1: Histogram of number of tornadoes per grid cell per convective day (12 UTC – 12 UTC) in SPC report data.

18-00 UTC), and a zero means that an outbreak tornado did not occur.".

C 1.2 — Line 120: I assume you mean you interpolate GEFS data to the 1x1 degree boxes? Native, raw grid spacing for some variables is 0.25deg x 0.25deg.

Reply: We mean interpolate GEFS data to 1°x1° resolution. We will change "aggregate" to "interpolate" on L120 for clarification.

C 1.3 — Line 127: Do you also aggregate all reports over the same daily, 1200-1200 UTC period? This isn't clear from previous paragraphs, which specifies a 6-hourly aggregation not 24 hours.

Reply: Yes, we aggregate the reports to daily (12-12 UTC period) to evaluate the (daily) tornado outbreak index. First, we aggregate to 6-hourly before applying part 1 of the Malloy and Tippett (2024) tornado outbreak index to GEFS (and NARR), which estimates 6-hourly likelihood. After that step, we aggregate the GEFS (and NARR) data to daily temporal resolution by taking the maximum likelihood map from 12, 18, 00, 06 UTC time steps to compute part 2 of index, the PDF of total U.S. tornado counts for the day.

To clarify the procedure, we will update the text starting around L123: "The first part provides a map of the *6-hourly* probability of outbreak tornado occurrence:

$$\log\left(\frac{p}{1-p}\right) = -20.2 + 0.76\log(\mathsf{CP}) + 1.82\log(\mathsf{SRH}) + 0.51\log(\mathsf{CAPE}) \tag{1}$$

"

And on L125: "The GEFS-based index is computed from the 6-hourly values of CP, SRH, and CAPE from individual GEFS ensemble members. In order to compute the second part of the index, the 6-hourly maps are aggregated to daily maps by taking the maximum value (probability/likelihood) at each grid cell over the convective day (12 UTC–12 UTC)."

C 1.4 — Lines 135-146: I think this section could use some rewording and/or word smithing. The reference to 'second part' of the index and equation two, which I originally thought was the second part of the index, is a bit confusing. I recommend adding that mu is computed across all daily maps to derive the PDF (as I understand it), which then allows you to randomly sample that PDF. The phrasing that you can generate random realizations of tornado occurrence based on "the same daily map" is inferring a random sample of the derived binomial PDF, but that is not explicit and can be confusing since mu doesn't vary as a function of a single map (equation 2).

Reply: Equation 2 as well as the equation on line 140 comprise the second part of the index. We do not compute mu for all daily maps to derive a PDF. The idea is that by relating mean (μ) to the variance (σ^2), we can generate a PDF and randomly draw from it. To make this clear, we will add the equation on line 140 as Equation 3 and state that together they are used to generate the PDF, i.e., the mean (and variance) in total U.S. is a function of a single map (aggregated from 6-hourly maps). Our reply to C 1.3 above should also clarify.

To clarify the procedure, we will update the text starting at L133: "The second part of the index calculates the probability distribution of the number of U.S. outbreak tornadoes using one daily probability maps at a time from above via negative binomial regression (Malloy and Tippett, 2024). We recalculated the coefficients for the second part of the index since the NARR-based probability maps are post-calibrated, though results are similar if using coefficients from the second part of the index from Malloy and Tippett (2024). The equation for the expected number of outbreak tornadoes based on a probability map for each day is:

$$\mu = \exp\{-1.14 + 2.16\log[\mathsf{sum}(\mathsf{P}_{\mathsf{CONUS}})] - 0.60\log[\mathsf{max}(\mathsf{P}_{\mathsf{CONUS}})]\} \tag{2}$$

where $\mathsf{sum}(\mathsf{P}_{\mathsf{CONUS}})$ is the index map sum and $\mathsf{max}(\mathsf{P}_{\mathsf{CONUS}})$ is the index map maximum. In the negative binomial regression, the variance (σ^2) is related to the mean (μ) via an overdispersion parameter:

$$\sigma^2 = \mu + 13.74\mu \tag{3}$$

Equation 3 makes it possible to generate random, or stochastic, realizations of tornado occurrence based on the same daily map of the environments/index. For each daily map of the index, we can generate as many outbreak events as desired, which we call $n_{\mathsf{realizations}}$, i.e., we can draw $n_{\mathsf{realizations}}$ samples of total U.S. outbreak tornadoes from the probability distribution of part 2 of the index (cf. Equations 2 and 3) which are all physically consistent with the large-scale environment. Thus, we further increase the sample size. Then, for each realization, we can *populate tornado locations based on map probabilities* from the first part of the index (cf. Eq. 1)."

