Manuscript ID: egusphere-2024-3216

We appreciate the comprehensive and constructive questions and suggestions from the reviewers. Below are our responses in blue italics. Any necessary revisions made to our manuscript are also indicated below.

Reviewer # 1:

The regime classification methodology differs from standard ones (EOFs are not used, there is no time filtering, once per day hourly snapshots of 500 hPa heights are used, the choice of cluster number is subjective, there is no normalization of variance), and there is no indication how these methodological choices impact the results. Variance normalization is important because 500 hPa height anomalies have substantially greater variance in April than in July. The k-means clustering method minimizes within-cluster variance, and seasonality in variance might bias the results. Consequently, it may well be the case that the weather regime frequencies vary with month (climatologically), which would confound any analysis with tornado frequency whose climatology also varies by month. Whether this is case or not is unclear because diagnostics such as the seasonality of regime frequency, variance explained, association with modes of large-scale variability, etc. are missing.

We thank the reviewer for raising this important question. "There is no unique or optimal way of classifying weather regimes" (Robertson and Ghil 1999). *In particular, Falkena et al. 2020 (https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3818) argued against the use of either EOFs or time filtering on top of K-means clustering because K-means clustering reduces the dimensions and temporal filtering changes the frequencies of weather regimes. In addition, our analysis shows that the application of the 5-day running mean or EOF dimension reduction prior to K-means does not qualitatively affect the regime patterns or the regime frequencies (Figs. R1-R3). We thus chose to use the simplest procedures for regime classification.*

Whether the data should be normalized is an interesting question for debate. We choose not to normalize the data for a couple of reasons. First, our weather regime analysis covers only four months, and seasonality is thus not as much of an issue, especially since we have removed the seasonal cycle (defined as the long-term daily mean on each calendar day). Second, our focus is on the link between weather regimes and tornado activity, which is quantified much differently than weather regimes. Tornado activity is not normalized for the same reason, given that we are working in one season where tornado activity is relatively common throughout the season, seasonality is not much of an issue. Since we do not normalize tornado activity (and the associated environmental parameters), we believe it is better not to normalize H500 for consistency.

*"*While many previous studies applied a low-pass filter or/and EOF dimension reduction prior to K-means clustering analysis (e.g., Robertson et al., 2020, Lee et al. 2023), Falkena et al. 2020 cautioned against the use of either EOFs or time filtering on top of K-means clustering. Our analysis shows that the application of the 5-day running mean or EOF dimension reduction prior

to K-means does not qualitatively affect the regime patterns or the regime frequencies (Figs. R1-R3). We thus chose to use the simplest procedures for regime classification."

The choice of k in K-means clustering is often somewhat subjective, because a metric does not always indicate an unambiguous optimal cluster number, and different metrics may yield different optimal cluster numbers (Dorrington and Strommen 2020, https://doi.org/10.1029/2020GL087907). This is a known limitation of K-means. We tried k=4 and 5. With k=4, 3 of the 4 regimes in our analysis are similar to those in Lee et al. (2023): WR-D and their Greenland High, WR-C and their Alaskan Ridge, and WR-A and their Pacific Trough, but it misses WR B in the k = 5 analysis, which is spatially similar to WR-A in Miller et al. 2020 and the Pacific Ridge in Lee et al. 2023 and is most favorable to tornado activity. We thus chose k=5 to incorporate this important regime.

Though some spatial similarities exist, there are some differences as well, which is not unexpected given our focus on one season. These similarities and differences have been further discussed in our study. Additionally, it is worth pointing out that the optimal number of WRs is higher likely because we used the total 500H anomalies and did not apply the EOF dimension reduction. (Falkena et al. 2020).

"Some WRs are similar to the year-round WRs in Lee et al., (2023), which were subsequently used by *Tippett et al. (2024*). More specifically WR-A features spatial similarities to a Pacific Trough, WR-B and WR-D show warm and cool phases of a Pacific Ridge and WR-E is characterized by an Alaskan Ridge. WR-C features spatial similarities to a Greenland High as well. It is worth mentioning our study focuses on a different region, a different season and chooses a different k value, and there are thus noticeable differences. WR-A features two anomalous highs over the two coasts as opposed to one anomalous high over the central-CONUS. The anomalous low in WR-B is more pronounced than in Lee et al., (2023). The anomalous high in WR-C is wavelike unlike the Greenland high in Lee et al., (2023). The dipoles in WR-E are further south than they are in the Alaskan Ridge in Lee et al., (2023)."

The 'once-per-day snapshot' approach was pursued because the chosen time (2100 UTC) of 500H represents a typical time of day when U.S. tornado outbreaks are ongoing (Cwik et al. 2022), thus potentially providing a more straightforward connection for WRs to serve as an intermediate between climate change and tornado activity.

Note, Figures R1-R3 have been added into the supplemental information as Figures S2-S4. The manuscript has been updated accordingly.

