9.2 PREDICTABILITY OF LARGE-SCALE PATTERNS THAT LEAD TO TORNADO OUTBREAKS AT LEAD TIMES OUT TO 14 DAYS

Allison Brannan*, Skylar Gertonson, Israel Jirak, Brett Roberts, Andrew Lyons, Jared Guyer

Cooperative Institute for Severe and High-Impact Weather Research and Operations, Norman, Oklahoma; NOAA Storm Prediction Center, Norman, Oklahoma; University of Oklahoma, Norman, Oklahoma

1. INTRODUCTION

For much of the continental United States, especially east of the Rocky Mountains, severe thunderstorms capable of producing tornadoes occur throughout the year. The general public relies on forecasts from operational meteorologists to make decisions and keep themselves safe during days with potential for severe weather threats. Unfortunately, the exact set of meteorological conditions needed to produce these hazardous storms are challenging to predict more than a day to a week in advance, and often their exact timing and location cannot be anticipated even then. Therefore, it is not currently possible to predict precisely on sub-seasonal timescales exactly when, where, or how many tornadoes will occur, but given the increasing ability to predict various climate patterns such as ENSO and the MJO, it may be possible to predict environmental conditions conducive to severe weather activity over a sufficiently broad geographical region.

Currently, forecasters at NOAA's Storm Prediction Center (SPC) issue convective outlooks out to 8 days in advance, specifying individual hazards (tornado, hail, wind) at days 1 and 2. However, the objective limit on the predictability of tornado outbreaks within current operational models remains uncertain. Given the importance of maximizing the lead time for potential tornado threats to better inform the public, the goals of this project include (1) discovering the lead time at which models can skillfully predict the large-scale pattern present on tornado outbreak days and (2) investigating whether the extent of the model skill is related to the size, longevity, or season of the tornado outbreak.

Objectively defining a "tornado outbreak" is not a straight-forward task. In fact, the definition of a tornado outbreak varies across previous studies and ranges from a vague idea to specific requirements. The AMS Glossary of Meteorology begins with a broad definition, stating that a tornado outbreak is defined as "multiple tornado occurrences associated with a particular synoptic-scale system." In attempt to objectively quantify these events, previous studies have used various thresholds for the minimum number of reports, intensity constraints, and/or temporal and spatial limitations. For example, studies have defined tornado outbreaks as events with 10 or more tornadoes from a single weather system (Galway 1977), 6 or more EF1+ tornadoes with no more than a 6 hour period between consecutive tornadoes (Grazulis 1993; Fuhrmann et al. 2014), at least 20 tornadoes over a 200 nautical mile radius (Maddox and Gray 1973), 4 or more tornadoes that occur within a four hour timespan (Hagemeyer 1997), 45 tornadoes occurring in adjoining states and within a six hour period (Forbes 2006), at least 15 tornadoes over a single day and located east of the Rocky Mountains (Gensini and Marinaro 2016), and 6 or more EF2+ tornadoes within a convective day (Grams et al. 2012).

Given the interest of this study to investigate the predictability at sub-seasonal timescales (out to two weeks), a novel technique involving the practically perfect hindcast (PPH) method is used to define a set of sufficiently large

* Corresponding author address: Allison Brannan, Cooperative Institute for Severe and High-Impact Weather Research and Operations, NOAA/NWS Storm Prediction Center, Norman, OK, 73072; e-mail: Allison.Brannan@noaa.gov

tornado outbreak days between 2000 and 2019. The predictability of the 500 hPa height pattern present on these outbreak days is analyzed using the Global Ensemble Forecast System's (GEFSv12) reforecast and reanalysis data. The limit of predictability is determined by analyzing anomaly correlation coefficients from 1-day to 16-day lead times for each tornado outbreak and noting when model skill is no better than a climatological average.

