11B.6

ANALYSIS OF TORNADO COUNTS WITH HIERARCHICAL BAYESIAN SPATIO-TEMPORAL MODELS

Christopher J. Anderson*
Department of Agronomy, Iowa State University, Ames, IA
and
Christopher K. Wikle
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), 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.

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 is 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).

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 an hierarchy of conditional probability models that are linked by applying Bayes theorem, a fundamental rule of probability calculus.

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), which provide measures of northern Atlantic and northern Pacific SST.

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 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

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

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 the 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:

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

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:

$$\log(\lambda(s_i,t))|\beta_i,\sigma^2 \sim N(\mathbf{x}_t\beta_i,\sigma^2_\eta)$$

where β_i is a 4x1 (i=1..4) vector of regression coefficients, $\mathbf{x_t}$ is a 4x1 vector of covariates (Nino3.4 SST, NAO, NPI, and time) that vary over time, and σ^2_η is a spatial covariance matrix that represents random error. That is, the log of the Poisson mean is modeled by a time-dependent linear regression with normally distributed errors that contain spatial dependence. Geographical sampling biases, such as those related to demographic characteristics, are partially accounted for σ^2_η, while λ ($\mathbf{x}_t\beta_i$) includes linear dependence on time to partially account for temporal sampling biases.

The *parameter model* is given by:

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

where Σ_i is a spatial covariance matrix. Distributions are assigned to Σ_i and σ^2_η 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

Results from this preliminary model indicate that NPI and NAO more strongly influence λ than Nino3.4 SST. The spatial pattern of coefficients for Nino3.4 SST is consistent with Marzban and Schaefer (2001) in that positive dependence occurs in the southeast U.S. However, the magnitudes of Nino3.4 SST coefficients are much less than those of NPI and NAO, which exhibit negative dependence in the eastern and central U.S., respectively.

Previous research has found an order of magnitude increase from the 1950s to the 1990s of annual U.S. tornado report counts (Brooks 2000). It is suspected that this increase is largely attributable to changes in public awareness and in warning verification by the NWS (Brooks 2000). Our results indicate that trend over time has large spatial variability, associated with population density. Positive coefficients (increasing trend) are evident in a corridor from the Ohio Valley into New York and in large metropolitan areas (Denver, Dallas-Fort Worth metro area, Oklahoma City, Minneapolis, New York, Philadelphia). In contrast, rural areas in the central U.S. show decreasing trend.

Future model development will concentrate on explicitly modeling sampling biases, introducing interaction terms for the climate indices, allowing Σ_i to vary over time, and designing separate models for F0-F5 and F2-F5 tornadoes.

5. ACKNOWLEDGEMENTS

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

6. 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.

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.

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