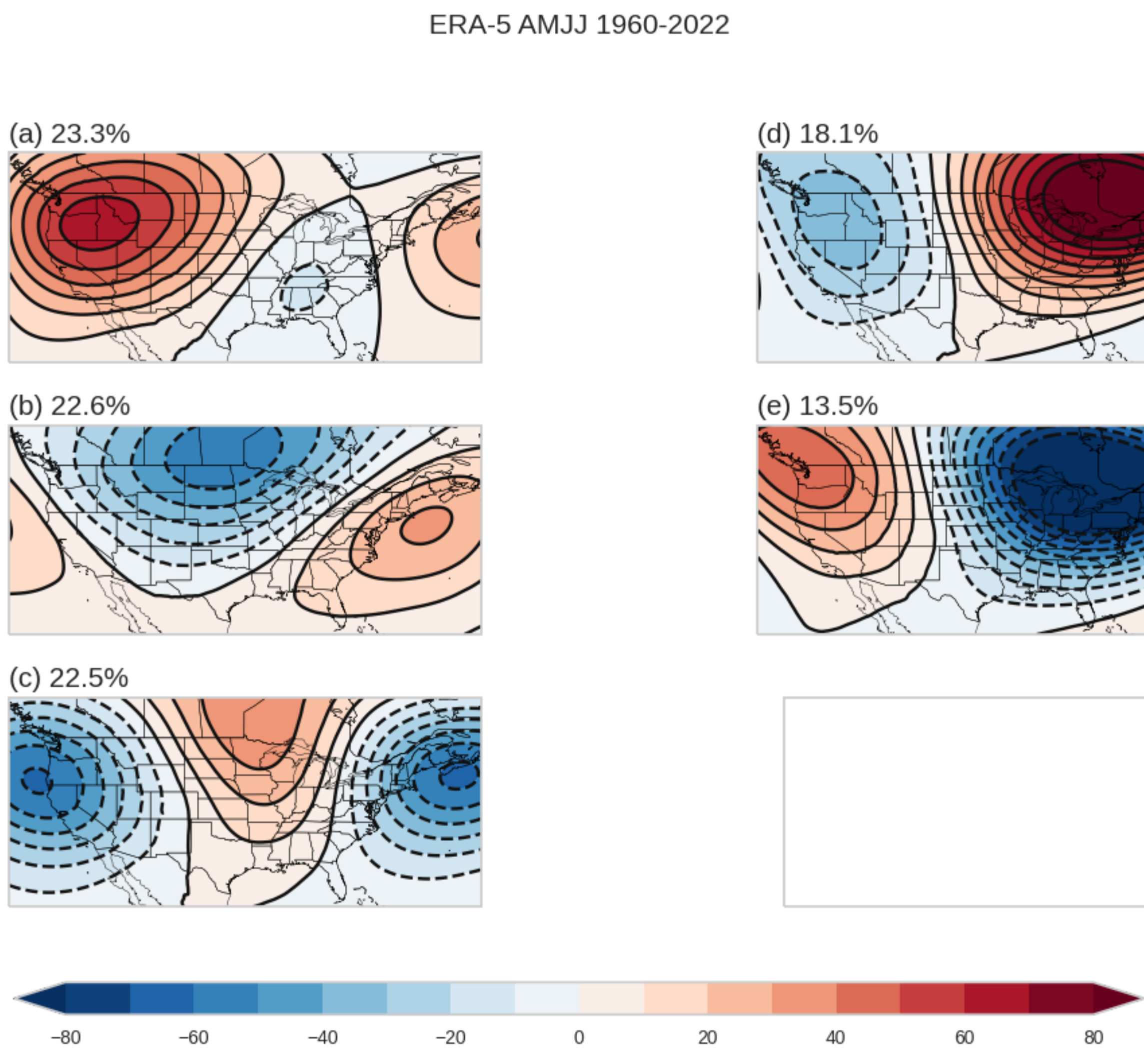

Figure R1: Original 500H weather regimes in the manuscript.

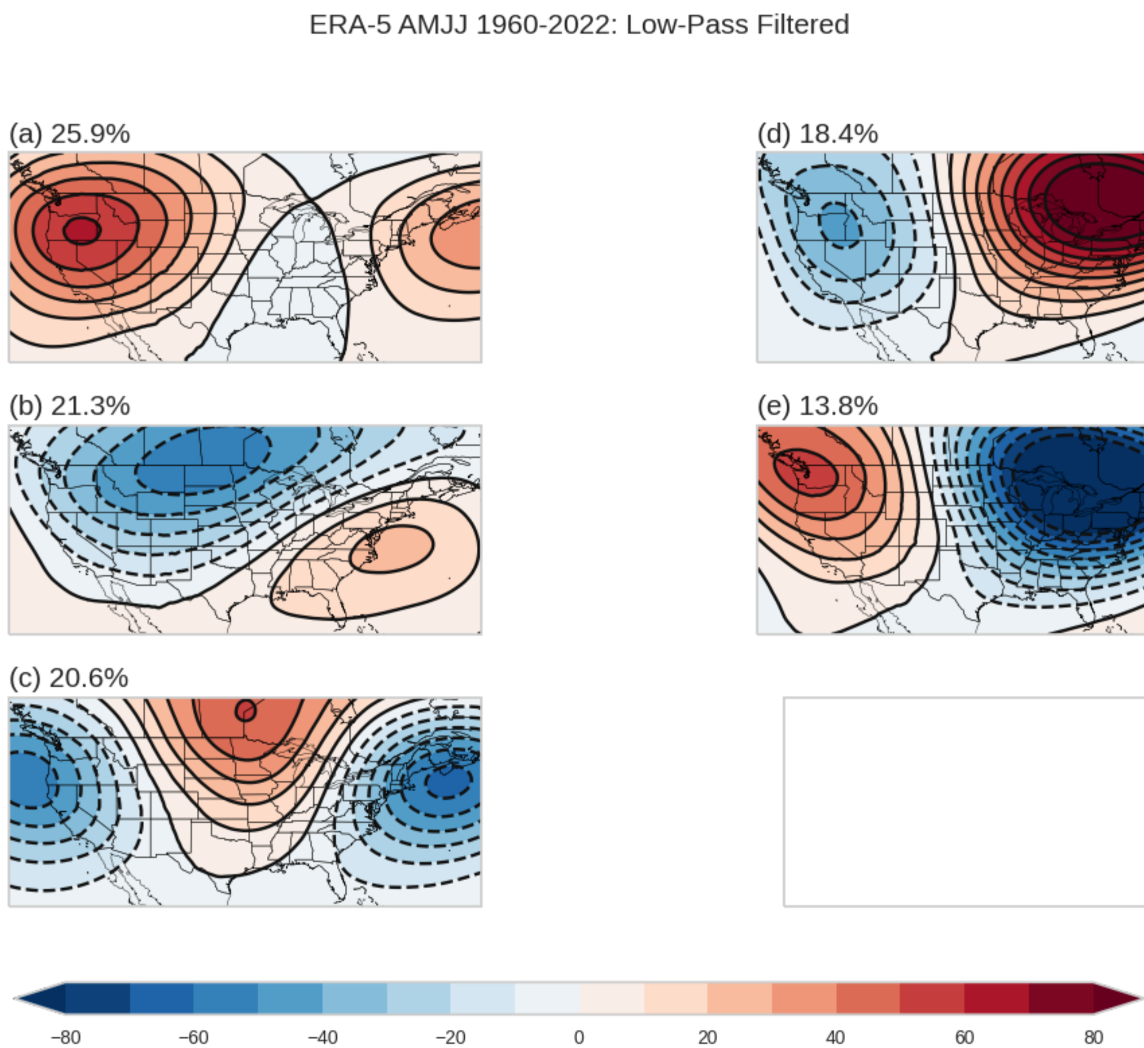

Figure R2: 500H weather regimes created applying a 5-day low-pass filter prior to K-means clustering analysis.

