egusphere-2026-536

Seasonal prediction of springtime tornado activity in the United States using a hybrid model

Graber et al.

Authors: We appreciate the reviewer's comments and suggestions, and that they believe our work has valuable contribution to the scientific community. Our responses to their comments and suggestions our below in blue.

Recommendation: major revisions

General Comments

The purpose of this study was to evaluate spring prediction skill of U.S. tornado activity by using April-initialized forecasts of weather regime (WR) frequency to model U.S. tornado and tornado outbreak activity in April-May. Prediction skill is explained and interpreted by stratifying forecast skill by modes of climate variability, linking modes of climate variability to favorability of tornado-related environments, and linking WR frequency to the modes of favorability. Finally, the authors examine SST patterns for WR phases. In general, seasonal WR information was most skillful for tornado outbreak predictions, and, when stratifying by climate mode, was more skillful during AO+, NAO+, PNA+, and El Niño years. The WRs associated with greater tornado outbreak activity was associated with respective climate mode that was linked to favorable environments, and vice versa. This is valuable research for the community. However, the authors need to address several concerns before publication. I outline them below.

Major comments

1. The delineation between what information comes from WRs and what information comes from the AO/NAO/PNA climate modes is unclear and arbitrary. I would like the authors to give more thought into how to present the "pathway" or route from more low-frequency modes (ENSO, SSTs) to daily synoptic setup for a tornado (outbreak) day.

   a. The authors aggregate AO/NAO/PNA over April-May to define a low-frequency climate mode. However, AO/NAO/PNA have daily indices too that, similar to WR, explain more synoptic-scale atmospheric variability. Therefore, are the WRs even necessary, i.e., do they give redundant information that forecasts of daily index of AO/NAO/PNA would provide?

This is a good point by the reviewer. First, weather regimes allow us to zoom in on the regional circulation patterns the US CONUS, not all of which are strongly related to AO/NAO/PNA. Second, we focus on the variability of the climate modes on the seasonal and longer time scales, which provide insights about the predictability of seasonal tornado activity. To clarify this, we have pointed out in the revised manuscript, "Since climate modes evolve on the seasonal and longer timescales, they may serve as sources of predictability for seasonal tornado activity. However, they do not fully capture the synoptic-scale variability of atmospheric circulation. Weather regimes (WRs) can effectively bridge the gap." (section 1), and "The impacts of some climate modes, ENSO (via Nino3.4), PNA, NAO, and AO, on tornado activity and WR counts were investigated to provide a physical basis for TD and TO predictability. We focus on their low-frequency variability on the seasonal and longer time scales" (section 2.4). This is why we did not use daily values of climate mode indices. Figure 6 further shows that low-frequency variability in climate modes can modulate WR frequency.

b. If the WRs provide synoptic-scale information, the authors should composite the CAPE and shear environments for the different WRs (rather than composite over seasonal AO/NAO/PNA phase) to compare to Figure S2 anomalous tornado (outbreak) days. Results like L230-232 (low CAPE + high VWS = active TO) is not clear/confusing and I think it might be from the way these are composited.

This is a good point made by the reviewer. Such analysis has been done in Graber et al., (2025). Figure 1 in Graber et al. (2025) shows composite CAPE and VWS anomalies for all 5 WRs over April-July, which was done to investigate the potential links between WRs and warm-season tornado activity. This study did not do any prediction tests, unlike this current study being reviewed. WR-A had anomalously low MUCAPE and VWS over the central CONUS, while WR-B had anomalously high MUCAPE and VWS over the central CONUS and Southeast. As expected, WR-A was the most unfavorable WR for tornado activity while WR-B was the most favorable. WR-C had anomalously high CAPE and anomalously low VWS over the Northern Great Plains. WR-D had anomalously high CAPE and anomalously low VWS throughout the Great Plains. WR-E had anomalously low CAPE and anomalously high VWS throughout the central CONUS. This analysis was vital in that manuscript for finding the links between WRs and tornado activity which supported the purpose of the analysis done in the present study.

