A Markov Chain Model of Tornadic Activity

MATHIAS DRTON

Department of Statistics, University of Washington, Seattle, Washington

CAREN MARZBAN

Department of Statistics, University of Washington, Seattle, Washington, and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

PETER GUTTORP

Department of Statistics, University of Washington, Seattle, Washington

JOSEPH T. SCHAEFER

Storm Prediction Center, Norman, Oklahoma

(Manuscript received 17 October 2002, in final form 13 May 2003)

ABSTRACT

Tornadic activity in four U.S. regions is stochastically modeled based on data on tornado counts over the years 1953–98. It is shown that tornadic activity on a given day is mostly affected by the activity on the previous day. Hence, the process can be modeled as a Markov chain. A parametric nonhomogenous Markov chain model is developed based on the well-known increase of tornadic activity in the spring and summer months. This model, with only eight parameters, describes tornadic activity quite well. The interpretability of the estimated parameters allows a diagnosis of the regional differences in tornadic activity. For instance, a comparison of the values of the parameters for the four regions suggests that in the South tornado persistence is specific mostly to the early part of the year. Finally, within the framework of probabilistic forecast verification, it is shown that the Markov chain model outperforms the climatological model, even though the former is far simpler in terms of the number of parameters (8 and 366, respectively). The superior performance of the model is confirmed in terms of several measures of performance in all four regions. The exception is the southern Tornado Alley, where the reliability of the model forecasts is nonsignificantly inferior to that of the climatological ones.

1. Introduction

Tornadoes are among the most damaging of severe atmospheric phenomena. Nearly 40 000 tornadoes were reported in the United States in the period 1953–98, causing billions of dollars in damages and thousands of deaths. It is therefore important to understand better the underlying process and to attempt prediction. The climatology of tornadoes has been well examined (Brooks and Doswell 2001; Brooks et al. 2003; Grazulis et al. 1993) and their relationship to other climatic forces has been investigated (Marzban and Schaefer 2001; Anderson and Wikle 2002). Furthermore, models have been developed to attempt the prediction of tornadic activity (Hamill and Church 2000; Colquhoun and Riley 1996; Reap and Foster 1979) from information on thermodynamic and other variables. However, it is possible to develop a model in which future tornadic activity is predicted only from past values of the same. To that end, in this paper a stochastic model for tornadic activity—namely, a Markov chain model—is developed, not only for enhancing the understanding of the underlying process but also for prediction purposes.

A Markov chain is a stochastic process, that is, a collection of random variables, indexed by discrete time (in the current case, by days) such that the present observation depends on the past observations only through the most recent observation, that is, today's observation depends only on yesterday's observation (section 2). Markov chains have been steadily gaining popularity in meteorological circles. An ''exact phrase'' search for ''Markov chain'' in the online journals of the American Meteorological Society suggests an increasing trend over the past 30 years. The interest, however, appears to be focused on precipitation, since of the 165 matches, 112 deal with precipitation, while the remaining few

Corresponding author address: Dr. Caren Marzban, 2819 W. Blaine Street, Seattle, WA 98199.
E-mail: marzban@stat.washington.edu

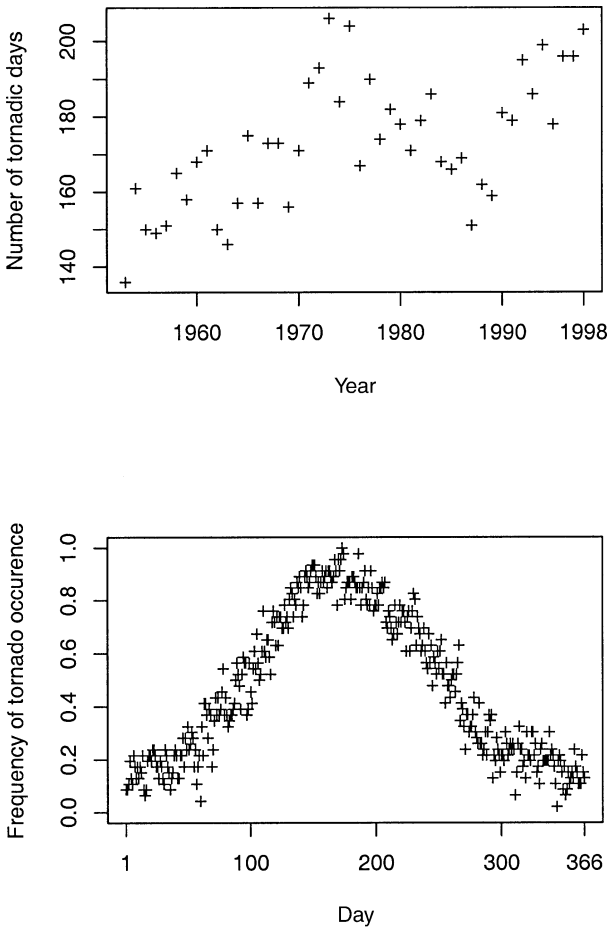


FIG. 1. (top) Number of days with tornado occurrence in the entire United States in a given year and (bottom) frequency of tornadic activity in the entire United States on a given day.

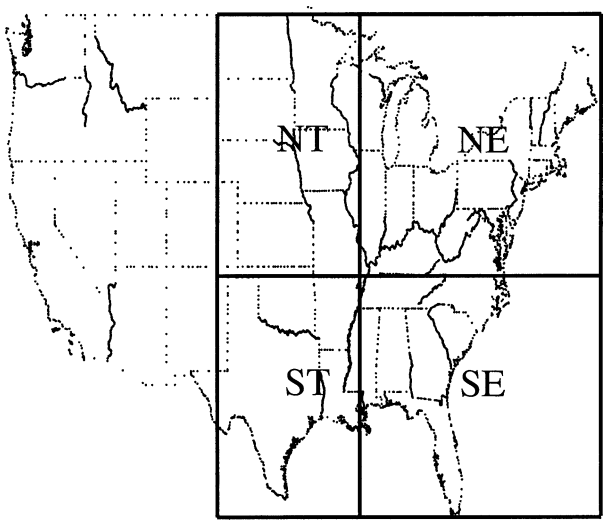


FIG. 2. The four considered U.S. regions: northern and southern Tornado Alley (NT and ST) and Northeast and Southeast (NE and SE).

cover a somewhat wide range of applications. A few examples of the former are Gates and Tong (1976), Hughes and Guttorp (1994), Katz (1974), Stern (1982), and Valdez and Young (1985), and an example of the latter is Lakshmanan (2001), in the context of data compression. The number of matches drops to zero if the word ''tornado'' is included in the search.

This work is based on daily tornado counts in the entire United States as well as in four subregions, spanning the period 1953–98. No attempt is made to model the exact counts of tornadoes; instead, only the occurrence or nonoccurrence of a tornado on a given calendar day within an area is modeled. The features of the ''tornado day'' data are illustrated in Fig. 1.

The top panel in Fig. 1 suggests an increasing trend.[1] In the current analysis, this possible trend is neglected. Instead, emphasis is placed on the within-year change in tornadic activity, assuming that the different years constitute identical and independent replications. It is this within-year change that is modeled with Markov chains. The assumption of identical and independent replications will be revisited and justified in the discussion section.

The occurrence of tornadoes follows a distinct spatial and temporal pattern (see http://www.nssl.noaa.gov/hazard for interesting illustrations). Typically, there are three regions of interest: the Northeast (NE), Southeast (SE), and central plains (also called Tornado Alley). It is well known (e.g., Kelly et al. 1978) that the NE and SE behave differently in regard to tornadic activity. As such, it is natural to search for a north–south difference in the central region as well. Therefore, in this article, Tornado Alley is subdivided into a northern part (NT) and a southern part (ST), to allow for differences between northern and southern U.S. regions. Figure 2 exhibits the locations of the four regions.