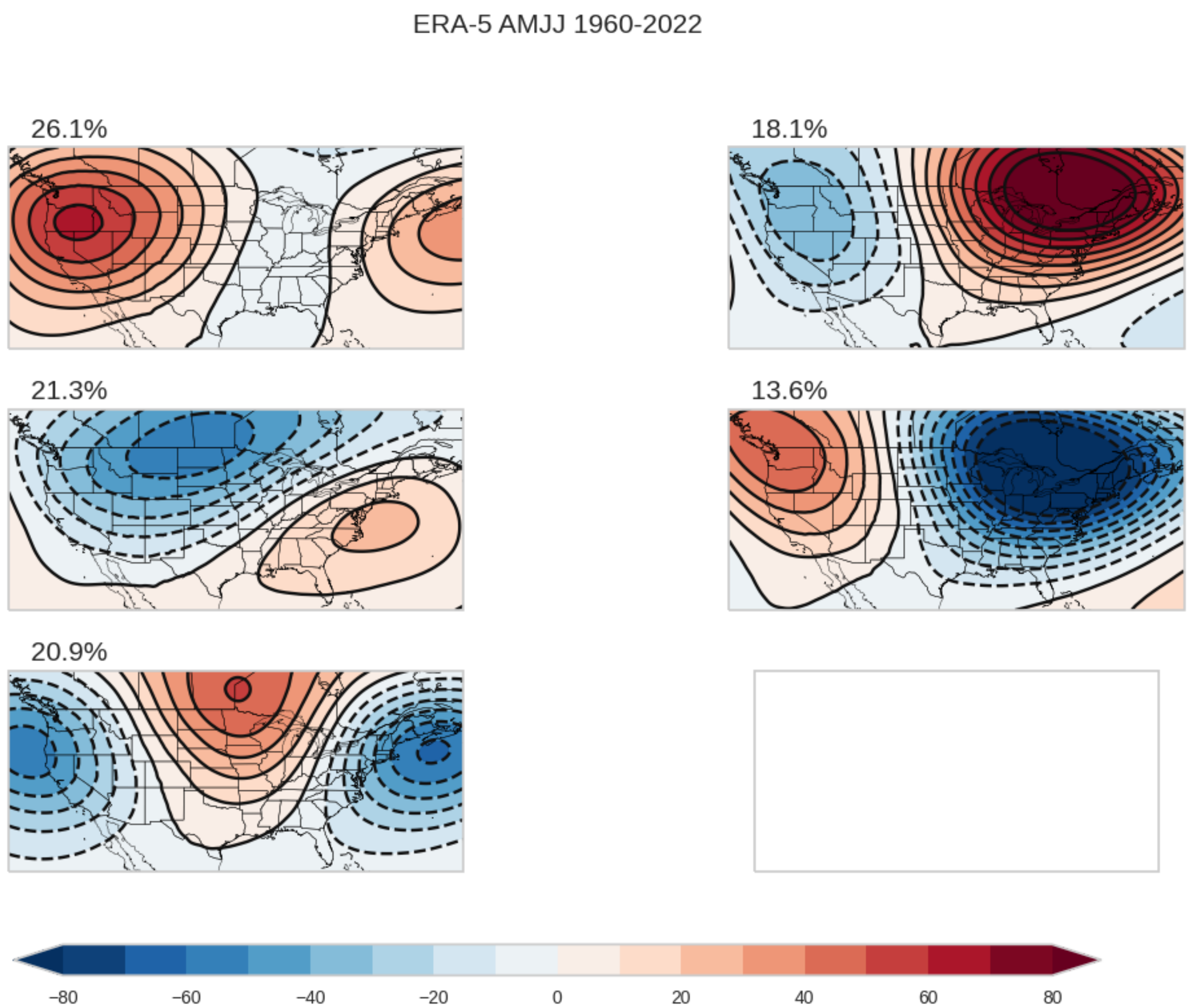

Figure R3: As in Fig. R1 except created using the first 5eofs for K-means clustering analysis and ordered by spatial structure as in Fig. R1 & R2.

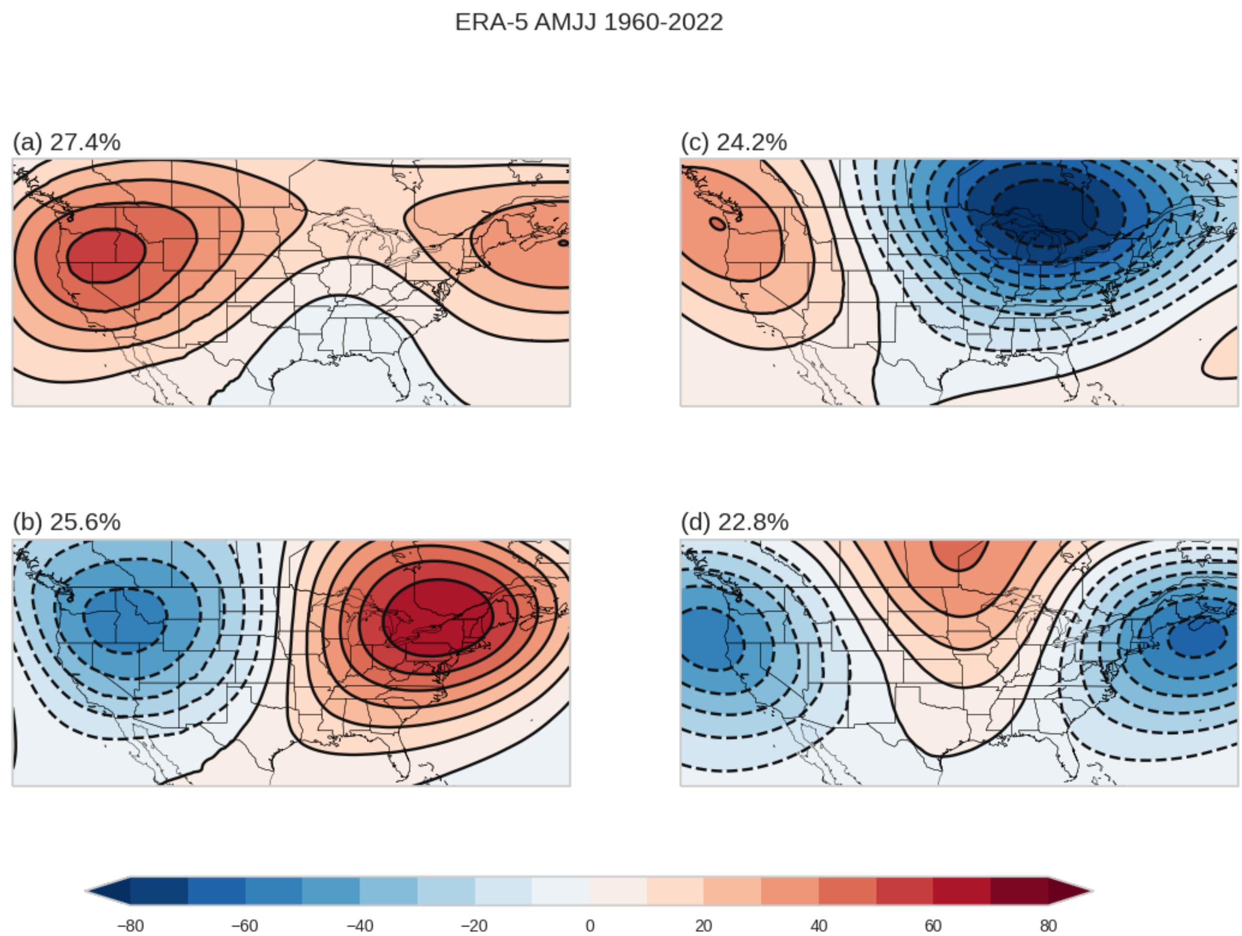

Figure R4: As in Fig. R1 except created for k = 4 clusters.

The new regime classification is not compared with previous ones from the same authors for April and May and with year-round regime classifications from Lee et al., (2023) [data is in Zenodo for download]. Making connections to previous work would increase the value of the current work. The classification data (data needed to classify independent data and classification of the days in the study) should be provided.