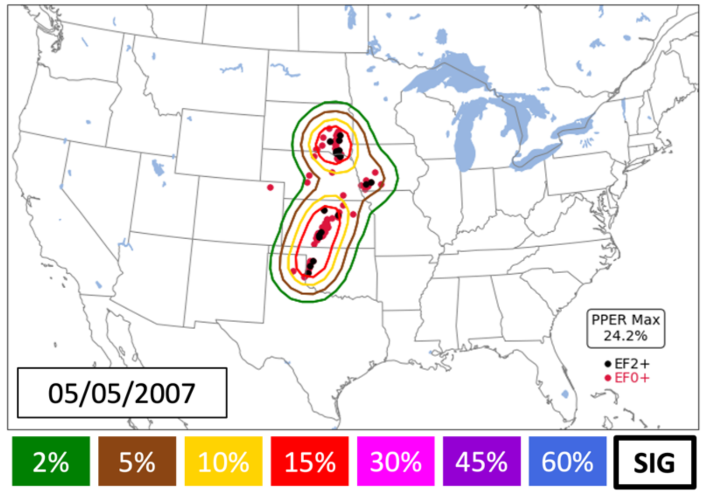

Fig 1 Practically perfect hindcast contours (colored lines) calculated from only EF2 and greater reports (black dots) on 5 May 2007. Red dots show EF0 and EF1 reports from the same day, though they are not used in the PPH calculation.

2. DATA AND METHODS

2.1 Tornado Data and Methods

This study uses tornado reports from 2000-2019 from NOAA's *Storm Data* database which includes information on each report's date, time, location, and magnitude (Schaefer and Edwards 1999). Given the majority of tornado reports occur between 20Z and 2Z and our interest in retaining 24-hour tornado outbreak periods, the time and date of each report is converted to align according to a 12Z-12Z convective day.

The practically perfect hindcast (PPH) technique is applied to all reports across an individual convective day. The first step in this calculation is to re-grid the reports onto NCEP's 211 Lambert Conformal Grid which has approximately 80-kilometer grid spacing. This grid is chosen to be roughly consistent with SPC's forecast definition of the probability of an event occurring within 25 miles of a point. After the reports are re-gridded, the algorithm searches each grid box and converts the field to a series of zeros and ones—zero if there are no reports in a specific grid box and one if there is at least one report in a given grid box. Note that if there are multiple reports in a single grid box, they are not double counted—the box still receives a value of one. A Gaussian filter is then applied to spatially smooth the field. This final step involves a smoothing parameter (sigma) that can be used to fine-tune the uncertainty of the resulting field. This study follows the advice of Hitchens et al. (2013) and implements sigma=1.5 (equivalent to 120-km on the 80-km grid) in order to produce contours similar to those drawn in outlooks issued by SPC. The output is displayed with individual probability contours corresponding to the coverage of tornado reports, which is essentially an objective representation of what a perfect SPC forecast would have looked like if the forecaster had knowledge of all future reports.

The PPH method is necessary for this study in order to objectively focus on sufficiently large tornado outbreak days that are the most likely to be skillfully predicted at sub-seasonal timescales. With that stipulation in mind, an iterative process in collaboration with SPC forecasters led to a definition of a tornado outbreak day as one with either (1) a 45% PPH contour calculated from all tornado reports, regardless of intensity (F/EF0+), across a given convective day or (2) a 15% PPH contour calculated from F/EF2 and greater tornado reports, across a given convective day. An example of an outbreak day qualifying through option (2) is shown in Fig. 1. These criteria yield a total of 181 tornado outbreak days between 2000 and 2019 (~9 per year on average).

2.2 Predictability Data and Methods

This study will be analyzing anomaly correlation coefficients (ACC) to evaluate the evolution of forecast skill across lead times. The ACC is a spatial correlation between the average anomalies of a forecast relative to climatology and the average anomalies of observations relative to climatology. Values of ACC are bounded by -1 and +1, where values near +1 denote perfect agreement between a forecast and the observations, values near 0 show poor agreement between the forecast and observations, and values near -1 indicate opposite phases between the forecast and observations. Following the study by Zhang et al. (2019), this investigation will use an ACC of 0.6 or higher to indicate forecasts that provide more skill than a climatological average.

