A Statistical–Climatological Tropical Cyclone Track Prediction Technique Using an EOF Representation of the Synoptic Forcing

ALAN R. SHAFFER AND RUSSELL L. ELSBERRY

Department of Meteorology, Naval Postgraduate School, Monterey, CA 93940

(Manuscript received 15 April 1982, in final form 13 August 1982)

ABSTRACT

An empirical orthogonal function (EOF) analysis is performed on a series of 504 normalized *D*-value fields at 500 mb. Each field is defined on a 120-point grid that is centered on the geographical position of a tropical cyclone. The EOF representation (eigenvectors) of the environmental flow field requires less than 10% of the storage needed for the original grid points. Each eigenvector represents a distinctly different synoptic-scale forcing pattern for the tropical cyclone motion.

Statistical regression equations are developed to predict the zonal and meridional displacements of the tropical cyclone at 12 h intervals to 84 h. The EOF coefficients are used to represent the synoptic-scale forcing in the equations. The track forecast errors are competitive with other statistical schemes. The average displacement errors in an independent sample were ~15% smaller than the official Joint Typhoon Warning Center forecasts. Thus, the EOF-based regression approach provides a simple and low cost technique for predicting tropical cyclone motion.

1. Introduction

Over the years, a number of statistical–climatological techniques have been developed to forecast tropical cyclone tracks (see review by Neumann and Pelissier, 1981). The original techniques (e.g., Riehl *et al.*, 1956) used latitudinal and longitudinal differences of geopotential values to estimate the steering components of the storm. Past motion of the storm is also typically included as a predictor in these statistical techniques. For example, the eastern North Pacific climatology and persistence model of Leftwich and Neumann (1977) uses only the motion of the storm during the past 24 h to estimate the future displacement.

Other statistical track forecast techniques have made direct use of the synoptic fields in the environment of the tropical cyclone. A typical grid arrangement used for this purpose (Neumann, 1979) is illustrated in Fig. 1. The grid is relocated each forecast period so that the storm center is always at the same gridpoint. Values of sea-level pressure or geopotential at upper levels are then extracted at each gridpoint from the appropriate analyses. These gridpoint values are entered as potential predictors in the statistical regression equations for the meridional and zonal displacement of the storm (e.g., Veigas *et al.*, 1959; Miller *et al.*, 1968). Although this approach incorporates synoptic forcing of the tropical cyclone motion, it is difficult to physically interpret the effect the individual gridpoint values have on the prediction. In the data-sparse tropics, there is also likely to be considerable uncertainty as to the actual value, regardless of whether the analysis is done subjectively or objectively. Neumann and Pelissier (1981) have attributed some of the deterioration in the performance of the statistical techniques to changes in the objective analysis schemes that are used to provide the gridpoint data.

The incorporation of synoptic-scale data into statistical forecast techniques has not been limited to just the analyses at the forecast time. Neumann and Lawrence (1975) have included the numerically predicted geopotential values on the grid as potential predictors. This technique (NHC73) consistently has been one of the best methods in the Atlantic region (Neumann and Pelissier, 1981). More recently, the dynamical models have produced competitive forecasts (Elsberry, 1979). These numerical models explicitly include the interaction between the tropical cyclone and its environment. Nested grid models have been recently applied (Harrison, 1981) to improve the resolution near the cyclone center. These dynamical models are more complex and expensive to run than are the statistical techniques.

Some of the analogue-type prediction schemes have also included a representation of the synoptic-scale field. Jarrell and Somervell (1970) used the latitude of the subtropical ridge and the longitude of the midlatitude trough as similarity factors to select potential analogues. In this technique and the others discussed above, the synoptic field has not been represented completely because of the amount of data (number of gridpoints) that would be required. One

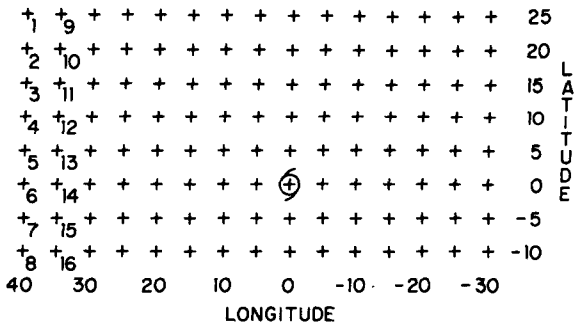

FIG. 1. Grid arrangement used in statistical–climatological track prediction after Neumann (1979). *D*-values at 500 mb were extracted relative to the storm center which is always at gridpoint 70. The latitude and longitude intervals are 5°.

then is reduced to using a few selected points from within the field, or to a characterization of the field such as the latitude of the ridge line. The requirement for a representation of the entire synoptic field has been highlighted by the recent work of Chan *et al.* (1980). They suggest that major changes in the direction of tropical cyclone motion may be anticipated from the environmental flow fields existing 24–36 h previously.

The purposes of this research are two-fold. First, we demonstrate that empirical orthogonal functions (EOF) may be used to represent the synoptic-scale flow on a grid centered on the tropical cyclone. The EOF approach reduces by an order of magnitude the amount of storage necessary to represent the synoptic forcing. Second, we use the coefficients from the EOF analysis as synoptic forcing predictors in a statistical–climatological track prediction scheme. These coefficients represent the entire environmental flow field surrounding the tropical cyclone. When combined with predictors describing the recent past motion of the storm, the EOF statistical–climatological technique shows promise for a simple, low cost and efficient track prediction technique.

