GEOPHYSICAL RESEARCH LETTERS, VOL. 39, L02801, doi:10.1029/2011GL050368, 2012

Association of U.S. tornado occurrence with monthly environmental parameters

Michael K. Tippett,[1] Adam H. Sobel,[2,3] and Suzana J. Camargo[3]

Received 15 November 2011; revised 12 December 2011; accepted 13 December 2011; published 18 January 2012.

[1] Monthly U.S. tornado numbers are here related to observation-based monthly averaged atmospheric parameters. Poisson regression is used to form an index which captures the climatological spatial distribution and seasonal variation of tornado occurrence, as well as year-to-year variability, and provides a framework for extended range forecasts of tornado activity. Computing the same index with predicted atmospheric parameters from a comprehensive forecast model gives some evidence of the predictability of monthly tornado activity. **Citation:** Tippett, M. K., A. H. Sobel, and S. J. Camargo (2012), Association of U.S. tornado occurrence with monthly environmental parameters, *Geophys. Res. Lett.*, *39*, L02801, doi:10.1029/2011GL050368.

1. Introduction

[2] There is a substantial body of work relating severe thunderstorms and tornadoes to contemporaneous, or nearly contemporaneous, observed environmental parameters [*Brooks et al.*, 1994; *Rasmussen and Blanchard*, 1998; *Brooks et al.*, 2003a]. Observations and short-term predictions of these environmental parameters provide guidance for forecasters who currently issue severe thunderstorm outlooks up to one week in advance, tornado watches when environmental conditions favor the development of rotating thunderstorms, and tornado warnings when tornadoes appear imminent [*Hamill and Church*, 2000; *Shafer et al.*, 2010]. A key goal in these studies is the identification of usable associations between atmospheric parameters and tornado activity. An analogous inquiry in the study of tropical cyclones (TCs) is the question of how environmental parameters are related to TC formation. Recognition that large-scale environmental parameters influence TC activity on monthly and seasonal time-scales has led to TC genesis indices [*Gray*, 1979; *Camargo et al.*, 2007; *Tippett et al.*, 2011] and seasonal TC forecasts based on the seasonal prediction of environmental parameters, notably in the Atlantic [*Vecchi et al.*, 2010]. Our goal here is to apply the same approach to tornadoes: identify environmental parameters associated with monthly tornado activity, combine them in an index and, to the extent that the index is predictable, provide a basis for extended range forecasts of monthly tornado activity.

[1]International Research Institute for Climate and Society, Columbia University, Palisades, New York, USA.

[2]Department of Applied Physics and Applied Mathematics and Department of Earth and Environmental Sciences, Columbia University, New York, New York, USA.

[3]Lamont-Doherty Earth Observatory, Earth Institute at Columbia University, Palisades, New York, USA.

0094-8276/12/2011GL050368

2. Data and Methods

2.1. Tornado Data

[3] U.S. tornado data covering the period 1979–2010 are taken from the Storm Prediction Center Tornado, Hail and Wind Database [*Schaefer and Edwards*, 1999]. There are substantial inhomogeneities in the data. For instance, the annual number of reported U.S. tornadoes increases by 14 tornadoes per year during the period 1954–2003 [*Verbout et al.*, 2006]. This trend is primarily due to changes in the number of reported weak (F0) tornadoes and is likely related to changes in reporting methods and the introduction of Doppler radar in the 1990's [*Brooks and Doswell*, 2001; *Verbout et al.*, 2006; *Brooks and Dotzek*, 2007]. An adjusted (http://www.spc.noaa.gov/wcm/adj.html) annual number of U.S. tornadoes can be computed using a trend line for the period 1954–2007 and taking 2007 as a baseline. We compute a gridded monthly tornado climatology by counting for each calendar month the number of reported tornadoes in each grid box of a $1° \times 1°$ latitude-longitude grid. Similar to *Brooks et al.* [2003b], there is no adjustment for trends in the computation of the gridded monthly tornado climatology, presumably leading to an underestimate of the climatological number of tornadoes.

