NOTES AND CORRESPONDENCE

Empirical Orthogonal Function Expansion Applied to Progressive Tornado Outbreaks

Joseph T. Schaefer

NOAA National Weather Service, Scientific Services Division
Central Region Headquarters, Kansas City, MO 64106

Charles A. Doswell, III

NOAA Environmental Research Laboratories Weather Research
Project Boulder, CO 80303

(manuscript received 6 July 1984, in revised form 30 September 1984)

Abstract

Synoptic data from progressive tornado outbreak days are examined using an Empirical Orthogonical Function (EOF) expansion. "Composite" tornado outbreak charts are created. A methodology for using EOF correlations for statistical forecasting is presented.

1. Progressive tornado outbreaks

As noted by Galway (1977), three distinct types of outbreaks of ten or more tornadoes occur. In increasing order of frequency, these types are local (all tornadoes within a circular region of about 1° in radius), line (limited eastwest progression, tornadoes form along an approximate north-south axis), and progressive (outbreak advances from west to east with time). Moeller (1979) independently found a somewhat similar breakdown of outbreaks. While these climatological facts are now known, details of the meteorological processes which differentiate among outbreak types remain unidentified.

As a first step towards such an identification, synoptic charts associated with historical progressive tornado outbreaks which occurred from 1950 through 1968 within a quasihomogenous spatially limited geographic area (Kansas, Oklahoma, and northern Texas) have been examined. For this period, 14 progressive outbreaks limited themselves to the region of interest. Rawinsonde data for seven different isobaric levels and the surface have been analyzed for the data time immediately preceding the outbreak.

One method of assimilating such large amounts of data is by constructing mean or composite charts (e. g., Beebe, 1956). For the sample of 14 progressive outbreaks, the mean 500 mb geopotential and thermal fields are shown in Fig. 1. It is interesting to note that the thermal trough lies slightly eastward of the geopotential trough. This configuration is the direct opposite of one which is baroclinically unstable (Charney, 1947). The baroclinic zone associated with the composite pattern is essentially a low- to mid-tropospheric phenomenon. By 250 mb, the isotherms and contours are essentially parallel. Although no comments are made, a similar configuration is apparent in the composite charts for Beebe's (1956) areas I and II which straddle the Kansas-Oklahoma region.

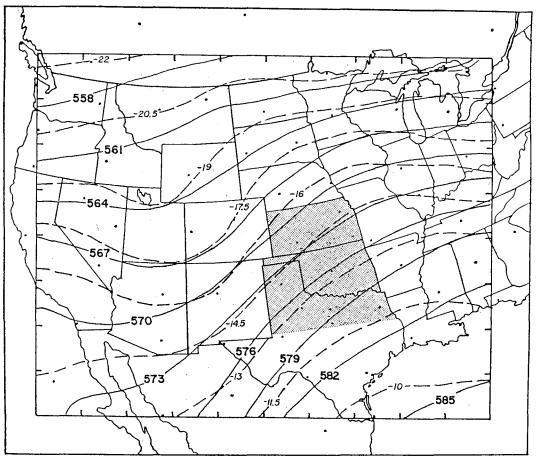

Fig. 1 Mean 500 mb geopotential and thermal fields for 14 progressive outbreaks. Contours (solid) are labelled in decameters, isotherms (dashed) in °C. All outbreaks are contained in the geographic region indicated by stippling. Small, unlabelled dots locate the current operational network rawinsonde sites, in order to give an idea of the data density. Tick marks on the border of the interior rectangle are grid point coordinates used in analysis-these are omitted in subsequent figures.

The present composite charts are very similar to Beebe's. This is not surprising since his data is based upon "family-type" outbreaks. However, a problem exists with the use of composite charts. The process of averaging smooths out details, allowing composite charts to contain only large scale features. Fine structure, which can be used to discriminate among weather patterns, is not present. This forced Beebe to observe that "in spite of seasonal and geographical differences in data, the similarity of patterns is striking."