The original motivation for this work arose from the results of Woodcock (1980), who demonstrated that regression equation sets derived separately for different synoptic situations provided better predictions than did a single set that did not distinguish among synoptic situations. Because tropical cyclone motion is so closely related to the environmental synoptic-scale forcing, it was thought that separate track prediction regression equations might be derived for different synoptic situations. In a first attempt, Brown (1981) used a map typing approach to characterize the synoptic-scale flow on the grid in Fig. 1. Brown introduced the EOF analysis as a method to determine similar map types. He was able to demonstrate that the official track forecast errors were different for different map types. However, his attempt to use an analogue-type track prediction technique based on similarity in the EOF coefficients did not produce uniformly good results in each map type.

In this paper, we first describe the environmental fields and the EOF representation. The relation of the EOF coefficients to different types of storm tracks is then demonstrated. This relationship is explored further within the context of regression analysis for track prediction. Finally, some results from this simple track prediction method will be presented.

2. Environmental fields

Nearly all of the previous applications of EOF analysis have been on fixed geographical grids. In this case, we want to represent the environmental flow field on a grid (Fig. 1) that is displaced with the storm. Northern Hemisphere analyses of the 500 mb heights are available at 12 h intervals from the Fleet Numerical Oceanography Center (FNOC) archives. These heights are converted to *D*-values by subtracting the standard atmosphere height. The 700 and 850 mb height analyses were also tested, but these levels produced similar results and are not described here. One might expect that a deep layer mean height would produce a more consistent result than any of the cases using a single level. A Bessel linear interpolation scheme was used to extract the gridpoint values in Fig. 1 from the FNOC 63 × 63 hemispheric analyses.

Several criteria were required for a tropical cyclone to be included in the sample. Only cyclones located between 10 and 25°N, and west of 180° were included. The southern boundary was selected to insure that the grid did not extend into the Southern Hemisphere. The northern boundary was chosen to minimize possible distortions associated with compositing on a latitude–longitude grid over a wide range of storm latitudes. The tropical cyclone had to be of at least tropical storm intensity and have persisted at least 30 h. Subsequent forecasts for the same storm were required to be separated by at least 36 h to assure independence in forecast situations. A total of 504 cases met the selection criteria during the 1967–76 period for which the FNOC geopotential analyses were available.

Warning positions and intensities during the prior 36 h were extracted for each storm from the *Annual Typhoon Reports* of the Joint Typhoon Warning Center (JTWC), Guam, Mariannas. These warning positions represent the information available to the forecaster in an operational situation. However, the predictions of the storm motion are compared with the "best-track" positions based on post-season analysis. The maximum forecast interval is 84 h to allow for the delay of 3–4 h after synoptic time that is necessary to receive and analyze the upper air data, and yet issue forecasts that will be valid for at least 72 h. Only 185 of the 504 cases had a complete history from the previous 36 to 84 h following forecast time.

The composite D-value field at 500 mb using all 504 cases is shown in Fig. 2a. As expected for tropical regions, all of the D-values are positive. The major features are the low values in the region of the tropical cyclone, and the strong height gradients to the north that are associated with the midlatitude westerlies. At the 700 and 850 mb levels (not shown), the circulation associated with the tropical cyclone is more intense, whereas the midlatitude gradients are diminished. The standard deviations of the D-values about the mean values in Fig. 2a are shown in Fig. 2b. There is a secondary maximum in the region of the tropical cyclone. However, the primary feature in Fig. 2b is the seven-fold increase in D-value fluctuations from the tropical regions to the northern boundary of the grid. These means and standard deviations are used to normalize the D-value fields prior to the EOF analysis. After normalization, the master set of 504 values will have the same variance at every gridpoint. Thus, each gridpoint will contribute equally to the EOF analysis. Without normalization, the greater variance associated with the midlatitude gridpoints would dominate the analysis.

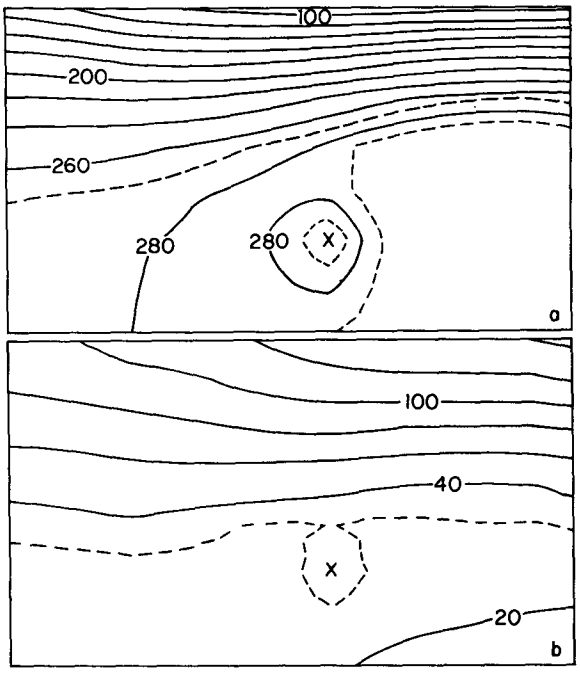

FIG. 2. (a) Mean D-value (m) field at 500 mb on the grid in Fig. 1. (b) standard deviations of the 500 mb D-values (m). The interval in these diagrams is 20 m, except near the tropical cyclone where the interval is 10 m (dashed).

3. Empirical orthogonal function analysis

The terminology "empirical orthogonal function" was introduced by Lorenz (1956), who made the first applications to atmospheric sciences. Since that time, EOF analysis has been widely used as a tool for explaining the variance of geophysical fields in terms of a minimum number of orthogonal components. For example, Rinne and Karhila (1979) have demonstrated that EOF analysis can efficiently represent the planetary and synoptic scale features over the entire hemisphere.