Graber, M., Wang, Z., and Trapp, R. J.: Linking weather regimes to the variability of warm-season tornado activity over the United States, Weather Clim. Dyn., 6, 807–816, https://doi.org/10.5194/wcd-6-807-2025, 2025.

c. ENSO modulates PNA too. How to detangle? This should at least be discussed in Summary section.

We added vast new discussions on each climate mode in the introduction section including a statement on how ENSO modulates the PNA:

Lines 71-75 now say: "Another important mode is the Pacific North American (PNA) pattern, which is a leading mode of low-frequency variability over the North Pacific and North America (Phillips et al., 2014)*. Although the PNA is influenced by ENSO* (Wallace and Gutzler, 1981)*, it exhibits some independent variability* (Li et al., 2019) *and can exist in the absence of interannual SST variability (Lau 1981)."*

In addition, both ENSO and PNA peak in the boreal winter which is where this relationship has been studied most extensively (Soulard et al., 2019)*. In Figure S3, composite VWS anomalies between PNA and ENSO phases are very similar but distinct changes exist for composite anomalies of CAPE, 500H, and tornado day probabilities. We did a correlation analysis between normalized indices of PNA and ENSO for December (Figure 1a) and April-May (Figure 1b) which shows that the relationship between PNA and ENSO is much weaker in the boreal spring. This helps explain the differences in tornado day/tornado outbreak probability anomalies. This is further supported by SST anomalies in ENSO and PNA years being similar but not identical (Fig. 2). This figure has been added to the supplemental information, and a statement is now included in the results section leading up to figure 5.*

Lines 281-285 now say: "Previous studies have found that ENSO modulates the phase of PNA during boreal winter (Li et al., 2019; Soulard et al., 2019)*; however, these spring results are not consistent with those findings. Accordingly, the correlation between the ENSO and PNA indices are much stronger during the peak stage of ENSO in December than during the ENSO decay stage in April-May (Fig. S5)."*

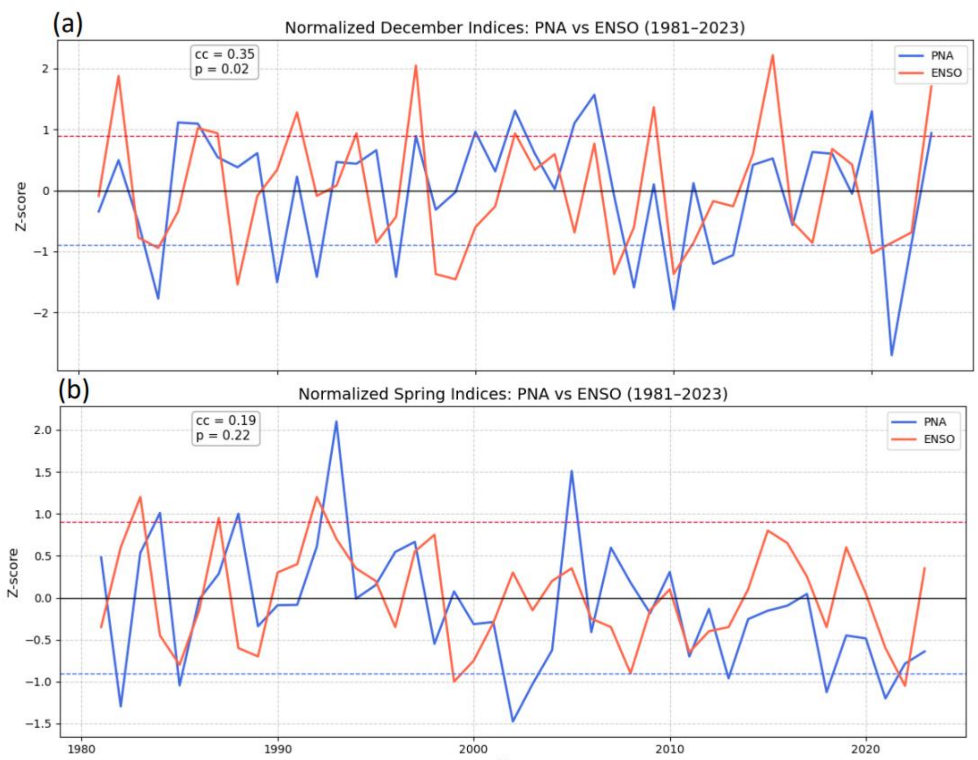

Figure 1: *Normalized indices of ENSO (red) and PNA (blue) in December (a) and April-May (b) with Spearman Rank correlation and p-value. Dashed red and blue lines represent the thresholds for positive and negative phases, respectively.*

