3.9 COVARIABILITY OF ANNUAL UNITED STATES TORNADO REPORT COUNTS AND CLIMATE INDICES

Christopher J. Anderson[1]*, Christopher K. Wikle[2], and Raymond W. Arritt[1]

[1]Department of Agronomy, Iowa State University, Ames, IA
[2]Department of Statistics, University of Missouri, Columbia, MO

1. INTRODUCTION

It is presumed that tornado report counts for monthly or longer periods may be directly related to climate indices, since tornadoes tend to occur during certain weather patterns (Fawbush and Miller 1954, Brooks and Craven 2002), and the frequency of weather patterns may be directly related to climate indices (Bjerknes 1969). In support of this supposition, Marzban and Schaefer (2001) have found statistically significant sample correlation between Pacific sea surface temperature (SST) and tornado counts in the southeast U. S. The research presented herein differs from previous work in that we simultaneously consider covariability of monthly tornado counts with multiple climate indices.

We have developed a hierarchical stochastic model to examine dependence between tornado report counts and climate indices. There are a number of reasons that standard modeling approaches, such as developing independent least squares regression equations with Normal errors for a field of observations, may be inadequate for examining dependence between tornado report counts and climate indices. From a meteorological perspective, frequency of weather patterns may differ in separate years and at separate locations even though climate indices are nearly identical, due to internal variability of the atmosphere. This means that dependence between weather patterns and climate indices may be non-stationary in space and over time. Additionally, tornado reports are likely to be correlated in space when summed over monthly or seasonal periods. From a statistical perspective, tornado reports are rare, discrete and non-negative. Thus, it is expected that these data do not follow a Normal distribution. Finally, societal changes introduce non-stationary reporting biases in both time and space (Doswell and Burgess 1988, Brooks and Craven 2002).

Typically, tornado report counts have been preprocessed to remove non-stationary behavior before applying a stochastic model to infer the significance of certain sample statistics. We have taken an alternative approach in which a stochastic model has been designed that explicitly models non-stationary behavior by constructing a hierarchy of conditional probability models that are linked by applying Bayes theorem, a fundamental rule of probability calculus.

* *Corresponding author address:*
Christopher J. Anderson, Iowa State University, 3010 Agronomy Hall, Ames, IA 50011-1010
e-mail: candersn@iastate.edu

2. DATA

Tornado reports from 1953-1995 were obtained from the Storm Prediction Center archive of severe weather reports (www.spc.noaa.gov/climo). A grid of 50-km boxes was overlaid on the U. S., and the number of tornadoes was tallied monthly in each box. Thus, time series spanning 1953-1995 of monthly tornado report counts was generated for each box.

Any number of climate indices may be considered as predictors in a stochastic model. At this preliminary stage, we have included the Nino3.4 SST index (since it is known that tornado report counts are significantly correlated with equatorial Pacific SST), the North Atlantic Oscillation (NAO), and the North Pacific Index (NPI).

3. STOCHASTIC MODEL

Hierarchical stochastic models attempt to decompose observed data into a series of conditional probability models. In this way, one can build separate models for the observations (*data model)*, the stochastic process describing the statistical behavior of the observations (*process model*), and the parameter uncertainty (*parameter model)*. The general hierarchical model has three components:

Data Model: Pr[data | process, parameters]
Process Model: Pr[process | parameters]
Parameter Model: Pr[parameters]

where Pr[ ] denotes that a probability distribution has been assigned, and the vertical line indicates the probability distribution is conditional.

- The *data model* assigns a theoretical conditional probability distribution to the tornado report counts. This provides the necessary flexibility to use probability distributions other than the Normal distribution. The parameters of the data model depend on underlying process and parameter models. Thus, the characteristics of the data model reflect uncertainty not only of the observations but also

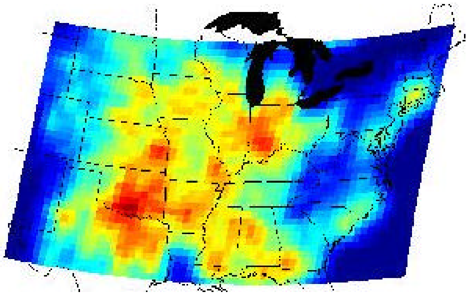

Figure 1. Posterior mean of β_1 (intercept) parameters. Blue (red) represents large negative (near zero) parameters.