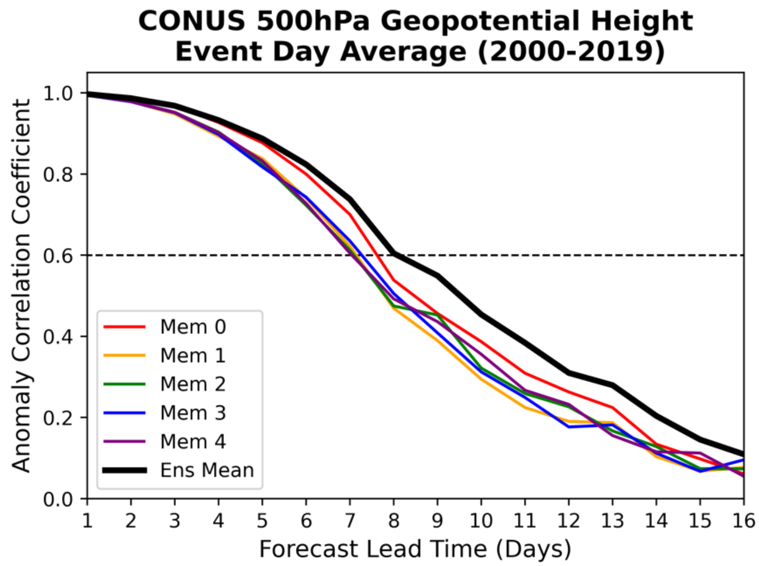

Fig 2 Anomaly correlation coefficient as a function of lead time, averaged over all 181 tornado outbreak events. Colored lines correspond to individual ensemble members, with the black line denoting the ensemble mean. Dashed line marks the skill threshold, ACC=0.6.

Given that we are interested in analyzing the predictability of the large-scale pattern, this study focuses on analyzing 500 hPa geopotential height fields from the Global Ensemble Forecast System (GEFSv12). In order to calculate the ACC, three main pieces of data are needed. The forecast data comes from GEFSv12 reforecast data where forecasts are generated once per day with 0Z initial conditions, across 5 ensemble members. Given the interest in sub-seasonal lead times, 24-hour timesteps are adequate, which extend out to 384 hours (16 days). The data is re-gridded to a 0.5 degree grid for consistency through all lead times. The observational data comes from the GEFS reanalysis which is valid at 0Z of each convective day and also has 0.5 degree grid spacing. Finally, daily mean 500 hPa heights from GEFS reanalysis are used to calculate a daily climatology from 2000-2019.

3. PREDICTABILITY

As expected, at short lead times, both the individual ensemble members and the ensemble mean exhibit ACC scores close to +1, indicating very good agreement between the forecasts and the reanalysis. As the lead time increases, naturally the forecast skill gradually decreases. When averaged over the 181 outbreak events, the total ensemble average shows skillful 500 hPa height forecasts out to lead times of 8.1 days (194 hours). It should be noted that this is not a suggestion that tornado outbreaks should be skillfully predicted out to 8.1 days. This analysis only focuses on the predictability of the large-scale geopotential height pattern, which is only one aspect of the synoptic pattern utilized in forecasting tornado outbreaks at sub-seasonal timescales. The remaining subsections focus on exploring several subsets of the 181 outbreaks to determine if certain descriptive characteristics can distinguish between outbreaks with better or worse predictability.

3.1 Spatial Extent of Tornado Outbreaks

The size of the outbreak is defined as the area encompassed by the PPH contour that qualified it to be included in this study—either the 45% contour calculated from all tornado reports or the 15% contour calculated from F/EF2 and greater reports. Focusing on the events qualifying under the EF2 criteria (right boxplot in Fig. 3a),