C 1.5 — Line 145: When you say "populate random locations", don't you mean populate based on weighted locations? It's not purely random if you have a probability field to weight where the sample of outbreak tornadoes (i.e., a pure count) should be placed.

Reply: Good point, it is weighted by the probabilities. We will revise to "populate tornado locations based on map probabilities" in the text, also seen in C 1.4 response above.

C 1.6 — Lines 178-193: "total U.S. tornadoes" appears to be a separate designation from "total U.S. outbreak tornadoes" – is this purposeful by the authors? I was under the assumption that mu is the expected number of outbreak tornadoes, but in Line 178 mu is referenced as the total U.S. tornadoes. Based on my understanding, all references to "total U.S. tornadoes" are really "total U.S. outbreak tornadoes" and if this is true, the text should be revised for clarity.

Reply: It is correct that μ refers to expected number of total U.S. outbreak tornadoes. Part 2 of the index actually predicts the PDF. The definition of tornado outbreak we use to train for the index is 6 tornadoes that happen in short time span (Malloy and Tippett, 2024; Fuhrmann et al., 2014). For the vast majority of cases, this means 6 tornadoes per convective day. Based on the way the index is designed, the expected number per day could be fewer than 6 U.S. tornadoes per day (i.e., non-outbreak) but there are possible outcomes from the PDF where it could be 6 or more U.S. tornadoes per day (outbreak).

For sake of clarity, we will update all instances of "total U.S. tornadoes" to "total U.S. outbreak tornadoes" for L170-199. It is technically correct, and our example subsampling procedure ignores days with $\mu < 6$ tornadoes anyway.

C 1.7 — Figure 2: The dots displaying observed reports are rather small and get obscured by the contour lines surrounding them. Moreover, the colors denoting number of reports can only be seen when zooming into the manuscript .pdf at extreme percentages. The depiction of storm reports should be reimagined to better convey these observations. One recommendation is to contour only (i.e., no dots) to remove one overlapping piece of information. Alternatively, the color shading of number of reports could be used alone, although I would recommend a different color scheme so even single reports are visible (i.e., not near white shading).

One other recommendation for Figure 2 is to add the total number of observed reports in the top left along with the mu parameter. It would be a good piece of information to include for comparing expected report numbers to true observations on the figure itself without having to refer back to the text (line 223).

Reply: We agree that visualization of Figure 2 could be improved, and we appreciate the suggestions by reviewer. We removed the report dots, made the observed report contour line thicker and blue, and adjusted the color shading scheme for the reports to contrast against the index shading scheme (making it 50% transparent for discerning their overlap). We also zoomed into CONUS more for this figure. We added the total report number to the top left panel (which would be the same for all panels so we did not want to duplicate the info) and moved the index μ to the bottom-left of all panels. See Figure R2 below for revised Figure 2.

C 1.8 — Line 225: The authors suggest that 1-day and 6-day forecasts are relatively skillful in predicting tornado outbreaks, but it would appear for this example case (Figure 2) that none of