d. Figure 7 shows the (potential) modulation of seasonal WR via SSTs. Would it be more appropriate/accurate that the pathway is that the SSTs are modulating the seasonal AO/NAO/PNA which are influencing the daily WR frequency? What would Figure 7 look like for AO/NAO/PNA phases?

This is an interesting point brought up by the reviewer. Although some specific SST signals are closely related to climate modes and may therefore modulate WR frequency through them, SST variability is not limited to signals coupled with climate modes. Figure 7 is intended for a more general examination of such linages. Figure 2 shows composite SST anomalies for the years associated with the phase of each climate mode, and ENSO was included for completeness which had the strongest SST signal which makes sense given that index is built from eastern equatorial Pacific SSTs. Composite 500H anomalies have some clear patterns in a few instances. +AO, +NAO, -NAO, and +PNA circulation patterns are quite clear which goes with 3 of the phases that have higher skill in our model. +NAO years have a pattern that is comparable to WR-A and +PNA years are comparable to WR-E. The patterns over the CONUS are what is plotted in Figure S3 in the main manuscript. This gives a look at the whole hemisphere.

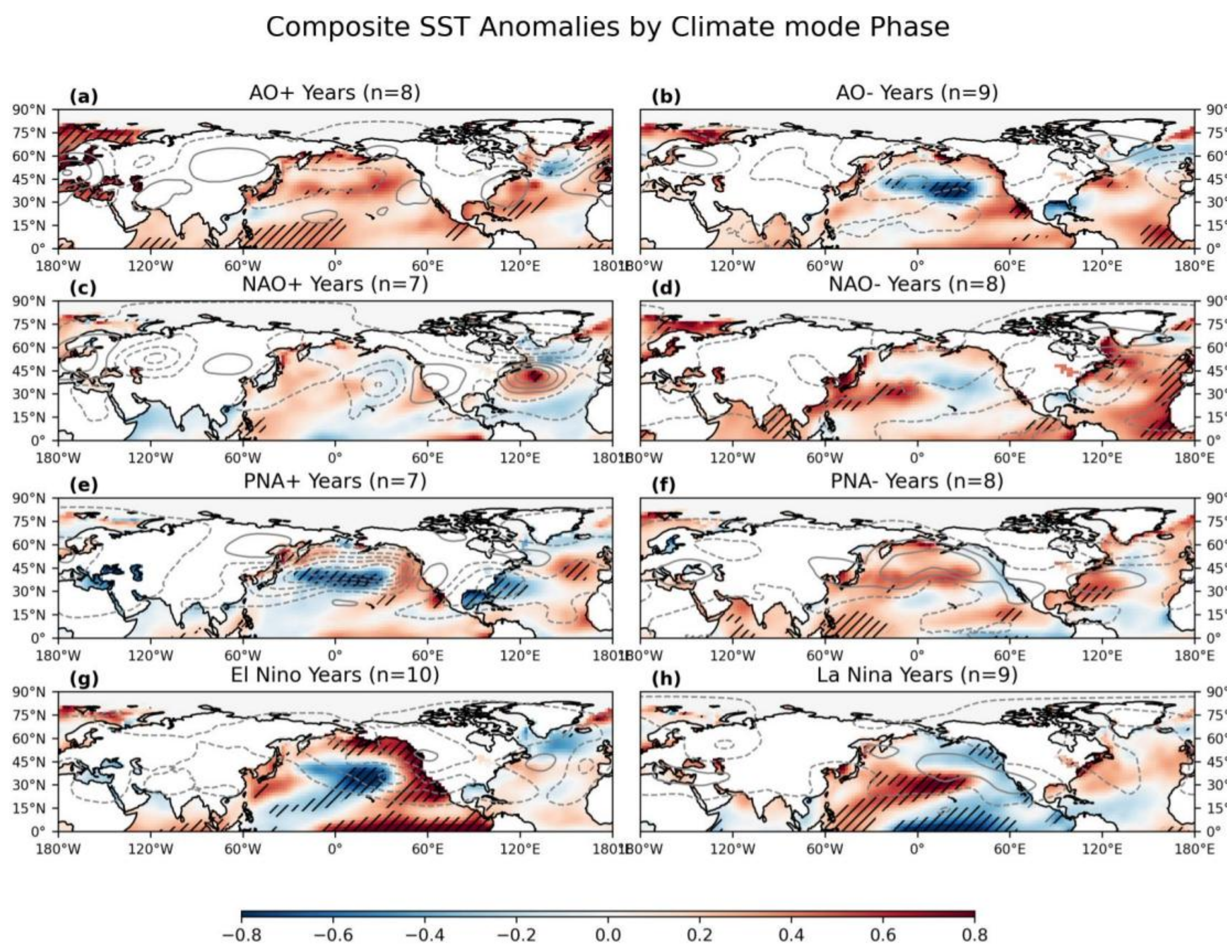

Figure 2: *Composite SST anomalies for positive and negative phases of AO (a-b), NAO (c-d), PNA (e-f), and ENSO (g-h). Hatching indicates significant SST anomalies ($p \leq 0.05$) using a one-sample, two-sided t-test. 500H anomaly contours are gray (+/- 10 m). Sample sizes are included in each panel title.*

The main purpose of the analysis as it is in the manuscript would be to use SSTs as sources of predictability for WRs which are connected to tornado activity.

2. The most skillful forecasts correspond to the phases that are unfavorable to tornado outbreaks (El Niño and PNA+), which suggests that the hybrid model performs better during inactive outbreak years, contrary to studies like Lepore et al. (2017). Could this be an artifact of using PC = correct predictions/total predictions? Or could it be that the variance is lower during inactive years, so greater signal-to-noise ratios? The authors should justify their skill metric or discuss this result more.