In the revised version of the manuscript, we now compare to more previous studies such as Lee et al. 2023, Robertson and Ghil 1999 for example. Lines 162-165 already make connections to Lee et al. 2023.

Weather regime identification code is available on github at:

https://github.com/Matt0604/Kmeans

The ERA-5 reanalysis data and the tornado report data are both publicly available.

ERA-5: https://doi.org/10.24381/cds.bd0915c6

Tornado report: https://www.spc.noaa.gov/wcm/#data

"The WR framework thus has a strong dynamic basis and have been used to reliably detect changes in regional temperature and precipitation previously (e.g., Robertson and Ghil, 1999)."

"Some WRs are similar to the year-round WRs in Lee et al., (2023), which were subsequently used by Tippett et al., (2024). More specifically, WR-A features spatial similarities to a Pacific Trough, WR-B and WR-D show warm and cool phases of a Pacific Ridge associated with ENSO, and WR-E is characterized by an Alaskan Ridge. WR-C features spatial similarities to a Greenland High as well. It is worth mentioning that our study focuses on a different region, a specific season and chooses a different k value, and there are thus noticeable differences. WR-A features two anomalous highs over the two coasts as opposed to one anomalous high over the central-CONUS. The anomalous low in WR-B is more pronounced than in Lee et al., (2023). The anomalous high in WR-C is wavelike unlike the Greenland high in Lee et al., (2023). The dipoles in WR-E are further south than they are in the Alaskan Ridge in Lee et al., (2023) …The favorable anomalies presented in WR-B agree with the Pacific Ridge findings in Tippett et al. (2024)."

The dependence of tornado activity as well as the dependence of CAPE and shear on month does not seem to have been accounted for in the analysis. In both cases anomalies are computed with respect to the April–July average. Using anomalies with respect to the April–July average means that the anomalies of quantities with a seasonal cycle will appear correlated but might actually be unrelated after accounting for seasonality.

As stated in L115-116 of the original manuscript, the seasonal cycle is defined as the long-term mean at each grid-point for each calendar day. The anomalies in our study are calculated with respect to the daily climatology, instead of April-July average, as shown in Eq. 1, which helps remove the seasonal cycle.

$$H'(d,y) = H(d,y) - \bar{H}(d) \qquad (1)$$

where d denotes calendar day and y denotes year. This has been further clarified in the revised manuscript. "Daily anomalies of MUCAPE, S06, and CP were calculated by removing the daily climatology on each calendar day, following Eq. 1."

We chose not to normalize the tornado activity anomalies because we are working across one season where seasonality is not as big of an issue as it would be for a year-round study (as in Tippett et al. 2024). To maintain data consistency, we therefore chose not normalize CAPE or shear.

Regarding the dependence of tornado activity as well as the dependence of CAPE and shear on month, the coincidence of higher variance of tornado activity and CAPE in April-May might lead to an artificial correlation due to the seasonal cycle if we had used ***monthly*** *mean data. However, this is not a concern for* ***daily*** *data because the strong coincidence of environmental condition* *anomalies and tornado activity anomalies on a daily time scale likely indicates a physical relationship.*

Along with seasonality, ENSO may be another factor/alternative hypothesis to consider.

We thank the reviewer for pointing this out. While ENSO was not explicitly explored in this manuscript, it is implied that the frequency of weather regime occurrence, CAPE, shear, and tornado activity may all be modulated by large-scale climate modes such as ENSO and MJO (Vigaud et al. 2018), with weather regimes serving as the intermediate piece between large-scale climate modes and tornado activity. Such relationships are outside the scope of the present study, but we have briefly discussed the potential role of low-frequency climate modes in the last section of the revised manuscript as it may help improve our prediction:

The following statement was added at the end of the conclusion:

"Furthermore, although not explored in this study, WRs and tornado activity may both be modulated by large-scale, low-frequency climate modes, with WRs potentially serving as the intermediate piece between large-scale climate modes and tornado activity, and the low-frequency modes may be important sources of predictability for the interannual variability of tornado activity."

Abstract. "Our study highlights the potential application of WRs for better seasonal prediction of tornado activity." The authors' previous regime/tornado study examine subseasonal prediction of weekly regimes and found that forecast skill was lost at about Days 7–13. Is there evidence that these regimes are predictable on seasonal time scales?

First, Miller et al. (2020) showed that the hybrid model has skill better than climatology out to Week 3. Second, with increasing forecast lead times, the information of the predictand will be less specific. Miller et al. focused on weekly mean tornado activity, but one may focus on seasonal mean tornado indices for seasonal prediction. Applying these regimes to seasonal prediction is our ongoing research, which shows promising results and we hope to publish in due time.

The Weather regime methodology differs substantially from that used commonly in the literature. There are no explanations provided why. The weather regime classification method lacks standard diagnostics and assessments of robustness. The classification data is unavailable which means the classification cannot be applied by others to independent data and cannot be compared with other classifications (e.g., Lee et al., 2023 which provides the data)

As explained above, what accounts for the "standard" regime classification is controversial. In particular, Falkena et al. (2020) argued against the use of either EOFs or time filtering on top of K-means clustering because K-means clustering reduces the dimensions and is a form of data filtration.

We have made our weather regime methodology code available at:

https://github.com/Matt0604/Kmeans

The ERA-5 and tornado report data are publicly available.

ERA-5: https://doi.org/10.24381/cds.bd0915c6

Tornado report: https://www.spc.noaa.gov/wcm/#data

Line 106. "500H at 21 UTC was used to represent the daily circulation patterns." Previous weather regime classifications have used daily means and subsequently smoothed those in time, e.g., 10-day low-pass-filtered (Grams et al., 2017, 2020, Lee et al., 2023)

The chosen time (2100 UTC) for 500H analysis represents a typical time of day when U.S. tornado outbreaks are ongoing (Cwik et al, 2022), thus potentially providing a more straightforward connection between WRs and tornado activity. In contrast, the use of 500H at times of day when tornado activity is much likely could result in misleading connections. Again, there is no standard data smoothing prior to K-means. Although Grams et al. (2017, 2020*), Lee et al. (2023) applied a 10-day low-pass filter, some studies used a 5-day low-pass filter (Robertson and Ghil, 1999), and Falkena et al. (2020) cautioned against smoothing. Our analysis (Fig. R1) showed that applying a 5-day low-pass filtering does not qualitatively affect WR patterns.*

There is no EOF filtering which differs from previous work (Michelangeli et al., 1995, Grams et al., 2017, 2020, Lee et al., 2023)

While many previous studies applied EOF and low-pass filtering prior to the clustering analysis, some studies chose to omit such procedures (Miller et al. 2020). In particular, Falkena et al. (2020; ***https://doi.org/10.1002/qj.3818****) argued against the application of these procedures and advocated the use of the full field.*

Line 121. "the number of clusters was determined as five using the elbow method."