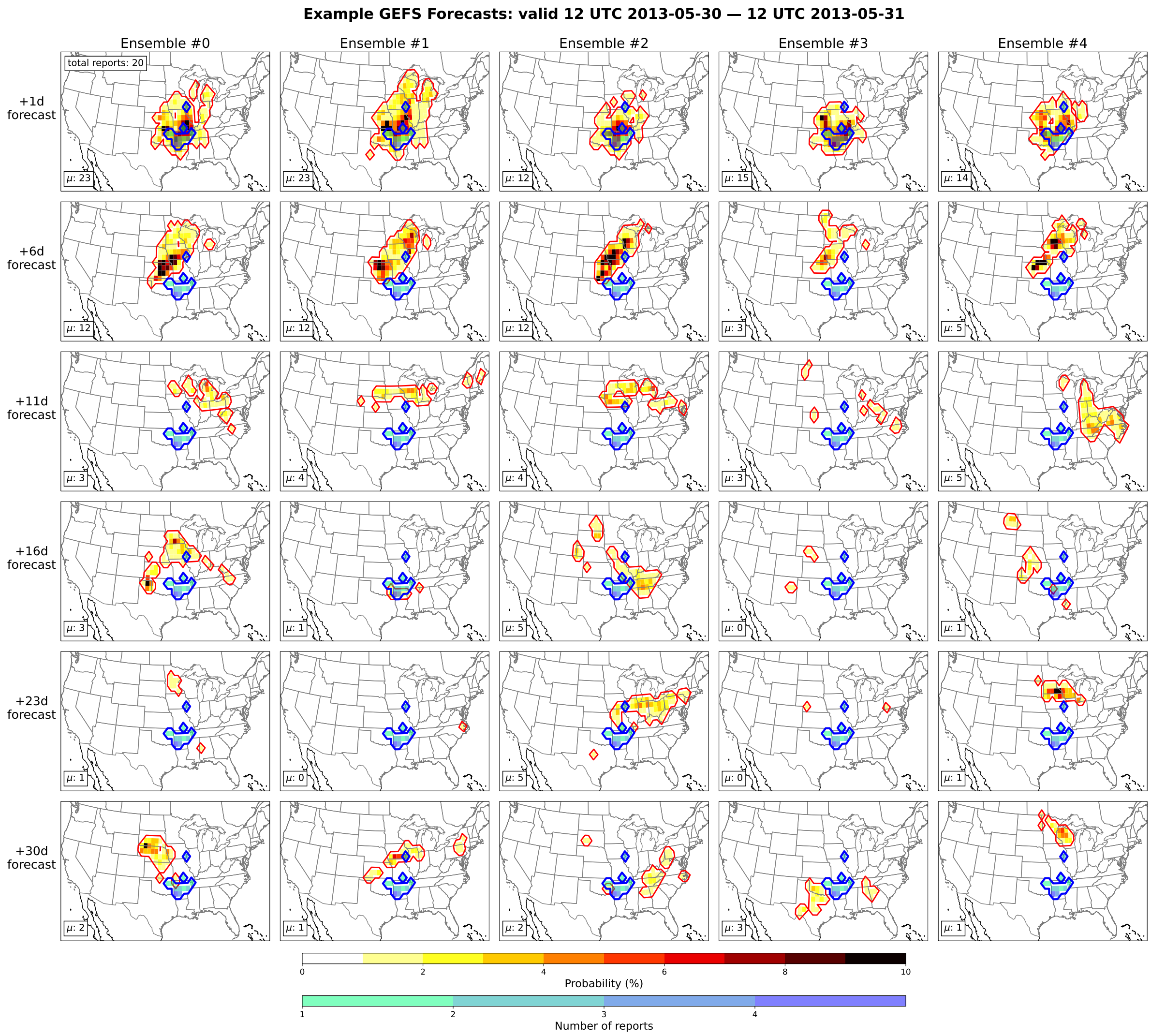

Figure R2: Example of GEFS synthetic events via its forecasts for 2013 12 UTC May 30 through 12 UTC May 31, where rows indicate corresponding forecast lead for event, and columns indicate corresponding GEFS ensemble member (only up to 5 members): tornado outbreak index (yellow-red-black shading) versus observed reports (green-blue-purple shading). Expected number (μ) of tornadoes based on outbreak index part 2 given in bottom-left of all panels. Total number of reports for observed event (20 reports) also in top left panel.

the observation locations verify as tornado outbreaks (i.e., > 6 tornadoes, Line 103); southwest Arkansas point perhaps has this criteria met but it's not discernable from the data provided in the figure. So how do the authors arrive at this conclusion that these are skillful forecasts of tornado outbreak potential? I see a generally skillful forecast in tornado location, regardless of count, but I don't believe that is what this outbreak parameter is truly identifying.

Reply: We agree that this might not be the perfect example of high skill for day 6, but this is just one case/event and should not be used to make broad conclusions about skill. In general, it is a good example of (1) 6-day forecast showing relatively high CONUS-wide outbreak risk and (2) high predictability since the ensemble members resemble each other, both of which are not seen for the longer lead forecasts. In addition, we do calculate skill metrics in Figure 1 and in the Malloy and Tippett (2025) study that we can reference here to provide evidence that tornado outbreak skill is relatively high at 6-day forecast leads.

We will add this nuance to the text starting around L123: "For the 1-day forecasts, the tornado outbreak index likelihood well matches the observed reports in regards to location and extent of event. For the 6-day forecasts, the tornado outbreak index likelihood also matches the observed reports in terms of predicting elevated CONUS outbreak risk, though the risk is shifted slightly more north compared to the observed event. In general, these shorter-lead forecasts demonstrate relatively high prediction skill."

