

RESEARCH ARTICLE

Statistical Models for Tornado Climatology: Long and Short-Term Views

James B. Elsner*, Thomas H. Jagger, Tyler Fricker

Department of Geography, Florida State University, Tallahassee, Florida, United States of America

* jelsner@fsu.edu

Abstract

This paper estimates regional tornado risk from records of past events using statistical models. First, a spatial model is fit to the tornado counts aggregated in counties with terms that control for changes in observational practices over time. Results provide a long-term view of risk that delineates the main tornado corridors in the United States where the expected annual rate exceeds two tornadoes per 10,000 square km. A few counties in the Texas Panhandle and central Kansas have annual rates that exceed four tornadoes per 10,000 square km. Refitting the model after removing the least damaging tornadoes from the data (EF0) produces a similar map but with the greatest tornado risk shifted south and eastward. Second, a space-time model is fit to the counts aggregated in raster cells with terms that control for changes in climate factors. Results provide a short-term view of risk. The short-term view identifies a shift of tornado activity away from the Ohio Valley under El Niño conditions and away from the Southeast under positive North Atlantic oscillation conditions. The combined predictor effects on the local rates is quantified by fitting the model after leaving out the year to be predicted from the data. The models provide state-of-the-art views of tornado risk that can be used by government agencies, the insurance industry, and the general public.



OPEN ACCESS

Citation: Elsner JB, Jagger TH, Fricker T (2016) Statistical Models for Tornado Climatology: Long and Short-Term Views. PLoS ONE 11(11): e0166895. doi:10.1371/journal.pone.0166895

Editor: Juan A. Añel, Universidade de Vigo, SPAIN

Received: July 13, 2016

Accepted: November 4, 2016

Published: November 22, 2016

Copyright: © 2016 Elsner et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data are available from the GitHub repository at the following URL: <https://github.com/jelsner/SeasonalPrediction>.

Funding: This work is supported by the Risk Prediction Initiative of the Bermuda Institute for Ocean Studies (BIOS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Seasonal climate forecasts are issued routinely. Each spring, for example, weather agencies in nations across the world make predictions for how hot and dry the summer is likely to be. And predictions for hurricane activity along the coast are typically accurate enough to warrant attention by the property insurance industry. Yet there remains no regularly issued forecasts of tornado activity months in advance despite demonstrated useful skill (accuracy above random guess) at predicting tornado activity prior to the start of the season [1] and initial forays into public dissemination of a forecast system [2]. The absence of routine seasonal tornado forecasts is due to large gaps in the understanding of how climate affects severe weather and to the limited methods to forecast activity on this time scale.

Dynamical models are used as guidance to make seasonal forecasts of temperature and precipitation anomalies but tornadoes are too small to be resolved in them. Dynamical guidance on environmental conditions favorable for tornadoes is available out to about two weeks [3] though there is a high false alarm rate on the predictions with increasing lead time. The

necessary conditions are not sufficient to completely distinguish between days with and without tornadoes. An alternative approach is to fit statistical models to historical observed data. Climate patterns related to active and inactive tornado seasons provide information to make predictions but population growth and changes to procedures for rating tornadoes result in a heterogeneous database.

Various methods for dealing with artifacts in the tornado data have been proposed [4–6] with most procedures assuming a uniform region of activity and estimating occurrence rates within a subset of the region likely to be most accurate. For example, tornado reports are often aggregated using kernel smoothing [7–9]. The resulting spatial density maps show regions of higher and lower tornado frequency which is useful for exploratory analysis and hypothesis generation. However, correctly interpreting the patterns is a problem since there is no control for environmental factors. Another drawback is the assumption (implicit) that tornadoes occur randomly (not clustered). This is not generally the case as a single thunderstorm can spawn a family of tornadoes within a relatively compact area [10]. Also, tornado reports tend to be more numerous near cities compared to rural areas confounding attempts to properly assess the risk over large regions [11]. Improvements in observing practices tend to result in more tornado reports, especially reports of weak tornadoes [12, 13] and ones occurring over remote areas. Finally, natural climate variations make some seasons more active than others. For instance, variations in sea-surface temperature and atmospheric convection in the tropical Pacific associated with the El Niño/Southern Oscillation (ENSO) modulate global weather and climate patterns including the threat of tornadoes [2, 14–17]. In short, a statistical model capable of controlling for these various factors and data artifacts provides a convenient way to ‘smooth’ the tornado risk.

The purpose of this paper is to describe a strategy for the development of a comprehensive seasonal tornado risk assessment system. It follows the methodology of [18] but significantly expands the scope with a larger spatial domain, more predictors, and more complete validation. Part one fits a climatology model to data aggregated by county in states across the Midwest, Plains, and Southeast (long-term view of risk). The model uses annual population to control for changes in observational practices over time. Results quantify a long-term view of risk independent of climate variability. Part two fits a conditional climatology model to data aggregated in grid cells that predicts how the rates should be adjusted given current (or projected) climate conditions. The model controls for changes in observational practices over time using a trend term. Results quantify a short-term view of risk that depends on current climate conditions. A discussion of the model results and the potential utility of the models follows.

Materials and Methods

Data

A key variable in the long-term model is annual population by county that serves as a proxy for changes in observational practices. These values are available as archives by the U.S. Census Bureau and available from www.nber.org/data/census-intercensal-county-population.html. The latest cleaned population values are available for 2012. Population density is computed by dividing the population by county area and expressing the values in persons per square kilometers. On average, the greater the population density the greater the chance that a tornado gets in the record. Annual population values vary over space and time but we set the population density for each county to be equal to its maximum density over the 46-year period. Other proxies for changes in observational practices are possible (see [19]).

The U.S. Census Bureau map boundaries are available from www.census.gov/geo/maps-data/data/cbf/cbf_counties.html. Here the 5m = 1:5,000,000 scale is used. State boundaries are

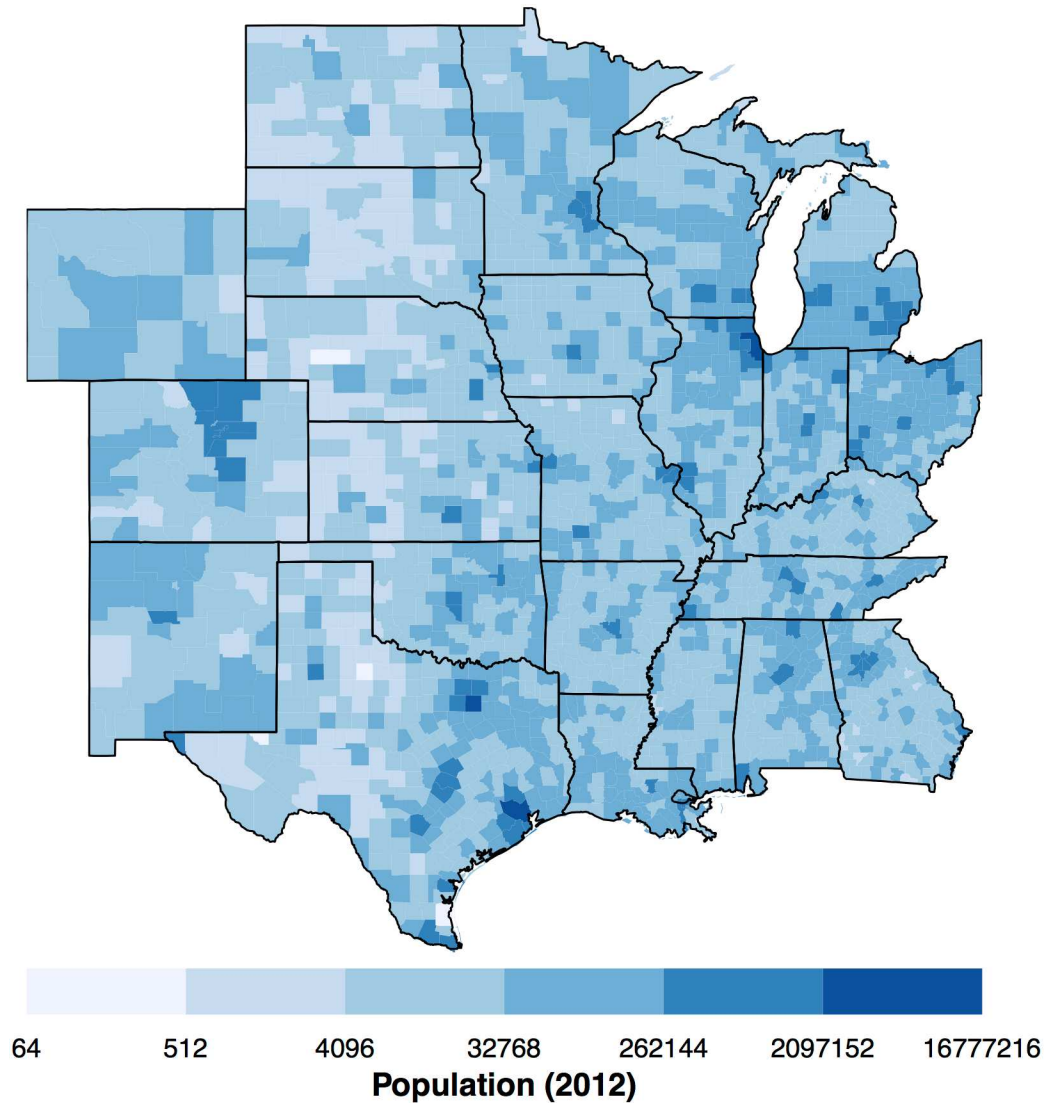


Fig 1. County population in 2012.

doi:10.1371/journal.pone.0166895.g001

extracted using the state Federal Information Processing Standard (FIPS). The set of 24 contiguous states including Wyoming, Colorado, New Mexico, North and South Dakota, Nebraska, Kansas, Oklahoma, Texas, Louisiana, Arkansas, Missouri, Iowa, Minnesota, Wisconsin, Michigan, Illinois, Indiana, Ohio, Kentucky, Tennessee, Georgia, Alabama, and Mississippi define the study area. There are 2168 counties across the two dozen states covering 7.6 million square kilometers. Combining the population data with the map boundaries the county-level population for 2012 is mapped in Fig 1. A north-south region of low population density extends across the western portions of the high Plains.

