Reviewer #2 Evaluations:

Science Category: Science Category 2

Presentation Category: Presentation Category A

Reviewer #2 (Comments to Author):

Overall Comments: The paper represents a brave effort to handle a difficult problem and the model simulations are of significant interest. My problem has to do with the assessment of the tornado portion of the record. I'm not satisfied that the detrending has been done in an appropriate way and it is extremely challenging to know what the most appropriate statistic to define "big" or "small" tornado years is.

We would like to thank the reviewer for the extremely useful comments and suggestions. The manuscript is now revised substantially based on these comments. As discussed in Verbout et al. [2006] and Dowell et al. [2009], there are numerous known deficiencies in the SWD. These limitations include inherent errors in the structural damage-wind speed relationship and the associated tornado ratings, changes in damage survey procedure, and population increase. However, we also acknowledge that, currently, there is no completely objective or straightforward way to correct the SWD. Therefore, we decided to use the number of intense (F3 and above) tornadoes as the primary tornado index because intense tornadoes are much more likely to be detected and reported, and thus relatively more consistent over time. Then, to remove any remaining long-term trend, regardless of whether it is real or not real, we used a simple least squares linear regression to detrend the tornado data. In the revised manuscript, we state clearly that until the SWD quality issues are fully resolved, any tornado related climate research, including this study, is subject to strong caveats, referencing Verbout et al. [2006] and Dowell et al. [2009]. Additionally, we also attempted two different methods to remove any nonlinear long-term trend in the tornado dataset, if there is any. These are described in our reply to the reviwer's comment #2.

As the reviewer points out, the number of intense tornadoes used in this study may not be the most objective metric for representing tornado years because some years with a large number of tornadoes are not qualified as outbreak years if the single day with the largest number of tornadoes in each year is taken out. Due to this limitation in the tornado metric used in this study, we test our main conclusions using different tornado indices. Another widely used metric is the intense U.S. tornado-days, which is obtained by counting the number of days in which more than a threshold number of intense (F3 and above) tornadoes occurred [e.g., Verbout et al. 2006]. The threshold number selected in this case is three and above, which roughly represents the upper 25% in the number of intense U.S. tornadoes in a given day of AM during 1950-2010. The time series of intense U.S. tornado-days in AM for 1950-2010 is shown in Figure S10. Table 1 is reproduced using this new metric. As shown in Table S4, the TNI is still significantly correlated (above 95% significance level) with the intense U.S. tornado-days in AM, supporting the overall conclusions of this study.

Major Comments:

1. *p. 2, line 2: The "1243" is a preliminary number of reports. The final number of tornadoes will be much lower. It really can't be compared to previous preliminary report counts. It is also important to note that, when the final count for May is determined, it is highly likely that it will be below or near normal. April was very high, but May was ordinary in terms of the meteorological events.*

This sentence is revised to add that “1243” is a preliminary number of reports.

2. *p. 2, lines 14-16: The 1950-1953 data shouldn't really be included in the analysis. They're significantly different than the 1954 and later data. In addition, there is a large body of evidence that changes in damage estimation procedures have led to recent tornadoes being rated less intense than the older tornadoes. There's no description of the detrending procedure, but there are definite steps in the damage estimation time series around 1975 and 2001. See Verbout et al. (2006) and Doswell et al. (2009) for discussion on both of these issues.*

As discussed in Verbout et al. [2006] and Dowell et al. [2009], there are numerous known deficiencies in the SWD. These limitations include inherent errors in the structural damage-wind speed relationship and the associated tornado ratings, changes in damage survey procedure, and population increase. We absolutely agree with the reviewer that until such issues in the SWD are resolved, any tornado related climate research is subject to strong caveats. This is point is now discussed in the discussion section referencing Verbout et al. [2006] and Dowell et al. [2009].

Acknowledging that, currently, there is no completely objective way to correct the SWD, we decided to use the number of intense (F3 and above) tornadoes as the primary tornado index because intense tornadoes are much more likely to be detected and reported, and thus relatively more consistent over time. Then, to remove any remaining long-term trend, regardless of whether it is real or not real, we used a simple least squares linear regression to detrend the tornado data.

We acknowledge the overrating issue of F3-and-greater time series after 1974 as discussed in Verbout et al. [2006] and the references therein. Obviously, linear regression is not a very effective way to remove abrupt changes in a time series. Therefore, we attempt to remove a sudden drop in F3-and-greater time series after 1974, which is arguably due to the overrating issue after around that time, by assuming that the long-term averaged number of F3-and-greater tornadoes during 1975-2010 is the same as that during 1950-1974. This is achieved by performing the following simple procedure to the 1975-2010 portion of the dataset:

$$\text{F3+(1975-2010)} = \text{F3+(1975-2010)} - \text{Avg[F3+(1975-2010)]} + \text{Avg[F3+(1950-1974)]}, \quad (1)$$

where F3+ is the number of F3-and-greater tornadoes in Apr-May, and Avg[F] represents a time average of the function F. Table 1 is reproduced using this new tornado dataset:

Index	DJF	FMA	AM
Gulf-to-U.S. moisture transport	0.08	0.19	**0.39**
Lower-level vertical wind shear	0.05	0.15	**0.33**
GoM SST	0.14	0.20	0.20
Niño-4	-0.21	-0.19	-0.18
Niño-3.4	-0.12	-0.13	-0.11
Niño-1+2	0.02	0.10	0.15
TNI	**0.27**	**0.28**	**0.32**
PNA	-0.05	-0.10	-0.20
PDO	-0.12	-0.09	-0.14
NAO	-0.01	-0.09	-0.17

As shown above, the TNI is still significantly correlated with the revised tornado index.