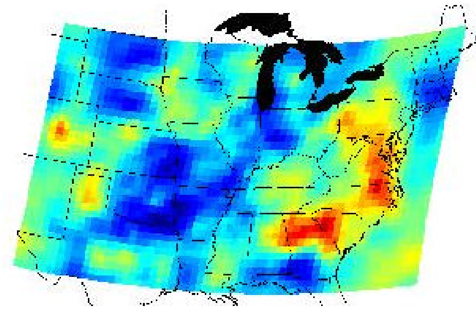

Figure 2. Posterior mean of β_2 (time) parameters. Color scheme is identical to figure 1.

of process and parameter model assumptions. It is advantageous to use scientific reasoning and knowledge in accordance with data analysis when selecting a distribution for the data model. With this approach, all available knowledge is formally incorporated into the analysis.

- The *process model* specifies a stochastic process that relates tornado occurrence to climate indices, with estimated parameter values describing the degree of association. It is possible that a number of stochastic processes might adequately reproduce the statistical behavior of tornado report counts. In the hierarchical framework, it is possible to systematically compare alternative process models.
- The *parameter model* assigns a theoretical probability distribution to the parameters of the process model. Generally, point estimates of parameters are used when predicting observations. For example, coefficients in linear regression models are considered constant when generating estimates of a predictand. In the hierarchical framework, the coefficients are considered to be random variables.

In this preliminary study, we let $Y(s_i;t)$ be the number of tornado reports in some geographical region indexed by s_i=1,...,n at times t=1,...,T (n is the number of boxes and T is the number of months). Thus, $Y(1,t)$ corresponds to the monthly time series of tornado report counts for box 1, as described in Section 2.

The *data model* is given by:

$Y(s_i,t)|\lambda(s_i,t) \sim \text{Poisson}(\lambda(s_i,t))$ for all s_i,t

That is, conditioned upon the Poisson mean (λ), tornado report counts are independent and follow a Poisson distribution. This does not suggest the counts are marginally independent. Instead, marginal spatio-temporal correlation is generated by an underlying process model rather than incorporated directly in the data model.

The *process model* is given by:

$\log(\lambda(s_i,t))|b_i, s^2 \sim N(\mathbf{x}_t b_i, s^2_h)$

where b_i is a 7x1 (i=1..7) vector of regression coefficients, $\mathbf{x}_t$ is a 7x1 vector of covariates (intercept, time trend, NAO, NPI, Nino3.4 SST, NAO cross Nino3.4, NPI cross Nino3.4) that vary over time, and s^2_h is site-specific variance that represents random error. That is, the log of the Poisson mean is modeled by a time-dependent linear regression with normally distributed, uncorrelated errors indexed in space. Geographical sampling biases, such as those related to demographic characteristics, are partially accounted for by s^2_h, while λ ($\mathbf{x}_t b_i$) includes linear dependence on time to partially account for temporal sampling biases.

The *parameter model* is given by:

$b_i \sim N(0,\Sigma_i)$

where Σ_i is a spatial covariance matrix. Distributions are assigned to Σ_i and s^2_h as well.

We then evaluate the joint distribution of all parameters given the observations using Bayes' Theorem. Markov Chain Monte Carlo (MCMC) methods are used to generate realizations of this joint distribution. See Wikle et al. (1998) for examples of MCMC applied to problems in atmospheric science.

4. RESULTS AND DISCUSSION

The spatial field for the posterior mean of β_1 can be interpreted as a representation of relative frequency of tornado reports as reproduced by the model, absent the effects of climate and population fluctuations (Figure 1). (Recall these parameters are projecting onto the logarithm of λ, so that parameters near zero are representative greater tornado report frequency.) Relatively large β_1 covers northern Texas, southern Arkansas, eastern Kansas, and much of Oklahoma. This pattern strongly resembles raw tornado counts over the U. S., although it is somewhat smoother, suggesting the model capably captures regional differences of tornado report frequency.