The tornado data are from NOAA’s Storm Prediction Center (SPC) and available from www.spc.noaa.gov/gis/svrgis/zippped/tornado.zip. The SPC maintains a comprehensive and up-to-date tornado database. Records extend back to 1950 and up through 2015 and include occurrence time, location, magnitude, track length and width, fatalities, and injuries for tornadoes in the United States. The version of the SPC database used in this study is in shapefile

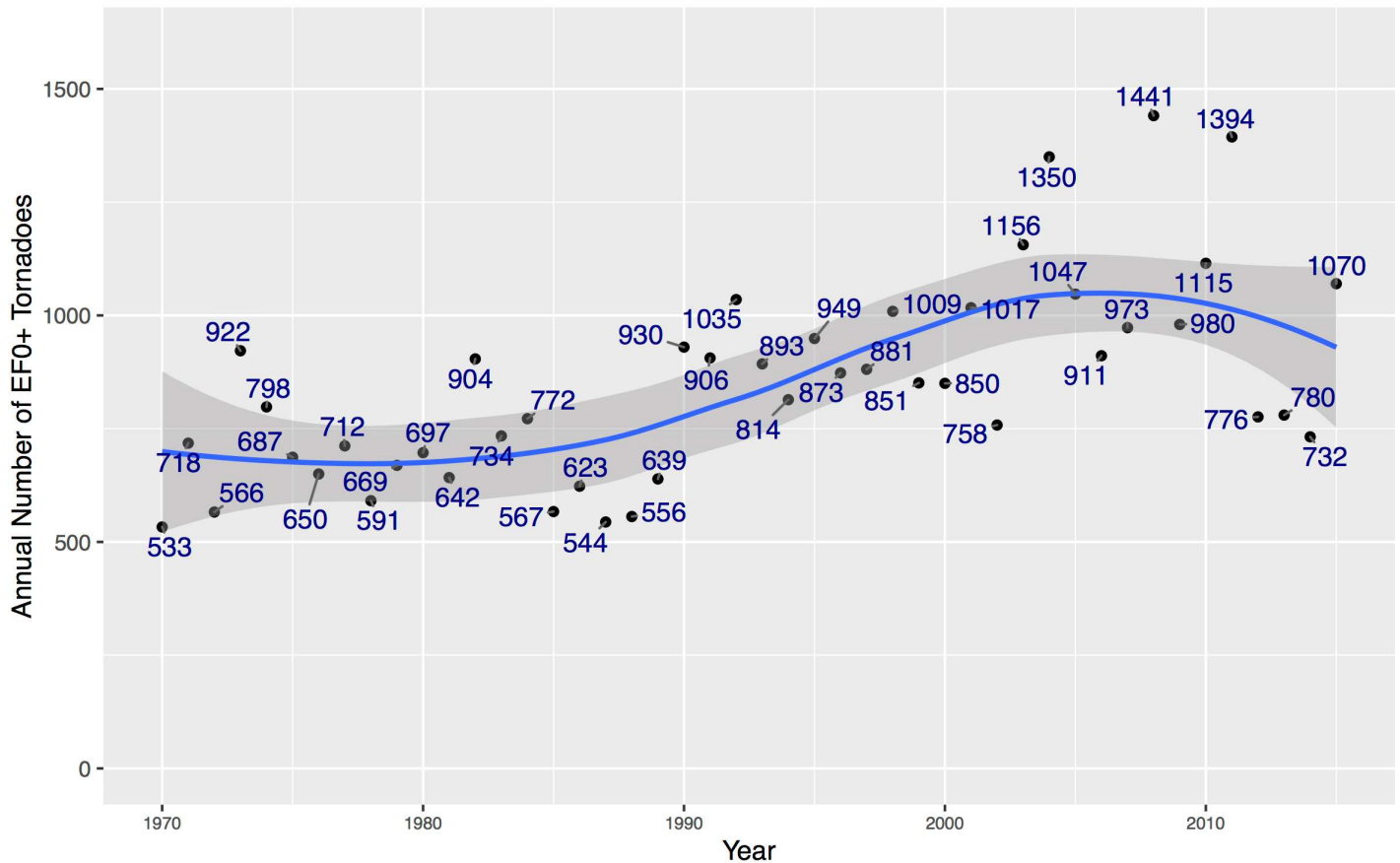


Fig 2. Annual number of tornadoes (1970–2015) over the states used in the long-term view model. The blue line is a 2nd-order local polynomial regression smoother (span = .75) and the gray band is the 95% uncertainty range on the smoother.

doi:10.1371/journal.pone.0166895.g002

format with each tornado represented as a straight-line track in a Lambert conformal conic (LCC) projection centered on 96° W longitude and parallels at 33° and 45° N latitudes. The native coordinates are transformed to a Mercator projection. The data are quality controlled for errors prior to inclusion in the database.

Tornado tracks are buffered to create damage-path polygons. The buffer is one half the value of the width variable specified in the attribute table. A flat cap on the buffer is used so the damage path is the same length as the track. The polygon paths are laid on top of the domain and a vector is returned indicating either NA (no portion of the damage path is inside the study area) or county numbers indicating which counties were affected. Duplicate paths (1.03% of all paths) are removed by checking for exact matches in width, length, date, time, and start location. For the long-term view all tornadoes starting with 1970 are used to create a 46-year climatology. There are a total of 39,015 tornadoes over this period and region. The annual number of tornadoes are plotted in Fig 2. There is an upward trend in the annual counts that appears to have leveled out around 2005. A trend term is included in the models.

Paths are laid over the county boundaries to get a per-county tornado count. The result is a list with the number of items equal to the number of counties. Each item contains a subset of the track attribute table. Padding with zeros is needed for the 19 counties without tornadoes (less than 1% of all counties). Then for each county the number of tornadoes and tornado days are tabulated and plotted (Fig 3). The maps show a tendency for larger counties to have more

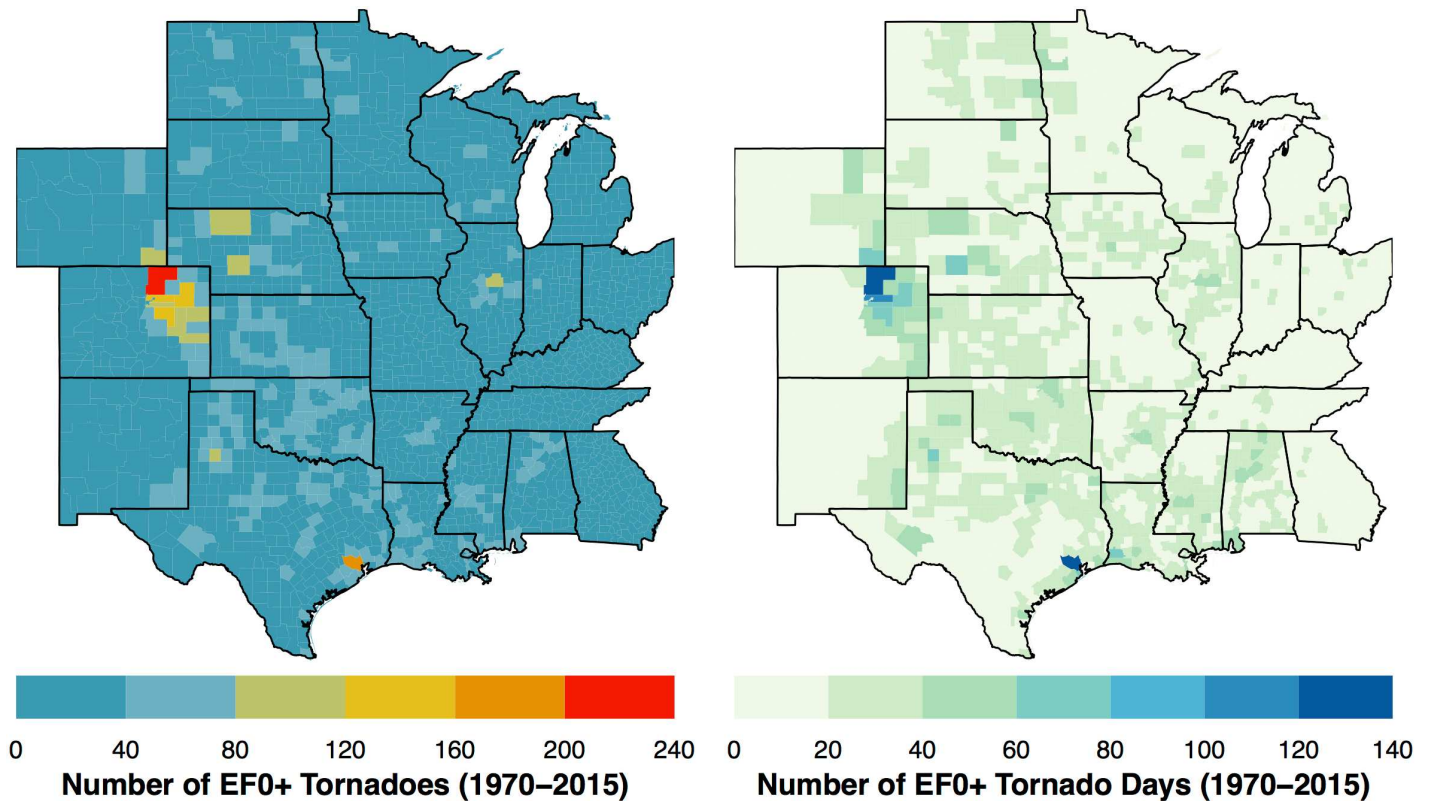


Fig 3. Number of tornadoes 1970–2015 (left) and number of days with at least one tornado.

doi:10.1371/journal.pone.0166895.g003

tornadoes and more tornado days. Leading the list is Weld County, Colorado with a total of 219 tornado reports falling on 137 tornado days but spread over more than 18,000 square km. Next is Harris County, Texas with a total of 192 tornadoes falling on 124 days spread over more than 6,000 square km. County areas are included in the model (see next subsection) as a part of the exposure term.

A short-term view quantifies how much the tornado risk changes with a unit change in the climate variable (predictor). If the influence varies regionally then quantification is done spatially. To manage spatial variation, the short-term view model uses a raster of grid cells. The raster has uniform size and shape contiguous cells making computations and comparisons easier. Annual tornado occurrences are counted in two degree cells based on track intersections (Fig 4). To simplify the overlay operation straight-line tracks are used instead of paths. The result is a space-time data set with cell area as a constant-time attribute and tornado count as a variable-time attribute. The study area extends from eastern Colorado to western Virginia and from the Mexican Gulf coast to southern Minnesota. The period of record runs from 1954 to 2015 for a total of 48,200 tornadoes representing 81.7% of all tornadoes and 91.6% of all violent tornadoes (EF4+). Data earlier than 1970 are included in the short-term view since the model has a trend term for the increasing probability of tornado detection with increasing modernization. It is clear that at the cell level there is large variability in counts from one year to the next. Variability across space is also large in some years.