C 1.9 — Figure 3: Recommend changing the color scheme so the smallest identified tornado per season value can still be visually seen – the light yellow blends into the white background so it is indistinguishable.

Reply: We will update color scheme to improve visualization of Figure 3. See revised Figure R3 below.

C 1.10 — Figure 4: Can you add vertical dashed lines at the return period thresholds (nominally just 1, 10, 100, and 1000) so it's easier to see the return period thresholds for the report/GEFS return curves, like you do in Figure 7? This could help when interpreting the graphics. (Same recommendation for Figure 5)

Reply: We will add dashed lines (as in Figure 7) to represent 1-, 10-, 100-, and 1000-year return levels for GEFS in Figure 4 and 5. See Figures R4 and R5 below.

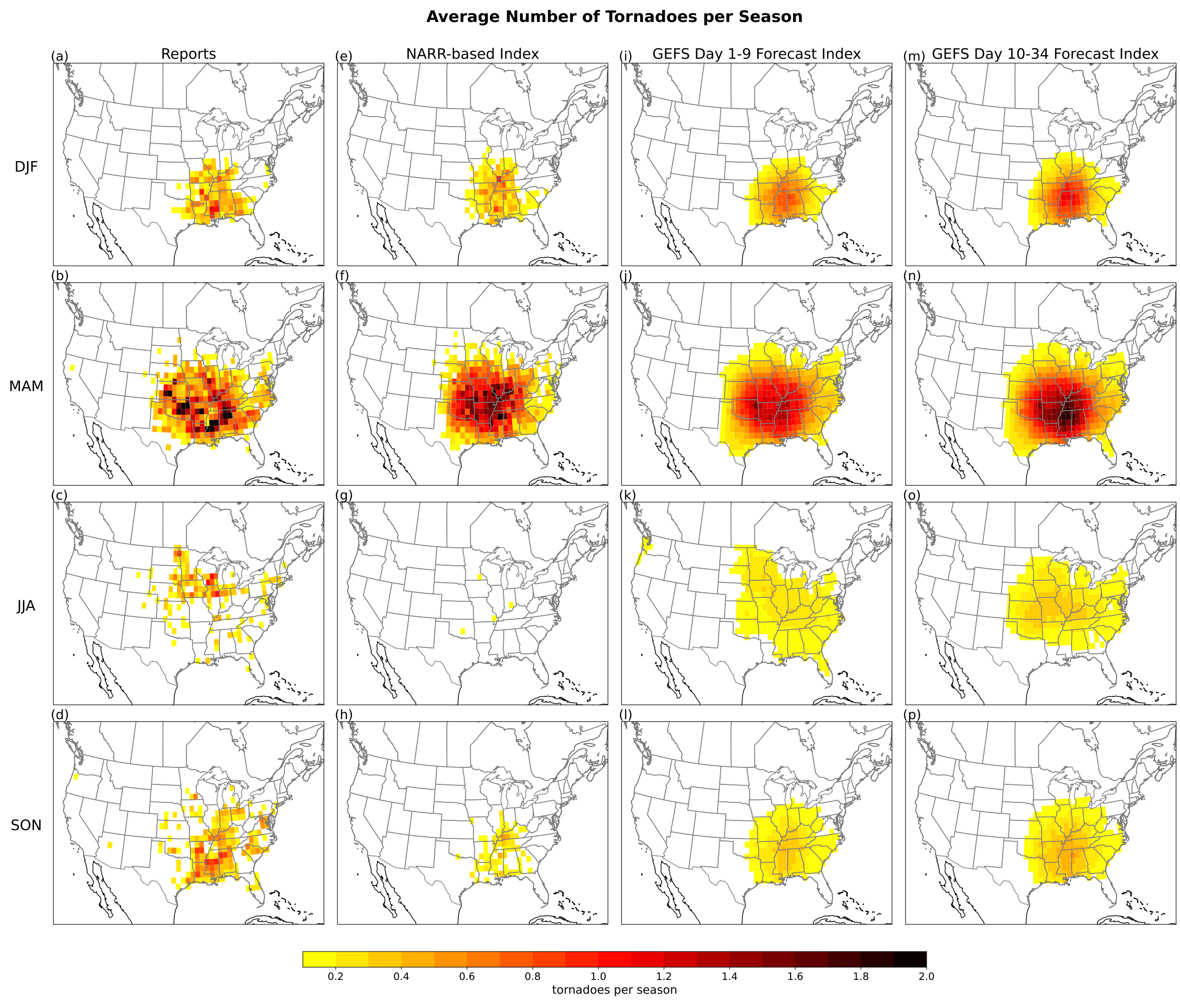

Figure R3: Average expected (μ) number of outbreak tornadoes during (top row) Dec-Feb, (second row) Mar-May, (third row) Jun-Aug, and (last row) Sept-Nov, calculated from (a-d) reports, (e-h) NARR-based index, (i-l) GEFS short-lead (day 1-9 forecasts) index, and (m-p) GEFS long-lead (day 10-34 forecasts) index.