From the reference cited, the elbow method "is a visual method. The idea is that Start with K=2, and keep increasing it in each step by 1, calculating your clusters and the cost that comes with the training. At some value for K the cost drops dramatically, and after that it reaches a plateau when you increase it further. This is the K value you want." This is not really an objective method. Lee et al., 2023 apply four objective, data-driven methods for determining the best number of clusters, including the classifiability and reproducibility indices of Michelangeli et al. (1995).

The choice of k is often somewhat subjective, because a metric does not always indicate an unambiguous optimal cluster number and different metrics may yield different optimal cluster numbers (Dorrington and Strommen 2020, https://doi.org/ 10.1029/2020GL087907). This is a known limitation of K-means. We tried k=4 given that Lee et al. (2023) chose k=4 in their analysis. As shown in Fig. R4, k =4 yields four of the same WR patterns as k=5 but misses WR B in the k = 5 analysis, which is spatially similar to WR-A in Miller et al. 2020 and the Pacific Ridge in Lee et *al. 2023. This regime is favorable for tornadoes and tornado outbreaks, so k=5 serves best for our purposes. Miller et al. 2020 also used k=5 based on the elbow method, which is a commonly used method for choosing k.*

There is no variance normalization to account for seasonality of variance (Grams et al., 2017 Lee et al., 2023). During the April–July period examined, Lee et al., (2023) found that the domain averaged Z500 std varied from 80 m in April to 50 m in July. Removing the daily climatology does not account for seasonality of variance.

We agree that normalization is beneficial when examining weather regimes throughout the entire year, as it helps account for seasonality. However, most studies do not apply normalization when dealing with a specific season as opposed to the whole year. Since we are only looking at AMJJ, we chose not to normalize H500, which also helps maintain the consistency of our data analysis as explained before.

Because k-means cluster analysis minimizes the total within-cluster variance, seasonality in the variance of the data means that clusters might be biased toward the later months of June and July when variance is small and consequently within-cluster variance is easier to minimize. Consequently the resulting cluster centroids are likely to be skewed toward patterns that best represent June/July variability at the expense of other months. And indeed, cluster A shows reduced activity which would be typical of the June/July period. This bias is potentially a serious flaw for the application here since US tornado activity is much higher in April–May than in June–July. In other words, tornado activity might be substantially higher in a particular weather regime simply because that regime is more frequent during calendar months when tornado activity is climatologically higher. Whether this is the case, and the association between tornado activity and regime frequency is simply due to their having similar seasonal cycles, is impossible to say because the authors have as far as I can see failed to provide any indication of the seasonality of cluster frequency or how the variance explained depends on month.

We thank the reviewer for this comment, which brings up an interesting point. We agree that it is possible that cluster centroids might be skewed toward patterns that best represent June/July variability. However, given that tornado activity is climatologically higher in April and May, such skewness would disrupt the weather regime-tornado correlation instead of amplifying the correlation. In addition, even if a strong skewness exists and reflects in the seasonality of cluster frequency, it would not explain the link between the interannual variability of tornado activity and weather regime frequency as we demonstrated using our empirical model (Fig. 5 in the manuscript).

The seasonality of the WRs (Fig. R5) shows that WR-A and WR-B both occur more frequently later in the season. Given that WR-A is associated with reduced tornado activity and WR-B with enhanced tornado activity, this seasonality does not support the reviewer's speculation that "tornado activity might be substantially higher in a particular weather regime simply because that *regime is more frequent during calendar months when tornado activity is climatologically higher." Additionally, although tornado activity decreases towards the end of July, it remains high during June and July. As shown in Graber et al. (2024), the peak day for tornado days from 1960-1979 was June 14th.*

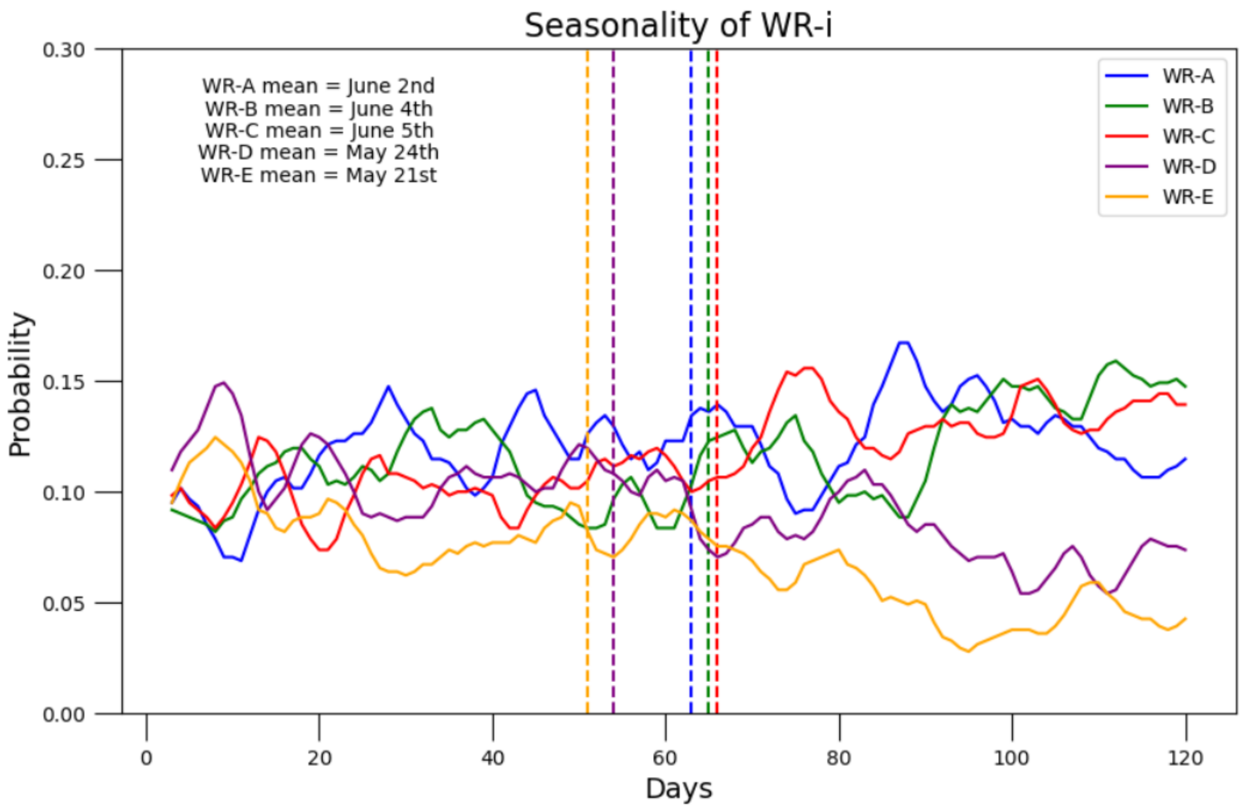

Fig. R5: Seasonality of WR frequency plotted as a 5-day running mean with dashed vertical lines at the mean date.

Overall there are essentially no diagnostics of the WRs such as variance explained. Also there is no assessment of how these regimes vary with large-scale modes of variability such as ENSO (known to be important for tornado activity), NAO, etc.