Key variables in the short-term view model are the climate predictors. Predictors are chosen based on literature research. They include indexes for ENSO and for the North Atlantic Oscillation (NAO) and sea-surface temperatures (SST) from the Gulf of Alaska and Western

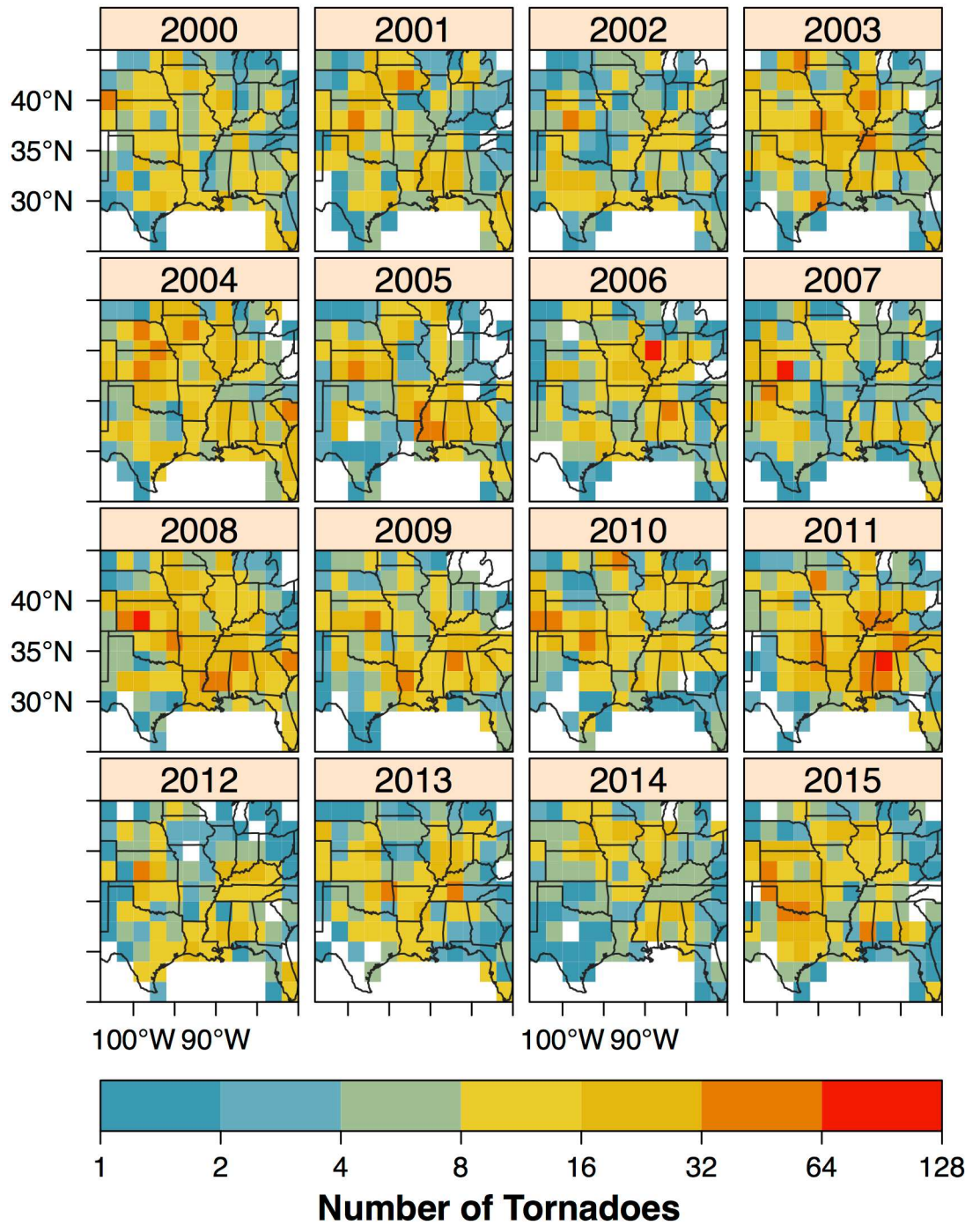


Fig 4. Annual tornado counts over the period 2000–2015 in two degree raster cells. A count is added to a cell if a tornado track intersects it.

doi:10.1371/journal.pone.0166895.g004

Caribbean Sea. The ENSO index, in units of standard deviation, is the bi-variate ENSO time series averaged from March through May. The monthly series combines a standardized Southern Oscillation Index with a standardized Niño3.4 SST series computed from the Hadley Centre’s data [20]. The monthly values from March through May are averaged to obtain the index.

Monthly NAO values, in units of standard deviation, are constructed from a rotated principal component analysis of the 500 hPa heights across the Northern Hemisphere. Higher than average heights over the subtropical Atlantic combined with lower than average heights in the vicinity of Iceland result in a strongly positive values of the NAO. Details of the procedure are given in [21]. Monthly values from April through June are averaged to obtain the NAO index. The ENSO and NAO indices are obtained from the Climate Prediction Center.

The SST regions are selected based on [1] showing a tendency for a combination of colder than average ocean waters in the Gulf of Alaska and warmer than average ocean waters in the Western Caribbean Sea to favor tornado activity across the central United States during spring. The monthly SST values are spatially averaged from the NCEP/NCAR reanalysis grids [22]. The Gulf of Alaska (GAK) region is bounded by 60 and 50.4° N latitudes and 136 and 153.6° W longitudes. Following [1] the GAK is the spatial average over the region for the month of April. The Western Caribbean Sea (WCA) region is bounded by 25 and 15° N latitudes and 70 and 90° W longitudes. The WCA is the spatial average over the region for the month of February. The indexes and SST data are obtained from the Physical Science Division of the Earth System Research Laboratory. Time series plots (Fig 5) of the four predictors used in the short-term view model indicate year-to-year variation but no significant upward or downward trends.

Models

Raw counts are problematic for directly assessing tornado rates because counties vary in size and population. To control for this a statistical model is fit to the counts. The long-term view model includes population density as a fixed effect. To account for improvements in the procedures to rank tornadoes by the amount of damage, the calendar year and an interaction term of year with population are included. Mathematically, the number of tornadoes in each county s (T_s) is assumed to be described by a negative binomial distribution with parameters probability p and size r [1]. If X is a random sample from this distribution, then the probability that $X = k$ is $P(k|r, p) = \binom{k+r-1}{k} (1-p)^r p^k$, for $k \in 0, \dots, \infty$, $p \in (0, 1)$ and $r > 0$. This relates the probability of observing k successes before the r -th failure of a series of independent events with probability of success equal to p . The distribution is generalized by allowing r to be any positive real number and it arises from a Poisson distribution whose rate parameter can be described by a gamma distribution [23].

The negative binomial distribution is re-formulated using the mean $\mu = r \frac{p}{1-p}$ and the size r . This separates the mean from the dispersion parameter. The mean μ_s is linked to a linear combination of the predictors and random terms, v_s , through the exponential function and the area of the cell, A_s (exposure). The dispersion is modeled with a scaled size parameter n where $n = r_s/A_s$ giving a dispersion of $1/p_s = 1 + \mu_s/n = 1 + \exp(v_s)/n$ that depends only on the tornado rate and n . More concisely the model is:

$$T_s | \mu_s, r_s \sim \text{NegBin}(\mu_s, r_s) \tag{1}$$

$$\mu_s = A_s \exp(v_s) \tag{2}$$

$$v_s = \beta_0 + \beta_1 \text{lpd}_s + \beta_2 (t - t_0) + \beta_3 \text{lpd}_s (t - t_0) + u_s + v_t \tag{3}$$

$$r_s = A_s n \tag{4}$$

where $\text{NegBin}(\mu_s, r_s)$ indicates that the conditional tornado counts ($T_s | \mu_s, r_s$) are described by negative binomial distributions with mean μ_s and size r_s , lpd_s represents the base two logarithm of the maximum population density over the 46-year period for each county, and t_0 is the base

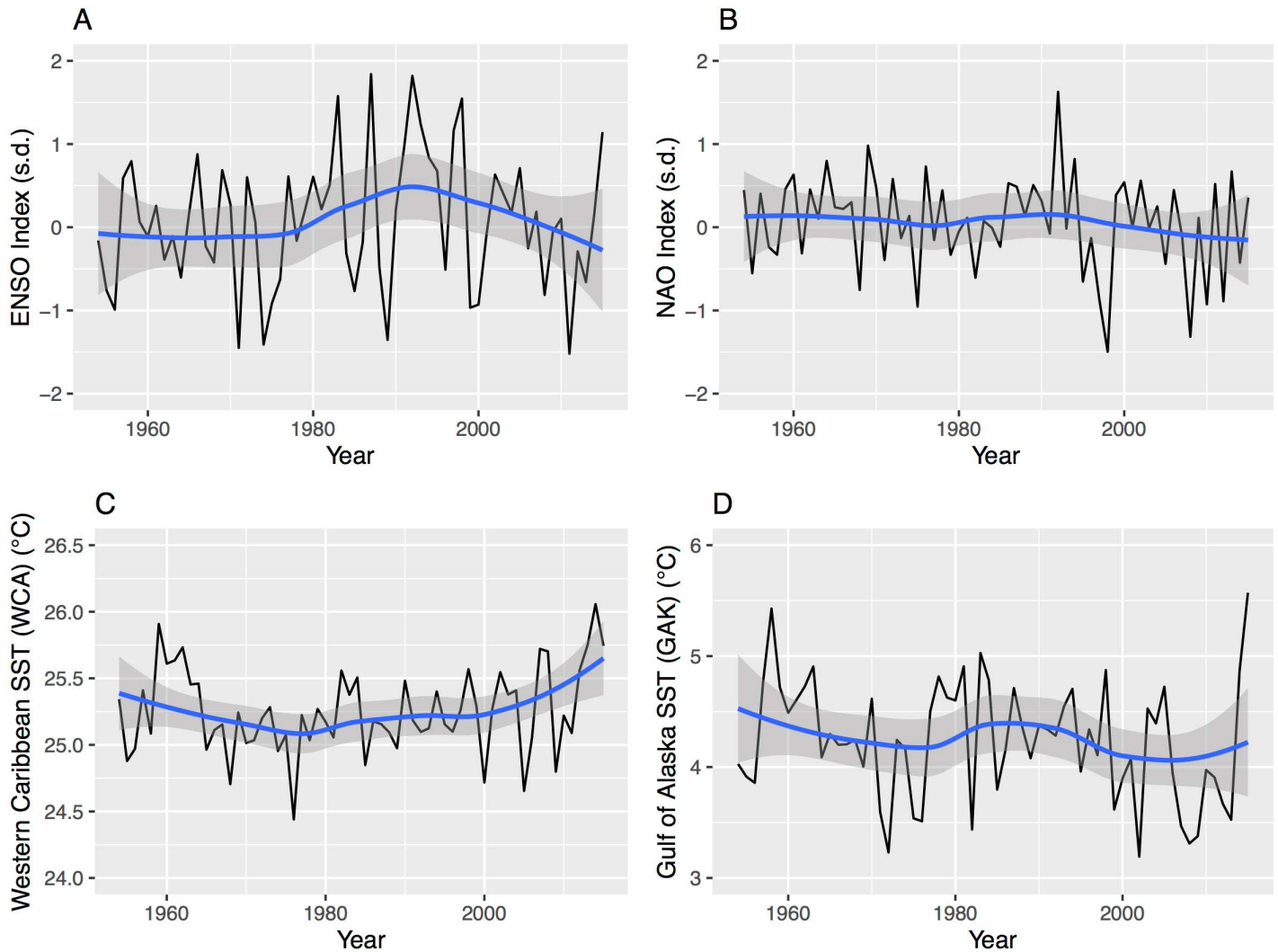


Fig 5. Time series (1954–2015) of the four predictors used in the short-term view model. The blue line is a 2nd-order local polynomial regression smoother (span = .75) and the gray band is the 95% uncertainty range on the smoother.

doi:10.1371/journal.pone.0166895.g005

year set to 1991 (middle year of the record). The spatially correlated random term u_s follows an intrinsic Besag formula with a sum-to-zero constraint [24].

$$u_i | \{u_{j \neq i}, \tau\} \sim \mathcal{N}\left(\frac{1}{m_i} \sum_{i \sim j} u_j, \frac{1}{m_i \tau}\right), \tag{5}$$

where \mathcal{N} is the normal distribution with mean $1/m_i \cdot \sum_{i \sim j} u_j$ and variance $1/m_i \cdot 1/\tau$ where m_i is the number of neighbors of cell i and τ is the precision; $i \sim j$ indicates cells i and j are neighbors. Neighboring cells are determined by contiguity (queen’s rule—eight nearest cells). The annual uncorrelated random term, v_b , is modeled as a sequence of normally distributed random variables, with mean zero and variance $1/\tau'$. The prior on the spatial random term is statistically independent from the annual random term.