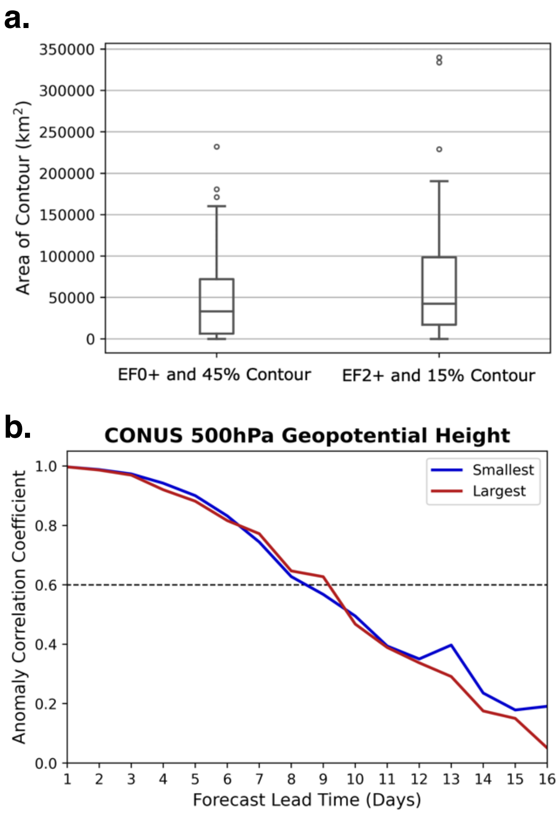

Fig 3 a. Spatial distributions of tornado outbreaks. Size is determined by the PPH contour listed on the x-axis based on each event's qualifying criteria. **b.** Anomaly correlation coefficient as a function of lead time for the ensemble average of the smallest events (blue) and largest events (red). Dashed line marks the skill threshold, ACC=0.6.

the "largest" tornado outbreaks are those 37 events in the upper quartile whose 15% contour covered greater than 98,557 km^2 and the "smallest" tornado outbreaks are those 49 days in the lower quartile whose 15% contour encompassed less than 17,069 km^2. Again focusing on the ACC from the 500 hPa heights, Fig. 3b shows that the forecasts for the largest outbreaks were skillful out to 9.2 days (220 hours) whereas the smallest outbreaks were only skillful out to 8.5 days (203 hours). It is curious that the skill for the smallest outbreaks extends to further lead times in comparison to the skill for the overall average in Fig. 2 which loses skill at 8.1 days. This fact, in addition to the Mann Whitney U Statistical Test, suggests that the predictability between these subgroups based on outbreak size are not significantly different. A potential explanation for this lack of significance lies within the methodology of selecting tornado outbreaks to be included in this study. Given the long-term goal of striving to create a sub-seasonal forecast for tornado outbreaks, the methodology of selecting tornado outbreaks to be analyzed in this study limits the cases to fairly large events, which have a chance at showing skill at week 2 lead times. For this reason, even the cases contained in the "smallest" category of outbreaks are still non-trivial and spatially expansive events. Potentially, there would a more significant difference in predictability based on spatial extent if the outbreaks included in this study were compared to tornado days that were not large enough to qualify as outbreaks, based on our definition.

3.2 Longevity of Tornado Outbreaks

Outbreaks are also categorized based on longevity. Multi-day outbreaks are those with two

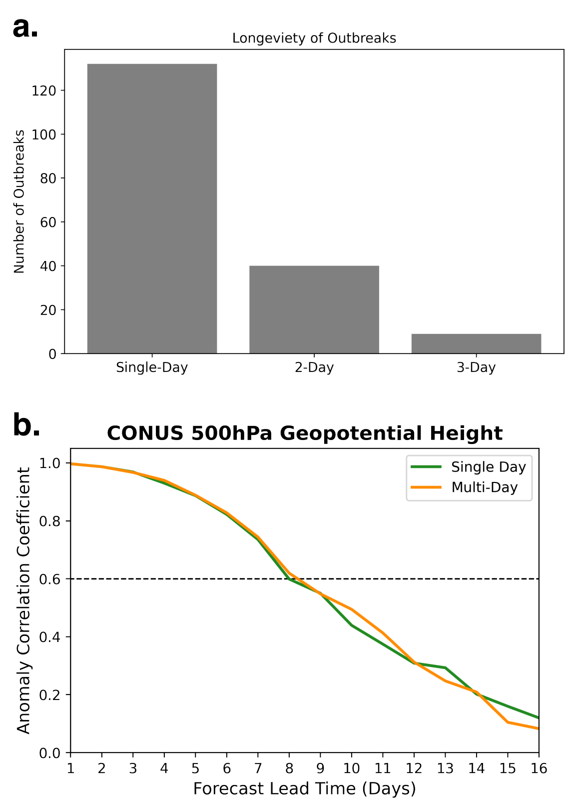

Fig 4 a. Distribution of the duration of tornado outbreaks included within this study. **b.** Anomaly correlation coefficient as a function of lead time for the ensemble average of the single-day events (green) and multi-day events (orange). Dashed line marks the skill threshold, ACC=0.6.