There is a strong seasonal factor; the midportion of the year is typically considered tornadic season for most of the country (see the bottom panel in Fig. 1).[2] Because of this seasonal factor, time-homogeneous Markov chains are not appropriate. For a first understanding of the memory of the process, that is, how many past days are relevant for future tornadic activity, a nonparametric nonhomogeneous Markov chain is considered (section 3). However, the large number of parameters in such a model renders it prone to overfitting, which results in transition probabilities that vary strongly from one day to the other. For that reason, a parametric nonhomogeneous Markov chain model is developed that, in particular, guarantees time-continuous transition probabilities (section 4). A

[1] The trend is most likely an artifact of the underreporting of tornadoes during the early portion of the period. Ray et al. (2003) argue that a significant underreporting problem still exists.

[2] There also seems to be a much smaller secondary peak in the fall, which is discussed in section 8.

comparison of the estimated parameters of the models provides insight into the differences in tornadic activity in the different regions (section 5 and 6).

In addition to offering insight into the underlying processes, the parametric model can be utilized for forecasting purposes. Since the model produces probabilistic forecasts, its performance is assessed within a probabilistic framework (Murphy and Winkler 1987, 1992). The performance of the models is compared to that of a forecast model based on climatological probabilities (section 7).[3] The conclusions are summarized in section 8.

2. Markov chains

Consider a collection of 366 binary random variables X_j, indexed by the days $j = 1, 2, \ldots, 366$. Interpret the event $X_j = 1$ as tornado occurrence and $X_j = 0$ as nonoccurrence. Let $x_j \in \{0, 1\}$, $j = 1, 2, \ldots, 366$, be the observed occurrences during some year. Then let $P_1(j) \equiv P(X_j = 1)$ be the climatological probability of a tornado on day j, $P_{11}(j) \equiv P(X_{j+1} = 1 | X_j = 1)$ the conditional probability of a tornado on day $j + 1$, given that there is a tornado on day j, and $P_{111}(j) \equiv P(X_{j+2} = 1 | X_{j+1} = 1, X_j = 1)$ the conditional probability of a tornado on day $j + 2$, given that there are tornadoes on days $j + 1$ and j. The obvious analogue denotes the case of nonoccurrences or combinations of occurrences and nonoccurrences. Note that $P_0(j) + P_1(j) = 1$, $P_{00}(j) + P_{01}(j) = 1$, etc.

The process $(X_j)_j$ is called a Markov chain if for all days j the probability of a tornado on day $j + 1$, given all past days, equals the probability of a tornado on day $j + 1$, given the state on day j only, that is,

$$P(X_{j+1} = 1 | X_i = x_i, i \le j) = P(X_{j+1} = 1 | X_j = x_j) = P_{x_j 1}(j).$$

If the so-called transition probabilities $P_{01}(j)$ and $P_{11}(j)$ do not depend on j, the Markov chain is said to be (time) homogeneous, otherwise it is nonhomogeneous. A straightforward generalization of Markov chains are mth order Markov chains, $m = 1, 2, \ldots$, defined through

$$P(X_{j+1} = 1 | X_i = x_i, i \le j) = P(X_{j+1} = 1 | X_i = x_i, j - m < i \le j).$$

Further details can be found in Guttorp (1995, chapter 2), Kao (1997), and Wilks (1995).

The data at hand covers 46 yr (1953–98) and can be represented as a matrix with 366 rows and 46 columns. Each element of the matrix, x_j^I, where $j = 1, \ldots, 366$, and $I = 1, \ldots, 46$, is a 1 or a 0, corresponding to the occurrence or nonoccurrence of a tornado.[4] As mentioned above, it is assumed that the different years are independent and identical copies; that is, the 46 random vectors $(x_1^I, \ldots, x_{366}^I)$, for $I = 1, \ldots, 46$, have the same distribution and are independent. The tornadic activity within 1 yr [i.e., $(x_1^I, \ldots, x_{366}^I)$] is modeled by a Markov chain.

Given the data, one can estimate the transition probabilities of such a Markov chain according to the principle of maximum likelihood estimation. The likelihood is the probability of observing the data as a function of the parameters of the underlying stochastic model. The likelihood of the Markov chain model is

$$L_1 = P_0(1)^{n_0(1)} P_1(1)^{n_1(1)} \times \prod_{j=1}^{365} P_{00}(j)^{n_{00}(j)} P_{01}(j)^{n_{01}(j)} P_{10}(j)^{n_{10}(j)} P_{11}(j)^{n_{11}(j)}. \quad (1)$$

Here $n_1(j)$ and $n_0(j)$ are the numbers of years for which day j is tornadic or nontornadic, respectively. Similarly, $n_{11}(j)$ is the number of years for which day j and $j + 1$ are both tornadic; $n_{00}(j)$, $n_{01}(j)$, and $n_{10}(j)$ are defined similarly. The parameters $P_k(j)$ and $P_k(1)$, $k, \ell = 0, 1$ are then estimated by their maximum likelihood estimators (MLE) $\hat{P}_{k\ell}(j) = n_{k\ell}(j)/n_k(j)$ and $\hat{P}_k(1) = n_k(1)/46$, respectively.

In order to justify a first-order Markov chain model, one compares the first-order model to a zeroth- and a second-order Markov chain model. In a zeroth-order model, all observations X_j are assumed to be independent. As such, only the climatological probabilities enter the model. The zeroth-order model has likelihood

$$L_0 = \prod_{j=1}^{366} P_0(j)^{n_0(j)} P_1(j)^{n_1(j)}. \quad (2)$$

The MLEs are $\hat{P}_k(j) = n_k(j)/46$.

The second-order Markov chain model has the likelihood

$$L_2 = P_0(1)^{n_0(1)} P_1(1)^{n_1(1)} P_{00}(1)^{n_{00}(1)} P_{01}(1)^{n_{01}(1)} \times P_{10}(1)^{n_{10}(1)} P_{11}(1)^{n_{11}(1)} \times \prod_{j=1}^{364} [P_{000}(j)^{n_{000}(j)} P_{001}(j)^{n_{001}(j)} P_{010}(j)^{n_{010}(j)} \times P_{100}(j)^{n_{100}(j)} P_{011}(j)^{n_{011}(j)} P_{101}(j)^{n_{101}(j)} \times P_{110}(j)^{n_{110}(j)} P_{111}(j)^{n_{111}(j)}], \quad (3)$$

where the $n_{k\ell i}(j)$ count how many years have tornadic activity scheme $k\ell i$ on days j through $j + 2$. The MLEs are $\hat{P}_k(1) = n_k(1)/46$, $\hat{P}_{k\ell}(1) = n_{k\ell}(1)/n_k(1)$, and $\hat{P}_{k\ell i}(j) = n_{k\ell i}(j)/n_{k\ell}(j)$.

[3] From a practical point of view, it would be beneficial to compare the performance of the models to other incumbent models (Hamill and Church 2000; Colquhoun and Riley 1996; Reap and Foster 1979). However, such a comparison would not contribute to the development of the former; as mentioned in the first paragraph of the introduction, the models developed here utilize autocorrelations, whereas the latter are based on cross correlations.

[4] In this paper, day 366 of a non–leap year is coded as a nontornadic day. This is not expected to adversely affect the results.

TABLE 1. Maximized log-likelihood and BIC for the three nonparametric Markov chain models.

	NE		SE		ST		NT	
i	$\log \hat{L}_i$	BIC_i	$\log \hat{L}_i$	BIC_i	$\log \hat{L}_i$	BIC_i	$\log \hat{L}_i$	BIC_i
0	−5597	−14 756	−7084	−17 730	−7274	−18 110	−5884	−15 330
1	−5304	−17 721	−6690	−20 493	−6853	−20 820	−5528	−18 170
2	−5026	−24 250	−6290	−26 779	−6483	−27 164	−5240	−24 678

The Bayes information criterion (BIC) (Guttorp 1995, chapter 2.8) provides a way to compare how well different models describe the data. Let $\hat{L}$ be the maximized likelihood of a model with p parameters for a dataset of N observations, then the BIC of the model is defined to be

$$\text{BIC} = 2 \log\hat{L} - p \log N.$$

The model with the higher BIC can be deemed as the better model.