2.2. Analysis and Forecast Data

[4] Monthly averaged environmental parameters are taken from the North American Regional Reanalysis (NARR) [*Mesinger et al.*, 2006]. NARR data are provided on a 32-km Lambert conformal grid which we interpolate to a $1° \times 1°$ latitude-longitude grid. We primarily select environmental parameters recognized as being relevant to tornado formation [*Brooks et al.*, 1994; *Rasmussen and Blanchard*, 1998; *Brooks et al.*, 2003a] and use monthly averages of the following NARR variables: surface convective available potential energy (CAPE), surface convective inhibition (CIN), best (4-layer) lifted index (4LFTX), the difference in temperature at the 700 hPa and 500 hPa levels divided by the corresponding difference in geopotential height (lapse rate), the average specific humidity between 1000 hPa and 900 hPa (mixing ratio), 3000-0 m storm relative helicity (SRH), the magnitude of the vector difference of the 500 hPa and 1000 hPa winds (vertical shear), precipitation, convective precipitation and elevation. Lapse rate and vertical shear are computed using monthly averages of the constituent variables.

[5] Forecast data comes from reforecasts of the Climate Forecast System version 2 (CFSv2), the successor of the CFS version 1 [*Saha et al.*, 2006] with improved physics,

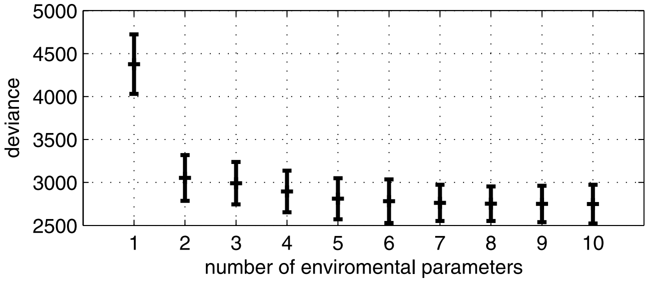

Figure 1. Deviance as a function of the number of environmental parameters used in the Poisson regression. Error bars indicate ± 1 standard deviation.

increased resolution and overall improved skill [*Yuan et al.*, 2011]. The CFSv2 is a comprehensive earth model with coupled atmosphere, ocean, and ice components. The atmospheric component is the NCEP Global Forecast System at T126L64 ($\sim$0.937°, 64 vertical levels) resolution. CFSv2 reforecasts are initialized using the Climate Forecast System Reanalysis (CFSR), a coupled data assimilation system [*Saha et al.*, 2010]. Over the 29-year reforecast period 1982–2010, single member ensemble forecasts are started every 5 days (without accounting for leap year days) at 0, 6, 12 and 18Z and integrated for 9 full months. Zero-lead forecasts are initialized in the month prior to the month being predicted or in the first pentad of the month being predicted. We average the most recent 16 ensemble members. For instance, a zero-lead forecast of the June monthly average consists of ensemble members from May 21, May 26, May 31 and June 5; the zero-lead June forecast started on June 5 is a partial average.

2.3. Poisson Regression

[6] We relate the *climatological* monthly number of U.S. tornadoes to *climatological* monthly averages of NARR atmospheric parameters using a Poisson regression (PR), the standard statistical method for modeling of count data. A similar method was used to develop a TC genesis index [*Tippett et al.*, 2011]. Fitting climatological data avoids many of the issues related to observational inhomogeneities. For instance, tornado record trends are not spuriously associated with climate trends. Moreover, fitting the PR with climatological data means that the yearly varying data provide an independent test of the PR. The probability that a Poisson distributed random variable N takes on the values n = 0, 1, 2,... is