or more consecutive dates within the full list of 181 outbreaks. These multi-day events were also visually inspected to be sure they were consistent with coherent eastward progression of the synoptic pattern. As shown in Fig. 4a, the majority of the outbreak days included in this study are single-day events. The 40 two-day and 9 three-day events are combined to make up the multi-day category. A visual inspection of the evolution of ACC with lead time in Fig. 4b shows extremely similar trends between the skill of single vs. multi-day events. Single-day events are associated with a skillful prediction of 500 hPa heights out to 8.0 days (192 hours), compared to multi-day events whose skill extends just a bit further to 8.3 days (199 hours). Given the extreme similarities in the skill degradation between these two groups, it is not surprising that there is not a statistically significant difference in the extent of

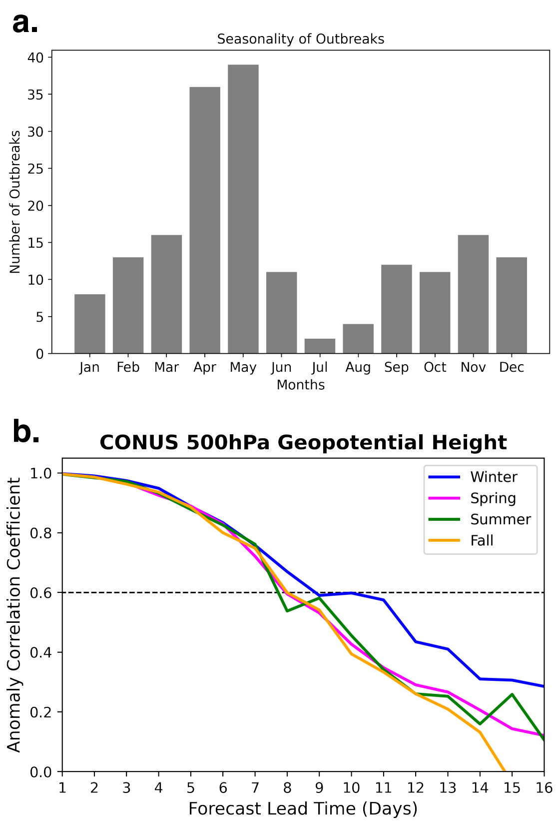

Fig 5 a. Distribution of tornado outbreaks by month. **b.** Anomaly correlation coefficient as a function of lead time for the ensemble average of winter (blue), spring (pink), summer (green), and fall (orange) outbreaks. Dashed line marks the skill threshold, ACC=0.6.

forecast skill between this set of single versus multi-day tornado outbreaks.

3.3 Seasonality of Tornado Outbreaks

Lastly, the seasonality of tornado outbreak events is plotted in Fig. 5a. The bimodal distribution shows the highest number of tornado outbreaks occurring in the spring months, with a peak in May and April. The subdued, secondary peak occurs in the fall months, while the fewest number of tornado outbreaks occur in the summer. Analyzing the predictability of the 500 hPa height field for the tornado outbreaks in each season shows a very similar evolution in skill degradation for the first 7 days. Beyond day 7, however, the differences in predictability become more significant. The forecast skill degrades fastest for summer events (17 events), showing no skill at 7.7 days (185 hours). The rate of skill-loss during spring (91 events) and fall (39 events) months is extremely similar and exhibits skill in predicting the large-scale height pattern out to 8.0 days (191 hours). The higher predictive skill for 500 hPa height patterns associated with winter outbreaks (34 events) becomes apparent at lead times greater than 7 days. In fact, winter-time forecasts qualify as skillful as far as 8.9 days (213 hours) before the event. Additionally, statistical tests confirm ($p < 0.10$; Mann-Whitney U test) that the predictability of the 500 hPa height field is statistically significantly different between winter-time outbreak days and fall and spring outbreak days. This finding is important when considering a long-term goal of issuing severe weather forecasts at sub-seasonal timescales. Based on the predictability of the 500 hPa height field alone, predicting tornado outbreaks in week 2 may find greater success when focusing on winter-time (i.e., Dec/Jan/Feb) outbreaks.

4. SUMMARY AND CONCLUSIONS

With a long-term goal of creating experimental sub-seasonal forecasts for tornado outbreaks, it must first be discovered if current operational models possess skill in predicting the synoptic-scale features present on tornado outbreak days at sub-seasonal lead times. This complicated question is also tangled in the challenges regarding an agreed-upon definition of which events qualify as a "tornado outbreak." This study settled on an objective method of defining sufficiently large tornado outbreaks in which the final set closely matched the subjective impression of what qualifies as a tornado "outbreak" based on the expertise from SPC forecasters. A tornado outbreak day in this study is defined as a convective day (12Z-12Z) with either (1) a 45% PPH contour calculated from all tornado reports, regardless of intensity (i.e., F/EF0+) or (2) a 15% PPH contour calculated from F/EF2 and greater tornado reports. These criteria create a set of 181 tornado outbreak events between 2000-2019, whose large-scale predictability was further analyzed.