Alternatively, nested models can be compared by the likelihood ratio test (Gutorp 1995, chapter 2.7). For the test of the hypothesis that the data come from the model of order i_1 with p_{i1} parameters versus the hypothesis that the data come from model i_2 with p_{i_2} parameters, where $i_2 > i_1$ and $p_{i_2} > p_{i_1}$, the likelihood ratio test statistic is $\lambda \equiv 2(\log\hat{L}_{i_2} - \log\hat{L}_{i_1})$. The statistic λ has asymptotically a chi-square distribution with $p_{i_2} - p_{i_1}$ degrees of freedom, which is used to determine the p value of the test. The p value is the probability that λ takes on a value larger than or equal to the observed value. Given a significance level $1 - \alpha$ (e.g., 95%), the likelihood ratio test rejects the model of order i_1 if the p value is smaller than α (e.g., 5%).

In general, the BIC and the likelihood ratio test can suggest different models. This is simply a consequence of the fact that the choice of the ''best model'' depends on the choice of the criterion (BIC, likelihood ratio, etc.), and that different criteria gauge different facets of the goodness of a model. In particular, note that BIC tends to prefer more parsimonious models, that is, models with less parameters, than the likelihood ratio test.

3. Nonparametric results

For the Markov chain models of order $i = 0, 1, 2$, the number of parameters are $P_0 = 366$, $P_1 = 2 \times 365 + 1 = 731$, and $P_2 = 4 \times 364 + 2 + 1 = 1459$. The number of observations is $N = 46 \times 366 = 17\,394$. Table 1 displays the maximized log-likelihood $\log\hat{L}_i$, $i = 0, 1, 2$ and the corresponding BIC for the four geographic regions. Table 2 shows the resulting values of the likelihood ratio test statistic and the corresponding p values. Except for the Southeast, the likelihood tests suggest that a first-order approach is appropriate. These results suggest that the process of tornadic activity can be well described by a Markov chain of order no higher than one.

However, all three models are unrealistic since they allow the transition probabilities on different days to vary freely. It may in particular happen that for some day j the estimated transition probability is $\hat{P}_{11}(j) = 0$, but $\hat{P}_{11}(j + 1) = 1$ on the next day. The extreme values 0 and 1 do not occur in the current data, but there are still unrealistically abrupt changes in transition probabilities. In the southern Tornado Alley, for example, $\hat{P}_{11}(j) = 0.125$ on day $j = 291$, but $\hat{P}_{11}(j + 1) = 0.9$ on day $j + 1 = 292$. Another feature of the data is that some consecutive days are always nontornadic, and thus no data is available to estimate the probability P_{11} for those days. In conclusion, all three nonparametric models simply have too many parameters and overfit the data.

There are at least two approaches to resolving this issue. The first is to define seasons, that is, stretches of consecutive days, in which the transition probabilities are assumed to be homogeneous. The second approach is to assume that the function that maps the days to the corresponding transition probabilities has a specific parametric form depending only on a few parameters. It turns out (details omitted) that the latter approach yields the best model in the BIC sense.[5] This is the model presented herein (section 4).

4. A parametric Markov chain model

A first-order Markov chain model for binary data is fully specified through the two transition probabilities $P_{01}(j)$ and $P_{11}(j)$ at all the days $j = 1, \ldots, 365$. Motivated through the bell shape of the occurrence fre-

[5] Likelihood ratio testing is no longer appropriate when comparing two models that are not nested, in the sense that none is a submodel of the other.

TABLE 2. Likelihood ratio test statistics λ and associated p values $P(\lambda)$ for testing the three nonparametric Markov chain models.

	NE		SE		ST		NT	
	λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$
0 vs 1	587	0.00	788	0.00	842	0.00	713	0.00
1 vs 2	555	1.00	799	0.03	741	0.36	576	1.00

quencies in the bottom panel in Fig. 1, the transition probabilities are modeled with trigonometric functions. Specifically, $P_{01}(j) = P_{01}(j, a_0, b_0, c_0, d_0)$ depends on four parameters:

$$P_{01}(j) = \begin{cases} b_0 + \frac{a_0}{2}\left[\cos\left(2\pi\frac{(j - d_0)}{c_0}\right) + 1\right] & \text{if } j \in \left(d_0 - \frac{c_0}{2}, d_0 + \frac{c_0}{2}\right), \\ b_0 & \text{otherwise.} \end{cases} \quad (4)$$

The other transition probability $P_{11}(j) = P_{11}(j, a_1, b_1, c_1, d_1)$ is defined analogously.

The range of the parameters is $a_i \in [0, 1]$, $b_i \in [0, 1]$, $c_i \in [0, 366]$, and $d_i \in [0, 366]$, $i = 0, 1$. To ensure that the defined probabilities indeed fall in the interval [0, 1] it must further be true that $a + b \leq 1$. Hence, the parameter space for the parameter vector $\theta \equiv (a_0, \ldots, d_1)$ is

$$\Theta = \{\theta \in R^8 : (c_i, d_i) \in [0, 366]^2, (a_i, b_i) \in [0, 1]^2, \; a_i + b_i \leq 1, i = 0, 1\}.$$

The interpretation of the parameters is as follows:

- b_i represents a baseline (throughout year) probability of $(i, 1)$ transition;
- d_i is the day of maximal $(i, 1)$-transition probability;
- c_i is the length of the period of increased $(i, 1)$-transition probability (as compared to the baseline b_i; this period is symmetric around d_i); and
- a_i stands for the magnitude of increase in $(i, 1)$-transition probability from baseline b_i to peak at day d_i.

Figure 3 displays the functional form of the transition probabilities for some values of the parameters.

This model can be fitted by numerically maximizing the likelihood yielding the MLE $\hat{\theta}$. By standard statistical theory (Guttorp 1995, chapter 2.7), the MLE has an asymptotic normal (Gaussian) distribution.

In evaluating the forecast performance of the parametric model, it is necessary to estimate functions $g(\theta)$ of the model parameters θ. Here, these functions are estimated by $g(\hat{\theta})$, that is, the value of the function at the MLE. It is then possible to employ parametric bootstrapping (Efron and Tibshirani 1993) to obtain confidence intervals for $g(\theta)$. Using the asymptotic normal distribution of the MLE $\hat{\theta}$, the bootstrap consists of simulating a sample of parameter values $\theta_1, \ldots, \theta_s$ from an eight-dimensional normal distribution some number of times (here, 10 000). The sample quantiles of $g(\theta_1), \ldots, g(\theta_s)$ provide the desired confidence intervals for $g(\theta)$. This bootstrap procedure is employed to obtain all confidence intervals reported henceforth.

5. Parametric results

One can compare the maximized log-likelihood of the parametric first-order model to that of the nonparametric zeroth-order model. Since the specification of the parametric first-order model does not include a specification of the (unconditional) distribution of tornado occurrence on the first day of the year, one compares a conditional log-likelihood, conditional on all the first days. Then it is found that the maximized conditional log-likelihood of the parametric model is -5661, -7084, -7281, and -5894 for the regions NE, SE, ST, and NT, respectively. Conditional on the first days, the nonparametric model of order zero has maximized log-likelihood -5597, -7073, -7270, and -5884 for the respective regions. Note that especially in the regions SE, ST, and NT the 8-parameter first-order model has a maximized log likelihood very close to the one of the 365-parameter zeroth-order model. Since the BIC takes the number of parameters into account as well, the parametric first-order model has a much larger BIC than the nonparametric zeroth-order model.

