

NOTES AND CORRESPONDENCE

**Empirical Orthogonal Function Expansion Applied
to Progressive Tornado Outbreaks**

Joseph T. Schaefer and Charles A. Doswell, III

REPRINTED FROM THE JOURNAL OF THE METEOROLOGICAL SOCIETY OF JAPAN
Vol. 62, No. 6, December 28, 1984

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 Orthogonal 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 progres-

sive 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.

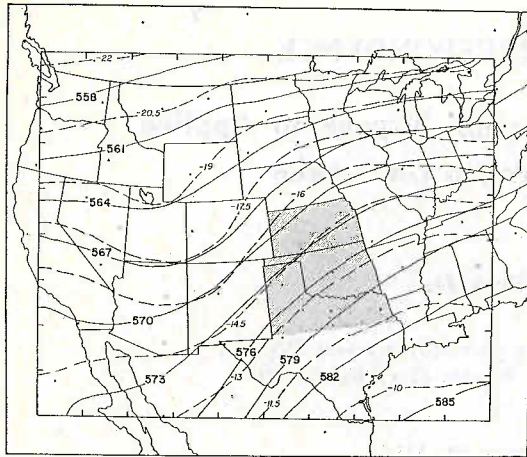


Fig. 1 Mean 500 mb geopotential and thermal fields for 14 progressive outbreaks. Contours (solid) are labelled in decameters, isotherms (dashed) in $^{\circ}\text{C}$. All outbreaks are contained in the geographic region indicated by stippling. Small, unlabelled dots locate the current operational network 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

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

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 successively 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 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 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 Q . For each case, the deviation vector is calculated by subtracting the field mean \bar{A}_f from the observation vector (F_f). 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

3). If the spatial data dis- examined, the EOFs are the temporal variation of the or temporal distributions, are examined). The first which, in the mean, makes tion to the spatial analysis on time (i.e., it explains Subsequent EOFs make er contributions.

xpansion is similar to other thogonal functions (e.g., etc.), it has several ad- the amount of explained EOF is known, the mean g the expansion is also the EOFs are determined tion, boundary effects are d for by the mathematics o not have to be examined it can be shown that an uires the fewest terms of ansion to represent a given (Holmström, 1963).

er the data set as a $p \times n$ F the n columns is a vector berved data (ordered by ific case and each of the ta from a specific location Thus, there are p observa- The EOFs are the eigen- iance matrix formed from e associated eigenvalue to eigenvalues gives the per- ed by any one EOF. It rable to use deviations as aw observations when de- a case, the deviation vector racting the field mean A_j on vector (F_j). Derivation can be found in the litera- r, 1963).

ive tornado outbreak study, a study involving measured al the number of observa- row is uniquely associated from one site. However, when the data span many release locations change, ns released vary from case

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 60°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 $(n-3)$ is less than p , typically there will only be $(n-3)$ 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.

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 (E_j) is proposed. R_{ij} is simply an inner product (dot product) of the j -th column of the *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)

$$F_j = M_j \sum_1^p R_{ij} E_i + A_j. \quad (1)$$

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

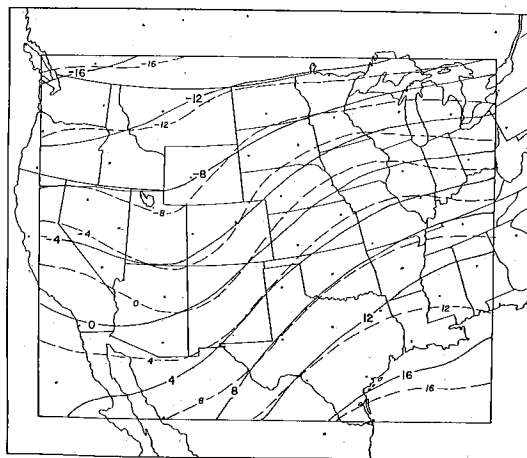


Fig. 2 Patterns of the primary EOFs for 500 mb geopotential (solid) and temperature (dashed). Isopleths are labelled in percent and are non-dimensional, with ambiguous sign.

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 correla-

tions, 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. 500mb 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.

$-R_{2j}$ (Correlation)	EOF 3 (5% Variance Reduction)	$-R_{3j}$ (Correlation)
	.09	
	-.19	
	-.24	
	-.02	
	-.26	
	.04	
	-.25	
	.21	
	.20	
	-.04	
	.03	
	.38	
	-.16	
	-.14	

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

500 mb geopotential height field progressive tornado outbreaks in the Oklahoma region typically fall into three categories (map types). This analysis can be performed for other map types. 500 mb temperature sorts into three categories as geopotential height fields. These fields are classed independently. This congruence implies that a consistency exists in the typing. This analysis by 500 mb characteristics shows vertical consistency. When other levels are examined, similar relationships are implied.

To reconstruct composite charts for the southwest flow situation, the categorization is done by the 500 mb geopotential height field for all fields. Category-specific constants in equation (1) are computed. Mean fields are reconstructed for each category using the constants. Since the charts are built from the normal expansion, considerably more detail is preserved than in the mean fields. If higher order EOFs were used, the reconstructed fields would have even more detail.

Figure 3a shows the normal 500 mb pattern with progressive tornado outbreaks in the Oklahoma region (Fig. 3a). It exhibits a negatively tilted trough

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

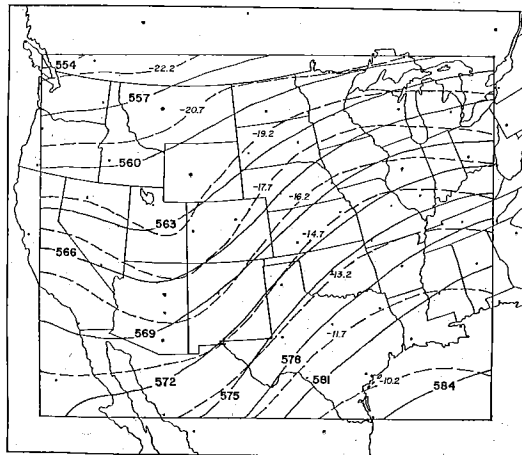


Fig. 3a Reconstructed 500 mb fields, as described in text, for the "southwest flow" pattern. Contours and isotherms are labelled and denoted as Fig. 1.