The mechanics of the EOF analysis presented here follows an elegant treatment by Kutzbach (1967). A matrix $\mathbf{A}$ is constructed from the normalized D-values. Each column of $\mathbf{A}$ consists of the M (=120) gridpoint observations for a particular case, and each row contains the sequence of N (=504) values at a single gridpoint. In the simplest sense, the objective of EOF analysis is to determine a vector $\mathbf{e}$ in M dimensions that best represents all of the N observation vectors. That is, we seek to maximize the quantity

$$(\mathbf{e}'\mathbf{A})N^{-1}/\mathbf{e}'\mathbf{e} \tag{1}$$

under the constraint that $\mathbf{e}'\mathbf{e} = 1$. It is important to notice that the vector $\mathbf{e}$ is defined in terms of a particular set of observations. In a more general case of M dimensions, we seek to maximize the explained variance in each of the dimensions.

The empirical orthogonal functions are computed as eigenvectors from the correlation matrix which contains the spatial relationships between the 504 observed D-value fields on the grid in Fig. 1. Each eigenvector is representative of a particular spatial pattern for this set of observations. These functions, which are constrained to be orthogonal, may be used to represent the fields. For example, the nth case in $\mathbf{A}$ may be represented as

$$\mathbf{a}_n = \sum_{i=1}^{M} c_{in}\mathbf{e}_i, \quad n = 1, 2, \ldots, N, \tag{2}$$

which is a linear combination of the orthogonal coefficients (c_{in}) and the elements of the eigenvectors.

The first k eigenvectors ($k \ll M$) generally represent a large fraction of the total variance in $\mathbf{a}_n$. Thus, the nth case may be approximated as

$$\mathbf{a}_n = \sum_{i=1}^{k} c_{in}\mathbf{e}_i, \quad n = 1, 2, \ldots, N. \tag{3}$$

The eigenvalues for the 500 mb D-value fields in this study are listed in Table 1. One can see from the percent explained variance that the first eigenvector represents 33% of the variance, and the second eigenvector includes an additional 18%. As indicated by (2), inclusion of all 120 eigenvectors results in a complete representation of the field.

There are several methods for determining the optimal number of eigenvectors (k) to retain (e.g., Cattell, 1958; Morrison, 1967; Preisendorfer and Barnett, 1977). Rinne and Karhila (1979) used the logarithmic eigenvalue diagram method, which led them to retain a relatively large number of eigenvalues. Richman (1981) examines this question, and the use of rotation

TABLE 1. Eigenvalues and cumulative percent explained variance of the normalized D-value fields at 500 mb.

	Eigenvalue	Explained variance
1	39.86	33
2	21.60	51
3	9.29	59
4	7.43	65
5	6.08	70
6	5.29	75
7	4.00	78
8	3.13	81
9	2.49	83
10	2.11	85
...		
15	1.03	91
...		
20	0.58	94
...		
40	0.11	98
...		
60	0.04	99
...		
120	0.00	100

of the components to improve the representation of the fields. Here, we adopt the Monte Carlo selection method of Preisendorfer and Barnett (1977).

A series of 100 simulations of point-normalized random data (zero mean and standard deviation of 1.0) is used to calculate Monte Carlo eigenvalues from 120 × 504 element matrices. The eigenvalues from the observed data are compared with the random Monte Carlo eigenvalues in Fig. 3. The standard deviations of the simulated eigenvalues are not shown in Fig. 3 because they are relatively small. For modes 9, 10 and 11 these standard deviations are only 0.023. For modes in which the physical eigenvalues deviate from the simulated eigenvalues by more than two standard deviations, one is 95% confident that the field is significantly different from a random field. That is, one is reasonably sure that the eigenvector is describing signal rather than noise. The 10 eigenvectors suggested for retention by the Preisendorfer and Barnett (1977) method contrasts with a value of 40 modes that would be required by the Cattell (1958) 99% retention rule. These first 10 eigenvectors only explain ~85% of the variance (Table 1). The remaining 15% of the variance may be associated with short-wave features, and with observational and analysis errors. The latter errors may be important in the tropics where the standard deviations in the fields are rather small (Fig. 2b). At any rate, our objective is to relate the motion of tropical cyclones to the synoptic-scale features in the environment, which evidently are well represented by only the largest 10 of the 120 eigenvectors. This represents a reduction of more than 90% of the storage that would normally be required to represent the synoptic-scale forcing of the tropical cyclone.

The first nine eigenvectors are illustrated in Fig. 4. Each eigenvector actually represents the pattern shown and the inverse of the pattern. Perhaps the most important features to be obtained from these eigenvectors are the changes in spatial patterns with increasing mode number. Eigenvector 1 has a single high–low couplet with a large D-value gradient across the center of the domain. In Eigenvector 2, there is still only a single high–low couplet, but the gradient is considerably weaker than for Eigenvector 1, and is displaced farther north of the storm. These two modes account for more than 50% of the variance across all of the D-value fields. Eigenvectors 3 and 4 are dominated by large amplitude lows to the west and northeast of the tropical cyclone, respectively. There tend to be increasing numbers of centers as the number of modes increases. However, Eigenvector 5 contains only a high center just north of the storm location, with low values to the east and west. It is not difficult to associate these patterns with physically significant, synoptic-scale features. Higher eigenvectors appear to account for tilts in these features (e.g., 7 and 8) or phase shifts (compare Eigenvector 9 with 5).