The MLEs for the parametric model in the four geographic regions are listed in Table 3. The corresponding transition probabilities are depicted in Fig. 4. These figures display a variety of behavior that can be explained in terms of known meteorology. First note the main characteristics of the various figures: 1) in all four regions, P_{11} is larger than P_{01} throughout the year; 2) in all four regions, both P_{01} and P_{11} peak in the middle portion of the year; 3) in all four regions, the rise in P_{11} (i.e., from base level to the peak) is about 10% larger than that of P_{01}; 4) the rise in P_{11} and P_{01} are largest in the northern Tornado Alley ($\sim$47%) and smallest in the Southeast ($\sim$12%), while the corresponding values in the Northeast and southern Tornado Alley are comparable ($\sim$25%); 5) the peak in P_{11} is somewhat earlier than that of P_{01} in all regions, except in the Northeast, where the peaks occur at the same time; 6) in the southern Tornado Alley, the P_{11} curve is sharper than the P_{01} curve, while in the remaining regions the widths of the two curves are comparable.

6. Meteorological discussion

The MLEs of the transition probabilities for the four regions (Fig. 4) are consistent with many of the heuristic notions of the regional and seasonal variability of tornado occurrence. They provide statistical confirmation and validation for the existence of temporal and spatial

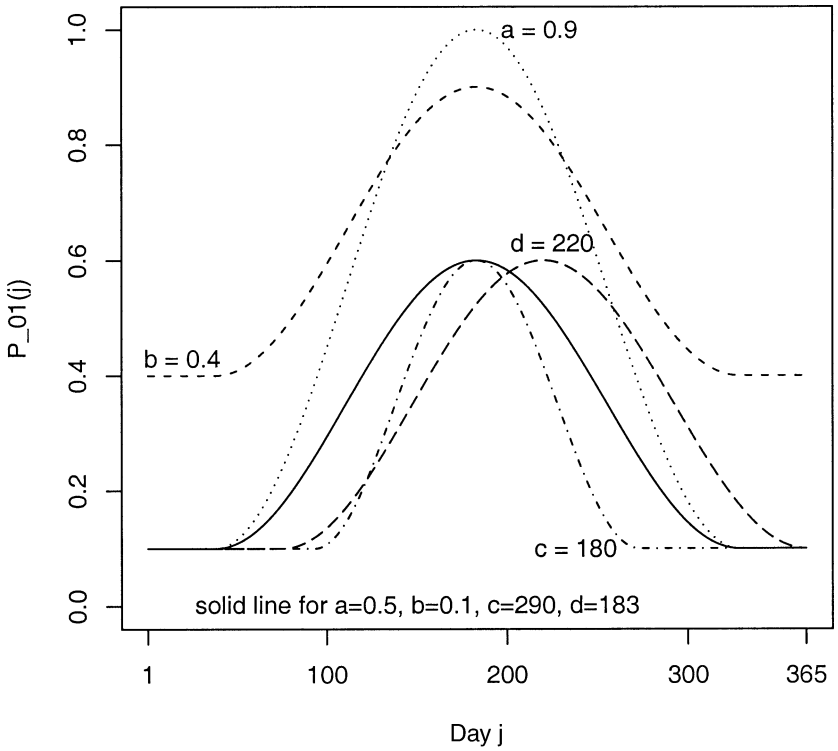


FIG. 3. The parameterization of the transition probabilities. Parameter b is a baseline—its variation shifts the curve vertically; d is the peak day—its change shifts the curve horizontally; c determines the length of the deviation from the baseline; and a scales the deviation's magnitude.

variations in tornado frequency. Some of these patterns have been explored in previous climatologies (e.g., Kelly et al. 1978), but others are more anecdotal in nature and have either not been documented or have been discussed only in semitechnical articles (e.g., Schaefer et al. 1980).

Conceptually, P_{11} is the probability that two tornado days in a row will occur in the region of concern. Conversely, P_{01} is the probability that a tornado day in the region will follow a nontornado day. In any of the regions studied and for any time of the year, the probability of a tornado day following a nontornado day is always less than that of having two tornado days in a row. This reflects the observation by Fawbush et al. (1951), in their seminal work on tornado forecasting, that synoptic-scale meteorological systems with strong fronts and associated deep-surface low pressure systems precondition the atmosphere for organized tornado activity. Such ''frontal systems'' typically remain in regions as large as those studied here for periods of days. After such a system migrates out of the region, tornadoes do not develop until the next system arrives. Accordingly, ''0–1'' days in a given region are much less common than ''1–1'' days.

The annual cycle of tornado activity and its geographic variation can be inferred from the MLE plots. In the Texas, Louisiana, Arkansas, Oklahoma area (southern Tornado Alley; bottom-left panel of Fig. 4) during the winter, there is little moisture in the lower atmosphere, and probabilities are at their baseline value.

TABLE 3. MLEs $\hat{\theta}$ in the four regions and their standard errors. Recall that $i = 0$ indicates a transition from a ''no tornado day'' to a ''tornado day,'' and $i = 1$ indicates a transition from a ''tornado day'' to a ''tornado day.''

Region	i	$\hat{a}_i$	$\hat{b}_i$	$\hat{c}_i$	$\hat{d}_i$
NE	0	0.25 ± 0.008	0.02 ± 0.002	278.3 ± 6.2	177.7 ± 1.5
	1	0.30 ± 0.03	0.16 ± 0.02	245.9 ± 22.9	176.5 ± 3.8
SE	0	0.17 ± 0.008	0.07 ± 0.004	292.7 ± 13.9	163.3 ± 2.8
	1	0.16 ± 0.02	0.26 ± 0.02	223.2 ± 40.8	138.3 ± 6.4
ST	0	0.25 ± 0.01	0.07 ± 0.004	272.3 ± 8.4	156.4 ± 2.0
	1	0.34 ± 0.02	0.29 ± 0.01	131.4 ± 12.1	138.9 ± 1.9
NT	0	0.41 ± 0.01	0.01 ± 0.002	257.1 ± 3.3	178.9 ± 1.0
	1	0.48 ± 0.03	0.20 ± 0.02	207.1 ± 11.5	170.4 ± 1.7