2. Empirical Orthogonal Function (EOF) expansion

A rapidly expanding body of meteorological literature on EOF analysis and its applications exists*. Essentially, EOF analysis is a mathematical artifice used to reduce a collection of data sets to a much smaller set of "characteristic" patterns. These patterns are the EOFs (Lorenz, 1956). If the spatial data distribution is being examined, the EOFs are determined from the temporal variation of the spatial pattern (for temporal distributions, spatial variations are examined). The first EOF is the pattern which, in the mean, makes the largest contribution to the spatial analysis at every observation time (i.e., it explains the most variance). Subsequent EOFs make successivelly smaller contributions.

While an EOF expansion is similar to other expansions via orthogonal functions (e.g., Fourier, Legendre, etc.), it has several advantages. Since the amount of explained variance by each EOF is known, the mean effect of truncating the expansion is also known. Because the EOFs are determined by the data distribution, boundary effects are implicitly accounted for by the mathematics (Buell, 1978) and do not have to be examined explicitly. Finally, it can be shown that an EOF expansion requires the fewest terms of any orthogonal expansion to represent a given amount of variance (Holmström, 1963).

Formally, consider the data set as a $p \times n$ matrix $\underline{Q}$. Each of the n columns is a vector composed of the observed data (ordered by location) for a specific case and each of the p rows contains data from a specific location (sorted by case). Thus, there are p observations and n cases. The EOFs are the eigenvectors of the covariance matrix formed from $\underline{Q}$. The ratio on the associated eigenvalue to the sum of all the eigenvalues gives the percent variance reduced by any one EOF. It is numerically desirable to use deviations as input rather than raw observations when defining $\underline{Q}$. For each case, the deviation vector is calculated by subtracting the field mean $\underline{A}_j$ from the observation vector (F_j). Derivation of the mathematics can be found in the literature (e.g., Grimmer, 1963).

For this progressive tornado outbreak study, $n=14$. Ideally, for a study involving measured data, p should equal the number of observations so that each row is uniquely associated with observations from one site. However, this is not possible when the data span many years. Rawinsonde release locations change, the number of balloons released vary from case

* For an extensive annotated bibliography see Richman (1980).

to case, and observations are missing. To compensate for this, the data are interpolated to a grid and each *grid point* is treated as an observation.

The grid has 99 points with a spacing of 381 km at 600°N (Fig. 1). This only allows for a synoptic scale analysis. Interpolation is performed via a one-pass Gaussian routine tuned to give a 99% suppression of 4 grid length waves (Barnes, 1964). Solution of the eigenvalue problem is via the QL method. If $(n\text{-}3)$ is less than p, typically there will only be $(n\text{-}3)$ nondegenerate EOFs for each data set analyzed (Preisendorfer and Barnett, 1977). Thus, for this study there are approximately 11 nondegenerate EOFs. Degenerate vectors are theoretically null vectors but in practice they may have non-zero components forced by imprecise calculations. However, such degenerate vectors do not meet the mathematical criteria required for an eigenvector and are eliminated from the set of EOFs.

EOFs were computed for 40 different fields (5 parameters at 8 different pressure levels). For the 14 progressive outbreaks studied, the first EOF always accounts for more than a 75% reduction of variance. Examples of the primary EOF for both 500 mb geopotential (80% variance reduction) and temperature (78% variance reduction) are given in Fig. 2. Contours on an EOF field are non-dimensional and only indicate pattern. Not only are magnitudes unimportant but the signs are also ambiguous. The contours show relative changes. While it is tempting to try to give a physical interpretation to the individual EOFs, there is no more a priori reason to expect a physical basis for the first EOF than there would be to expect one from the first Fourier component.