An example of how these eigenvectors may be combined to represent an actual map is given in Fig. 5. As the input data have been normalized, the field for the jth case must be reconstructed using

$$d_{ij} = \bar{d}_i + s_i \sum_{n=1}^{10} c_{nj}\mathbf{e}_{in}, \quad i = 1, 2, \ldots, 120, \quad (4)$$

where $\bar{d}_i$ and s_i are the means and standard deviations at the 120 gridpoints (Fig. 2) and $\mathbf{e}_{in}$ is the ith component of the nth eigenvector. The case of Typhoon Marge on 0000 GMT, 27 August 1967 was selected to demonstrate the reconstruction for a particular value of j. Typhoon Marge was near 18°N, 125°E at

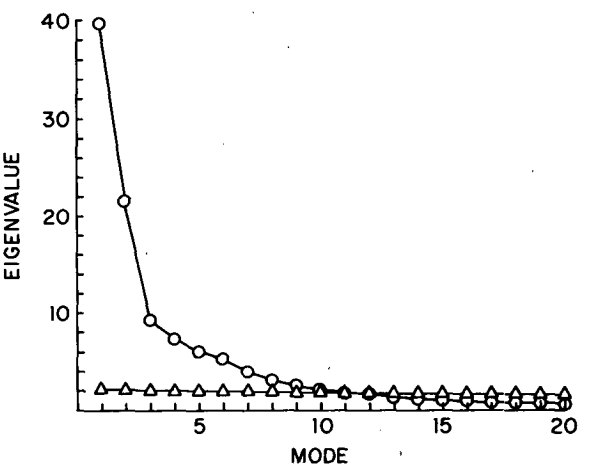

FIG. 3. Eigenvalues of the 500 mb D-value fields (circles) and the Monte Carlo generated eigenvalues (triangles) based on 100 simulations of random data (see text for explanation).

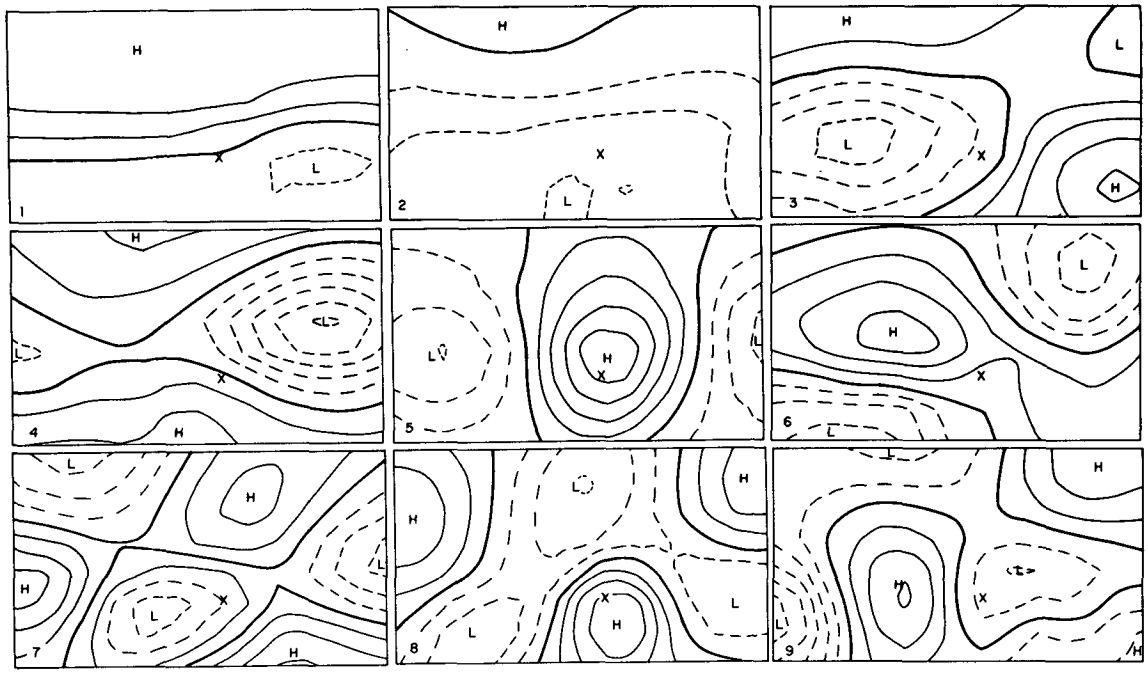

FIG. 4. Eigenvector elements multiplied by 100 on the grid in Fig. 1 for the first nine modes. The interval is five units with positive (negative) values indicated by solid (dashed) lines.

this time. Notice that the grid encompasses both tropical and extratropical features (Fig. 5a). The orthogonal coefficients associated with this case are listed in Table 2. The largest coefficients are associated with Eigenvectors 1 and 9, with the next largest contributions associated with Eigenvectors 2, 3 and 5.

Using only Eigenvector 1 in the reconstruction (Fig. 5b) produces the broad-scale features of a tropical cyclone and a belt of midlatitude westerlies separated by a ridge of high pressure. The high pressure center to the northeast of the storm is indicated fairly well. Neither the amplitude of the high pressure center to the northwest, nor the amplitude of the short wave to the north of the tropical cyclone, are well represented. Inclusion of Eigenvectors 2 through 4 (not shown) mainly results in the improvement of the representation of the midlatitude westerlies and in the position of the high center to the northeast. Inclusion of Eigenvector 5 (Fig. 5c) produces a better representation of the high center to the northwest and the midlatitude trough. The final diagram (Fig. 5d) indicates that the inclusion of only nine eigenvectors reconstructs the essential features of the original map (Fig. 5a). The intensity and shape of the flow around the northeastern high cell appears to be quite good. There is also an indication of a low center in the southwest corner, which did not appear with fewer modes. The major differences are in the location and intensity of the center to the northwest, and consequently also in the strength of the westerlies to the north of this high center. One has to include approximately 40 eigenvectors to represent the features in the northwest portion of the grid. Based on the Preisendorfer and Barnett (1977) selection rule, one would expect the contributions of these higher mode eigenvectors to contribute noise in most cases. Including a large number of terms would be contrary to the goal of representing efficiently the synoptic-scale forcing of the tropical cyclone motion.