Similar to seasonality, because we are dealing with one season, the variance would be expected to be more uniform and less spread out around the mean. This is especially true since we are dealing with the warm season when the jet stream is less active. Accounting for variance would certainly be more important if this was a year-round analysis where there would be more spread in the data, as was done in Lee et al. (2023); Lee and Messori (2024) and Tippett et al. (2024).

By performing the low-pass filter and EOF filter analysis in figs R2 and R3, there is evidence that the regimes we have created are capturing distinct modes of variability. In addition, Figure S4 in the supplemental information illustrates the year-to-year and decadal variability of the 5 WRs in *addition to the persistence over time. The latter is an important point considering that multiple studies have shown that WR persistence may lead to more tornado outbreaks.*

We agree that WR frequency may be modulated by large-scale climate modes such as ENSO and NAO which are relevant in tornado climatology, but how each WR varies with such modes goes beyond the scope of the current study.

Line 134. The tornado probability anomalies fail to account for seasonality because they are with respect to the April–July frequency. Consequently, substantial anomalies may occur simply because some regime are more frequent during April–May rather than June–July. I see no exploration in the text of this hypothesis. This issue applies to environments as well reports.

The tornado probability anomalies shown in Figure 2 were defined with respect to the daily climatology at each grid point, instead of with respect to the April–July average. We did not discuss this point because our data are properly deseasonalized and we do not think it would artificially amplify the WR-tornado correlations. As shown in Fig. R5, WR-B, which is associated with enhanced tornado activity, actually occurs more frequently during June-July, contrary to what the reviewer expected.

Line 138 now reads: "The TD probability anomalies (P_a) were calculated at each grid-point for each WR as follows:"

Line 160. "These WRs have some spatial similarities to the year-round WRs found by Lee and Messori, (2024)." The more appropriate citation is Lee et al., (2023) which describes the classification in detail, and is not cited here. Also Lee et al., (2023) provide that classification data which means that authors here can make a more precision statement regarding the similarity of the classification. That is, with what frequency are the classifications the same. Also applying the diagnotic methods of Lee and Messori, (2024) to the regimes here would provide some evidence that regimes here are physically or dynamically meaningful.

Lee et al. (2023) has been replaced by Lee and Messori (2024). We would like to point out that the region and season(s) examined in Lee et al. (2023) are different from those in our study, and some differences are thus expected.

Lines 97-100 now read: Year-round WRs (Lee et al. 2023) have also been used and found to have statistically significant relationships with tornado activity in all months except June-August (Tippett et al. 2024) although without any consideration of WR persistency."

Line 165. "Composite anomalies of these parameters were calculated by subtracting the corresponding climatological mean." The same issue of using an inappropriate climatology applies to the MUCAPE and S06 anomalies, i.e., they will have seasonality both in their mean and variance. Because the April–July climatology is used, MUCAPE anomalies will tend to be positive in later months and negative in earlier months, and the opposite for S06. This means that any seasonality in the regime frequencies will project onto these anomalies, even if there is no relation when seasonality is taken into account. Moreover statistical significance tests used (e.g., Fig. 2) are inappropriate because they assume identically distributed but the variance of the data depends on month.

Same as the other variable, the anomalies here are defined with respect to daily climatology, and seasonality is thus removed. We have made this clearer in the revised manuscript,

"Daily anomalies of MUCAPE, S06, and CP were calculated by subtracting each calendar day's mean from every calendar day, following

$$\boldsymbol{H}^{\prime}\ (\boldsymbol{d},\boldsymbol{y}) = \boldsymbol{H}(\boldsymbol{d},\boldsymbol{y}) - \overline{\boldsymbol{H}}(\boldsymbol{d})$$

where y is year, d is calendar day, H is the variable or parameter of consideration, and the overbar denotes the long-term mean."

Line 136. Tornado data from period 1960–2022 is used and it is claimed (line 132) regarding well-known report trends that "this trend is not reflected in TDs" (tornado days). However, Miller et al., (2020) with two of the same three authors conclude that the period 1990–2019 "represents a compromise between data set length and an allowance for a significant fraction of the reports to have occurred during the Next-Generation Radar era and thus have undergone some quality control." Moreover, Fig. 5a shows a very large, presumably secular shift in the number of tornado days, which is as large or larger than the year-to-year variability. The authors state later (line 294) that "modelled TDs are nearly out of phase with observations in the 1960s, when tornado reports are less reliable" which supports analysis on a shorter period.

The tornado dataset is imperfect even in the current era, and therefore judicious choices must be made as a function of the particular application. The use of the term "compromise" by Miller et al. (2020) was not meant to suggest that tornado reports prior to 1990 were unusable, but rather that their particular application – hybrid S2S prediction–required relatively more certainty. Moreover, Miller et al. (2020)'s focus on hybrid prediction limited their maximum period length to 1990-2019 due to data availability in the ECMWF reforecasts. Since our study's focus was not prediction, we used the longer period length to match Graber et al. (2024) to look for ties between already observed tornado activity and seasonal weather regimes. The longer period also works *better with the results found by both Graber et al. (2024) and Brooks et al. (2014) since we are interested in providing physical explanations for the TD and TO trends that they found. Figs. 1 and 2 in Graber et al. 2024 demonstrate the "secular shift in the number of tornado days" in Figure 5a of the present study. April-July all shows large drop-offs in the 1980s, particularly April which has a few low outliers. One of the takeaways from that study was that the warm-season months were responsible for decreasing TD trend. The same shift appears in Figure 5a which is expected for the observations. Fig. S6 does provide the additional support that a shorter training period would be ideal for prediction studies, but the model performing adequately with the full period in spite of the report discrepancies supports the potential that seasonal weather regimes have to improve seasonal tornado prediction.*

Line 210. "may be possibly linked to tropical cyclones (Figs. 1e and 2e)." This is an interesting point and should be verified using data here https://www.spc.noaa.gov/exper/tctor/ and https://www.spc.noaa.gov/publications/edwards/tctor.xls

Thank the reviewer for pointing out the dataset. Using this dataset, we found that there are 98 such TC-induced EF-1+ tornadoes during June and July but only 8 occurred during a WR-E. The result doesn't support our speculation, and the statement has been removed in the revised manuscript.

Figs. 3 and 4 mention resampling to assess statistical significance without details. Depending on the design of the resampling procedure, the results may be incorrect if seasonality is not accounted for. For instance, in the case of a permutation test a possible way of taking seasonality into account is to compare tornado day frequency in say regime A with tornado day frequency on exactly the same calendar days when regime A did not occur.

Significance in Figs. 3 & 4 was calculated using a Monte Carlo simulation test with 10000 resamples. The p-value was then calculated based on the proportion of simulations that were more extreme than the observation. More information has been added in the revised manuscript.

Lines 145-148 now read: "A Monte Carlo simulation test with 10000 resamples was used to test for significance of the anomalies. The number of WR-i days was multiplied by the climatological mean TD probability to get an expected number of tornado days. The p-value was calculated based on the proportion of simulations that were more extreme than the observations."