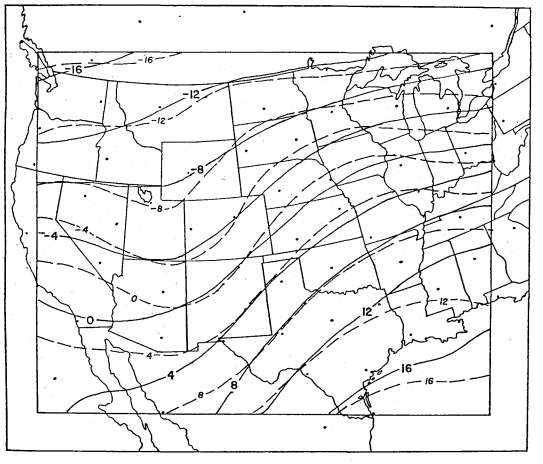

Fig. 2 Patterns of the primary EOFs for 500 mb geopotential (solid) and temperature (dashed). Isopleths are labelled in percent and are non-dimensional, with ambiguous sign.

The actual meteorological pattern is derived from the EOF via multiplication by a number (positive or negative) which is a constant for a given case $(n=j)$ but can vary from case to case. Because the principal EOF accounts for a large majority of the variance in this study, it does bear a marked similarity to the field obtained by compositing the data (Fig. 1).

3. EOF combinations to produce map types

Even though an individual EOF does not necessarily have a physical meaning, the set of EOFs spans the observations and forms an efficient basis for orthogonal expansion of the observed data. Many purely objective techniques for combining EOFs into rotated components (e.g., Wash and Richman, 1981) have been proposed. However, these methods all suffer from lack of a comprehensive statistical theory dictating a rationale for rotation. No fully satisfactory objective criteria exist for determining the optimum rotation.

A physically based, quasi-subjective method of combining EOFs based upon the correlation (R_{ij}) between data from a particular case and an individual EOF $(\underline{E}_j)$ is proposed. R_{ij} is simply an inner product (dot product) of the j-th column of the $\underline{Q}$ matrix and the i-th EOF, divided by the magnitude of the column (M_j). The correlations are also the case-dependent constants needed to perform an EOF expansion of the original data (F_j)

$$\underline{F}_j = M_j \sum_{1}^{p} R_{ij}\underline{E}_i + \underline{A}_j . \qquad (1)$$

Table 1 shows the correlations obtained for the first 3 EOFs of the 500 mb geopotential height data. The first 3 EOFs explain 95% of the variance in the 500 mb geopotential height field. Thus, a small and known amount of mean information (5%) would be lost by

Table 1 Correlation between first three three EOFs and observed data for 500 mb geopotential height.

Case (j)	Date	EOF 1 Correlation-R_{1j} (80% Variance Reduction)	EOF 2 Correlation-R_{2j} (10% Variance Reduction)	EOF 3 Correlation-R_{3j} (5% Variance Reduction)
1	12Z 25 Mar 59	.87	−.05	.09
2	12Z 4 May 61	.95	−.09	−.19
3	12Z 21 May 61	.77	.53	−.24
4	12Z 5 Jun 61	.95	.06	−.02
5	12Z 27 May 62	.80	−.50	−.26
6	12Z 26 May 63	.96	.13	.04
7	12Z 13 May 65	.93	.06	−.25
8	12Z 4 May 60	.92	−.26	.21
9	12Z 29 May 59	.86	−.25	.20
10	12Z 30 Apr 61	.72	.68	−.04
11	00Z 8 Apr 65	.96	.08	.03
12	12Z 14 Apr 65	.91	.04	.38
13	12Z 8 Jun 66	.97	−.12	−.16
14	12Z 27 Dec 68	.96	−.14	−.14

truncating the expansion after only three terms.

Examination of the correlations reveals an essentially bimodal distribution of R_{1j}. The correlation between the first EOF of 500 mb geopotential height and the observed data is significantly different for cases 3, 5 and 10 than for the other 11 cases. Effectively, the distribution of R_{1j} sorts the cases into two types. The primary category (11 cases) has a mean first EOF correlation of 0.93 with a standard deviation of 0.04. The other category has a mean $R_{1j}(\bar{R}_{1j})$ of 0.76 with a 0.04 standard deviation.