The short-term view model extends the long-term view model. Subscripts on parameters and variables indicate a space (s) and a time (t) component. Mathematically, the raw tornado

count in grid cell s for year t is given as:

$$T_{s,t} | \mu_{s,t}, r_{s,t} \sim \text{NegBin}(\mu_{s,t}, r_{s,t}) \tag{6}$$

$$\mu_{s,t} = A_s \exp(v_{s,t}) \tag{7}$$

$$v_{s,t} = \beta_1 (t - t_0) + \beta_2 \text{ID}_s + \beta_3 \text{GAK}_t + \beta_{4,s} \text{ENSO}_t + \beta_{5,s} \text{NAO}_t + \beta_{6,s} \text{WCA}_t \tag{8}$$

$$r_{s,t} = A_s n \tag{9}$$

where again the conditional tornado count in each cell is described by a negative binomial distribution with mean $\mu_{s,t}$ and size $r_{s,t}$.

The annual rate in each cell $\mu_{s,t}$ is linked to a linear combination of the predictors ($v_{s,t}$) through the exponential function and the area of the cell, A_s . The predictors include the GAK as a spatially-uniform effect and ENSO, NAO, and the WCA as spatially-varying effects. The spatially-varying effects have an intrinsic Besag formula (Eq 5). Each cell is allowed to have unique variability through the ID term. The cell area times the number of years is the square-meter-years exposed to tornadoes and is normalized to have a mean of one over the domain.

Gaussian priors with low precision are assigned to the β 's. To complete the models, the scaled size (n) is assigned a log-gamma prior and the precision parameters (τ and τ') are assigned a log-Gaussian prior [23]. Application of Bayes rule using the method of integrated nested Laplace approximation (INLA) [25, 26] results in posterior densities for the model parameters.

Results

Long-Term View

Climatology. By specifying a 'future' year (here 2016) the long-term model predicts a distribution for the counts in each county. The mean value of the predictive distribution is the expected annual occurrence rate conditional on the historical raw counts and the model. Values are standardized by dividing by county area and expressing them as a rate per 100 km square region (Fig 6). This allows comparisons that are independent of county area. Note that the average county area is approximately 3500 km² so the per county rate in many cases is smaller. The color ramp is on a logarithmic scale so that each level of color saturation indicates a doubling of the occurrence rate.

As anticipated the broad-scale pattern matches the well-known tornado climatology. Highest rates are found from the northern High Plains into the Mid South. Lowest rates are found over the Rockies, the Rio Grande Valley, and the northern Great Lakes. Regional features include lower rates over the Appalachian Mountains and a local minimum to the northeast of the Ozark Mountains. Occurrence rates are consistently above two tornadoes per year across the central and southern Great Plains. Rates drop off rapidly moving westward with values generally less than one tornado every eight years across the Rockies. Rates drop off less rapidly moving eastward with most regions exceeding one tornado per year with the exceptions of the northern parts of Minnesota, Wisconsin, and Ohio, and over the Appalachian regions of Ohio and Kentucky. Uncertainty about the predicted risk is metered by the standard deviation of the predictive distribution (Fig 7). In general, areas with the largest standard errors are found in regions with the highest rates. For most of the study domain the uncertainty amounts to less

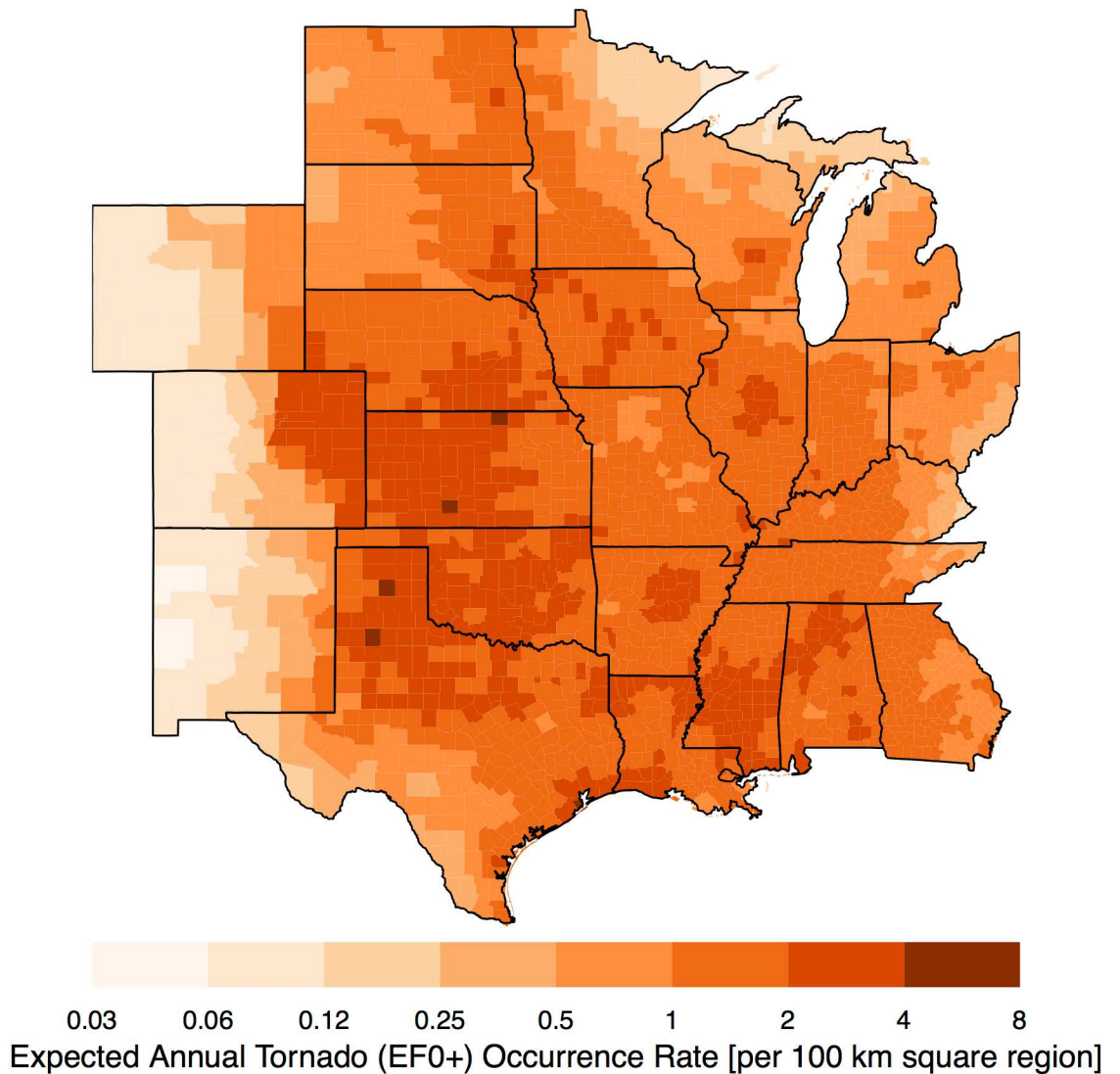


Fig 6. Long-term view. Expected annual tornado (EF0+) occurrence rates per 100 square kilometers. The data cover the period from 1970–2015.

doi:10.1371/journal.pone.0166895.g006

than .4 tornadoes per 100 square kilometers. Galveston County Texas has the largest uncertainty in excess of .9 tornadoes per 100 square kilometers.

Model Quality. Quality of the model is assessed by the cross-validated log score, the correlation between counts and expected rates, and the distribution of probability integral transform values. The log score is equivalent to a mean square error with smaller values indicating better prediction quality [27]. For all tornadoes the log score is .83 and the correlation between observed counts and expected rates is +.24. While the value is small relative to a perfect correlation the observations are counts while the predictions are rates so the upper limit is much less than unity especially given the over-dispersed counts. The distribution of the modified probability integral transform values is nearly uniform (Fig 8) indicating that the model fits the data well and that there is no model bias.