As explained before, the data are deseasonalized by removing the daily climatology, and we do not think seasonality of variance would induce an artificial link between WRs and tornado activity. To further address the reviewer's concern, we replotted Fig 4 for April-May and June-July (Fig R6), separately. Keep in mind that while tornado days still occur at a high rate in June and July, *tornado outbreaks peak during April and May, so the June-July Tornado outbreaks panel has a low sample size. The significant anomalies shown in Fig. 4 remain the same sign (except for persistent WR-E in June-July) with quantitative differences.*

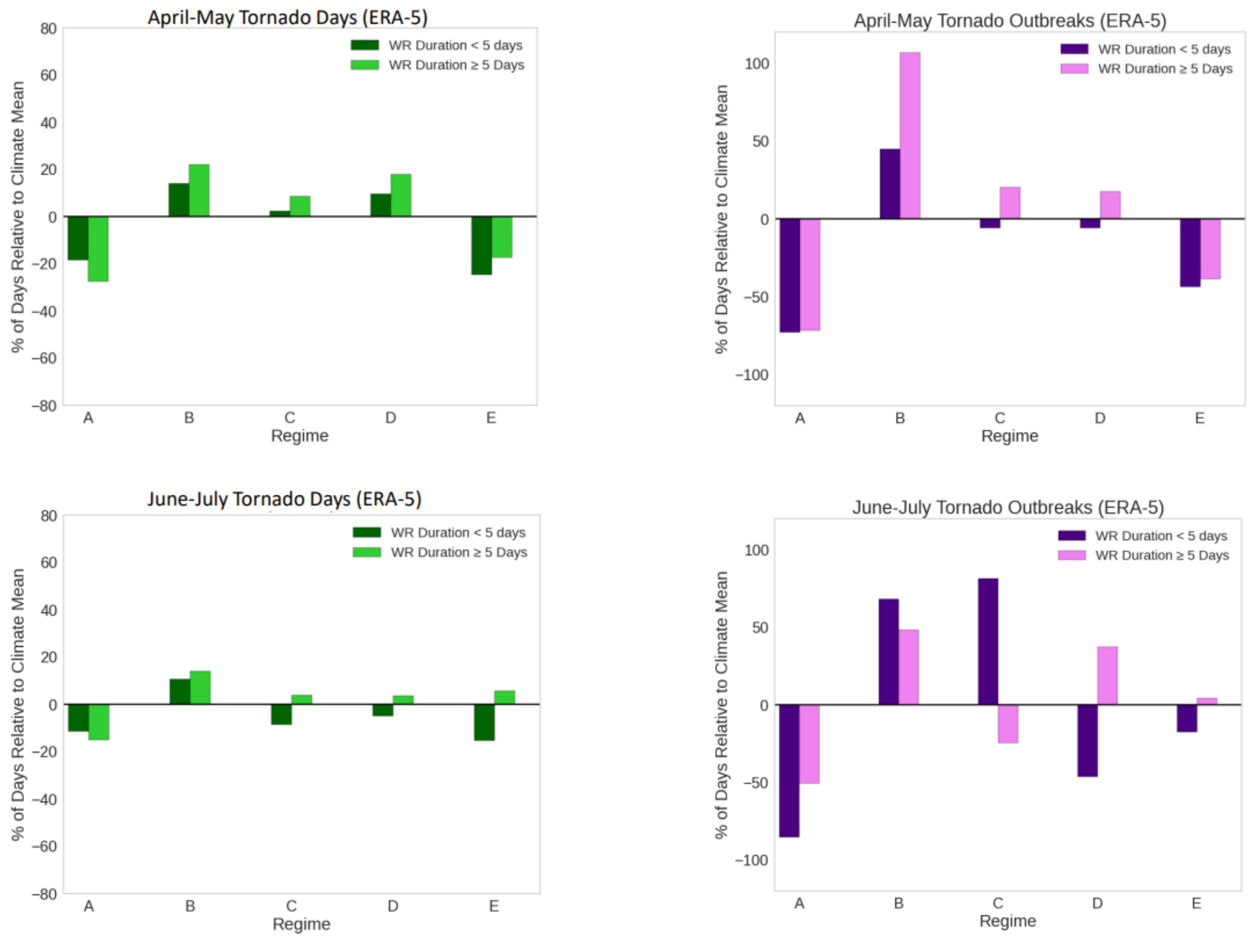

Fig. R6: Tornado probability anomaly plots for persistent and nonpersistent TDs/TOs in April-May and June-July.

Fig. S4 "WR TD time series" it is unclear what this quantity is. If it is the number of tornado days in that weather regimen, then of course, there are fewer tornado days in that weather regime in years when that weather regime is less frequent (the very high correspondence between the blue and green curves). This would be the case even when there is no relation between tornado days and weather regimes. That being the case, reporting the correlation coefficient does not seem informative and might confuse some readers. (I think my interpretation of the green curve is correct because it is different in panels S4a-e.)

Yes, your interpretation of the blue and green curves in Fig. S4 is correct. The correlation is expectedly high as pointed out by the reviewer, but we feel this figure is useful as it demonstrates the strong variability of weather regimes and its influence on tornado activity.

Line 290 and Fig. 5. "the empirical model fails to capture the observed decreasing trend or the decadal shift in the 1980s." The use of Spearman correlation here may obscure the extent to which empirical model fails to capture observed variability. A scatterplot would likely give a much more accurate and pessimistic picture. The association is stated to be statistically significant at the 5% level but visually is hard to see. Perhaps a bootstrap test might give a more credible assessment of statistical significance.

The poor visual agreement is mainly due to the underestimated tornado variance represented by the empirical model. We now show the observed and predicted time series with two separate y-axes (Fig. R7) so that it is easier to visualize the year-to-year fluctuations that lead to the Spearman correlation.