Separating the second correlation coefficients (R_{2j}) into the two categories shows that the correlations within the primary category are not clustered. For the prime category $\bar{R}_{2j}$ is −0.05 with a standard deviation of 0.13. Even though the second broad category only consists of three cases, R_{2j} indicates that it should be subdivided. Cases 3 and 10 have second correlations of 0.53 and 0.68, respectively, while case 5 has one of −0.50. Accordingly, this category is further subdivided, with R_{1j} being recomputed for the newly created divisions. Further divisions on the third correlation are possible, but since the associated EOF explains only 5% of the sample variance, this step was deemed unnecessary.

Based upon this examination of the correlations, the 500 mb geopotential height field associated with progressive tornado outbreaks over the Kansas-Oklahoma region typically fits into one of three categories (map types). Similar typing can be performed for other meteorological fields. 500 mb temperature sorts into the exact same categories as geopotential height. Since these fields are classed independently, this congruence implies that a physical consistency exists in the typing. However, typing by 500 mb characteristics does not insure vertical consistency. When the correlations for other levels are examined, different categorizations are implied.

In order to reconstruct composite charts showing the synoptic situation, the categorization determined by the 500 mb geopotential heights is used for all fields. Category-averaged values for the constants in equation 1 (Table 2) are computed. Mean fields are then constructed for each category using the first three EOFs. Since the charts are built from an orthogonal expansion, considerably more detail is preserved than in the mean charts. If higher order EOFs were used, the reconstructed fields would have even more detail.

Southwest flow is the normal 500 mb pattern associated with progressive tornado outbreaks in Kansas-Oklahoma (Fig. 3a). It exhibits a slightly negatively tilted trough

Table 2 Constants used to construct MAP types.

	850 mb Geopotential Height	850 mb Temperature	850 mb Dewpoint Temperature	500 mb Geopotential Height	500 mb Temperature
Southwest Flow (11 cases)					
$\overline{A}_j$	1476.2	9.8	5.3	5687	−16.2
$\overline{M}_j$	337.6	54.3	20.4	934.9	42.3
$\overline{R}_{1j}$	−.762	.851	.831	.933	.907
$\overline{R}_{2j}$	−.072	.063	−.062	−.050	−.042
$\overline{R}_{3j}$	.060	−.032	.032	.017	−.018
Northwest Flow (2 cases)					
$\overline{A}_j$	1494.2	11.3	5.4	5719	−15.5
$\overline{M}_j$	252.0	61.2	23.1	1057.5	46.5
$\overline{R}_{1j}$	−.612	.802	.897	.747	.799
$\overline{R}_{2j}$	.541	.482	.246	.604	.546
$\overline{R}_{3j}$	−.174	.173	.190	−.143	−.044
Deep Cyclone (1 case)					
$\overline{A}_j$	1469.8	11.0	6.0	5702	−15.3
$\overline{M}_j$	346.9	54.0	18.1	983.9	35.1
$\overline{R}_{1j}$	−.791	.882	.846	.805	.713
$\overline{R}_{2j}$	−.557	−.195	−.110	−.500	−.425
$\overline{R}_{3j}$	.097	−.349	−.252	−.263	−.446

running from western Arizona to western Washington. As in the composite analysis, there is a thermal trough lying slightly east of the geopotential trough. This allows cold air advection to precede the geopotential trough and to destabilize the atmosphere. At low levels (Fig. 3b), a closed low lies along the Rocky Mountains. Generally, strong south-southwesterly geostrophic flow prevails over the tornado prone area. A sharp

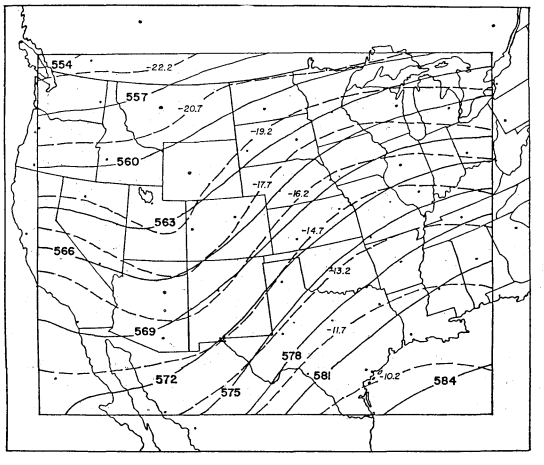

Fig. 3a Reconstructed 500 mb fields, as described in text, for the "southwest flow" pattern. Contours and isotherms are labelled and denoted as Fig. 1.