$$P(N = n) = \frac{e^{-\mu}\mu^n}{n!}, \qquad (1)$$

where μ is the expected value of N. In this study N is the number of observed tornadoes over some specified number of years, and we use a log-linear model for μ given by $\mu = \exp(\mathbf{b}^T\mathbf{x})$ where $\mathbf{x}$ is a vector of environmental parameters and $\mathbf{b}$ is a vector of regression coefficients. A constant term (intercept) is included in the model by taking one of the elements of $\mathbf{x}$ to be unity. An *offset* is added to account for the area associated with each grid box and the number of years. Thus the model for the expected value μ (our "index") becomes

$$\mu = \exp\left[\mathbf{b}^T\mathbf{x} + \log(\Delta x\ \Delta y\ T\cos\phi)\right], \qquad (2)$$

where ϕ is the latitude, Δx and Δy are the longitude and latitude spacings in degrees, respectively, and T is the number of years; T = 32 when applying the PR to 1979–2010 tornado counts, and T = 1 for predicting annual counts. The offset term makes the units of $\exp(\mathbf{b}^T\mathbf{x})$ be the number of tornadoes per unit area per year, and thus the values of regression coefficients $\mathbf{b}$ are independent of grid resolution and climatology length. The regression coefficients $\mathbf{b}$ are estimated from data by maximizing the log-likelihood of the observed numbers of tornadoes given the environmental parameters. A standard PR goodness of fit measure is the deviance.

[7] We begin with the ten candidate environmental parameters listed in the description of the NARR data. We take the logarithm of CAPE, SRH and shear, consistent with previous work [e.g., *Brooks et al.*, 2003a] and because doing so reduces the deviance of single parameter PRs; the same is done for precipitation and convective precipitation. To select the parameters to include in the PR, we first perform a forward selection procedure in which one variable is added at a time to the PR, and the variable whose addition most reduces the deviance is identified. The deviance is computed using 10-fold cross-validation in which the data is randomly separated in 10 subsets, 9 of which are used to estimate the regression coefficients, and one is used to compute the deviance. This procedure gives 10 estimates of the deviance

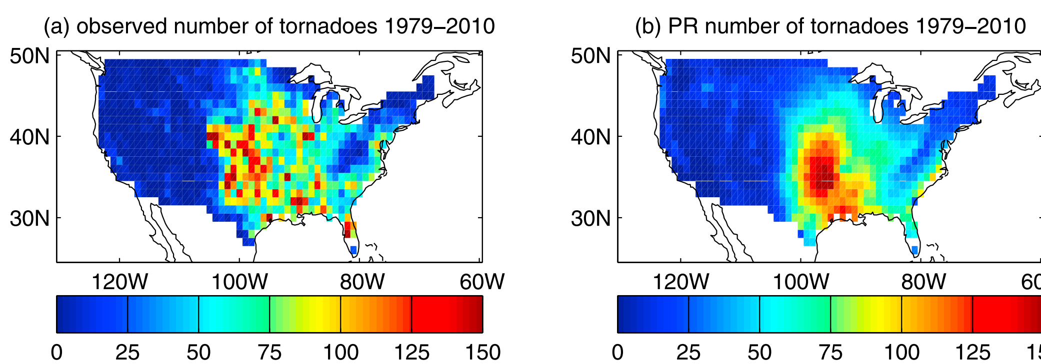

Figure 2. Colors indicate the (a) observed and (b) Poisson regression (PR)-fit number of tornadoes for the period 1979–2010.

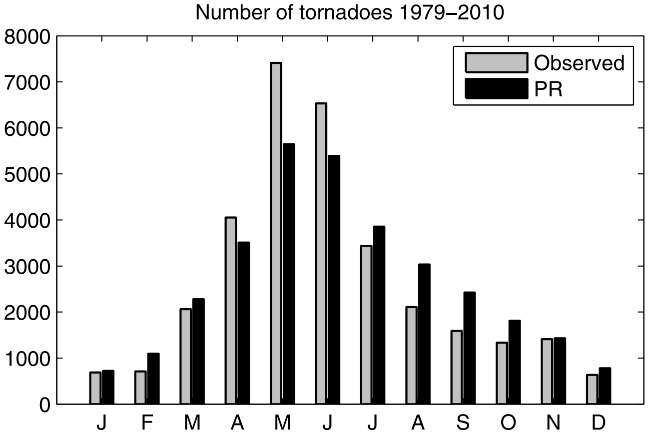

Figure 3. Observed (gray) and Poisson regression (PR)-fit (black) number of tornadoes per month during the period 1979–2010.