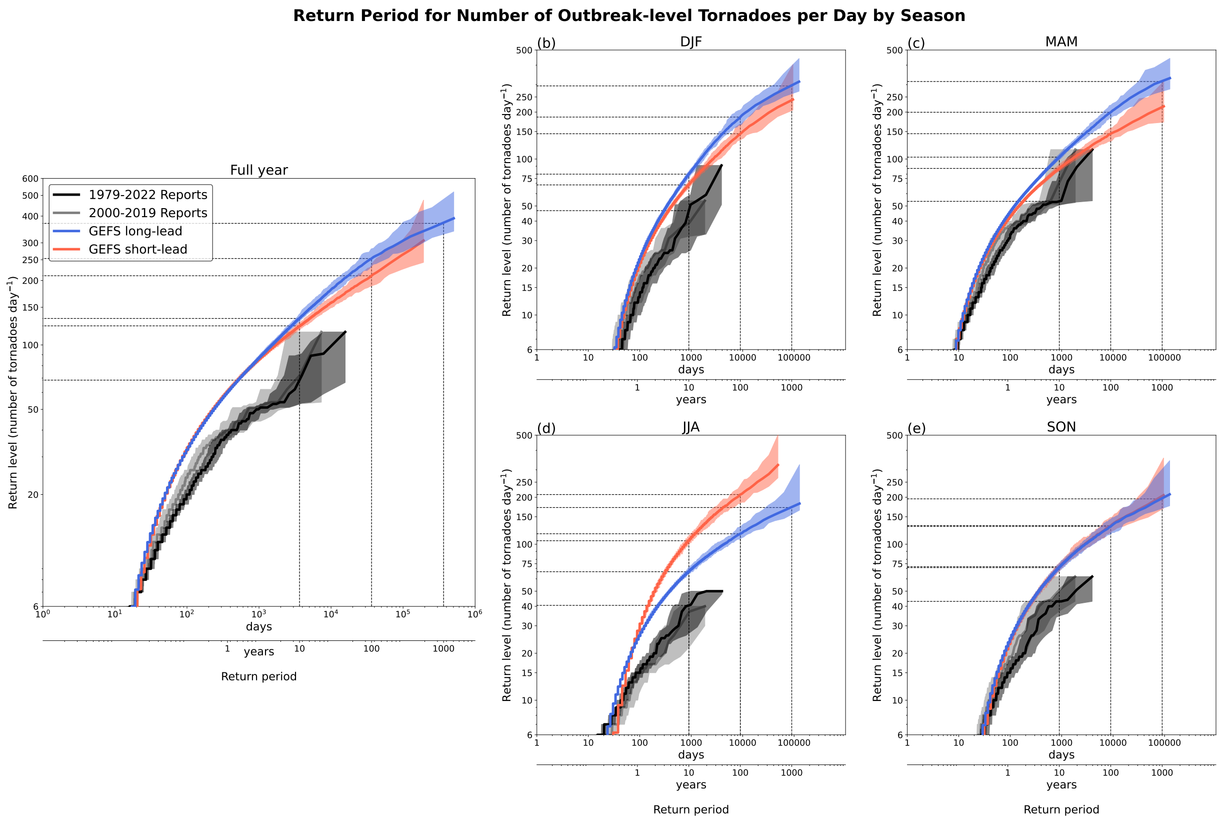

Figure R4: Return level curves for number of outbreak-level tornadoes per day from (black line) 1979–2022 reports, (gray line) 2000–2019 reports, to be consistent with GEFS time period, (blue line) GEFS long-lead forecast, and (orange line) GEFS short-lead forecasts, for (a) full year of data, and for (b) DJF, (c) MAM, (d) JJA, and (e) SON. Shading indicates sampling uncertainty. Dotted lines highlight the 10-, 100-, and 1000-year return period levels.