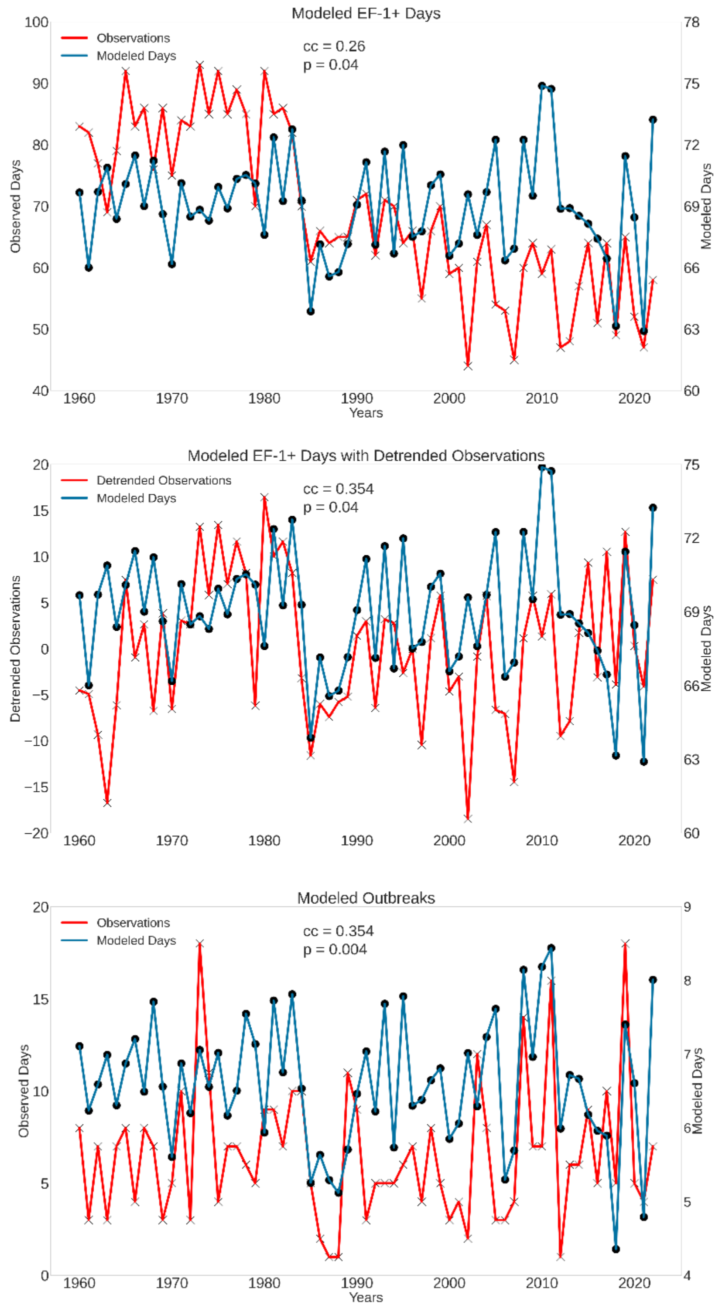

Fig. R7: (New Figure 5 in the manuscript) Empirically modeled TDs (blue with circles) per year overlaid with (a) observed TDs (red with crosses) and (b) detrended observed TDs (red with crosses) with spearman rank correlation coefficient (cc) and p-value; (c) empirically modeled (blue with circles) and observed (red with crosses) TOs per year with the spearman rank correlation coefficient and p-value.

Line 312 and also the abstract. "the empirical model captures the interannual variability of TDs reasonably well" this seems an overly generous description.

As stated in line 302, it is a common limitation of statistical modeling that the model curve will underestimate the magnitude of the year-to-year variability. We acknowledge the evidence of this in Fig. 5b & c. However, given the significance of the relationship between the observations and the model, we stand by this statement.

We have revised our statement to be more specific: The TD time series estimated by the empirical model shows a significant rank correlation (above the 95% confidence level) with the observed time series but underestimates the observed variance.

Conclusions. Line 343. "A year that includes a high number of WR-B days is likely to have an above average number of TDs and TOs." Is there analysis/figure in the manuscript that supports this statement?

Figures 1-4, along with Figure S2, all implicitly show this. The availability of more favorable environmental conditions in WR-B in Figs. 1 & 2 supports the higher WR-B tornado probability anomalies seen in Figs. 3 & 4. Since the probability values are used to create the empirical model, it is implied that a year with several WR-B days is likely to have an above average number of TDs and TOs.

Line 285. "The frequencies of persistent WRs also show changes across different multidecadal time periods (Fig. S4f" I really don't see any substantial changes in Fig. S4f and I question whether a t-test is suitable for a change in frequency, perhaps Fisher's exact test.

It is worth pointing out that the values are all normalized by the number of years in each period, and each period contains at least 20 years. The Fisher's Exact test is a good test to use for very small sample sizes (such as n<20), and while our sample size here is not that high, this test would not be the most useful for this figure.

References

Brooks, H. E., G. W. Carbin, and P. T. Marsh, 2014: Increased Variability of Tornado Occurrence in the United States. *Science*, **346**, 349–352, https://doi.org/10.1126/science.1257460.

Cwik, P., R. A. McPherson, M. B. Richman, and A. E. Mercer, 2022: Climatology of 500-hPa Geopotential Height Anomalies Associated with May Tornado Outbreaks in the United States. *Int. J. Climatol.*, **43**, 893–913, https://doi.org/10.1002/joc.7841.

Dorrington, J., and K. J. Strommen, 2020: Jet Speed Variability Obscures Euro-Atlantic Regime Structure. *Geophys. Res. Lett.*, **47**, https://doi.org/10.1029/2020GL087907.

Falkena, S. K. J., J. de Wiljes, A. Weisheimer, and T. G. Shepherd, 2020: Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector. *Q. J. R. Meteorol. Soc.*, **146**, 2801–2814, https://doi.org/10.1002/qj.3818.

Graber, M., R. J. Trapp, and Z. Wang, 2024: The Regionality and Seasonality of Tornado Trends in the United States. *Npj Clim. Atmospheric Sci.*, **7**, https://doi.org/10.1038/s41612-024-00698-y.

Grams, C. M., R. Beerli, S. Pfenninger, I. Staffell, and H. Wernli, 2017: Balancing Europe's Wind-power Output through spatial development informed by weather regimes. *Nat. Clim. Change*, **7**, 557–562, https://doi.org/10.1038/nclimate3338.

Grams, C. M., L. Ferranti, and L. Magnusson, 2020: How to make use of weather regimes in Extended-range Predictions for Europe. *ECMWF Newsl.*,.

Lee, S. H., and G. Messori, 2024: The Dynamical Footprint of Year-Round North American Weather Regimes. *Geophys. Res. Lett.*, **51**.

——, M. K. Tippett, and L. M. Polvani, 2023: A New Year-Round Weather Regime Classification for North America. *J. Clim.*, **36**, 7091–7108, https://doi.org/10.1175/JCLI-D-23-0214.1.

Miller, D., Z. Wang, 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**, https://doi.org/10.1029/2020GL087253.

Robertson, A. W., and M. Ghil, 1999: Large-Scale Weather Regimes and Local Climate over the Western United States. *J. Clim.*, **12**, 1796–1813, https://doi.org/10.1175/1520-0442(1999)012%3C1796:LSWRAL%3E2.0.CO;2.

——, N. Vigaud, J. Yuan, and M. K. Tippett, 2020: Toward Identifying Subseasonal Forecasts of Opportunity Using North American Weather Regimes. *Mon. Weather Rev.*, **148**, 1861–1875, https://doi.org/10.1175/MWR-D-19-0285.1.

Tippett, M. K., K. Malloy, and S. H. Lee, 2024: Modulation of U.S. Tornado Activity by year-round North American Weather Regimes. *Mon. Weather Rev.*, https://doi.org/10.1175/MWR-D-24-0016.1.

Vigaud, N., A. W. Robertson, and M. K. Tippett, 2018: Predictability of Recurrent Weather Regimes over North America during Winter from Submonthly Reforecasts. *Mon. Weather Rev.*, **146**, 2559–2577.