for each partition of the data. Here we use 10 partitions and obtain 100 estimates of the deviance. The mean and standard deviation of these 100 values of the deviance are shown in Figure 1 as a function of the number of environmental parameters used in the PR. There is a substantial decrease in the deviance as the number of environmental parameters is increased from 1 to 2, but further increases in the number of environmental parameters do not result in significant (95% level) decreases in deviance, in the sense that their one standard deviation error bars overlap with those of the 2 parameter regression.

[8] The two environmental parameters chosen by the forward selection are the logarithms of convective precipitation (CP) and SRH. Convective precipitation is precipitation associated with conditional instability. *Galway* [1979] related yearly and seasonal precipitation with tornado activity. SRH, a measure of the potential for rotational updrafts and is often used in tornado forecasting [*Davies-Jones*, 1993]. The PR based on NARR climatology data is (suppressing the offset)

$$\mu = \exp(-10.59 + 1.36\ \log(\mathrm{CP}) + 1.89\ \log(\mathrm{SRH})). \quad (3)$$

The units of SRH and CP are m^2/s^2 and kg/m^2/day, respectively. The bootstrap-estimated standard errors of the regression parameters are 0.25 for the intercept term and 0.02 and 0.05 for the convective precipitation and SRH coefficients, respectively. Using a $0.5° \times 0.5°$ grid gives as intercept, CP and SRH coefficients: -10.35, 1.36 and 1.8, respectively, showing that the regression coefficients are relatively insensitive to grid resolution and that the standard error estimates for the coefficients are reasonable measures of uncertainty. Excluding F0 tornadoes from the regression gives as intercept, CP and SRH coefficients: -12.00, 1.34 and 2.0, respectively, indicating similar sensitivities but fewer overall numbers.

3. Results

[9] The spatial distribution of the total number of reported tornadoes 1979–2010 and the corresponding PR-fit values are similar (Figure 2). The dominant feature is the so-called "Tornado Alley" running north-south in the Central U.S. Observations and PR-fit values show few tornadoes west of the Rocky Mountains and over the Appalachian Mountains. Relatively high tornado activity is observed in northeastern Colorado and Florida but is not seen in the PR-fit values. This difference may be due to non-supercell tornadoes being common in both of these area [*Brooks and Doswell*, 2001] while the PR focuses on quantities related to supercell dynamics. Tornado activity along the coasts of the Atlantic seaboard states are seen in both observations and PR-fit values.

[10] The observed and PR-fit seasonal cycle of tornado occurrence have similar phasing (Figure 3) with maximum values occurring in May, followed closely by those of June. There is no explicit accounting for seasonality in the PR; all seasonality comes from the environmental parameters. PR-fit values are too small during May and June (particularly in the High Plains and Upper Midwest regions, not shown) and too large in late summer and autumn (particularly in the Missouri-Iowa-Nebraska-Kansas area, not shown). A measure of the seasonal cycle of the spatial distribution is found by computing at each grid point the month with the maximum number of tornadoes (Figure 4). The PR-fit values capture the observed general northwest progression [*Brooks et al.*, 2003b].