Additionally, we performed a quadratic regression to remove nonlinear long-term trend in the tornado dataset. As shown in the following table, the TNI is still significantly correlated with the number of intense tornadoes in Apr-May:

Index	DJF	FMA	AM
Gulf-to-U.S. moisture transport	0.16	0.26	**0.46**
Lower-level vertical wind shear	0.04	0.17	**0.33**
GoM SST	0.20	**0.27**	**0.26**
Niño-4	-0.23	-0.21	-0.19
Niño-3.4	-0.15	-0.15	-0.13
Niño-1+2	0.01	0.10	0.14
TNI	**0.27**	**0.30**	**0.31**
PNA	-0.07	-0.12	-0.21
PDO	-0.18	-0.18	-0.22
NAO	-0.02	-0.15	-0.20

Regarding the potential issues with 1950-1953 data, we would like to point out that these three years are not identified as active tornado years (see Table S1). Therefore, removing these three years does not affect the main conclusions of this study. Table 1 is reproduced using only the tornado dataset for 1954-2010 to find that the TNI is still significantly correlated with the number of intense tornadoes in Apr-May:

Index	DJF	FMA	AM
Gulf-to-U.S. moisture transport	0.25	0.22	**0.40**
Lower-level vertical wind shear	0.05	0.13	**0.34**
GoM SST	**0.27**	0.23	0.16
Niño-4	**-0.27**	-0.24	-0.22
Niño-3.4	-0.18	-0.19	-0.15
Niño-1+2	0.02	0.09	0.15
TNI	**0.33**	**0.32**	**0.35**
PNA	-0.11	-0.08	-0.18
PDO	-0.18	-0.14	-0.18
NAO	0.06	-0.07	-0.20

3. *p. 2, line 22: Wind shear is not a "triggering mechanism." While "trigger" is a phrase I don't particularly care for, it is typically associated with some kind of storm initiation process. Wind shear is what organizes severe thunderstorms.*

We wish to thank the reviewer for pointing out this. As the reviewer points out, the vertical wind shear effect is not a triggering mechanism, but one of the most important environmental conditions needed for tornado formation. We have completely revised the manuscript to correct this. In the revised manuscript, both the differential advection and lower-level vertical wind shear are discussed and analyzed as important environmental factors for tornado activity. Therefore, Table 1, Figure 1, 3, 4, S2, S6, S7, S8 and the related discussions are all revised. The model results (EXP_TNI) indeed show that the lower-level vertical wind shear over the central and eastern U.S. is increased during a positive phase of TNI (Figure 3c in the revised manuscript), thus strengthening the overall conclusion of this study.

4. *Section 2: The proper metric for describing tornado years is unclear. 63 of the F3+ tornadoes in 1974 occurred on one convective day. 1974 is an ordinary April/May with one extraordinary day (that followed a well-below normal first 3 months of the year). If the metric chosen was something like a trimmed number of F3+ tornadoes where the largest single day each year is taken out, 1974 drops out of the top ten. 1973 would be a bigger year and 1957 is the biggest year in the record. The results are very sensitive to the one big day and, to a lesser extent, the big day in 1965. This doesn't mean that the relationship between the variables and the big outbreaks isn't important, but it's difficult to assess what a "big tornado year (or April/May)" is. Is the appropriate model a year with many days with tornadoes (e.g., 1957) or a ordinary year with a big day (e.g., 1974)?*

As the reviewer points out, the number of intense tornadoes used in this study may not be the most objective metric for representing tornado years because some years with a large number of tornadoes are not qualified as outbreak years if the single day with the largest number of tornadoes in each year is taken out. Due to this limitation in the tornado metric used in this study, we test our main conclusions using different tornado indices. Another widely used metric is the intense U.S. tornado-days, which is obtained by counting the number of days in which more than a threshold number of intense (F3 and above) tornadoes occurred [e.g., Verbout et al. 2006]. The threshold number selected in this case is three and above, which roughly represents the upper 25% in the number of intense U.S. tornadoes in a given day of AM during 1950-2010. The time series of intense U.S. tornado-days in AM for 1950-2010 is shown in Figure S10. Table 1 is reproduced using this new metric. As shown in Table S4, the TNI is still significantly correlated (above 95% significance level) with the intense U.S. tornado-days in AM, supporting the overall conclusions of this study.