Lepore, C., M. K.Tippett, and J. T.Allen (2017), ENSO-based probabilistic forecasts of March–May U.S. tornado and hail activity, Geophys. Res. Lett., 44, 9093–9101, doi:10.1002/2017GL074781.

This is a very interesting point, and we appreciate this comment from the reviewer. It is true that phases that are more unfavorable for tornado activity, such as PNA+ and NAO+ are the most skillful in our model, which goes against Lepore et al. (2017).

An additional skill metric that could be used would be the threat score (Schaefer, 1990)*, which in our case we would only measure the upper tier of events using the following equation:*

$$TS = \frac{hits}{hits + misses + false\ alarms}$$

*Hits would be when both the observed and predicted time series are in the upper third. Misses would be when the observed time series are in the upper third, but the predicted time series is not. False alarms would be when the predicted time series is in the upper third, but the observed time series is not. After calculating this skill score for tornado outbreaks, the TS is **0.375** with 6 hits, 2 misses, and 8 false alarms. This would say that the model may be overpredicting outbreaks just a bit. This skill score can also be applied to the bottom third of events. Here, the TS is **0.429** with 9 hits, 7 misses, and 5 false alarms. Here, the model does get more correct, but it also misses more. The middle third had a TS of **0.308** with 8 hits, 11 misses, and 7 false alarms which suggests the model performs worse during average years, which is expected. Therefore, the skill difference is probably not an artifact.*

In addition, we found more significant SST anomalies associated with the inactive years of tornado activity (Figure 3), which may explain the different predictability levels: "We also examined the composite SST and 500H anomalies for active and inactive TO years (Fig. S6). Interestingly, significant SST anomalies are found over equatorial eastern Pacific, North Pacific, and subtropical western North Atlantic during inactive TO years, while significant SST anomalies are nearly absent during active TO years. The lack of significant SST anomalies in the active tornado years are consistent with the relatively limited SST signals during the active WR-B and WR-D years and during the inactive WR-A and WR-E years, and it helps to explain the higher predictability of inactive tornado years discussed in the previous section."