[11] Although the PR fits well the climatological tornado data on which it was developed, there is no assurance that the same relations are relevant to year-to-year tornado variability. However, when the PR developed with climatological data is applied to yearly-varying monthly values (1979–2010), the PR-estimated values correlate well with the

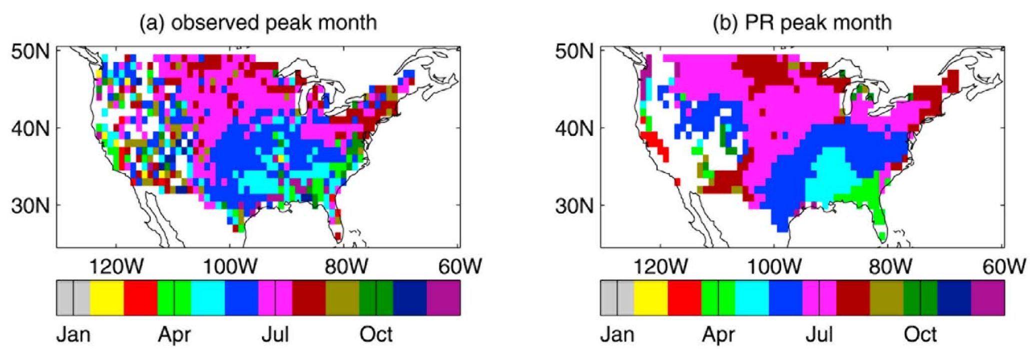

Figure 4. Colors indicate the calendar month with climatological maximum number of tornadoes according to (a) observations and (b) Poisson regression (PR)-fit values.

Table 1. Pearson and Rank Correlation (Spearman's Rho) Between Reported Number of Tornadoes and North American Regional Reanalysis (NARR) Poisson Regression Estimates 1979–2010, and CFSv2 Forecast Poisson Regression Estimates 1982–2010[a]

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
NARR PR Pearson corr.	**0.75**	**0.64**	**0.54**	**0.50**	**0.60**	**0.67**	**0.75**	**0.40**	0.15	0.25	**0.48**	**0.74**
NARR PR Rank corr.	**0.73**	**0.55**	**0.56**	**0.55**	**0.69**	**0.72**	**0.63**	**0.50**	0.25	**0.44**	**0.57**	**0.58**
CFSv2 PR Pearson corr.	0.36	**0.38**	0.3	0.35	0.31	**0.72**	**0.59**	**0.41**	−0.25	0.18	**0.41**	**0.37**
CFSv2 PR Rank corr.	**0.65**	0.19	0.28	**0.42**	0.31	**0.71**	**0.43**	**0.48**	−0.17	0.33	**0.40**	0.06

[a]Correlations significant at 95% level are in bold font.

observed monthly number of tornadoes (Table 1); the seasonal cycle is removed and does not contribute to the monthly correlations. We show both Pearson correlation and rank correlation (Spearman's rho). Pearson correlation measures linear association while rank correlation is a nonparametric association measure which is insensitive to outliers. The Pearson correlation (rank correlation) between the observed annual number of tornadoes and that given by the PR is 0.51 (0.28), and this value increases to 0.64 (0.48) when observations are adjusted to account for changes in observing system (Figure 5a). The rank correlation more harshly penalizes the fact that the observed annual counts have a trend while the PR values do not. The fact that the annual PR values have no obvious trend is further evidence for the observational trend being nonphysical. Notable also is the PR value of 501 for April 2011, the most active U.S. tornado month on record (Figure 5b).

[12] The demonstrated relation between monthly averaged environmental parameters and monthly tornado numbers provides a framework for extended-range forecasts of tornado activity. One first predicts the environmental parameters in the PR and then uses the PR to predict the impact of those parameters on tornado activity. The skill of the resulting tornado activity forecasts clearly cannot exceed the skill of the forecasts of the environmental parameters. Verification of CFSv2 reforecasts of CP and SRH with NARR data shows U.S. annually averaged correlations of 0.27 and 0.48 respectively. We compute the PR index using CFSv2