4. EOF coefficients and storm motion

Although the 10 EOF coefficients and the associated eigenvectors provide an efficient representation of the environmental flow fields, it is essential also that these coefficients be related to the *subsequent* storm motion. To examine this relationship, the magnitudes of the orthogonal coefficients were correlated with the meridional and zonal displacements (Table 3) for 454 cases at 24 h, and fewer cases at longer time intervals. Only the coefficients at 24, 48 and 72 h are shown, as all of the other correlations at 12-h intervals to 84 h were consistent with the magnitude and time variations shown in Table 3. The 95% significance levels for the appropriate sample sizes at 24-, 48- and 72-h are 0.10, 0.11 and 0.14, respectively. The most striking feature in Table 3 is that certain eigenvectors are more highly correlated with zonal motion, while other eigenvectors are related to meridional motion. For the zonal case, the correlations decrease in time in a systematic manner. Eigenvectors 1 and 5 are correlated in a positive sense with the zonal component, which indicates a tendency for motion toward the west when these coefficients are

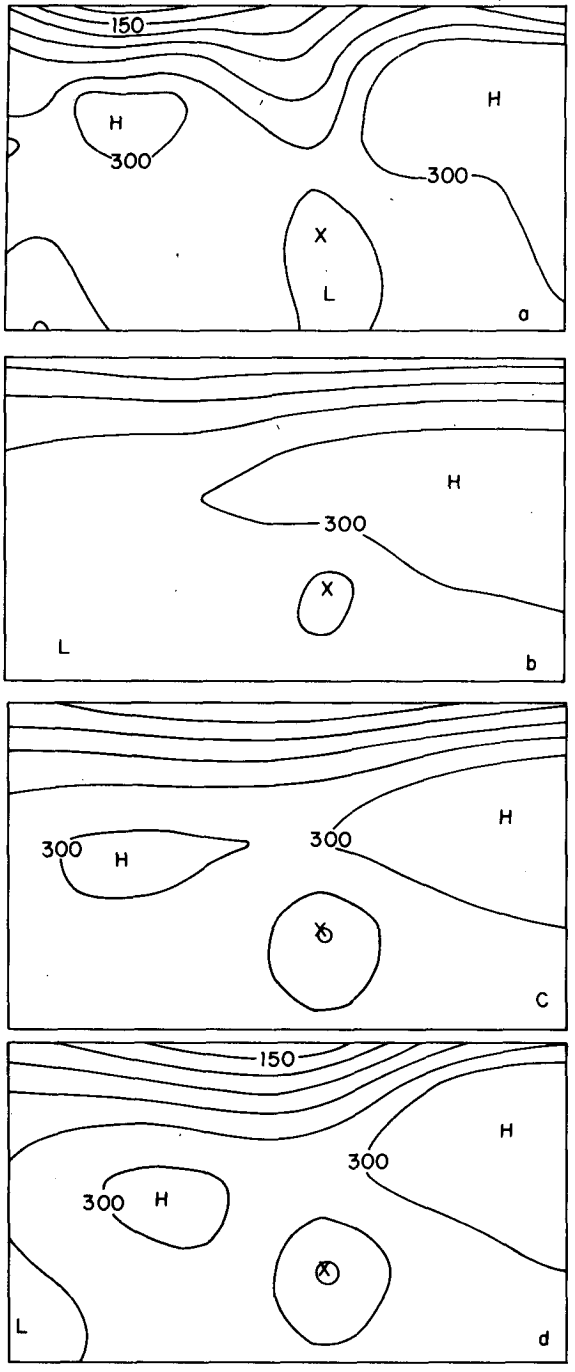

FIG. 5. (a) Original 500 mb *D*-value (m) field relative to typhoon Marge located at 18°N, 125°E on 0000 GMT, 27 August 1967. Reconstructed *D*-value fields using: (b) only the first; (c) the first five; and (d) the first nine eigenvectors and orthogonal coefficients. The contour interval is 30 m.

positive. Eigenvectors 4 and 8 contribute in the opposite sense. It is also clear that Eigenvectors 2, 3, 6, 7 and 10 are not related to the subsequent zonal motion of the storm. However, these eigenvectors have the largest correlations with the meridional motion. Eigenvectors 3 and 7 are correlated in a positive sense (northward motion), whereas the correlations are negative for Eigenvectors 2, 6 and 10. There is more time variation in the meridional motion correlations. Eigenvector 1 contributes to a southward component at 24 h and to a northward component at 72 h. The maximum correlations for Eigenvectors 9 and 10 occur at 72 and 48 h, respectively. Thus, the EOF coefficients have a physical significance in that they appear to be related to different types of synoptic forcing of the tropical cyclone motion.

Another demonstration of the relationship between the EOF coefficients and storm motion is given in Fig. 6. The storm tracks that are associated with the Eigenvector 1 coefficient being large and positive are contrasted with those with a large negative coefficient. This variation in coefficient 1 tends to distinguish between tracks that are predominately toward the northwest versus those toward the northeast. One can also see the potential use for analog-type forecasting, which has been another application of the EOF analysis. This distinction between types of tracks provides additional confidence that the EOF coefficients represent the synoptic-scale forcing of the tropical cyclone motion.