This is different from Lepore et al. (2017) probably because they used December-February ENSO state to predict March-May tornado and hail activity. Our skill test did find that El Niño springs were more skillful, but we found that El Niño springs are statistically significantly favorable for TDs in the Southern Great Plains and for TOs in the Northern Great Plains which differs from what winter ENSO phases tell us based on previous studies (Lepore et al. 2017; Allen et al. 2015; Cook and Schaefer 2008

A few statements on this are added in the section leading up to figure 5 since this is where the tornado probability anomalies for climate modes are introduced:

Lines 286-292 now state: "The model appears to perform better (Fig. 4) in years when climate modes unfavorable for TOs, such as +NAO and +PNA, are present. To further evaluate skill, the Threat Score (Schaefer, 1990) *was calculated for upper-tercile TO years. These years yield a Threat Score of* ***0.375*** *with 6 hits, 2 misses, and 8 false alarms, indicating a tendency for the model to overpredict TOs.*

For lower-tercile years, the Threat Score is slightly higher at ***0.429****, with 9 hits, 7 misses, and 5 false alarms. While lower-tercile years have greater overall skill, they also include more misses. Overall, the model demonstrates higher skill for lower-tercile years and modest skill for upper-tercile years."*

An additional statement was added in the summary as well:

Lines 427-431 now state: "These results suggest that the model performs better during years that are unfavorable for tornado activity, possibly due to a lower signal-to-noise ratio, and use of the Threat Score confirms the model is better in inactive TO years. However, this skill test still demonstrates modest skill in active TO years."

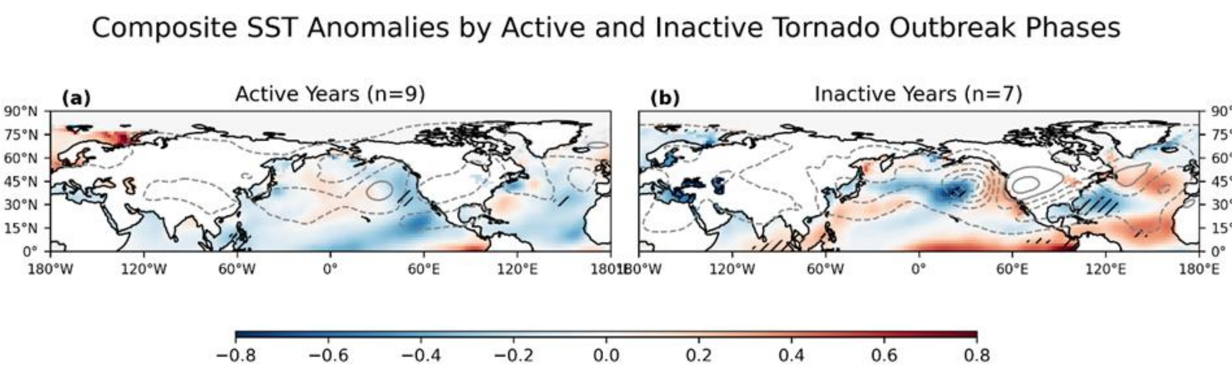

Figure 3: April composite SST anomalies for active and inactive tornado outbreak years, and hatching indicates significant SST anomalies ($p \leq 0.05$) using a one-sample, two-sided t-test. 500H anomalies are shown in gray contours. Sample sizes are included in each panel title.

3. Is all the skill coming from April WRs? I would be curious if skill is very low/negative for May WRs (initialized in April).

This is a valid question, and our study does not explicitly separate April vs May WR skill since we designed the model to assess seasonal predictability. We only use April 1st initialized forecasts for WR identification and the empirical model calculates a Tornado Index (TI) for each spring season, which does a good job at picking out some springs with several higher end events.