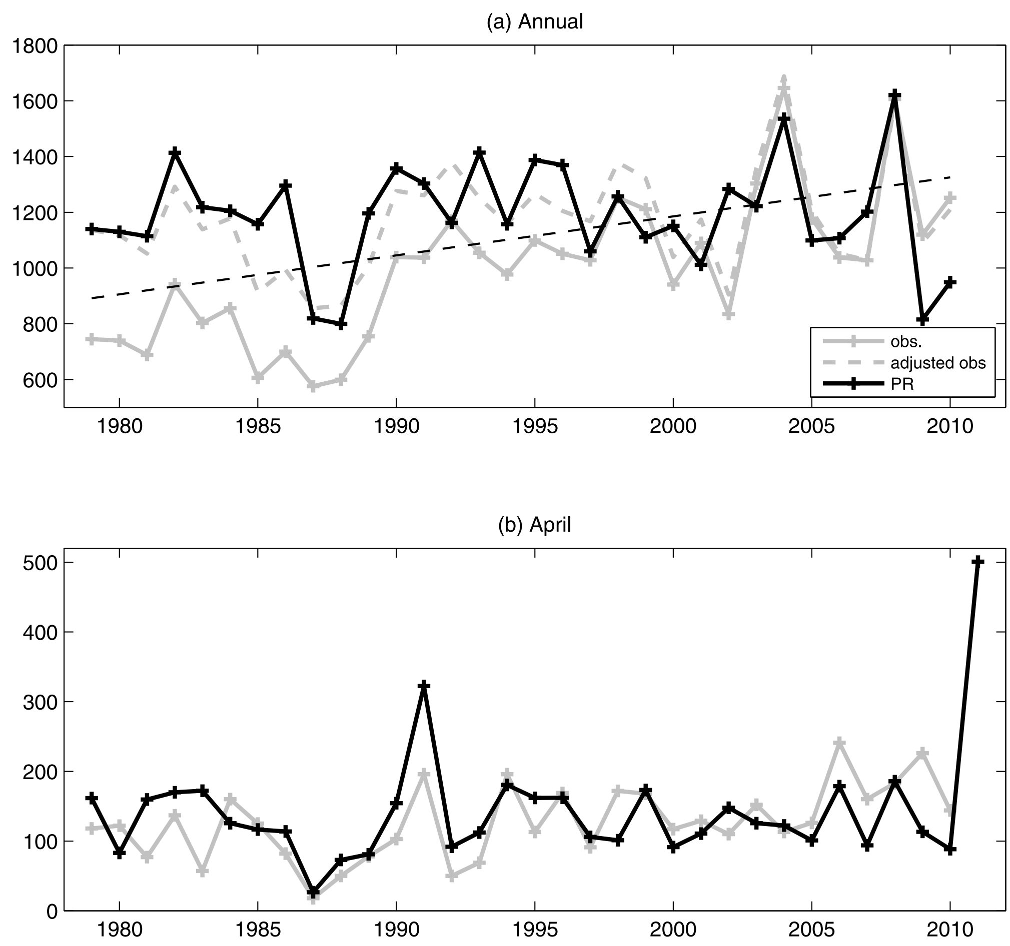

Figure 5. Observed (gray), trend adjusted observations (dashed gray) and Poisson regression (black) (a) annual and (b) April numbers of tornadoes during the period 1979–2010. The dashed black line in panel (a) is the 1954–2007 trend line used in the adjustment. Panel (b) includes the April 2011 Poisson regression value.

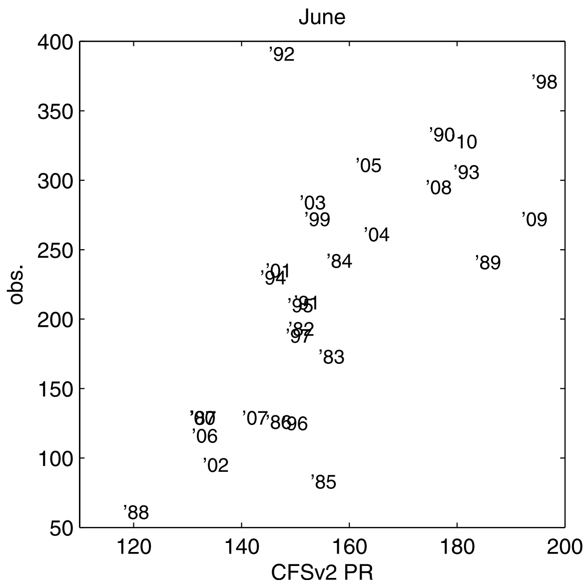

Figure 6. Scatter plot of observed with 0-month-lead CFSv2 forecast model-based predicted number of June tornadoes 1982–2010. The 2-digit number indicates the year.