A comparison of rates provides additional insight on the quality of the model. The raw rate is the number of tornadoes in the county divided by the number of years and scaled to have

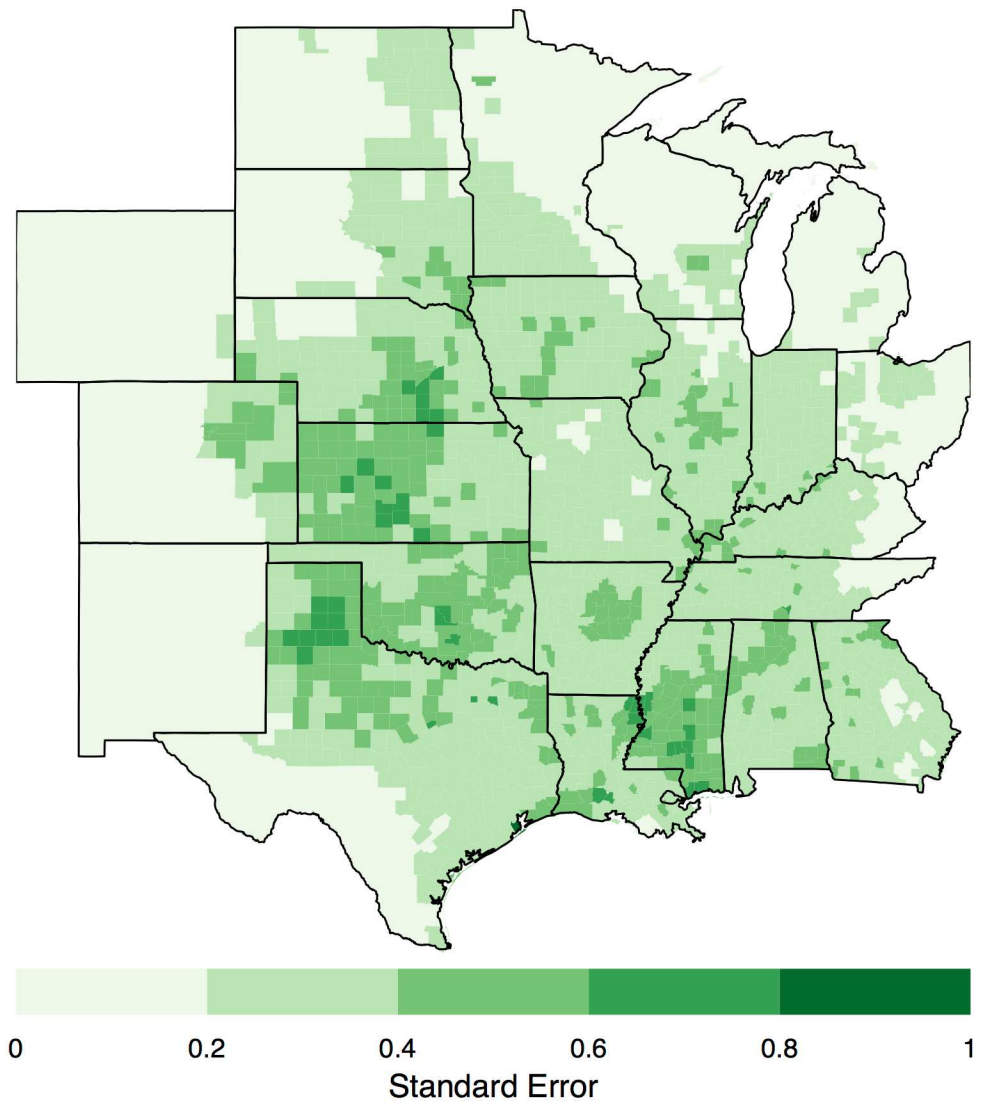


Fig 7. Variation in long-term rates. Standard deviation of tornado occurrence per 100 square kilometers.

doi:10.1371/journal.pone.0166895.g007

units of per 100 km squared area. The correlation between raw and expected rates is +.79. The relationship is shown in Fig 9. Each point on the graph indicates a county. Note that the raw rate does not control for trends or for the tendency of having fewer reports in less populated counties. Divergence between the raw and expected rates results from the smoothing imposed by the spatially correlated random term. Counties that have been exceptionally unlucky in terms of tornado strikes relative to neighboring counties are smoothed toward the neighborhood average. In the absence of evidence that a county has a unique risk relative to its neighbors, this is how it should be. The influence of multiple tornadoes intersecting a county on a given day is examined by refitting the model to tornado days rather than to tornado counts. The fit results in a correlation between raw and expected tornado-day rates of +.83 indicating a marginal improvement. The improvement is offset by the less intuitive interpretation of the expected rates (tornadoes per area vs tornado days per area).

A similar model is fit to a subset of data consisting only of tornadoes rated EF1 and higher. Just over two percent of the counties did not experience an EF1+ tornado over the period of

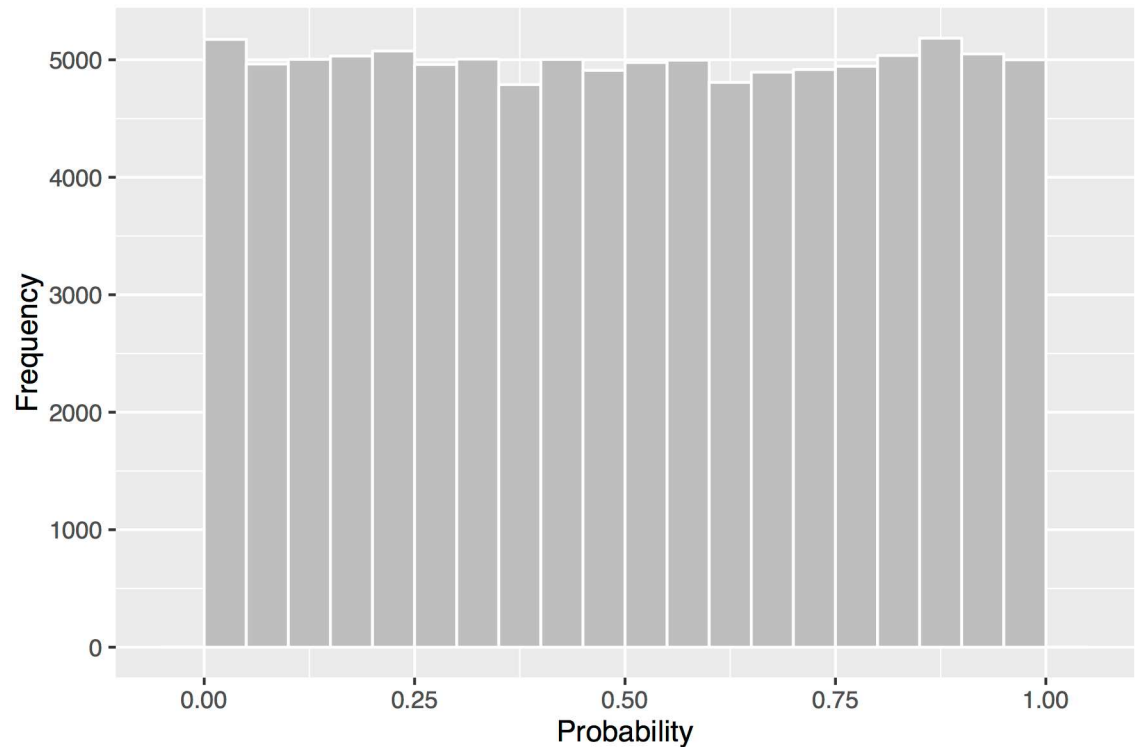


Fig 8. Distribution of the modified probability integral transform values. The values have a distribution that looks uniform.

doi:10.1371/journal.pone.0166895.g008

record. With this subset of data the interaction term between population and year is not included because it is not statistically significant. Occurrence rates are consistently above one tornado every other year across most of the region (Fig 10). Highest rates occur across the mid South and central Great Plains. Relative to the model fit on all the data, the log score is improved at .58 but the correlation between observed counts and predicted rates drops to .20 as expected with many counties having no tornadoes in a given year. The distribution of the modified probability integral transform values is again nearly uniform indicating that the model fits the data well and that there is no problem with model bias. The correlation between raw and predicted rates increases to +.84.

Short-Term View

Diagnostic Mode. Under the assumption of a stationary climate, long-term rates provide a background climatology from which losses can be projected given additional models for intensity and damages. The risk in a given year for a particular region however might be higher (or lower) than the long-term rate depending on climate factors so it is valuable to have a short-term view of the risk. The model described in Eqs (6–9) is fit to the annual tornado counts in raster cells. The predictors are left in native units (standard deviations for the indices and Kelvin for the temperatures) rather than normalized to have common variance. This facilitates physical interpretation of the effects but complicates a comparison of the relative strengths of the effects.

The linear trend term indicates an average annual increase in tornadoes over the domain of 1.2% [(1.1, 1.3)%, 95% credible interval (CI)] consistent with improvements in observing practices. The posterior mean of the spatially-uniform GAK term (β_3) on the short-term view

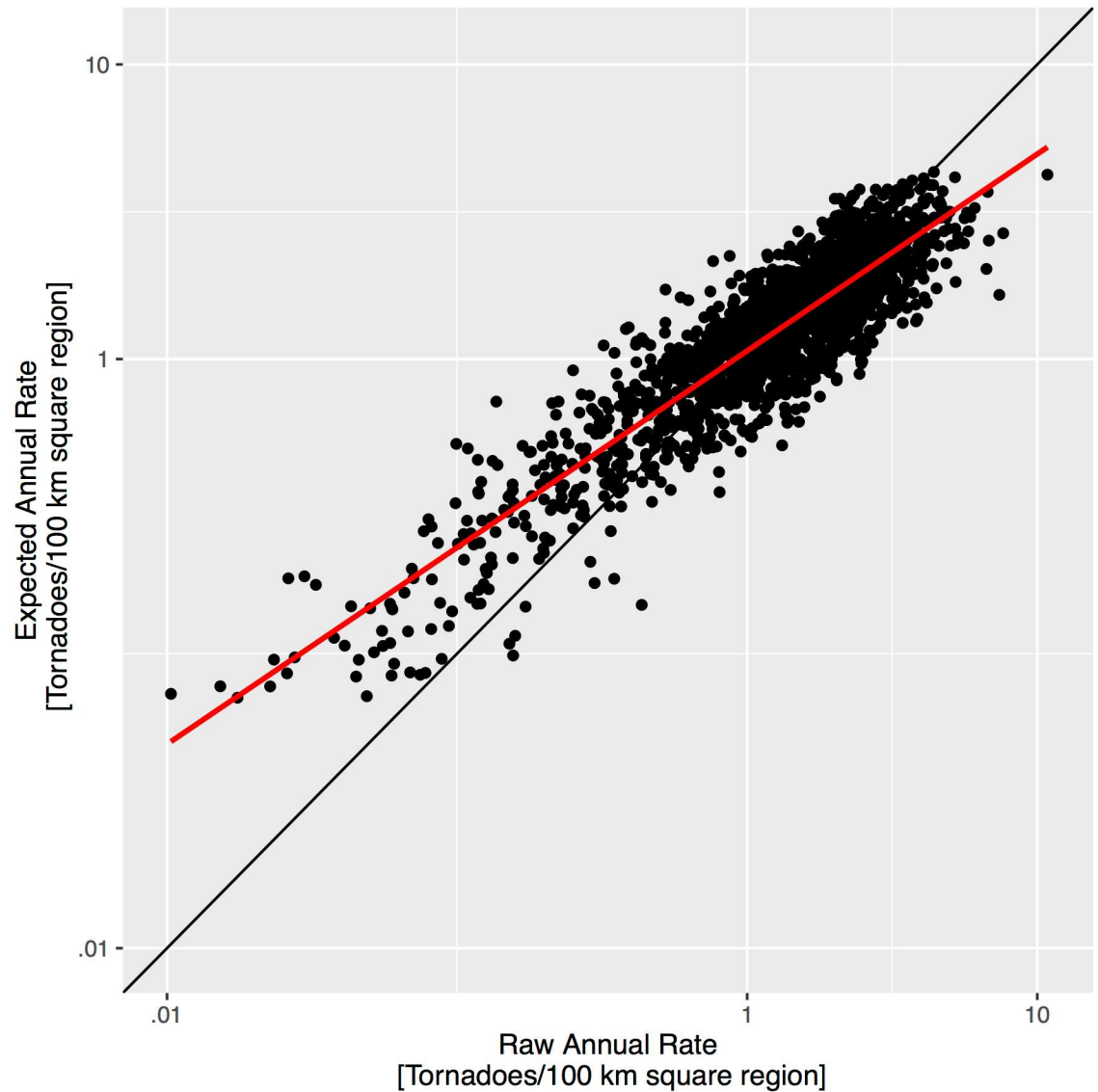


Fig 9. Raw versus expected tornado rates. The best-fit regression line is shown in red.

doi:10.1371/journal.pone.0166895.g009

model indicates a 4.3% [(0.6, 7.9)%, 95% CI] reduction in tornadoes across the domain for every one degree increase in SST in this region consistent with an earlier study over the Central Plains ([1]). An earlier model found that a spatially varying GAK term did not improve the model performance.