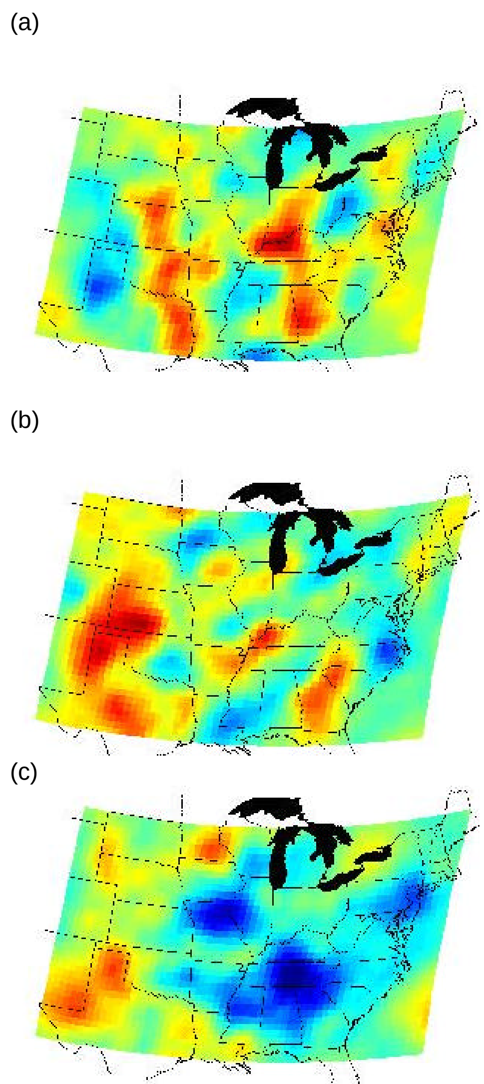

Figure 3. Posterior mean of (a) β_3 (NAO), (b) β_4 (NPI), and (c) β_5 (SST) parameters. Color scheme is identical to figure 1.

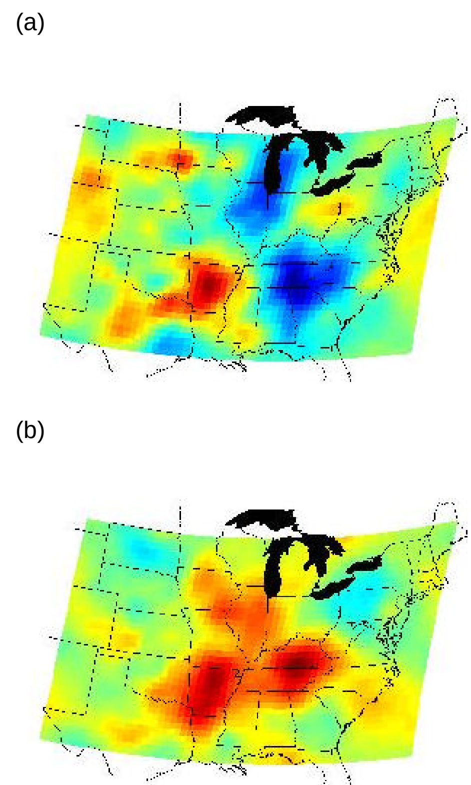

Figure 4. Posterior mean of (a) β_6 (NAO cross with Nino3.4) and (b) β_7 (NPI crossed with Nino3.4) parameters. Color scheme is identical to figure 1.

Linear trend of tornado report counts is analyzed independently with β_2 (Figure 2), so that changes in time of reporting biases are separated from dependence of tornado report counts and climate indices. The spatial field of β_2 shows large positive posterior mean values along the East Coast and the southeast United States in regions of relatively high population density; whereas, large negative values are evident over much of the central United States, where β_1 is large.

Previous research has found a slight reduction of annual U. S. F2-F5 tornado report counts from the 1950s to the 1990s (Brooks 2000). It is suspected that this decrease is largely attributable to changes of procedure for assigning Fujita scale ratings of tornado damage (Brooks and Craven 2002). Tornado damage ratings prior to 1973 were assigned from written eyewitness accounts reported by the media. After 1973, damage surveys have been conducted for tornadoes that have caused extensive damage. Results of Brooks and Craven (2002) suggest that tornado ratings prior to 1973 are systematically larger by about one half of a Fujita category. That is, a tornado rated F2 prior to 1973 may be given a rating of only F1 by today's rating standards. This is consistent with negative β_2 in less densely populated areas in the central United States in that prior to 1973 ratings for tornado damage in these regions would've been heavily dependent upon reports from news media. However, positive β_2 in densely populated areas contradicts this theory, suggesting more sophisticated theories are needed for explaining impacts of population shifts and severe weather awareness on reporting bias.