5. Regression analysis

The correlations in Table 3 of the EOF's with the subsequent storm motion suggest the possibility of using the EOF eigenvectors as predictors in a statistical model. The storm data base was divided into a dependent sample of 454 cases (of which 150 had a complete history from 36 h prior to 84 h after the forecast time) and 50 independent cases (35 with complete history). The selection of cases was done randomly. As there is at least 36 h between subsequent positions of the same storm, it is assumed that the cases are statistically independent.

The 14 predictands for this study are the 12-h zonal and meridional displacements out to 84 h. Positive motion is defined to be to the north and to the west, since the majority of the storm tracks are to the north-

TABLE 2. Orthogonal coefficients and correlation of the actual field on 27 August 1967 with the reconstructed field using the indicated number of eigenvectors (see text).

Number	Coefficient	Correlation
1	5.94	0.62
2	1.50	0.58
3	−1.70	0.66
4	−0.82	0.74
5	−1.85	0.75
6	−1.03	0.76
7	−0.75	0.73
8	0.26	0.73
9	2.56	0.88
10	−0.38	0.87

west. The potential predictors which are listed in Table 4 were selected to include the EOF representation of the synoptic forcing as well as indicators of the recent storm motion. Thus the 10 EOF coefficients are included along with the past 36-h motion and intensity data that would be available to the forecaster. A stepwise screening regression procedure (Dixon and Brown, 1979) is used with an F-ratio of 4.0.

An abbreviated summary of the statistical characteristics of the dependent sample and the explained variance of the regression equations is provided in Table 5. As noted above, the positive values for the mean displacements correspond to a northwesterly track. The standard deviations about the mean zonal displacements are much larger than for the meridional displacements. These standard deviations increase systematically with forecast interval. The regression equations explain nearly 75% of the variance in the 24-h zonal motion, although the amount of explained variance decreases with increasing time. Consistently smaller amounts of the variance in the meridional motion are explained. Considering the increase in standard deviations with forecast interval, we expect the regression equation forecast errors to increase rapidly with time. A separate set of regression equations was derived without including the EOF coefficients as predictors. As indicated in Table 5, the

TABLE 3. Correlations between the orthogonal coefficient associated with each eigenvector and the zonal and meridional displacements at 24, 48 and 72 h. A positive correlation implies westward and northward displacement.

Eigenvector	Time interval (h) 24	48	72
	Zonal		
1	0.53	0.48	0.36
2	−0.06	−0.05	−0.09
3	−0.10	−0.07	−0.01
4	−0.41	−0.37	−0.36
5	0.27	0.25	0.28
6	0.08	−0.04	−0.09
7	−0.08	−0.08	−0.06
8	−0.25	−0.21	−0.24
9	−0.10	−0.15	−0.12
10	0.02	0.03	0.09
	Meridional		
1	−0.21	0.02	0.19
2	−0.18	−0.18	−0.20
3	0.36	0.26	0.18
4	−0.18	−0.11	−0.04
5	0.03	0.01	0.04
6	−0.16	−0.07	−0.10
7	0.22	0.25	0.20
8	0.08	0.02	−0.05
9	−0.05	0.16	0.21
10	−0.18	−0.26	−0.14

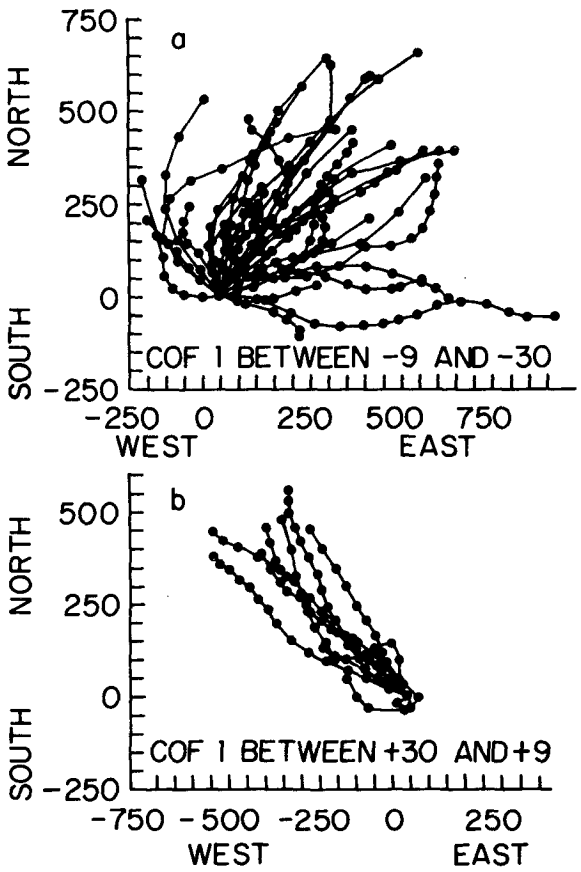

FIG. 6. Storm tracks relative to the initial position for all storms with an orthogonal coefficient 1 that is: (a) between −9 and −30; or (b) between +9 and +30. Each dot indicates a 12-h displacement.

EOF coefficients contribute relatively more to the meridional equations than to the zonal equations. For both the meridional and zonal components, the contribution of the synoptic-scale predictors increase with time relative to the persistence-type predictors.