The posterior mean of the spatial-varying ENSO term ($\beta_{4,s}$) answers the question: what is the geographic pattern of the ENSO effect on tornadoes? Spatially the ENSO effect implies fewer tornadoes over the Mid South and more across the High Plains during El Niño (Fig 11). The pattern is consistent with earlier results using observed tornado days and environmental conditions [2]. The reduction exceeds 15% over a large part of Tennessee and extends westward to northeastern Texas and southeastern Kansas and northward into eastern Wisconsin and Michigan. The enhancement in tornado activity across the Plains extends from western Texas northward into western South Dakota. The statistical significance of the effect is estimated by dividing the posterior mean by the posterior standard deviation (Fig 12). Areas of

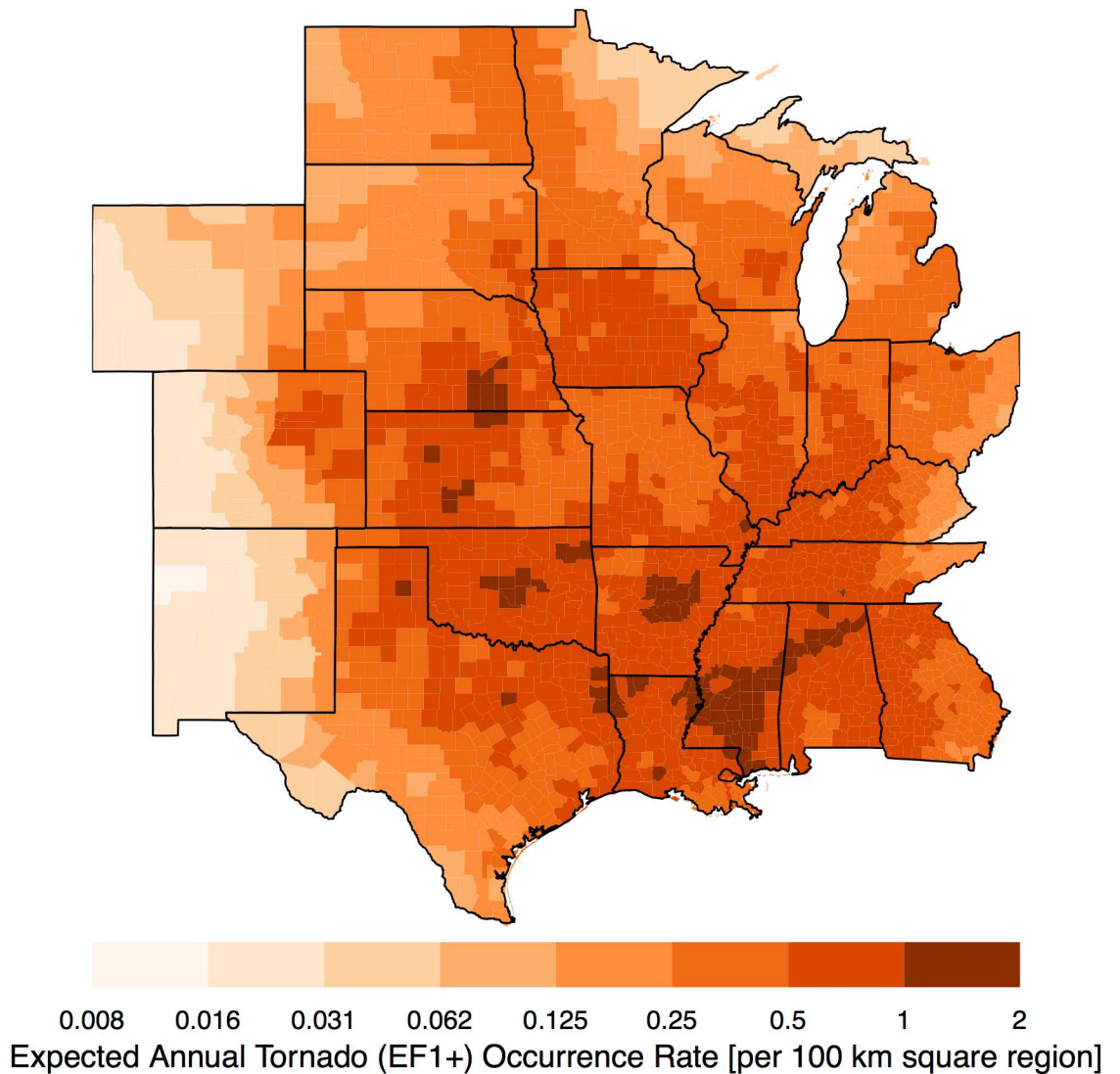


Fig 10. Expected annual tornado (EF1+) occurrence rates per 100 square kilometers.

doi:10.1371/journal.pone.0166895.g010

the Mid South centered on Tennessee show the most statistical significance. Areas across the central Great Plains that tend to get somewhat more tornadoes during El Niño have lower significance levels (generally less than 1.5) consistent with earlier results [1].

The NAO effect is similar to the ENSO effect in magnitude although not as symmetric (many more cells have negative values) with most of the Southeast extending into the Central Plains indicating a decrease in tornado occurrence with a positive NAO index (Fig 13). Only portions of West Texas indicate an increase in tornado frequency with a positive NAO. A positive NAO is marked by lower heights (or pressures) over Iceland and higher heights over the subtropics. Not surprisingly the magnitude of the effect is most pronounced over Georgia and South Carolina closest to the mean position of the subtropical high pressure area. The statistical significance of the NAO effect is confined to the Southeast. The WCA effect indicates a general increase in the rate of tornado occurrence especially over the normally drier regions of the central and northern Great Plains but the magnitude of the effect is considerably smaller than the ENSO and NAO effects.

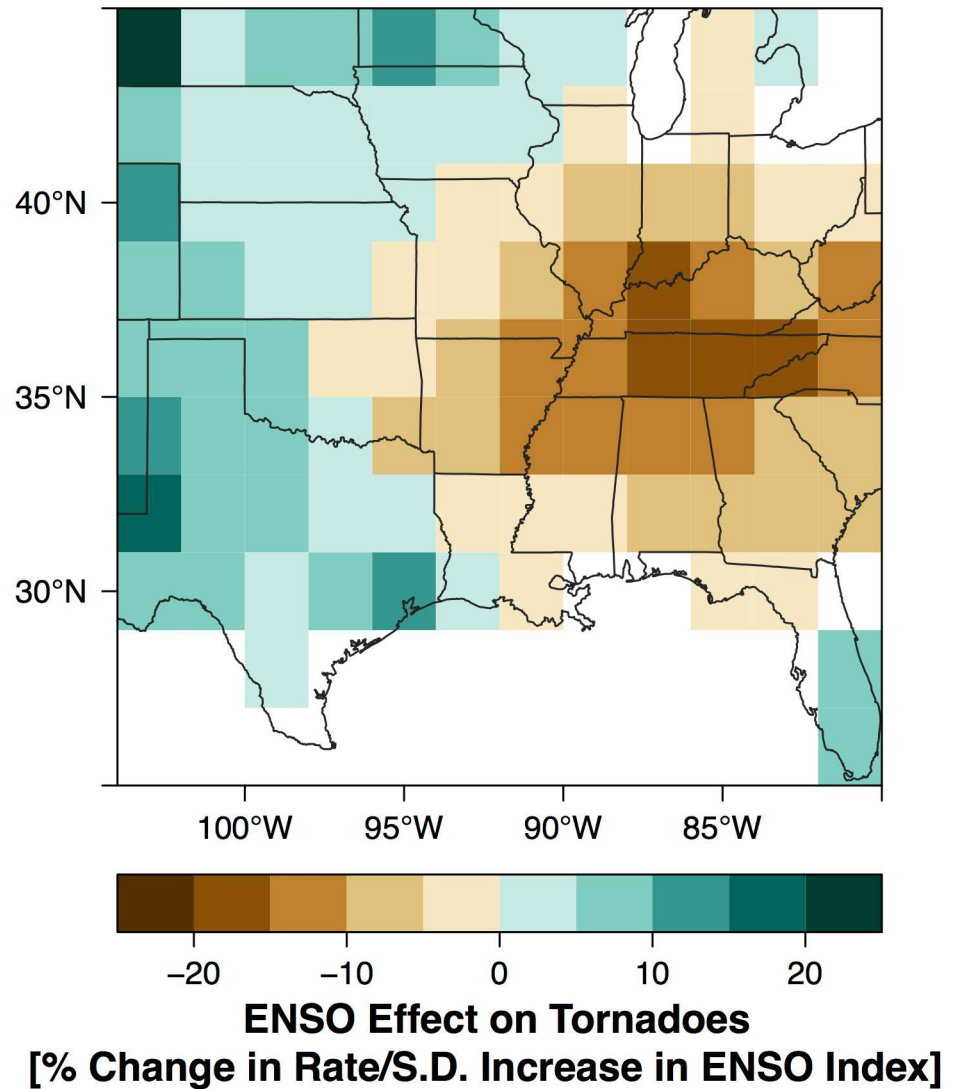


Fig 11. Short-term view. ENSO effect on tornadoes. Magnitude of the effect in units of percentage change in tornado rate per standard deviation (s.d.) increase in the springtime (Mar–May) value of the bi-variate ENSO index. Positive values indicate more tornadoes during El Niño.

doi:10.1371/journal.pone.0166895.g011

Predictive Mode. A short-term view provides a hedge against the long-term rates when climate signals are strong and the effects are aligned. Results from the short-term model run in a diagnostic mode indicate how to adjust rates, on average, during El Niño holding the other climate variables constant. To quantify the *combined* effect of the predictors the short-term view model is run in predictive mode. To predict the rate adjustment for a particular year, the cell counts for that year are left out of the model fit. The prediction is made using spring values of the climate variables but the forecast target is the entire year. Because the approach is Bayesian where all parameters and data are treated as random variables, the fitting procedure estimates the cell rates for the year removed. As an example, early in 2011 a La Niña event was occurring and the springtime NAO index was negative. Together with lower than normal Gulf of Alaska SST the stage was set for an increased threat of tornadoes across the Southeast. Removing the counts for 2011 the model predicts (hindcasts) increased tornado activity across