To partially address the concern, figure 4 shows the mean number of incorrect classifications by the 25 ensemble members relative to the ERA5 on each calendar day over the 43 years. Forecast skill of WR-classification decreases with lead-time and stabilizes around April 15th, consistent with known limits of *synoptic predictability (Lorenz 1969; Judt 2020). Likewise, individual ensemble members have limited skill in predicting tornado activity. Focusing on April and May as individual months was also a focus in* Miller et al., (2020) *which used ECMWF medium-range reforecasts and found model skill out well into 3 weeks.*

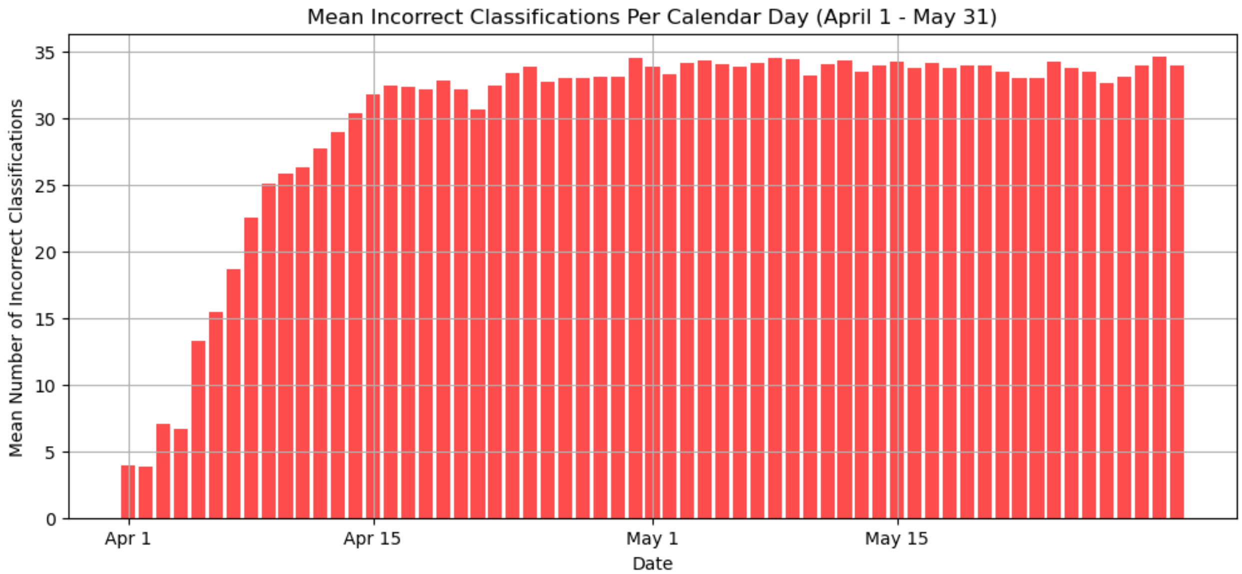

Figure 4: *Mean incorrect classifications by the 25 ECMWF seasonal forecasts ensemble members on each calendar day from April 1st-May 31st.*

The previous figure suggests that skill is higher in April, and predicting April and May separately confirms this (Figure 5). April vastly exceeds May in model skill, but when analyzing both months together we get a value that is well representative of the entire spring season. However, it is correct to say that the majority of model skill stems from April, as expected. The Threat Score for lower-tercile years in April was 0.304 with 7 hits, 9 misses, and 7 false alarms. Whereas the Threat Score for lower-tercile years in May was 0.16 with 4 hits, 11 misses, and 10 false alarms. As we showed earlier, however, the model when looking at April and May together is modestly skillful.