This study utilized anomaly correlation coefficients to demonstrate the objective limit on the predictability of the 500 hPa geopotential height pattern on tornado outbreak days using reforecasts from the Global Ensemble Forecast System (GEFSv12). Using the ensemble mean averaged over all 181 outbreaks, GEFSv12 shows skill in forecasting 500 hPa heights extending out to 8.1 days (194 hours) of lead time. This is not suggesting that tornado outbreaks are necessarily predictable at this lead time, as this is only representative of the predictability of the synoptic-scale height pattern, which is just one aspect examined in forecasting tornado outbreaks. Future work aims to analyze additional variables, such as 2-meter dew point temperatures and upper-level winds. This study also discovered that while the predictability is not statistically significantly different based on the outbreak size or duration, there is evidence of greater predictive skill in the 500 hPa height pattern present on winter-time (December-February) tornado outbreak days compared to the other seasons. Winter-time tornado outbreaks on average exhibited skillful 500hPa height forecasts out to 8.9 days (213 hours), compared to spring and fall outbreaks where the large-scale height forecasts were skillful out to 8.0 days (191 hours) and summer-time outbreaks with skill that degraded the quickest, with ACC falling below 0.6 by 7.7 days (185 hours). While the winter-time outbreaks are those with the most predictable large-scale patterns, it is interesting to note that even in those tornado outbreaks, skill does not even extend over the 9 day lead-time mark. With this being the case, further investigation is required to determine if other phenomena (e.g., MJO) and climate indices (e.g., ENSO) can extend sub-seasonal predictive skill of tornado outbreaks beyond 9 days.

Aside from performing a similar analysis on other large-scale parameters, an investigation of whether the translational speed of the mid-latitude jet has any impact on the extent of the forecast skill for tornado outbreaks. In addition, the skillful lead times of individual outbreak events in the GEFSv12 reforecasts will be compared to the timing that convective outlook probabilities were introduced by the Storm Prediction Center to evaluate the forecasters' confidence in the trends shown by the current operational models.

5. REFERENCES

Forbes, G. S., 2006: Meteorological aspects of high-impact tornado outbreaks. Preprints, Symp. on the Challenges of Severe Convective Storms, Atlanta, GA, Amer. Meteor. Soc., P1.12.

Fuhrmann, C. M., C. E. Konrad, M. M. Kovach, J. T. McLeod, W. G. Schmitz, & P. G. Dixon, 2014: Ranking of tornado outbreaks across the United States and their climatological characteristics. *Weather and Forecasting*, 29(3), 684-701.

Galway, J. G. 1977: Some climatological aspects of tornado outbreaks. *Mon. Wea. Rev.*. 105. 477–484.

Gensini, V. A., and A. Marinaro, 2016: Tornado frequency in the United States related to global relative angular momentum. *Mon. Wea. Rev.*, 144, 801–810.

Grams, J. S., R. L. Thompson, D. V. Snively, J. A. Prentice, G. M. Hodges, & L. J. Reames, 2012: A Climatology and comparison of parameters for significant tornado events in the United States, *Weather and Forecasting*, 27(1), 106-123.

Grazulis, T. P., 1993: Significant Tornadoes 1680–1991. Environmental Films, 1326 pp.

Hagemeyer, B. C., 1997: Peninsular Florida tornado outbreaks. *Wea. Forecasting*, 12, 399–427.

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

Maddox, R. A., and W. M. Gray, 1973: A study of tornado proximity data and an observationally derived model of tornado genesis. Atmospheric Science paper No. 212, Colorado State University, Fort Collins, CO, USA.

Schaefer, J. T., and R. Edwards, 1999: The SPC tornado/severe thunderstorm database. Preprints, 11th Conf. Applied Climatology, Amer. Meteor. Soc., Dallas, TX, 603-606.

Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, & K. Emanuel, 2019: What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences*, 76(4), 1077–1091.