reforecast CP and SRH and correlate it with the reported numbers of tornadoes by month (Table 1). The correlation of reported tornado numbers with CFSv2 PR estimates is generally lower than with NARR PR estimates, likely due to the overall low CP skill level. However, 6 of the 12 correlations are significant at the 95% level (with some disagreement between Pearson and rank correlation), and the June correlation is particularly strong, though the CFSv2 predicted values have low amplitudes compared to observations (Figure 6).

4. Summary and Discussion

[13] Ultimately, we would like to understand why some periods have more (or less) tornado activity than others. If this question can only be answered using high-frequency and high spatial resolution environmental information, the prospects would appear bleak for both extended-range forecasts and climate projections of tornado activity. Here we have demonstrated that although tornado formation directly depends on the immediate environment, monthly U.S. tornado activity can be related to observed monthly averaged environmental parameters and that Poisson regression can be used to construct an index that captures aspects of the climatological and year-to-year variability of tornado activity. The value of spatially averaged environmental data was previously demonstrated by *Brooks et al.* [2003a] and applied to climate change projections [*Trapp et al.*, 2007]. The utility of monthly averaged environment parameters in describing monthly tornado activity is new, but is consistent with previous studies which have considered the modulation of tornado activity by monthly and seasonal phenomena such as precipitation, ENSO and the Intra-Americas Sea low-level jet [*Galway*, 1979; *Cook and Schaefer*, 2008; *Muñoz and Enfield*, 2011].

[14] The predictability of such an index depends on the predictability of its constituent parameters, here, storm relative helicity and convective precipitation (CP). We find in the extended-range context that the predictability of CP appears to be the limiting factor, though the details of the joint spatial distributions of skill and tornado activity are likely important. Computing the index with parameters from an operational seasonal forecast model shows statistically significant skill in forecasting the tornado activity of the following month for some months of the year.

[15] **Acknowledgments.** The authors thank two anonymous reviewer for their useful comments. M.K.T. is supported by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration (NA05OAR4311004). A.H.S. and S.J.C. acknowledge support from NOAA grant NA08OAR4320912. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies.

[16] The Editor thanks the two anonymous reviewers for their assistance in evaluating this paper.