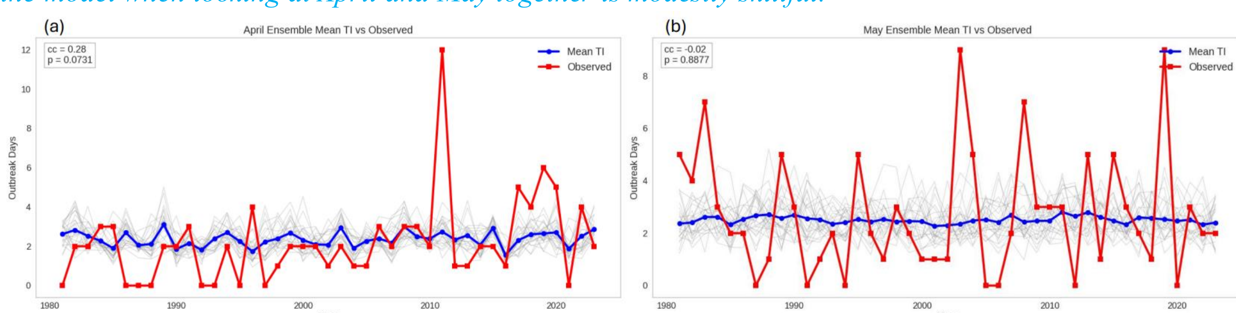

Figure 5: *April (a) and May (b) predictions vs observations with Pearson correlation coefficient and p-value.*

4. ENSO can still experience a transition in phase during April. Does it make sense to use a threshold for EN3.4 that is not fixed, e.g. 25th and 75th percentiles based on month?

This is a valid concern, and we agree that ENSO can experience phase transitions during April and May since it peaks in the boreal winter and weakens heading into Spring, which is why most studies *focus on winter ENSO phase* (Cook and Schaefer, 2008). *We want to maintain consistency among all 4 climate modes that were analyzed, and thus it makes sense to use a fixed threshold for ENSO.*

The main objective was to use low-frequency variability of climate modes to determine sources of predictability for springtime tornado activity, so we wanted one individual state for each climate mode to represent the whole season. Introducing subseasonal transitions in climate mode phase would complicate interpretation of the results when comparing to NAO, AO, and PNA given that the ECMWF seasonal forecasts are only initiated on April 1st and go out through May 31st.

Minor comments

The overuse of acronyms hurts its readability. TO vs. TD vs. TI specifically gets confusing.

This is a valid concern. Given that TI only appears in the methodology, this is spelled out in full each time it is used. We decided to keep TD and TO as they are for consistency with Graber et al. (2025). We hope this helps the readability a little bit.

L85-95: Are the anomalies detrended for k-means as in Graber et al. (2025) and Tippett et al. (2024) WR studies? I assume so based on L90 but should be stated anyway.

We chose not to detrend 500H from the ERA5 to be consistent with the ECMWF forecast data, since the latter is not detrended. We used the same weather regimes that are identified in Graber et al. (2025) as the reference regimes. We have revised the text in section 2.1, "Unlike Graber et al. (2025), we did not remove the long-term linear trend in 500H of ERA5 to be consistent with the ECMWF seasonal forecast data. Rather than re-deriving WRs for this period, we used the WRs derived by Graber et al. (2025) using the K-means clustering method for 1960-2022 April-July as the reference WR patterns. WRs were assigned by finding the reference WR pattern with the smallest Euclidean distance from the daily 500H anomalies."

Use of detrended anomalies produced similar results with worse model skill for TDs and the ECMWF forecast data had a poor time predicting the seasonal frequency of WR-C.

L221-222: Enhanced outbreak and AO- link different from other studies (Tippett et al. 2022), and the discussion of Fig 5b results does not match what is shown in Figure S2 in terms of environments.

The results shown in Figure S2 (now Figure S3) show Gaussian smoothed TD probability anomalies. Relative to other climate modes, any notable positive or negative anomalies are spatially small which would have a minor impact on TD probability anomalies that encompass an entire region, which is shown in Figure 5a. In Tippett et al. 2022, though the TEI signal is negative in April for the -AO, it is rather weak, which is consistent with our results in Figure 5a and figure S3b. The positive TO probability anomalies in figure 5b does not match Tippett et al. 2022 findings, however that study did not focus on tornado outbreaks and did not look at the month of May, either. Additionally, the anomalies are only significant in the Midwest and subsequently the CONUS, which we argue is supported by the positive VWS anomalies over this region as well as the anomalously low CAPE anomalies not covering major parts of Illinois, Indiana, and Missouri. In addition, figure S3b does show positive TD probability anomalies over a few small parts of the

Midwest where anomalously positive VWS anomalies are present, so it is not contradictory to the TO anomaly results. This is clarified in the manuscript.