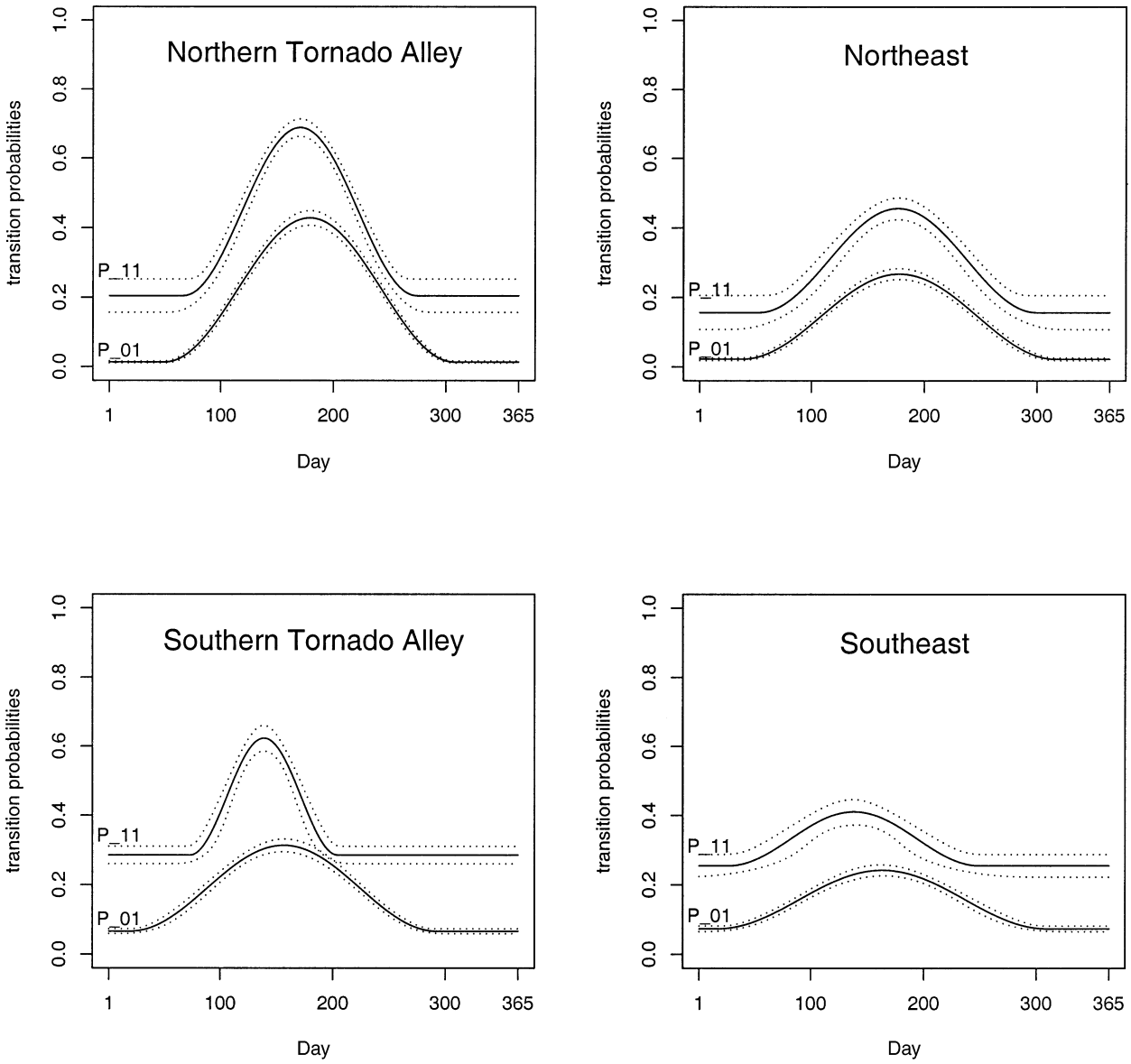


FIG. 4. The MLE of the transition probabilities in the four regions and pointwise, i.e., daywise, 95%-confidence intervals.

However, as Gulf moisture starts to return to the area in February, the probability of a tornado day following a nontornado day, P_{01}, starts to rise. In early April, surface temperatures over the southern Tornado Alley area get warm during the day, and relatively strong frontal systems frequently traverse the area. This causes a rapid rise in the probability of two successive tornado days, P_{11}. This probability continues to rise until mid-May when it reaches a value of about 0.6 in early May (i.e., a 60% chance that a tornado day today will be followed by another tornado day tomorrow). Starting in mid-May, both the number of frontal systems that propagate through the southern Tornado Alley and P_{11} rapidly decline. However, P_{01} continues to increase until mid-June, as the moisture and instability in the lower layers of the atmosphere increases and tornado activity can develop in association with weak upper systems rather than with strong frontal systems. This period between the start of the increase in P_{11} (early April) and the start of the decrease in P_{01} (mid-June) corresponds to the tornado season, with fewer nontornado days (high P_{01}) and more frequent successive tornado days (high P_{11}).

In the northern Tornado Alley (Missouri, Kansas, Nebraska, the Dakotas, Minnesota, western Wisconsin, and Iowa; upper-left panel of Fig. 4), the increased distance to the Gulf of Mexico allows the cold-season lower atmosphere to be both drier and more stable than it is over the southern Tornado Alley. This is reflected not only by the markedly lower baseline probabilities but by the delay in the increase of P_{01} until early March.

Interestingly, the increase in P_{11}, the probability of successive tornado days, starts about the same time (early April) over the northern Tornado Alley as it does over southern Tornado Alley. However, the maximum of both P_{11} (in late June) and P_{01} (in early July) occur later in the year in the northern area. These features argue that the strength and frequency of frontal systems is the dominant factor in the intensive spring activity over Tornado Alley. The maximum of P_{11} in the northern section is higher than it is over the southern section, but this is likely a factor of the difference in areas of the two regions.[6] In the Northern Tornado Alley, the tornado season extends from early April through early July.

The tornado day occurrence pattern in the Northeast (essentially all the United States from Kentucky and Virginia northward and east of the Mississippi River; upper-right panel of Fig. 4) is very similar to that in the northern Tornado Alley. However, in the Northeast both sets of probabilities are lower, and their peaks are much less pronounced (they have a lower ''kurtosis''). This makes the tornado season much less pronounced in the Northeast than in Tornado Alley. This is a reflection of the decreased availability of warm, moist unstable air over this region than in the one immediately to its west. Also in the Northeast, the return of P_{11} to its baseline value occurs in late September, much later in the year than in the other regions examined. This reflects that while frontal activity is diminished and a relatively strong thermal inversion present is typically over the Great Lakes area and the Ohio Valley, strong fronts that can focus tornadic storm development traverse New York and New England through the summer (Johns and Dorr 1996).

The region with the most unique tornado day pattern is the Southeast (the United States from Tennessee and the Carolinas southward and east of the Mississippi River; lower-right panel of Fig. 4). The kurtosis of the MLE pattern is so low that sharp peaks in probabilities have almost been wiped away. In late January, P_{11} starts to rise from its baseline value and does not drop back to it until August. The period of increased P_{01} extends from mid-January through late September. Part of this is due to tropical storm-spawned tornadoes. These tornadoes occur within 24 h of the tropical storm's landfall within a sector from 0° to 120° relative to true north from the storm center (Spratt et al. 1997). These tornadoes tend to increase P_{01} during the late summer. Peak values of both probabilities are much lower than in any other region, but the baseline values are comparatively high. Also, the P_{11} baseline value is only slightly lower than it is over the southern Tornado Alley, and the one for P_{01} is slightly higher. This combination indicates that throughout the year, there are frequent periods of consecutive tornado days in this part of the country. This implies that many of this region's tornadoes are associated with weak perturbations aloft that propagate across the moist unstable lower-layer air rather than with strong frontal systems. At one time, a distinction was made between ''cyclonic tornadoes'' (those associated with strong frontal systems) and ''convectional tornadoes'' (Brown 1933).

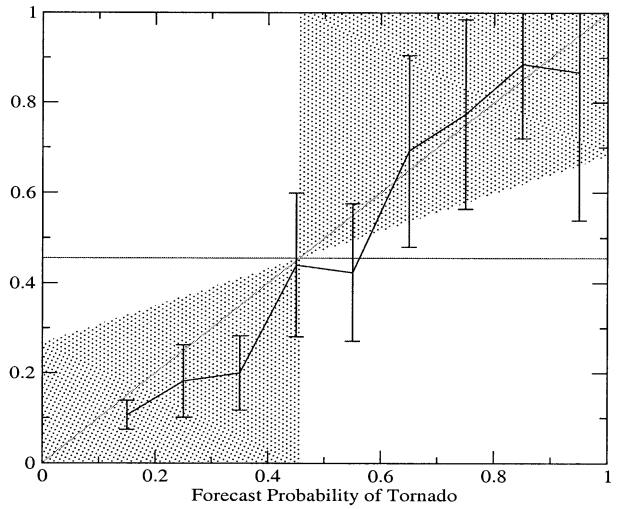
[6] Other things being equal, there would be a direct relationship between the probability of a tornado day and the size of the area being considered.