Each climate index projects significantly onto the logarithm of λ. Extensive areas of large magnitude for posterior means of β_4 and β_5 are evident (Figure 3). The spatial field of β_5 exhibits large negative values over much of the southeast United States. This pattern is consistent with results from Marzban and Schaefer (2001) and Monfredo (1999) both of which diagnosed significant negative correlation between Pacific SST and tornado reports in the southeast United States. In the spatial field of β_5, large positive values cover the western Plains of the central United States. The spatial coherence of these regions of negative and positive mean parameter values suggests that these signals are robust in the sense that similar spatial patterns of tornado counts may be produced in multiple years. In contrast, small-scale variability of the parameters might indicate a tornado outbreak in a particular year has forced the model to project strongly onto the parameters.

It is apparent from visual inspection of spatial fields for posterior mean values of β_4 and β_5 that these parameters contain much more small-scale spatial variability than β_4. Since NPI and NAO have lower frequency fluctuations compared to Nion3.4, it is possible that weather conditions associated with certain phases of NAO or NPI may have greater variability than those associated with fluctuations of Nino3.4, causing variability of the location of tornado outbreaks. This hypothesis is supported by spatial coherence of fields of posterior means for parameters of interaction terms in which the slower varying NAO and NPI indices are individually combined with the Nino3.4 index (Figure 4). Interestingly, parameters of opposite sign are produced in the southeast United States with negative values associated with interaction between NAO and Nino3.4.

5. FUTURE WORK

Future model development will concentrate on incorporating explicit models for sampling biases. In particular, the hierarchical model will be extended to include a sub-component that accounts for the probability that a tornado occurred but was unreported. Incorporating this conditional probability will further elucidate the extent to which fluctuations of tornado reports are explained by shifts of population rather than climate.

Meteorological analysis will concentrate on identifying favorable weather parameters that are associated with these climate indices. To this end, synthetic soundings will be created from reanalysis data and will be examined for common characteristics in regions of coherent positive and negative parameters following the results of Rasussen and Blanchard (1998) and Brooks and Craven (2002). Immediate interest will be given to negative dependence of Nino3.4 and tornado report counts. It is widely reported that storminess is increased over the southeast United States during the spring months of a decaying El Nino. That is, precipitation and frequency of thunderstorms is known to be greater while Nino3.4 is still positive in the springtime following a strong wintertime El Nino. This would seem to be an environment that would also support more frequent tornadoes. The analysis reported herein and elsewhere by other authors indicates this is not the case, although testable meteorological hypotheses have yet to be formed. In addition to meteorological analysis, spatial variability of parameters will be analyzed in order to quantify scales of spatial variability and to determine their causes.

6. ACKNOWLEDGEMENTS

This research was supported by NSF grant ATM-9911417 and U. S. EPA Science to Achieve Results (STAR) Assistance Agreement R827257-01-0.

7. REFERENCES

Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163-172.

Brooks, H. E., 2000: The climatology of severe thunderstorms: what we can know. *Preprints*, 20th Conference on Severe Local Storms, Orlando, Florida, American Meteorological Society, 126-129.

Brooks, H. E., and J. P. Craven, 2002: Database of proximity soundings for significant severe thunderstorms, 1957-1993. *Preprints*, 21st Conference on Severe Local Storms, San Antonio, Texas, American Meteorological Society, 639-642.

Doswell, C. A., III, and D. W. Burgess, 1988: On some issues of United States tornado climatology. *Mon. Wea. Rev*, **116**, 495-501.

Fawbush, E. J., and R. C. Miller, 1954: The types of air masses in which North American tornadoes form. Bull. *Amer. Meteor. Soc*., **35**, 154-165.

Marzban, C., J. T. Schaefer, 2001: The correlation between U.S. tornadoes and Pacific sea surface temperatures. *Mon. Wea. Soc.*, **129**, 884-895.

Monfredo, W., 1999: Relationships between phases of the El Nino-Southern oscillation and caharacter of the tornado season in the south-central United States. *Phys. Geog.*, **20**, 413-421.

Rasmussen, E. N., D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148-1164.

Wikle, C. K., Berliner, L. M., and N. Cressie, 1998: Hierarchical Bayesian space-time models. *Journal of Environmental and Ecological Statistics*, **5**, 117-154.