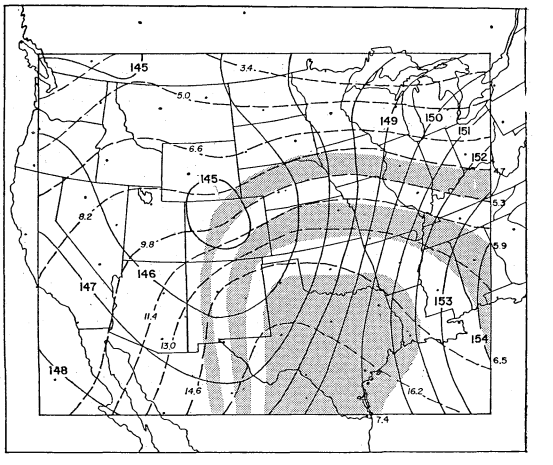

Fig. 3b Reconstructed 850 mb fields for the "southwest flow" pattern. Contours and isotherms are as in Fig. 1, while mixing ratio isopleths are suggested by the stippling and labelled (g/kg) outside the analysis grid on the right and lower margins.

moisture gradient is present in eastern New Mexico, while an axis of high temperatures extends northwestward from Texas and is positioned slightly west of the high dew point axis. This overall structure is consistent with both Miller's (1972) type A and type B synoptic pattern.

The second 500 mb flow pattern exhibits northwest flow over the area of interest (Fig. 4a). This flow is the result of a weak ridge over the lee slopes of the Rockies. Weak geopotential and thermal troughs are apparent to the west of this height ridge and cold air advection again prevails. At low levels (Fig. 4b) an open geopotential trough associated with sharp temperature and moisture gradients is present over the Texas panhandle. These conditions indicate the presence of a warm front positioned across southern Kansas, and are very reminiscent to those described by Johns (1984) as being associated with northwest flow outbreaks of severe weather.

Since the third category contains only one case, no chart is presented. Basically, on this day, a deep closed 500 mb cyclone existed over California. A short wave moving northward in this flow was associated with the tornadic activity. The 850 mb chart was quite similar to the mean chart for southwest flow (Fig. 3b).

4. Potential applications of map typing

Once average correlations for each category are obtained, they and their associated EOFs can be used to determine if any independent meteorological data set belongs to one of the map types. This is accomplished by first computing the correlations between the EOFs and the data in question and then comparing them to the category average values. To test this concept, correlations have been calculated for each 12 GMT data set from May 1961. This month was chosen because it contains two of the outbreak days (one southwest flow, one northwest flow). By requiring the first three correlations of both the 500 mb geopotential height and temperature to lie within two standard deviations of the southwest flow category average, four potential days were found. Even though only one of these had a progressive outbreak over the area of interest, severe thunderstorms and tornadoes occurred over the Kansas-Oklahoma region on all four days. Apparently, the southwest flow 500 mb map type occurs infrequently but is typically associated with very strong convective storms. This is consistent with the concept that the "textbook" synoptic conditions associated with tornadoes occur only a few times a year (Maddox and Doswell, 1982).

While examining the first correlations between the fourteen individual cases and the EOFs for all 40 fields, it was noted that the coefficients for five parameters classically associated with severe convective weather are quite stable on outbreak days. These five

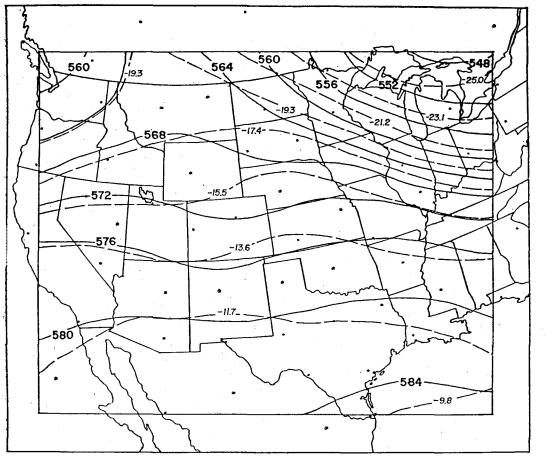

Fig. 4a As in Fig. 3a, for the "northwest flow" pattern.