7. Forecast performance

The question of performance (or forecast verification) is highly complex. One ingredient is the choice of performance measure. In this section, several measures are described. Although there exists a plethora of performance measures, many have undesirable properties (Marzban 1998). The measures selected herein are among the less problematic ones. Most importantly, they are multidimensional—as opposed to a single scalar measure—and as such convey a great deal of information about the quality of the forecasts. The measures will be computed for the parametric first-order Markov chain from section 4 (hereafter referred to as ''the model'') and compared with those computed from climatological forecasts (i.e., zeroth-order Markov chain).

The data are divided into a training set to estimate the transition probabilities and a validation set to obtain an unbiased measure of performance. Here, the method of leave-one-out cross-validation is adopted, whereby each of the 46 years serves as a validation set (Efron and Tibshirani 1993). Confidence intervals for the performance measures are derived from 10 000 bootstrap simulations (as described in section 4).

Attributes diagrams (Murphy and Winkler 1987, 1992) constitute one assessment of a model's performance in a probabilistic setting. An attributes diagram is a generalization of a reliability diagram (Wilks 1995; Murphy and Winkler 1992); in addition to displaying the reliability of the forecasts, it shows whether or not a given forecast contributes positively to the skill of the model. The ''skill,'' gauged in terms of the Brier skill score (Wilks 1995), is the mean square error (MSE) of forecasts and observations, with the forecasts ranging from 0 to 1, and the observations being 0 or 1. Two other facets of probabilistic forecasts are captured through discrimination and refinement diagrams. A scalar measure capturing the reliability of the forecasts is the mean square error of the reliability curve, with the ''correct value'' given by the diagonal line. Perfectly reliable forecasts would yield a reliability MSE equal to 0. Receiver operating characteristic (ROC) diagrams (Masters 1993; Marzban 2000) are also examined. Briefly, an ROC diagram is a parametric plot of the probability of detection (or true positive) versus the false alarm rate (or false positive) as a threshold on a probabilistic forecast is varied from zero to one. The amount of bowing of an ROC curve is a measure of performance. A diagonal ROC curve represents random forecasts, and the more the curve bows above and beyond



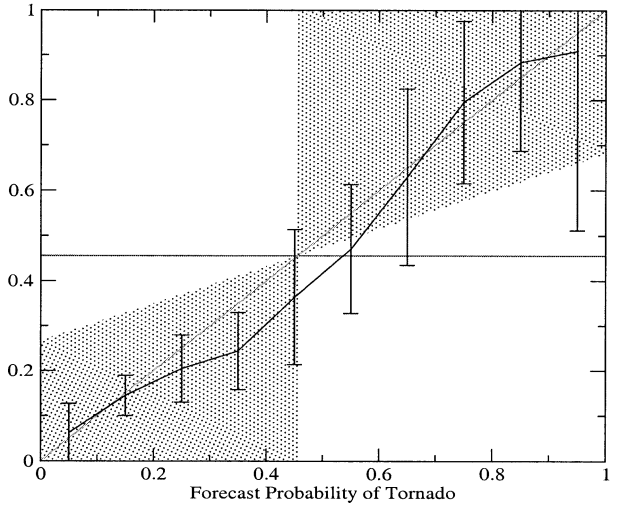
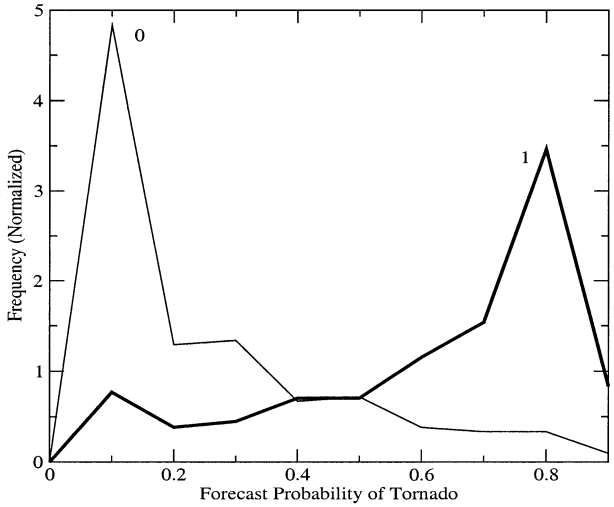


FIG. 5. The attributes diagram over the entire United States for the year 1979: (top) model forecasts and (bottom) climatological forecasts.



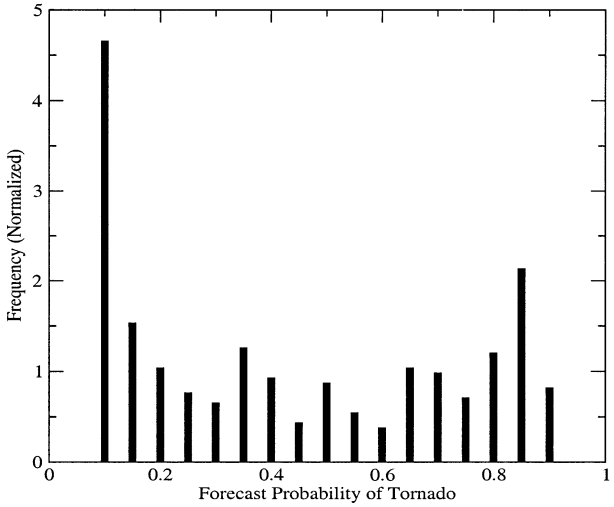


FIG. 6. (top) The discrimination and (bottom) refinement diagram for the model forecasts for the year 1979 over the entire United States. In the top panel, ''0'' indicates nonoccurrence, and ''1'' indicates occurrence of a tornado.

the diagonal line, the higher the performance. This allows use of the area under the ROC curve as a scalar measure of performance; an area equal to 1 implies perfect performance, and area of 0.5 corresponds to random forecasts.

In order to illustrate the graphical evaluation of the model, the year 1979 is chosen as the left-out (or validation) year. Aggregating tornadic activity in the entire United States, Fig. 5 displays the model's attributes diagrams, as computed for this validation set. It can be seen that not only are the forecasts highly reliable (perfect, within the standard errors), they are also all skillful because they reside in the shaded region of the diagram. For comparison, the attributes diagram for climatological forecasts is also shown in Fig. 5.

According to the discrimination diagram (Fig. 6, top), the model forecasts do clearly discriminate between tornadic and nontornadic events. The refinement diagram (Fig. 6, bottom) suggests that not only does the model produce forecasts spanning the full range from 0 to 1, but also that these forecasts are refined in that the diagram displays the desirable U-shape (Murphy and Winkler 1992).

The ROC diagram for the forecasts in 1979 over the entire United States is plotted in Fig. 7. For comparison, the ROC diagram of climatological forecasts (i.e., zerothorder Markov chain) is also shown. Since the ROC curve from the model forecast is much more convex than the one from the climatological forecast, it can be seen that the model forecasts outperform the climatological ones.

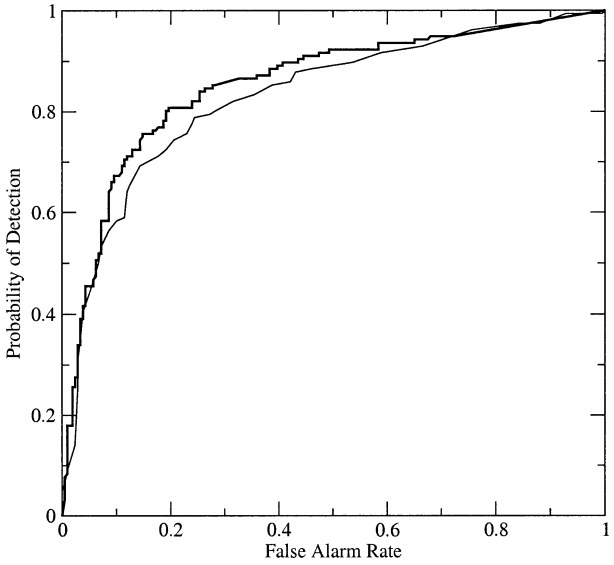


FIG. 7. The ROC diagram for the entire United States for the year 1979 (thick curve). The thin curve corresponds to climatological forecasts.