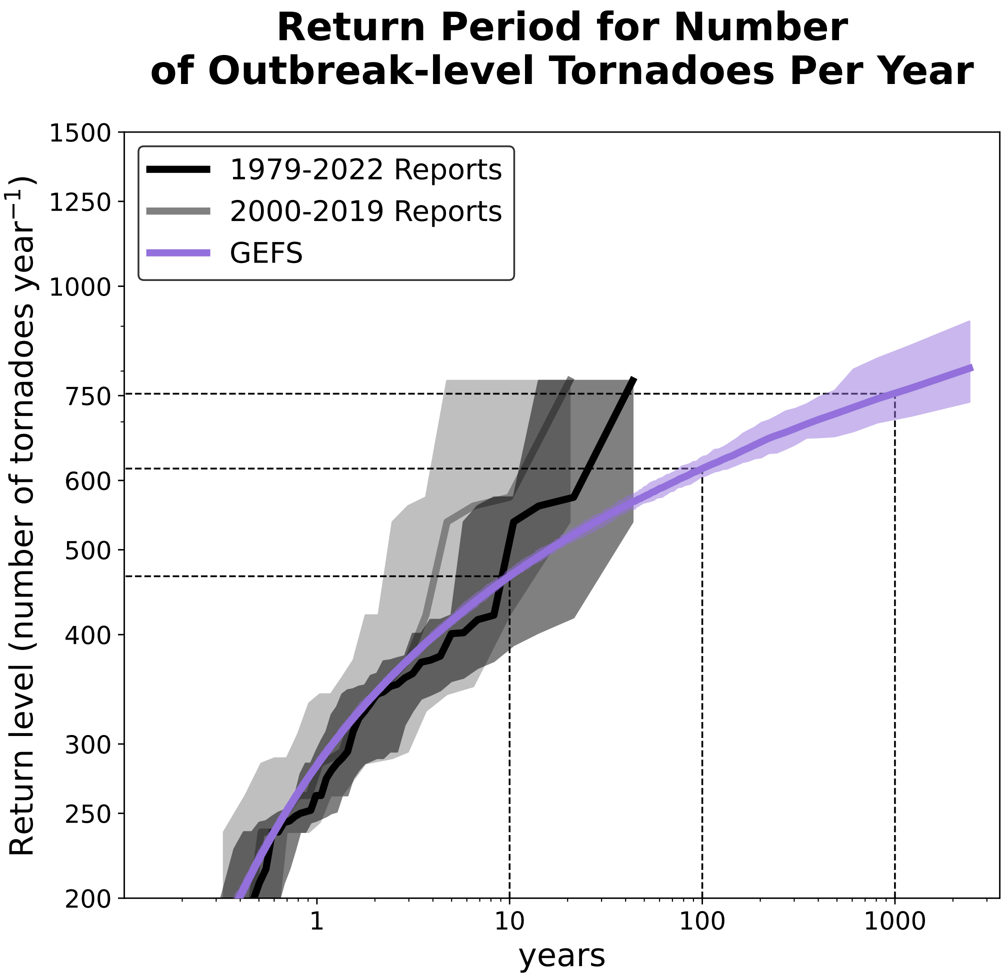

Figure R5: Return level curves for number of outbreak-level tornadoes per year for (black line) 1979–2022 reports, (gray line) 2000–2019 reports, and (purple line) GEFS synthetic events. Dotted lines highlight the 10-, 100-, and 1000-year return period levels.

Technical edits

C 1.11 — Line 143: "total U.S. outbreak tornadoes"

Line 160: "ensemble member j index value at grid cell"

Line 162: "Frequencies" changed to "tornado frequency" – mismatch between singular "an" and plural "frequencies"

Line 175: "of the original"

Line 215: "the 14-day forecast lead"

Line 229: "from the observed event"

Line 231: "11+ day forecasts"

Line 241: "the sporadic, rare nature"

Line 246: "might explain increased mean" – this phrase needs an article ("the" increased mean, "an" increased mean)

Lines 383-385: See Das and Allen (2024) for hail return interval estimation:
https://www.nature.com/articles/s44304-024-00052-5

Reply: We will make the necessary technical edits as suggested here.

We will also include the Das and Allen (2024) study to our Conclusions about generating a hail synthetic event set around L382: "In Das and Allen (2024) study, extreme hail likelihood was generated using fitted extreme value models, a pure statistical approach."

Thank you to the reviewer for their careful examination of the manuscript.