References

Brooks, H. E., and C. A. Doswell III (2001), Some aspects of the international climatology of tornadoes by damage classification, *Atmos. Res.*, *56*(1–4), 191–201.
Brooks, H. E., and N. Dotzek (2007), The spatial distribution of severe convective storms and an analysis of their secular changes, in *Climate Extremes and Society*, edited by H. F. Diaz and R. Murnane, pp. 35–54, Cambridge Univ. Press, New York.
Brooks, H. E., C. A. Doswell III, and J. Cooper (1994), On the environments of tornadic and nontornadic mesocyclones, *Weather Forecast.*, *9*, 606–618.
Brooks, H. E., J. W. Lee, and J. P. Craven (2003a), The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data, *Atmos. Res.*, *67–68*, 73–94.
Brooks, H. E., C. A. Doswell III, and M. P. Kay (2003b), Climatological estimates of local daily tornado probability for the United States, *Weather Forecast.*, *18*, 626–640.
Camargo, S. J., A. H. Sobel, A. G. Barnston, and K. A. Emanuel (2007), Tropical cyclone genesis potential index in climate models, *Tellus, Ser. A*, *59*, 428–443.
Cook, A. R., and J. T. Schaefer (2008), The relation of El Niño–Southern Oscillation (ENSO) to winter tornado outbreaks, *Mon. Weather Rev.*, *136*, 3121–3137, doi:10.1175/2007MWR2171.1.
Davies-Jones, R. P. (1993), Hourly helicity, instability, and EHI in forecasting supercell tornadoes, paper presented at 17th Conference on Severe Local Storms, Am. Meteorol. Soc., St. Louis, Mo.
Galway, J. G. (1979), Relationship between precipitation and tornado activity, *Water Resour. Res.*, *15*, 961–964.
Gray, W. M. (1979), Hurricanes: Their formation, structure and likely role in the tropical circulation, in *Meteorology Over the Tropical Oceans*, edited by D. B. Shaw, pp. 155–218, R. Meteorol. Soc., London.
Hamill, T. M., and A. T. Church (2000), Conditional probabilities of significant tornadoes from RUC-2 forecasts, *Weather Forecast.*, *15*, 461–475.
Mesinger, F., et al. (2006), North American regional reanalysis, *Bull. Am. Meteorol. Soc.*, *87*, 343–360.
Muñoz, E., and D. Enfield (2011), The boreal spring variability of the Intra-Americas low-level jet and its relation with precipitation and tornadoes in the eastern United States, *Clim. Dyn.*, *36*, 247–259, doi:10.1007/s00382-009-0688-3.
Rasmussen, E. N., and D. O. Blanchard (1998), A baseline climatology of sounding-derived supercell and tornado forecast parameters, *Weather Forecast.*, *13*, 1148–1164.
Saha, S., et al. (2006), The NCEP Climate Forecast System, *J. Clim.*, *19*, 3483–3517.
Saha, S., et al. (2010), The NCEP Climate Forecast System reanalysis, *Bull. Am. Meteorol. Soc.*, *91*, 1015–1057, doi:10.1175/2010BAMS3001.1.
Schaefer, J. T., and R. Edwards (1999), The SPC Tornado/Severe Thunderstorm Database, paper presented at 11th Conference on Applied Climatology, Am. Meteorol. Soc., Dallas, Tex.
Shafer, C. M., A. E. Mercer, L. M. Leslie, M. B. Richman, and C. A. Doswell III (2010), Evaluation of WRF model simulations of tornadic and nontornadic outbreaks occurring in the spring and fall, *Mon. Weather Rev.*, *138*, 4098–4119, doi:10.1175/2010MWR3269.1.
Tippett, M. K., S. J. Camargo, and A. H. Sobel (2011), A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis, *J. Clim.*, *24*, 2335–2357.
Trapp, R. J., N. S. Diffenbaugh, H. E. Brooks, M. E. Baldwin, E. D. Robinson, and J. S. Pal (2007), Changes in severe thunderstorm

environment frequency during the 21st century caused by anthropogenically enhanced global radiative forcing, *Proc. Natl. Acad. Sci. U. S. A.*, *104*, 19,719–19,723.
Vecchi, G. A., M. Zhao, H. Wang, G. Villarini, A. Rosati, A. Kumar, I. M. Held, and R. Gudgel (2010), Statistical-dynamical predictions of seasonal North Atlantic hurricane activity, *Mon. Weather Rev.*, *139*, 1070–1082, doi:10.1175/2010MWR3499.1.
Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz (2006), Evolution of the U.S. tornado database: 1954–2003, *Weather Forecast.*, *21*, 86–93.
Yuan, X., E. F. Wood, L. Luo, and M. Pan (2011), A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction, *Geophys. Res. Lett.*, *38*, L13402, doi:10.1029/2011GL047792.

S. J. Camargo, Lamont-Doherty Earth Observatory, Earth Institute at Columbia University, Palisades, NY 10964-8000, USA.

A. H. Sobel, Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027, USA.

M. K. Tippett, International Research Institute for Climate and Society, Columbia University, Palisades, NY 10964, USA. (tippett@iri.columbia.edu)