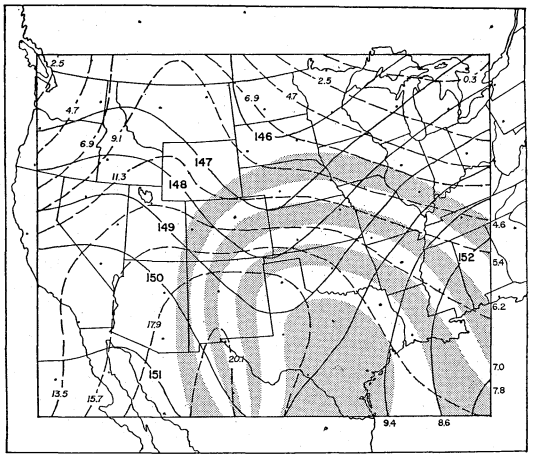

Fig. 4b As in Fig. 3b, for the "northwest flow" pattern.

Table 3 First correlation mean and standard deviation for selected parameters of tornado outbreak days.

Parameter	Level	Mean first correlation	Standard deviation
V wind Component	Surface	0.67(R_{V_s})	0.14
Dewpoint Temperature	850 mb	0.84(R_{D_8})	0.08
Geopotential Height	500 mb	0.90(R_{Z_5})	0.07
Dewpoint Temperature	500 mb	0.58(R_{D_5})	0.22
U Wind Component	250 mb	−0.60(R_{U_2})	0.27

parameters and the mean and standard deviation of the first correlation are given in Table 3. This suggested that a simple index, D, given by

$$D=R_{V_s}+R_{D_8}+R_{Z_5}+R_{D_5}-R_{U_2} \qquad (2)$$

(the notation is defined in Table 3) could be a useful prognostic tool.

To test this concept, D was calculated from the 12 GMT rawinsonde data for each day in a randomly selected month (May 1961). For the 31 days, the average value of D was 2.73, with a standard deviation of 1.07. However, for the 19 days during that month on which severe weather occurred over the Kansas-Oklahoma region D averaged 3.32 with a standard deviation of 0.50, while on days with no severe convection D averaged 1.84 with a standard deviation of 1.03. If a simple threshold for D of 2.8 is used, the index for this partially dependent data set discriminates quite well between days with and without severe convective activity. Eight-four percent of the severe weather days were properly predetected, while only 11% of the above-threshold days were false alarms. These combine to give a critical success index (Donaldson *et al.*, 1975) of 0.76.

5. Extensions

Even though both of the previous exercises are based upon limited data samples and are by no means conclusive, they serve to highlight potential operational applications of the map typing technique developed. The map types can be used to highlight days of very high tornado potential. Since the map types show mean tornado conditions, they can also be used to give an indication of how typical a given situation is.

Much further work needs to be done. Conditions attendant to local and line tornado outbreaks must be examined and compared to conditions found here. Also, a smaller grid spacing must be used so that more detail is preserved in the initial objective analysis. It is anticipated that this detail will require the inclusion of more than 3 EOFs in the typing process. These EOFs can then be used in multiple regression formulas (Mori, 1984) to forecast tornado occurrence.

Another modification of the present work entails expanding the $\underline{Q}$ matrix to include more than one data type and more than one data type and more than one pressure surface. As noted by Kutzbach (1967) using such a multi-field data matrix forces a consistency in EOF fields. Further, the physical connection between the meteorological data as represented by the EOF expansions and observed weather needs elucidation.

Another extension of this work involves incorporation of a movable grid so that tornadic conditions in other geographic areas can be studied. The grid should be on a pseudo-Mercator projection so that metric variations do not contaminate the analysis. This movable grid will also serve to increase the number of tornado outbreak cases available for analysis. After elimination of the geographic dependence, it will be possible to use the developed map types to classify historic rawinsonde data for different areas of the country. A regionalization of tornado risk will be developed using meteorological reasoning as simple damage assessments.