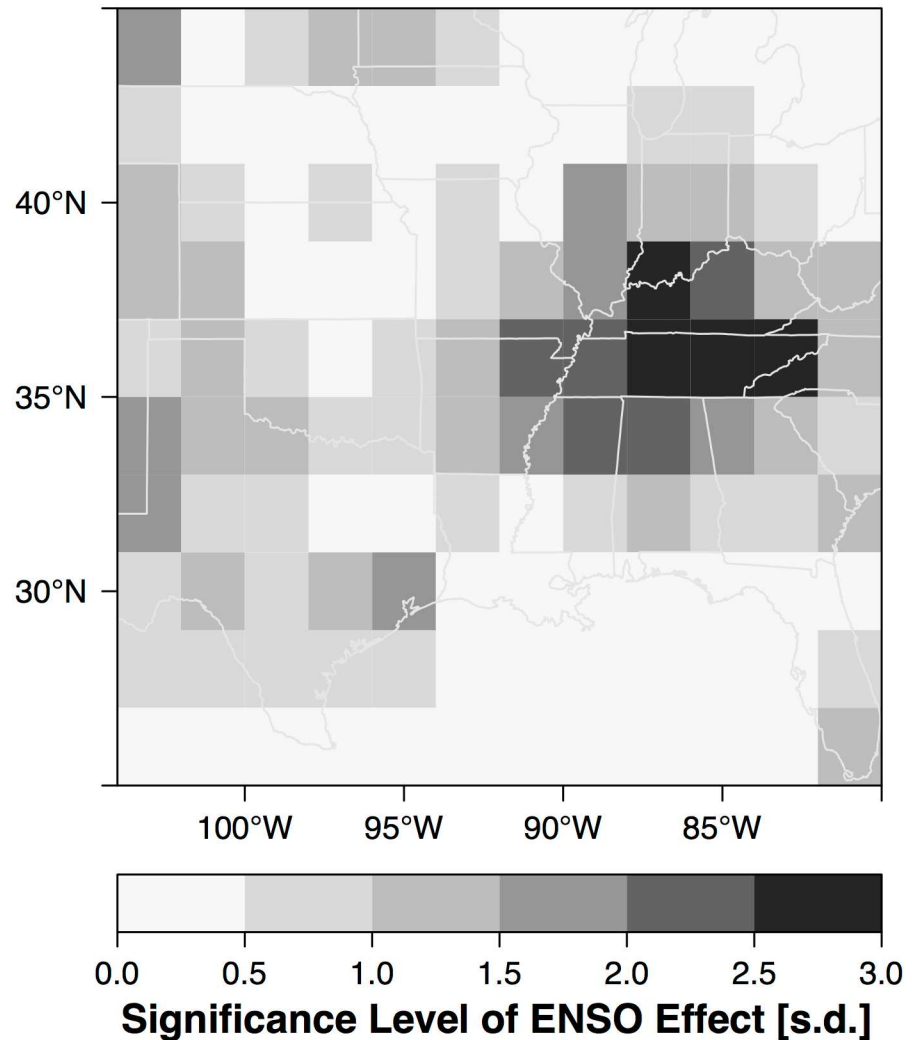


Fig 12. Significance of the ENSO effect on tornadoes in units of standard deviation.

doi:10.1371/journal.pone.0166895.g012

a large part of the region east of 97° W longitude with much of Tennessee expecting rates to be near 150% of the long-term rate (Fig 14). The straight-line tracks of all tornadoes during 2011 are shown as white lines clearly indicating the predicted preference for the activity across the Southeast. A portion of the 2011 tornadoes occurred before the months used to create the predictor indexes so the map does not represent a true forecast.

Discussion and Conclusion

This paper demonstrates that regionally specific tornado risk assessments are possible with spatial statistical models. The models are built from data in the historical tornado record. Modern fitting techniques are able to handle clustered and pathological records. The pathology associated with fewer tornado reports in rural area is handled by using population density at the county level. The models produce long- and short-term views of regional tornado risk. The most important predictor examined is ENSO. This finding is consistent with earlier work on this topic [2, 14, 17]. Accordingly, during spring in a La Niña phase of ENSO a strengthened Inter-American Seas (IAS) low-level jet enhances the spread of moisture across the Southeast

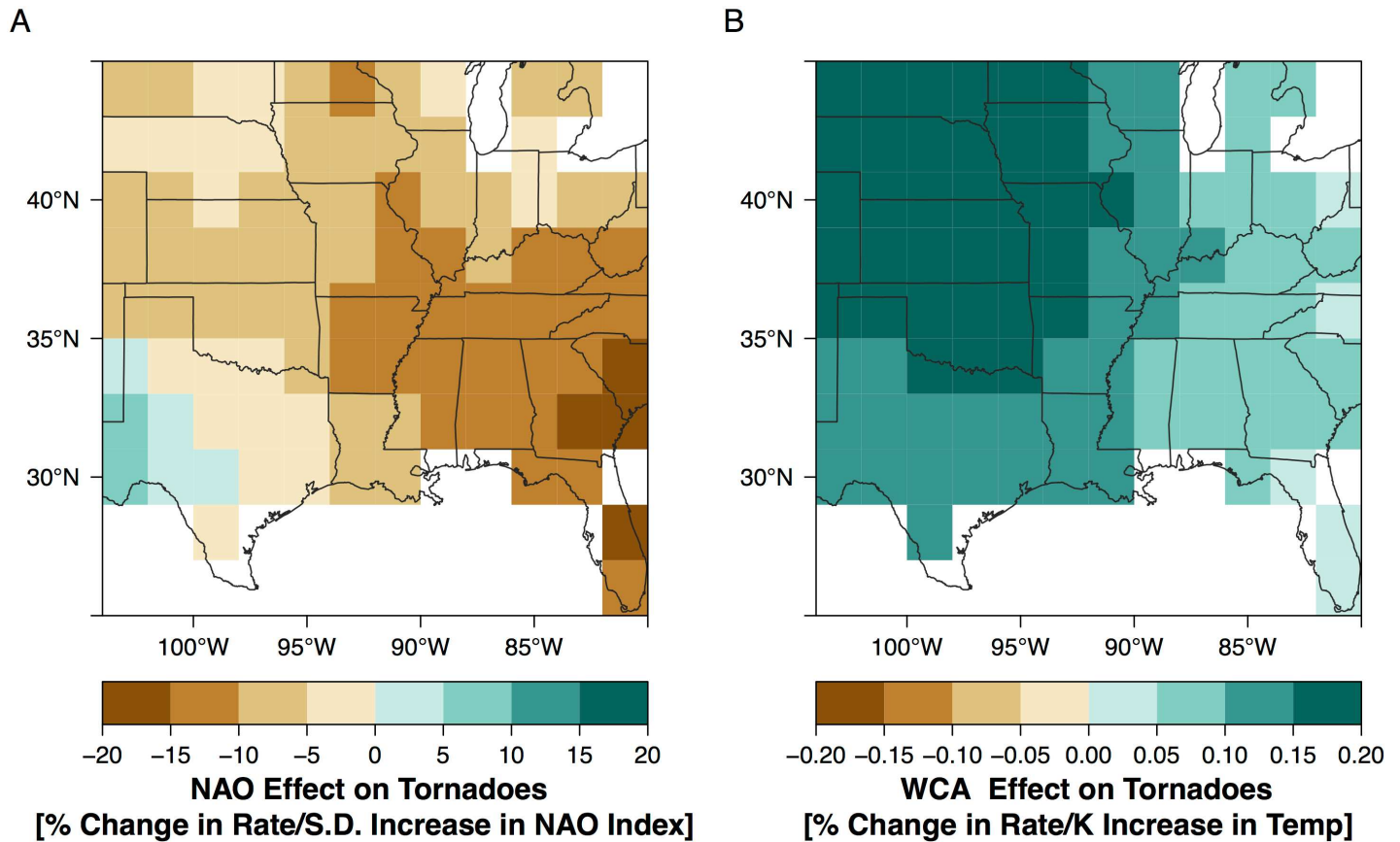


Fig 13. NAO and WCA effect on tornadoes. (A) Magnitude of the NAO effect in units of percentage change in tornado rate per s.d. increase in the springtime (Apr–Jun) value of the NAO index. Negative values indicate fewer tornadoes with a positive NAO. (B) Magnitude of the WCA effect in units of percentage change in tornado rate per °C increase in February SST in the Western Caribbean Sea.

doi:10.1371/journal.pone.0166895.g013

United States. Greater instability associated with more moisture is coupled with increased shear from a strengthened upper-level jet setting the stage for more tornadoes. Whether or not tornado activity actually increases above the long-term rate for a particular La Niña depends on additional unknown and unpredictable factors but given these conditions over many cases the data show statistical evidence for elevated risk.

When the climate exhibits a positive phase of the NAO the chance of tornadoes decreases across a large part of the study area but especially over the Southeast. A positive NAO is associated with a strong subtropical high pressure zone over the North Atlantic that inhibits deep convection across this part of the country, although more research is needed to connect this statistical result with a physical understanding. The Gulf of Alaska SST term indicates fewer tornadoes with a warmer ocean and the Western Caribbean Sea SST term indicates more tornadoes with a warmer ocean consistent with earlier research [1], although the size of this effect is at least an order of magnitude smaller than the size of the ENSO effect. This earlier work also found no statistically significant relationship between the frequency of tornadoes over the central Great Plains and either ENSO or the NAO; a result that is consistent with the weak (statistically insignificant) signals the present model shows for these two factors over this region.