The various diagrams for the year 1979 assess the forecasts without any loss of information. However, in order to illustrate the model's forecast performance across the whole validation span 1953–98, one must further distill the information. As a representative of the scalar measures of performance considered, Fig. 8 displays the area under the ROC curve of the model forecasts for each of the 46 validation years and each of the four geographic regions. Also shown are the corresponding climatological forecasts. The error bars are 95%-confidence intervals obtained from the 10 000 bootstrap simulations. Since the ROC area from the model forecasts is always much closer to 1 than the ROC area from the climatological forecasts, the model outperforms climatology for every one of the 46 validation sets.

The degree to which the model outperforms climatology depends on the geographic region. In order to numerically compare the performance over the different regions, we distill the information even further by reporting the results of a *t* test performed on the difference between the ROC area performance of the model forecasts and climatological forecasts (see Fig. 8). Those values are listed in the first row of Table 4, and all are statistically significant at the 95% level.

Table 4 also displays, in the second row, the values of the *t* statistic as obtained for the reliability MSE. They are derived from figures (not shown) analogous to Fig. 8 for ROC. The negative values indicate that the model outperforms climatology, because MSE is an inverse measure of performance. The Brier score of the models is tabulated in the third row.

Evidently, when performance is gauged in terms of the area under the ROC curve, the models perform better in the North (both Northeast and northern Tornado Alley) than in the South. This is the same pattern displayed according to the BIC (Table 1). In terms of the reliability MSE, the model performs well in the SE and NT, but its performance in the NE is only marginal. Furthermore, it performs poorly in the ST, where it does worse than climatology. However, the p values in Table 4 show that the difference in reliability MSE between the model and climatology is statistically significant at the 95% level only in the SE and NT, where the model outperforms climatology. The marginally superior reliability in the NE and the inferior reliability in the ST are not statistically significant. Moreover, it is important to note that reliability is a somewhat difficult measure to appease. This is especially true when model forecasts are compared to climatological forecasts, because the latter are perfectly reliable when assessed with the training set. In terms of the Brier score, the models perform comparably in the different regions, with the worse performance in the NE. In contrast to reliability MSE, the Brier scores of the models are superior to climatology in all regions.

It is not surprising that the comparison of the model and climatology forecasts depends on the choice of the performance measure. Each measure gauges a different facet of performance quality. However, the question then becomes whether the value added by the model is significantly higher than that already provided by climatological forecasts. To answer that question, one must examine a larger list of measures. As mentioned above, discrimination and refinement are two other facets of probabilistic forecasts. They are difficult to reduce to single (scalar) measures, and for that reason plots analogous to Fig. 8 are not available. However, an exhaustive examination of the discrimination and refinements plots for each of the 46 validation sets and in each of the four regions suggests that the model forecasts are significantly superior to climatological forecasts. As such, the model has value.

8. Summary and discussion

Based on a dataset consisting of 46 years of daily tornado counts, tornadic activity is modeled with Markov chains in four different geographic U.S. regions. From a meteorological point of view, the Markov chain model of tornado day activity agrees with a dynamic climatological interpretation of U.S. tornado occurrence. From a forecasting point of view, it can be added to the growing arsenal of forecasting tools. Unlike many other tools for tornado forecasting, it does not rely on any hardware, such as Doppler radar or satellites; it utilizes only the tornadic activity on the day(s) prior to the forecast day. As such, it can be used to a priori estimate the risk of tornadic activity before reliable meteorological data is available. This information is useful for preparedness work such as spotter training.

Statistically, the Markov chain model is nontrivial

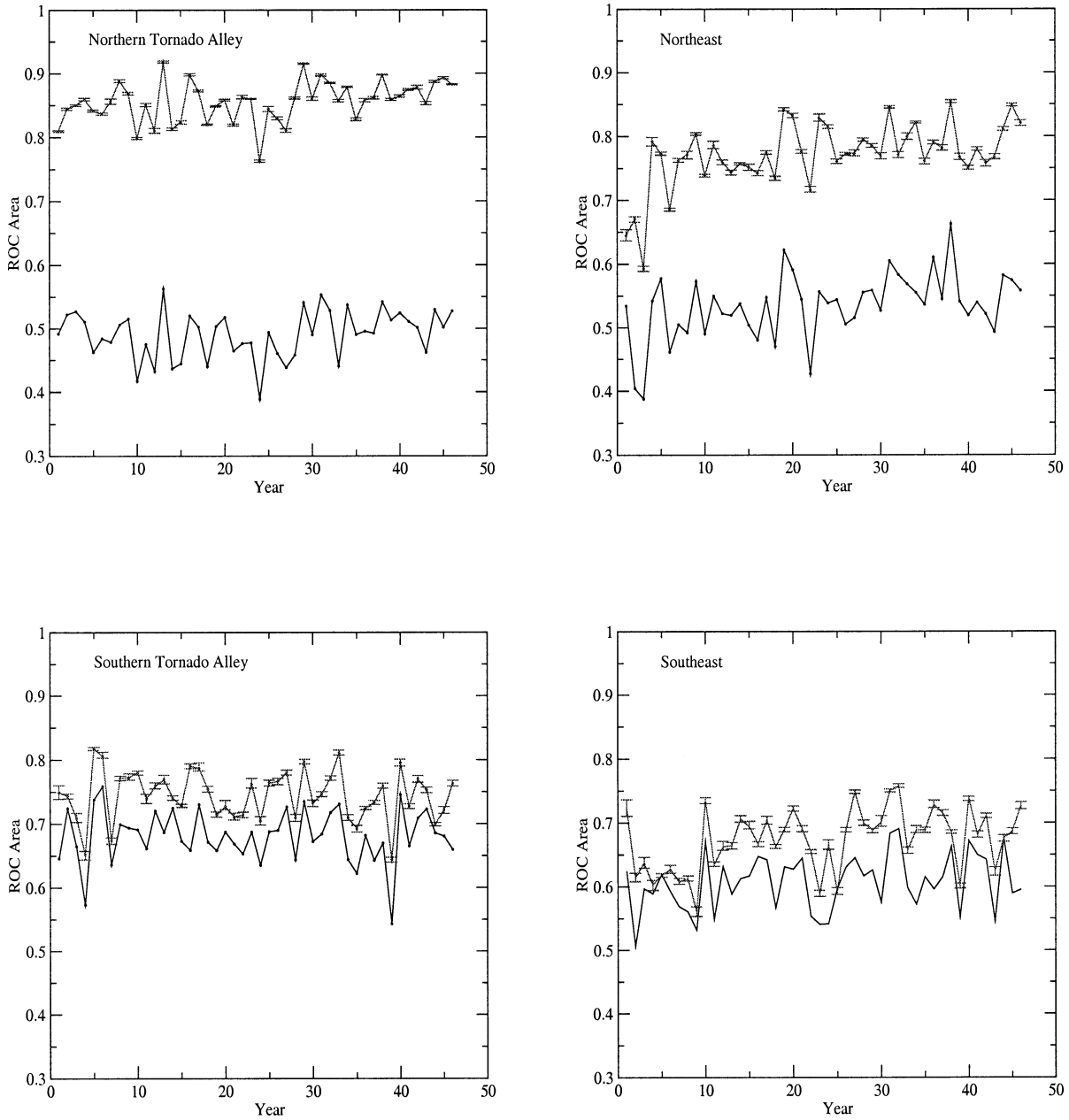


FIG. 8. The performance of the model for each of the 46 validation years and the four geographic regions. The curve with error bars showing yearwise 95%-confidence intervals corresponds to the model, while the remaining curve refers to climatology. The graphs are laid out in relation to their corresponding geographic locations.