Acknowledgements

This work was performed while the authors were with the National Severe Storms Forecast Center. We would like to express our gratitude to Dr. R.P. McNulty now at WSFO Topeka and to Dr. H.R. Glahn of TDL for their interest and help at various stages of this work. The eigenvector computer algorithms was supplied by Mr. H. Woolf of NESS

Development Laboratory, Madison. Typing is by Mrs. Beverly Lambert and Mrs. Dorothy L. Babich. This research was performed under Interagency Agreement RES-76-106 between the Office of Nuclear Regulatory Research, Nuclear Regulatory Commission and the National Severe Storms Forecast Center, National Weather Service, NOAA.

References

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.*, **3**, 396-409.

Beeber, R. G., 1956: Tornado composite charts. *Mon. Wea. Rev.*, **84**, 127-142.

Buell, C. E., 1978: The number of significant proper functions of two-dimensional fields. *J. Appl.. Meteor.*, **17**, 717-722.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. of Meteor.*, **4**, 135-162.

Donaldson, R. J., R. M. Dyer and M. J. Kraus, 1975: An objective evaluator of techniques for predicting severe weather events. Preprints, 9th Conf. on Severe Local Storms, Amer. Meteor. Soc., Norman, OK, 321-326.

Galway, J. G., 1977: Some climatological aspects of tornado outbreaks. *Mon. Wea. Rev.*, **105**, 477-484.

Grimmer, M., 1963: The space-filtering of monthly surface temperature anomaly data in terms of pattern, using empirical orthogonal functions. *Quart. J. of the Roy. Meteor., Soc.*, **89**, 395-408.

Holmström, I., 1963: On a method for parametric representation of the state of the atmosphere. *Tellus*, **15**, 127-144.

Johns, R. H., 1982: A synoptic climatology of northwest severe weather outbreaks Part II: Meteorological parameters and synoptic patterns. *Mon. Wea. Rev.*, **112**, 449-464.

Kutzbach, J. E., 1967: Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North Anerica. *J. Appl. Meteor.*, **6**, 791-802.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Sci. Rept. No. 1, Statistical Forecasting Proj. Dept. of Meteor., MIT Cambridge, 44 pp.

Maddox, R. A. and C. A. Doswell III, 1982: An examination of jetstream configurations, 500 mb vorticity advection and low level thermal advection patterns during extended periods of intense convection. *Mon. Wea. Rev.*, **110**, 184-197.

Miller, R. C., 1972: Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central. AWS Tech. Rep. 200 (Rev.), Headquarters Air Weather Service, Scott AFB, IL.

Moeller, A. R., 1979: The climatology and synoptic meteorology of Southern Plains' tornado outbreaks. M. S. Thesis, U. of Oklahoma, Norman, 70 pp.

Mori, N., 1984: 192-hour prognosis of precipitation by using numerical-prediction products: Application of EOF to multiple-regression formula. *J. Meteor. Soc. Japan*, **62**, 183-189.

Preisendorfer, R. W. and T. P. Barnett, 1977: Significance tests for empirical orthogonal functions. Preprints, 5th Conf. on Probability and Statistics in Atmospheric Science, Amer. Meteor. Soc., Las Vegas, NV, 169-172.

Richman, M. B., 1980: Map typing associated with potential urban enhanced precipitation. M. S. Thesis, U. of Illinois, Urbana, 123 pp.

Wash, J. E. and M. B. Richman, 1981: Seasonality in the associations between surface temperatures over the United States and the Pacific Ocean. *Mon. Wea. Rev.*, **109**, 767-783.

トルネード発生日の総観場の EOF 解析

J. T. Schaefer

Central Region Headquarters, NWS, NOAA

C. A. Doswell, III

Environmental Research Lab., NOAA

トルネード発生日の総観場を EOF 解析し，それにもとずいて composite charts を使用したトルネード発生の統計予報の可能性を示した。