Reviewer 2

In this paper, the authors provide and analyze a set of tornado outbreaks synthetically generated by the GEFS. This allows for the estimation of rare events, such as 1-in-100-year and even 1-in-1000-year tornado outbreaks, enabling the assessment of the most extreme events. Moreover, teleconnection ENSO influence on tornado outbreak activity is investigated.

This work is highly valuable and well-structured. The main objectives are clear, the methodology is robust, and the results are clearly presented and thoroughly discussed. The conclusions are consistent with the findings presented in the paper. Only some clarifications are needed.

Therefore, I recommend accepting this manuscript after the authors address a few minor revisions.

Reply: Thank you for the positive feedback. We have addressed each of your comments below.

Specific comments

C 2.1 — Line 65: for which period have these "upward trends in tornado (outbreak) activity" been detected? Please, specify it.

Reply: In the Malloy and Tippett (2024) study, the time period for calculating the tornado outbreak activity trends is 1979–2021. For the other studies, the time period for calculating tornado trends ranges, e.g., 1960–2022 for Graber et al. (2024) study and 1979–2017 for Gensini and Brooks (2018) study.

We will clarify in the text starting L63: "Malloy and Tippett (2024) detected upward trends in 1979–2021 tornado outbreak activity, especially for the winter and spring seasons, consistent with the upward trend in 1979–2015 tornado outbreak activity from Tippett et al. (2016), the upward trend in 1960–2022 tornado outbreak days from Graber et al. (2024), and the upward trend in 1979–2017 STP from Gensini and Brooks (2018).

C 2.2 — Line 103: are you only taking into account CONUS tornadoes in outbreak definition, or all the USA? On the other hand, are you working with "tornado outbreak days"? (e.g., if > 6 tornadoes occur one day and > 6 the day after with no more than 6 hours between consecutive tornadoes, you consider it as one or two tornado outbreaks?). It should be clarified.

Reply: We only consider CONUS tornadoes. We do define "outbreak-level" tornado as when there are no more than 6 hours between tornadoes, regardless of day, so that if the outbreak tornadoes happen on different outbreak days it does not affect the labeling. We do not count "tornado outbreak days" in this study, only consider shifts in number of outbreak tornadoes per day, outbreak tornadoes per season, or outbreak tornadoes per year. Though beyond the scope of this study, counting "outbreak days" could be included in future work.

We will update the text to clarify the definition of outbreak-level tornado around L103: "We define a tornado outbreak when six or more tornadoes occur *over the contiguous U.S. (CONUS)* with no more than 6 hours between consecutive tornadoes."

C 2.3 — Line 107 and 114: why the 6-hourly resample and 1 deg. resolution? It would be grateful to justify it in the text.

Reply: We resample to 6-hourly and 1x1° grid to match the spatial and temporal resolution of many forecast and reanalysis datasets. GEFS output is every 6 hours (00, 06, 12, 18 UTC) and and 1x1° grid.

We will update the text to justify our choice of 6-hourly and 1x1° resolution (around L107): "We construct a dataset of outbreak tornado occurrence at a 6-hourly and 1°x1° resolution from the SPC report data. The 6-hourly and 1°x1° resolution matches that of many weather forecast models (and reanalysis datasets), and this resolution is useful for simulating tornado outbreak activity from large-scale meteorological environments."

C 2.4 — Lines 27 to 29: in Brooks (2004) tornado path lengths and widths are compared to its F-scale rating, but no EF-scale. Please, change EF by F. Moreover, take into account that, as explained in that paper, "The mean width was reported prior to and including 1994 and the maximum width after 1994".

Reply: Thank you for catching this. We will update the text around L27: "The spatial extent of tornadoes is small: The average path length of a tornado ranges from 4-5 kilometers for F1 to 44-55 kilometers for F4/5, and the average width of a tornado ranges from 64 meters for F1 to 460-555 meters for F4/5 (though reports reflect mean width until 1994 and maximum width after 1994; Brooks, 2004)."

C 2.5 — Line 112-115: the use of NARR data is not clearly explained in the Data section. Please, consider adding a sentence here clarifying the specific purpose for which it is being used.

Reply: NARR is used to construct a reanalysis-based dataset of the tornado outbreak index calculated from 20 years of historical meteorological environments. This is to be compared with a GEFS-based dataset of the tornado outbreak index calculated from 2000 years of meteorological environments (including events resembling historical events and events that are realistic but unrealized). As described around L240, the GEFS-based dataset better resolves the climatological risk of tornado outbreaks, demonstrating the number of events needed to "converge" statistics of (outbreak) tornadoes.