For the short-term view to be used to make an actual prediction, a forecast of the relevant predictors must be made before the year begins. Seasonal forecasts of mean sea-level pressures and near-surface air temperatures with lead times of up to six months are routinely issued by

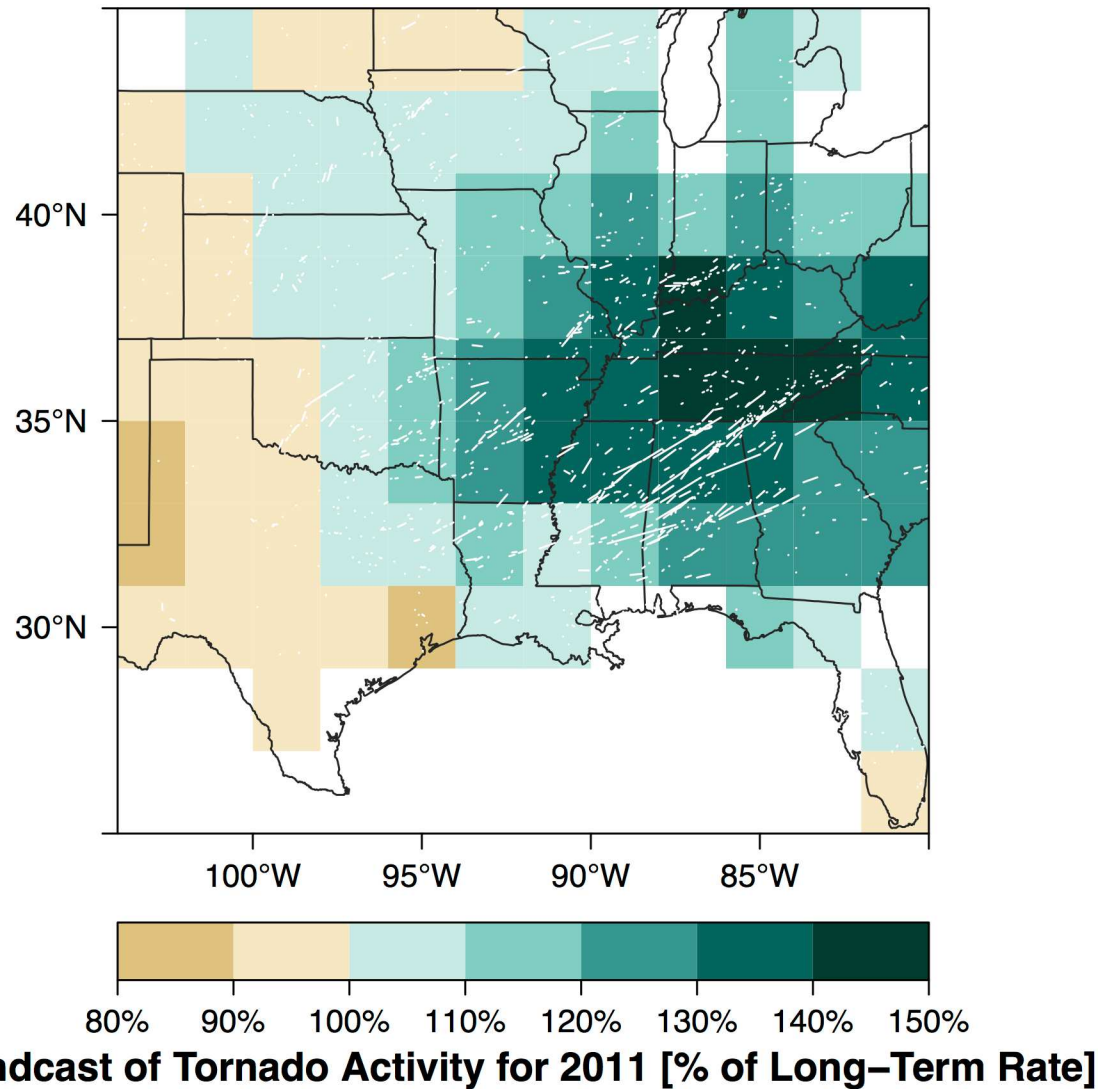


Fig 14. Hindcast of the 2011 tornado year. The straight-line tracks of all tornadoes during that year are shown in white.
 doi:10.1371/journal.pone.0166895.g014

the European Centre for Medium Range Weather Forecasting based on a global coupled ocean-atmosphere general circulation model to calculate the evolution of the ocean and atmosphere. The output from a post-processing of the raw numerical output can be used to create the relevant predictors.

The long-term rates can be used by property insurance companies to set policy rates and by emergency managers to allocate resources. The short-term rate adjustments can be used by reinsurance companies and hedge funds looking for ways to adjust a risk portfolio. Future work will focus on fitting the models to the historical database of Grazulis [28]. The additional data will enhance the precision on the long-term rate estimates and will better define the influence climate predictors have on the short-term rates. For example, population density might not be the best way to handle the under-reporting bias in the tornado records and additional variables like distance-to-nearest city [11] and distance-to-nearest roadway might improve the model. Finally, since the predictors in the short-term view model were based on previous studies, future work might focus on a more comprehensive examination of other climate variables.

The code for the long-term view model is available at rpubs.com/jelsner/tornadoRisk_longTermView and can be modified to estimate risk in other tornado-prone states and regions. The code for the short-term view model is available at rpubs.com/jelsner/tornadoRisk_shortTermView.

Acknowledgments

This work is supported by the Risk Prediction Initiative of the Bermuda Institute for Ocean Studies (BIOS) with special thanks to Mark Guishard and John Wardman. The authors declare no competing interests.

Author Contributions

Conceptualization: JBE THJ.

Data curation: JBE.

Formal analysis: JBE THJ TF.

Funding acquisition: JBE.

Investigation: JBE THJ TF.

Methodology: JBE THJ.

Project administration: JBE.

Resources: JBE.

Software: JBE THJ TF.

Supervision: JBE.

Validation: JBE THJ TF.

Visualization: JBE.

Writing – original draft: JBE.

Writing – review & editing: JBE TF.

References

1. Elsner JB, Widen HM. Predicting spring tornado activity in the central Great Plains by March 1st. *Mon Wea Rev.* 2014; 142:259–267.
2. Allen JT, Tippett MK, Sobel AH. Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nature Geosciences.* 2015; 8:278–283.
3. Carbin GW, Tippett MK, Lillo SP, Brooks HE. Visualizing long-range severe thunderstorm environment guidance from CFSv2. *Bulletin of the American Meteorological Society.* 2016; 97:1021–1031. doi: [10.1175/BAMS-D-14-00136.1](https://doi.org/10.1175/BAMS-D-14-00136.1)
4. King P. On the absence of population bias in the tornado climatology of southwestern Ontario. *Weather and Forecasting.* 1997; 12:939–946. doi: [10.1175/1520-0434\(1997\)012%3C0939:OTAOPB%3E2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012%3C0939:OTAOPB%3E2.0.CO;2)
5. Ray PS, Bieringer P, Niu X, Whissel B. An improved estimate of tornado occurrence in the central Plains of the United States. *Monthly Weather Review.* 2003; 131:1026–1031. doi: [10.1175/1520-0493\(2003\)131%3C1026:AIEOTO%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131%3C1026:AIEOTO%3E2.0.CO;2)
6. Anderson CJ, Wikle CK, Zhou Q. Population influences on tornado reports in the United States. *Weather and Forecasting.* 2007; 22:571–579. doi: [10.1175/WAF997.1](https://doi.org/10.1175/WAF997.1)

7. Brooks HE, Doswell CA, Kay MP. Climatological estimates of local daily tornado probability for the United States. *Wea Forecasting*. 2003; 18:626–640. doi: [10.1175/1520-0434\(2003\)018%3C0626:CEOLDT%3E2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018%3C0626:CEOLDT%3E2.0.CO;2)
8. Dixon PG, Mercer AE, Choi J, Allen JS. Tornado risk analysis: Is Dixie alley an extension of tornado alley? *Bull Amer Meteor Soc*. 2011; 92:433–441. doi: [10.1175/2010BAMS3102.1](https://doi.org/10.1175/2010BAMS3102.1)
9. Shafer CM, Doswell CA. Using kernel density estimation to identify, rank, and classify severe weather outbreak events. *Electronic Journal of Severe Storms Meteorology*. 2011; 6:1–28.
10. Doswell CA, Burgess DW. On some issues of United States tornado climatology. *Mon Wea Rev*. 1988; 116:495–501. doi: [10.1175/1520-0493\(1988\)116%3C0495:OSIOUS%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116%3C0495:OSIOUS%3E2.0.CO;2)
11. Elsner JB, Michaels LE, Scheitlin KN, Elsner IJ. The decreasing population bias in tornado reports. *Weather, Climate, and Society*. 2013; 5:221–232.
12. Doswell CA, Brooks HE, Kay MP. Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Weather and Forecasting*. 2005; 20:577–595. doi: [10.1175/WAF866.1](https://doi.org/10.1175/WAF866.1)
13. Verbot SM, Brooks HE, Leslie LM, Schultz DM. Evolution of the U.S. tornado database: 1954–2003. *Weather and Forecasting*. 2006; 21:86–93.
14. Marzban C, Schaefer JT. The correlation between US tornadoes and Pacific sea surface temperatures. *Monthly Weather Review*. 2001; 129(4):884–895.
15. Cook AR, Schaefer JT. The relation of El Niño–Southern Oscillation (ENSO) to winter tornado outbreaks. *Monthly Weather Review*. 2008; 136:3121–3137.
16. Muñoz E, Enfield D. The boreal spring variability of the Intra-Americas low-level jet and its relation with precipitation and tornadoes in the eastern United States. *Climate Dyn*. 2011; 36(1–2):247–259.
17. Lee SK, Atlas R, Enfield D, Wang C, Liu H. Is there an optimal ENSO pattern that enhances large-scale atmospheric processes conducive to tornado outbreaks in the United States? *Journal of Climate*. 2013; 26:1626–1642.
18. Elsner JB, Fricker T, Jagger TH, Mesev V. Statistical models for predicting tornado rates: Case studies from Oklahoma and the Mid South USA. *International Journal of Safety and Security Engineering*. 2016; 6:1–9. doi: [10.2495/SAFE-V6-N1-1-9](https://doi.org/10.2495/SAFE-V6-N1-1-9)
19. Elsner JB, Murnane RJ, Jagger TH, Widen HM. A spatial point process model for violent tornado occurrence in the U.S. Great Plains. *Mathematical Geosciences*. 2013; 45:667–679.
20. Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late Nineteenth Century. *Journal of Geophysical Research*. 2003 July; 108(D14):4407–4444. doi: [10.1029/2002JD002670](https://doi.org/10.1029/2002JD002670)
21. Barnston AG, Livezey RE. Classification, seasonality, and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*. 1987; 115:1083–1126. doi: [10.1175/1520-0493\(1987\)115%3C1083:CSAPOL%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115%3C1083:CSAPOL%3E2.0.CO;2)
22. Kalnay E, et al. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*. 1996; 77:437–470. doi: [10.1175/1520-0477\(1996\)077%3C0437:TNYRP%3E2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077%3C0437:TNYRP%3E2.0.CO;2)
23. Jagger TH, Elsner JB, Widen HM. A statistical model for regional tornado climate studies. *PLoS ONE*. 2015; 10(8):e0131876. doi: [10.1371/journal.pone.0131876](https://doi.org/10.1371/journal.pone.0131876) PMID: 26244881
24. Besag J. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1975; 24:179–195. doi: [10.2307/2987782](https://doi.org/10.2307/2987782)
25. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009; 71:319–392. doi: [10.1111/j.1467-9868.2008.00700.x](https://doi.org/10.1111/j.1467-9868.2008.00700.x)
26. Rue H, Martino S, Lindgren F, Simpson D, Riebler A, Krainski ET. INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation; 2014 +0100). R package version 0.0-1417182342.
27. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association, Series B*. 2007; 102:359–378. doi: [10.1198/016214506000001437](https://doi.org/10.1198/016214506000001437)
28. Grazulis TP. Significant Tornadoes, 1880–1989: Discussion and analysis. *Significant Tornadoes, 1880–1989*. Environmental Films; 1990. Available from: <http://books.google.com/books?id=E8hFAAAAYAAJ>.