Lines 252-256 now state: "Although reduced MUCAPE tends to suppress TD probability, the increased VWS supports positive TO probability anomalies when it coincides with sufficient instability (Diffenbaugh et al., 2013; Sherburn et al., 2016)*. These positive TO probability anomalies during -AO are only statistically significant over the Midwest, consistent with the region of anomalously positive VWS anomalies."*

In addition, figure 6a shows that WR-B is the most frequent WR to occur during a -AO phase, which is the most favorable WR for tornado activity. In addition, the WR spatial structure has strong resemblance to the negative phase pattern for winter surface pressure anomalies shown by the NOAA Climate Prediction Center (https://www.climate.gov/news-features/understanding-climate/climate-variability-arctic-oscillation).

L287: "WR-E occurs more frequently in the positive phase of ENSO…" I also note that WR-E is more frequent during La Niña (stat. sig. too) and should be mentioned.

A statement is added to address this with the clarification that WR-E's occurrence during La Nina is still less than any other WR.

Lines 326-328 now states: "Relative to the neutral phase, WR-E also occurs more frequently during the negative phase of ENSO, which is statistically significant. However, WR-E's occurrence during the negative phase of ENSO is still less than any other WR."

L397: Add "statistically" in front of "insignificant" and consider changing instances of "patchy" to "incoherent"/"inconsistent"

All instances where "insignificant" is said, the word "statistically" is placed in front. 'Patchy' is changed to "incoherent" in both instances.

Figure 1: Fix panels f and h legend.

Done

Figure 7: Perhaps show sample sizes on panels.

Yes, the SST figure on revised version of the manuscript will include samples.

Our References

Cook, A. R. and Schaefer, J. T.: The relation of El Nino-Southern Oscillation (ENSO) to winter tornado outbreaks, Mon. Weather Rev., 136, 3121–3137, https://doi.org/10.1175/2007MWR2171.1, 2008.

Diffenbaugh, N. S., Scherer, M., and Trapp, R. J.: Robust increases in severe thunderstorm environments in response to greenhouse forcing, Proc. Natl. Acad. Sci. U. S. A., 110, 16361–16366, https://doi.org/10.1073/pnas.1307758110, 2013.

Duan, W. and Wei, C.: The “spring predictability barrier” for ENSO predictions and its possible mechanism: Results from a fully coupled model, Int. J. Climatol., 33, 1280–1292, 2012.

Graber, M., Wang, Z., and Trapp, R. J.: Linking weather regimes to the variability of warm-season tornado activity over the United States, Weather Clim. Dyn., 6, 807–816, https://doi.org/10.5194/wcd-6-807-2025, 2025.

Li, X., Hu, Z.-Z., Liang, P., and Zhu, J.: Contrastive influence of ENSO and PNA on variability and predictability of North American winter precipitation, J. Clim., 32, 6271–6284, https://doi.org/10.1175/JCLI-D-19-0033.1, 2019.

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

Schaefer, J. T.: The critical success index as an indicator of warning skill, Weather Forecast., 5, 570–575, https://doi.org/10.1175/1520-0434(1990)005%3C0570:TCSIAA%3E2.0.CO;2, 1990.

Sherburn, K. D., Parker, M. D., King, J. R., and Lackmann, G.: Composite environments of severe and nonsevere high-shear, low-CAPE convective events, Weather Forecast., 31, 1899–1927, https://doi.org/10.1175/WAF-D-16-0086.1, 2016.

Soulard, N., Lin, H., and Yu, B.: The changing relationship between ENSO and its extratropical response patterns, Sci. Rep., 9, https://doi.org/10.1038/s41598-019-42922-3, 2019.

Wallace, J. M. and Gutzler, D. S.: Teleconnections in the geopotential height field during the Northern Hemisphere winter, Mon. Weather Rev., 109, 784–812, https://doi.org/10.1175/1520-0493(1981)109%3C0784:TITGHF%3E2.0.CO;2, 1981.