We will update the text to clarify use of NARR later around L125: "The GEFS-based index is computed from the 6-hourly values of CP, SRH, and CAPE from individual GEFS ensemble members, which is resampled to a daily resolution by taking the maximum value over the convective day (12 UTC–12 UTC). This results in a probability map for each day. The NARR-based index is similarly computed with 6-hourly values of observed/historical CP, SRH, and CAPE and is aggregated to a daily resolution. The NARR-based index represents tornado outbreak occurrence calculated from the 20 years of historical, realized meteorological environments." And L131: "With 20 years of reforecasts initialized daily, each with 5-11 ensemble members and being run out to 16-35 days, GEFS provides 889,514 daily maps (equates to over 2000 years) of outbreak tornado probabilities. Hence, the GEFS-based index represents tornado outbreak occurrence from meteorological environments closely resembling historical, realized meteorological environments *as well as* unrealized meteorological environments."

C 2.6 — Line 120: the data aggregation for GEFS is as for NARR (the sum for CP and 6-h average for SRH and CAPE)? Which is the original temporal and spatial resolution for GEFS? It should be stated in the main text.

Reply: The GEFS v12 reforecasts have an original spatial resolution of 0.25x0.25° for first 10 days of forecasts and 0.5x0.5° after 10 days. The original temporal resolution is 3-hourly output for first 10 days of forecasts and 6-hourly output after 10 days.

We will add this to text around L120: "Reforecasts are originally provided with 3-hourly output and 0.25°x0.25° spatial resolution for the first 10 days of a forecast, and with 6-hourly output and 0.5x0.5° after the first 10 days of a forecast. To keep a consistent resolution, and to match the outbreak tornado occurrence data from reports and NARR, GEFS data is aggregated to 6-hourly output and 1°x1° spatial resolution."

Technical edits

C 2.7 — Line 11: in the main text you refer to tornado intensity as F/EF. Please, replace EF/F1+ by F/EF1+ for consistency.

Line 25: it would be fine to provide a reference about F and EF scales, for example Fujita (1971) for F-scale and WSEC (2006) for EF-scale.

Fujita T.T. (1971): Proposed characterization of tornadoes and hurricanes by area and intensity. SMRP Research Paper, 91: 48.

WSEC, 2006. A Recommendation for an Enhanced Fujita Scale (EF-scale). http://www.spc.noaa.gov/faq/torna

Figure caption 3: (m) to (p) maps are GEFS extended (day 10-34 forecasts) index, not short-lead (day 1-9 forecasts) index.

Line 321: Fig. 9c-e does not exist (it is Fig. 9a-c)

Line 322: there is an extra space between "further into the extremes" and the dot.

Reply: We will change all instances of "EF/F+" to "F/EF1+."

We will also include the Fujita (1971) and WSEC (2006) studies around L25.

We will fix the Figure 3 caption as well as the typos on L321-322.

Thank you to the reviewer for their careful examination of the manuscript.

References

Das, S. and Allen, J. T.: Bayesian estimation of the likelihood of extreme hail sizes over the United States, npj Natural Hazards, 1, 47, 2024.

Fuhrmann, C. M., Konrad, C. E., Kovach, M. M., McLeod, J. T., Schmitz, W. G., and Dixon, P. G.: Ranking of tornado outbreaks across the United States and their climatological characteristics, Weather and Forecasting, 29, 684–701, 2014.

Gensini, V. A. and Brooks, H. E.: Spatial trends in United States tornado frequency, npj Climate and Atmospheric Science, 1, 38, 2018.

Graber, M., Trapp, R. J., and Wang, Z.: The regionality and seasonality of tornado trends in the United States, npj Climate and Atmospheric Science, 7, 144, 2024.

Malloy, K. and Tippett, M. K.: A Stochastic Statistical Model for U.S. Outbreak-level Tornado Occurrence based on the Large-scale Environment, Monthly Weather Review, https://doi.org/https://doi.org/10.1175/MWR-D-23-0219.1, 2024.

Malloy, K. and Tippett, M. K.: Forecasting U.S. Tornado Outbreak Activity and Associated Environments in the Global Ensemble Forecast System (GEFS), Weather and Forecasting, 2025.

Tippett, M. K., Lepore, C., and Cohen, J. E.: More tornadoes in the most extreme US tornado outbreaks, Science, 354, 1419–1423, 2016.