It is also useful to examine the contribution of the EOF coefficients to the separate regression equations in view of the discussion in the previous section. A summary of the regression equations for 24, 48 and 72 h is provided in Table 6. The final prediction equation is obtained by summing the products of these coefficients and the corresponding value of the variable. A past movement variable is always the first variable selected in the stepwise procedure. However, these are not simply persistence forecasts because four or five predictors related to the EOF coefficients are generally selected in each equation. The maximum wind speed predictor contributes little to the final equations. For the zonal motion equations, the coefficients associated with Eigenvectors 1, 4, 5 and 8 are consistently selected. Coefficients of Eigenvectors 2, 3 and 7 are generally included in all meridional equations. Coefficients of Eigenvectors 4, 9 and 10 also are selectively entered in the meridional equations for

TABLE 4. Potential predictors for use in regression equations for track prediction.

Predictors	Description
1–10	Orthogonal coefficients associated with eigenvectors 1–10
11–16	Latitudinal displacements between initial and prior positions at 00–12, 00–24, 00–36, 12–24, 12–36 and 24–36 h
17–22	Corresponding longitudinal displacements
23–26	Maximum wind speed observed at initial time and at 12, 24 and 36 h previously

different time intervals. The interpretation is again that the orthogonal coefficients represent distinctly different types of synoptic forcing of the tropical cyclone motion.

6. Track forecast errors

The ultimate test of the EOF regression equation approach must be in terms of forecast errors. The magnitude of the vector error is obtained by comparison with the best-track position determined by the Joint Typhoon Warning Center (JTWC), Guam. Both the dependent sample and independent sample errors will be described. In the latter case, the errors are compared with the official (JTWC) errors, where they are available for the same cases. It should be mentioned that JTWC issues forecasts at synoptic times, whereas the EOF regression equations could not be applied until the 500 mb analysis is completed approximately 3 h later.

The regression equations would be expected to produce the best results for the dependent sample. The forecast errors (Table 7) in this sample are ~10–15% smaller than the long-term mean JTWC errors of about 205, 405 and 610 km at the 24, 48 and 72 h forecast intervals, respectively. Of equal significance, the standard deviations of the forecast errors in the dependent sample grow slowly with time. Thus, the regression equation approach tends to produce a consistent forecast. The small error standard deviations may be due to the effective removal of "noise" in the synoptic forcing through the EOF selection procedure.

The sample size for the independent homogeneous set of JTWC and regression equation forecasts is quite small. The forecast error characteristics for the regression approach are generally similar to those for the larger dependent sample, except that the standard deviation at 72 h is significantly larger (notice that only 17 cases are available at this time). At each time interval, the mean forecast errors from the regression approach are smaller than for the JTWC errors. The standard deviations are also generally smaller. Based on this limited test, it appears that the regression approach including the EOF representation of the synoptic-scale forcing provides useful forecast guidance.

7. Concluding remarks

Various forms of empirical orthogonal function (EOF) analysis have been applied to geophysical fields since Lorenz (1956) introduced the EOF approach. Here we derive the EOF's from a series of 500 mb height fields on a grid that is centered on the tropical storm position. The EOF approach efficiently describes the D-value field surrounding the tropical cyclone with the desirable feature that small scale noise has been removed. Only ~10% of the original data storage is required by the EOF representation. An additional benefit is that the EOF coefficients represent the entire synoptic forcing rather than only selected point values. Because of the orthogonality property, each eigenvector represents a distinctly different forcing of the tropical cyclone motion. Those eigenvector coefficients that are well correlated with the subsequent zonal displacement of the storm tend to be uncorrelated with the meridional displacement, and vice versa. This independence of contributions of the different eigenvectors is further illustrated in the regression equations for track prediction. Although persistence is an important contributor to these regression equations, the synoptic forcing represented by the EOF coefficients also enters in a significant sense. The resultant track forecast errors are competitive with other statistical schemes, and are 10–15% better than the JTWC forecast errors for a limited set of independent data.

The advantage of the EOF-based regression ap-

TABLE 5. Mean values and standard deviations (km) of the zonal and meridional displacements for the dependent sample. The explained variance (correlation coefficient squared) in the predictand is also shown for a set of regression equations excluding the EOF coefficients.

	Forecast interval (h)		
	24	48	72
Sample size	351	256	163
Zonal motion			
Mean	172	361	568
Standard deviation	326	572	733
Explained variance:			
with EOF coefficients	0.725	0.613	0.556
without EOF coefficients	0.675	0.506	0.386
Meridional motion			
Mean	220	413	585
Standard deviation	185	306	426
Explained variance:			
with EOF coefficients	0.476	0.354	0.315
without EOF coefficients	0.318	0.217	0.133

TABLE 6. Regression equations for the zonal and meridional displacements (km) at 24, 48 and 72 h based on the 500 mb EOF coefficients 1–10, prior latitudinal (Y) and longitudinal (X) displacements observed over the time intervals indicated and the observed maximum wind speed (Z) at prior time intervals.

	Forecast interval (h)		
	24	48	72
	Zonal		
Intercept	68.7	196.1	312.1
COF 1	6.66	18.47	19.37
COF 4	−7.78	−21.93	−48.06
COF 5	8.46	11.07	24.25
COF 8	−12.33	−22.36	−41.32
COF 10			34.04
Y(00–12)		−0.41	
Y(00–24)			−0.36
Y(24–36)	−0.12		
X(00–12)	−2.66	−0.96	−0.97
Z(00)			2.18
Z(−36)			−2.11
	Meridional		
Intercept	130.3	217.2	338.0
COF 1			9.74
COF 2	−4.44	−5.65	−11.96
COF 3	8.78	12.64	11.92
COF 4	−5.80		
COF 6	−5.28		
COF 7	6.50	17.32	20.43
COF 9		12.29	24.46
COF 10	−8.98	−16.20	
Y(00–12)	0.63	1.07	1.25
X(00–12)	0.50		
X(00–24)	−0.16		
X(00–36)			0.20
Z(−12)	0.32	0.68	1.35

proach is in its efficiency, low operational cost and simplicity. Once the eigenvectors and regression equations have been generated from the dependent data, the application to an independent sample is trivial in terms of computer resources. The normalized D-value data on the 120 grid points, and the recent storm movement and intensity are all the information that is required. Determination of the 10 EOF coefficients for the new case, and the subsequent calculation of the regression equations, could be done on a hand-held programmable calculator.