TABLE 4. Student's t test statistics and associated p values $P(t)$. All p values are significant at the 95% level, except for the reliability MSE in the NE and ST.

	NE		SE		ST		NT	
	t	$P(t)$	t	$P(t)$	t	$P(t)$	t	$P(t)$
ROC area	51.82	0.00	12.84	0.00	18.23	0.00	120.35	0.00
Reliability MSE	−0.68	0.50	−18.87	0.00	1.70	0.10	−2.61	0.01
Brier score	−5.11	0.00	−7.28	0.00	−8.94	0.00	−8.98	0.00

because the process underlying tornadic activity is not homogeneous. Among the models examined, the one that best fits the data is first order, parametric, and nonhomogeneous. As such, only a knowledge of yesterday's tornadic activity is required to make an optimal forecast of today's tornadic activity. The developed parametric Markov chain model based on a simple trigonometric parameterization of the transition probabilities describes tornadic activity quite well. Fitting the model in subregions of the United States suggests, in particular, that in the South tornado persistence is specific mostly to the early part of the year.

This model's performance is evaluated in a realistic forecasting scheme in terms of attributes diagrams, discrimination and refinement plots, the Brier score, and receiver operating characteristic (ROC) curves. It is found that the Markov chain model outperforms climatological forecasts in each of the four regions and in terms of all the performance measures, with the exception of forecast reliability, which, in the southern Tornado Alley, is inferior to that of climatological forecasts (but only to a statistically nonsignificant degree).

The specific parametric form adopted in this study assumes only one peak in tornadic activity within a calendar year, namely, the one in the midportion of the year. However, there is some evidence for yet another peak in the latter portion of the year, albeit a much smaller one. This second peak is more apparent in the actual tornado counts than in the tornadic occurrences considered here. However, it is present only in the Northeast and Southeast and not in Tornado Alley. Alas, the current analysis can be viewed as a first-order approximation, but one that captures the bulk of the underlying process.

Another assumption made here has been that of identical and independent replication. One might question the validity of this assumption given the expectation that large-scale changes in climate (e.g., global warming, El Niño) affect small-scale weather phenomena such as severe storms and tornadoes. To test the validity of this assumption, the parametric Markov chain model was enlarged by two additional parameters. These allow the baseline transition probabilities b_0 and b_1 (see section 5) to exhibit a linear trend in time. In the spirit of the model selection approach employed previously, BIC was used to compare the resulting parametric Markov chain model incorporating trend to the analog without trend. The BIC values for the model with trend are $-11\,417$, $-14\,208$, $-14\,656$, and $-11\,879$ in the regions NE, SE, ST, and NT. The model without trend yields $-11\,401$, $-14\,246$, $-14\,640$, and $-11\,865$. Hence, the trend is of importance only in the Southeast, and even there it is of not much importance. Therefore, the assumption of identically distributed years is not violated to a significant degree. However, contrary to the tornado occurrences considered in this paper, the actual tornado counts do display a strong increasing trend over time (not shown). However, as discussed in footnote 1, this increase is believed to be an artifact of tornado reporting.

REFERENCES

Anderson, C. J., and C. K. Wikle, 2002: Analysis of tornado counts with hierarchical Bayesian spatio-temporal models. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 459–460.

Brooks, H. E., and C. A. Doswell III, 2001: Some aspects of the international climatology of tornadoes by damage classification. *Atmos. Res.*, **56,** 191–201.

——, ——, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting,* **18,** 626–640.

Brown, C. W., 1933: A study of the time, areal, and type of distribution of tornadoes in the United States. *Trans. Amer. Geophys. Union,* **14,** 100–106.

Colquhoun, J. R., and P. A. Riley, 1996: Relationships between tornado intensity and various wind and thermodynamic variables. *Wea. Forecasting,* **11,** 360–371.

Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap.* Chapman and Hall, 436 pp.

Fawbush, E. J., R. C. Miller, and L. G. Starrett, 1951: An empirical method of forecasting tornado development. *Bull. Amer. Meteor. Soc.,* **32,** 1–9.

Gates, P., and H. Tong, 1976: On Markov chain modeling to some weather data. *J. Appl. Meteor.,* **15,** 1145–1151.

Grazulis, T. P., J. T. Schaefer, and R. F. Abbey Jr., 1993: Advances in tornado climatology, hazards, and risk assessment since Tornado Symposium II. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr.*, No. 79, Amer. Geophys. Union, 409–426.

Guttorp, P., 1995: *Stochastic Modeling of Scientific Data.* Chapman and Hall, 384 pp.

Hamill, T. M., and A. T. Church, 2000: Conditional probabilities of significant tornadoes from RUC-2 forecasts. *Wea. Forecasting,* **15,** 461–475.

Hughes, J. P., and P. Guttorp, 1994: Incorporating spatial dependence and atmospheric data in a model of precipitation. *J. Appl. Meteor.,* **33,** 1503–1515.

Johns, R. H., and R. A. Dorr Jr., 1996: Some meteorological aspects of strong and violent tornado episodes in New England and eastern New York. *Nat. Wea. Dig.,* **20,** 2–12.

Kao, E. P. C., 1997: *An Introduction to Stochastic Processes.* Duxbury Press, 438 pp.

Katz, R. W., 1974: Computing probabilities associated with the Markov chain model for precipitation. *J. Appl. Meteor.,* **13,** 953–954.

Kelly, D. L., J. T. Schaefer, R. P. McNulty, C. A. Doswell III, and R. F. Abbey Jr., 1978: An augmented tornado climatology. *Mon. Wea. Rev.*, **106,** 1172–1183.

Lakshmanan, V., 2001: Lossless coding and compression of radar reflectivity data. Preprints, *30th Conf. on Radar* Meteorology, Munich, Germany, Amer. Meteor. Soc., 50–52.

Marzban, C., 1998: Scalar measures of performance in rare-event situations. *Wea. Forecasting,* **13,** 753–763.

——, 2000: A neural network for tornado diagnosis. *Neural Comput. Appl.,* **9,** 133–141.

——, and J. T. Schaefer, 2001: The correlation between U.S. tornados and Pacific sea surface temperature. *Mon. Wea. Rev.*, **129,** 884–895.

Masters, T., 1993: *Practical Neural Network Recipes in C++.* Academic Press, 493 pp.

Murphy, A. H., and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.,* **115,** 1330–1338.

——, and ——, 1992: Diagnostic verification of probability forecasts. *Int. J. Forecasting,* **7,** 435–455.

Ray, P. S., R. Bieringer, X. Niu, and B. Whissel, 2003: An improved estimate of tornado occurrence in the central plains of the United States. *Mon. Wea. Rev.*, **131,** 1026–1031.

Reap, R. M., and D. S. Foster, 1979: Automated 12–36 hour probability forecasts of thunderstorms and severe local storms. *J. Appl. Meteor.*, **18,** 1304–1315.

Schaefer, J. T., D. L. Kelly, C. A. Doswell III, J. G. Galway, R. J. Williams, R. P. McNulty, L. R. Lemon, and B. D. Lambert, 1980: Tornadoes, when, where, how often. *Weatherwise*, **33,** 52–59.

Spratt, S. M., D. W. Sharp, P. Walsh, A. Samdrik, F. Alsheimer, and C. Paxton, 1997: A WSR-88D assessment of tropical cyclone outer rainband tornadoes. *Wea. Forecasting*, **12,** 479–501.

Stern, R. D., 1982: Computing a probability distribution for the start of the rains from a Markov chain model for precipitation. *J. Appl. Meteor.*, **21,** 420–422.

Valdez, M. P., and K. C. Young, 1985: Number fluxes in equilibrium raindrop populations: A Markov chain analysis. *J. Atmos. Sci.*, **42,** 1024–1036.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.