TABLE 7. Track forecast error (km) summaries for the dependent and independent samples.

	Forecast interval (h)		
	24	48	72
	Dependent sample		
Sample size	351	255	164
Mean	170	376	554
Standard deviation	135	211	282
	Independent sample		
Sample size	45	31	17
Regression:			
mean	165	365	509
standard deviation	119	237	415
JTWC:			
mean	204	430	609
standard deviation	128	291	411

One disadvantage of the EOF-based approach that is shared with other statistical (and dynamical) techniques that incorporate synoptic data is the time delay that is necessary to receive and analyze the radiosonde data. Thus, a delay of 3–4 h beyond the synoptic time is anticipated. The delay would be longer if predicted synoptic fields are included, as has been done in NHC-73 by Neumann and Lawrence (1975). The inclusion of predicted fields is expected to improve further these track forecasts.

In summary, it appears that EOF analysis can be used effectively to represent the synoptic-scale forcing of tropical cyclone motion. There appears to be a number of applications in the tropics and extratropics where it is desirable to characterize the synoptic-scale influence on significant weather-producing events. Because of the low cost and simplicity of these statistical approaches, they should be tested as alternatives to more expensive and complex solutions.

Acknowledgments. This research was performed while the principal author (A.S.) was a graduate student with the support of the U.S. Air Force. The participation of the second author is sponsored by the Naval Air Systems Command through the Naval Environmental Prediction Research Facility, Monterey, California, under Program Element 62759N, Project Number WF59-551 entitled "Meteorology Models and Prediction." Constructive comments by R. J. Renard and R. L. Haney improved the manuscript, which was typed by Ms. M. N. Marks.

REFERENCES

Brown, D. W., 1981: Tropical storm movement based on synoptic map typing using empirical orthogonal functions. M.S. thesis, Naval Postgraduate School, Monterey, CA, 80 pp. [NTIS AD A104 591].

Cattell, R. B., 1958: Extracting the correct number of factors in factor analysis. *Educ. Psychol. Meas.*, **18,** 791–837.

Chan, J. C. L., W. M. Gray and S. Q. Kidder, 1980: Forecasting tropical cyclone turning motion from surrounding wind and temperature fields. *Mon. Wea. Rev.*, **108,** 778–792.

Dixon, W. J., and M. B. Brown, 1979: BMDP Biomedical Computer Programs P-Series. University of California Press, Berkeley, 367–460.

Elsberry, R. L., 1979: Applications of tropical cyclone models. *Bull. Amer. Meteor. Soc.*, **60,** 750–762.

Harrison, E. J., Jr., 1981: Initial results from the Navy two-way interactive nested tropical cyclone model. *Mon. Wea. Rev.*, **109,** 173–177.

Jarrell, J. D., and W. L. Somervell, Jr., 1970: A computer technique for using typhoon analogues as a forecast aid. NAVWEARSCHFAC Tech. Pap. 6-70, 47 pp. [NTIS AD A008860].

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

Leftwich, P. W., and C. J. Neumann, 1977: Statistical guidance for the prediction of eastern North Pacific tropical cyclone motion. Part 2. NOAA Tech. Memo WR-125, NWS, Salt Lake City, 32 pp. [NTIS PB-272 661 10GI].

Lorenz, E. N., 1956: Empirical orthogonal function and statistical weather prediction. Sci. Rep. 1, Statistical Forecasting Proj., Dept. Meteor., MIT, 48 pp.

Miller, B. I., E. C. Hill and P. P. Chase, 1968: A revised technique for forecasting hurricane movement by statistical methods. *Mon. Wea. Rev.*, **96,** 540–548.

Morrison, D. F., 1967: *Multivariate Statistical Methods*, 2nd ed. McGraw-Hill, 338 pp.

Neumann, C. J., 1979: Statistical techniques. World Weather Watch WMO Tropical Cyclone Project, Tech. Memo. WMO-28, II4, 1–23.

——, and M. B. Lawrence, 1975: An operational experiment in the statistical–dynamical prediction of tropical cyclone motion. *Mon. Wea. Rev.*, **103,** 665–673.

——, and J. M. Pelissier, 1981: Models for the prediction of tropical cyclone motion: An operational evaluation. *Mon. Wea. Rev.*, **109,** 522–538.

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

Richman, M. B., 1981: Obliquely rotated principal components: An improved map typing technique? *J. Appl. Meteor.*, **20,** 1145–1159.

Riehl, H., W. H. Haggard and R. W. Sanborn, 1956: On the prediction of 24-hour hurricane motion. *J. Meteor.*, **13,** 415–420.

Rinne, J., and V. Karhila, 1979: Empirical orthogonal functions of the 500 mb height in the Northern Hemisphere determined from a large data sample. *Quart. J. Roy. Meteor. Soc.*, **105,** 873–884.

Veigas, K. W., R. G. Miller and G. M. Howe, 1959: Probabilistic prediction of hurricane movements by synoptic climatology. *Occas. Pap. Meteor.*, No. 2, The Travelers Weather Research Center, Inc., 54 pp.

Woodcock, F., 1980: On the use of analogues to improve regression forecasts. *Mon. Wea. Rev.*, **108,** 252–297.