INTERNATIONAL JOURNAL OF CLIMATOLOGY, VOL. 15, 1011–1019 (1995)

PREDICTION OF ANNUAL TROPICAL CYCLONE ACTIVITY OVER THE WESTERN NORTH PACIFIC AND THE SOUTH CHINA SEA

JOHNNY C. L. CHAN
Department of Physics and Materials Science, City University of Hong Kong

Received 8 February 1994
Accepted 20 December 1994

ABSTRACT

This study investigates the relationship between the spatial distribution of sea-surface temperature (SST) and the annual tropical cyclone (TC) activity over the western North Pacific and the South China Sea. Monthly distributions of SST are represented by a set of empirical orthogonal functions (EOFs). The principal components (PCs) of the monthly SST associated with these EOFs are then correlated statistically with the annual number of TCs occurring over (i) the western North Pacific and (ii) the South China Sea using a forward stepwise linear multiple regression scheme.

It is found that four PCs of the monthly SST distribution together explain 89 per cent of the total variance of the annual number of TCs over the western North Pacific. Predictions of this number from both the dependent and independent (simulated using the jack-knife method) samples have errors of less than half a standard deviation of the actual number. Two of these PCs are from the September SST, whereas the other two are from May and June SSTs. If only the latter two PCs are used, the accuracies of the prediction are slightly reduced. Therefore, a preliminary prediction can be made using these two PCs and a modified prediction can be performed at the end of September.

Two PCs are found to correlate significantly with annual TC activity over the South China Sea, representing the SST distribution in May and July. Predictions using these two PCs also give very reasonable results for both the dependent and independent samples.

KEY WORDS: tropical cyclones; Western North Pacific; South China Sea; annual prediction; sea-surface temperatures

1. INTRODUCTION

Tropical cyclone (TC) formation, development, and movement are governed to a large extent by the synoptic-scale environment in which the cyclone is embedded (WMO, 1979). Therefore, it might be expected that any variation in the environment will result in a concomitant change in the number of TCs forming in a particular region. Such variations are in turn governed by both regional and planetary scale changes, which are related to phenomena such as the El Niño–Southern Oscillation (ENSO). Thus, if correlations can be identified between these phenomena and the subsequent variations in tropical cyclone activity, predictions of the latter can be made. For example, Gray (1984a, b) found that ENSO, the quasi-biennial oscillation in the stratosphere, and sea-level pressure anomalies over the Caribbean are all related to hurricane activity in the North Atlantic. Since then, he has made annual real-time forecasts of such activity with reasonable success (e.g., Gray *et al.*, 1992, 1993, 1994). In other ocean basins in which TCs occur, similar studies have also been made (Nicholls, 1984; Chan, 1985; Dong, 1988). Nicholls (1992) also developed a method for forecasting the annual Australian TC activity.

The above studies concentrated generally on the relationship between TC activity over the whole ocean basin and ENSO. However, various other factors (which may or may not be related to the ENSO) also could be important. For example, Nitta (1987) found that variations in the distribution of sea-surface temperature (SST) over the western North Pacific (WNP) are correlated with the locations of TC formation in the same ocean basin. The present study extends such a finding by investigating the annual TC activity over the entire WNP and the South China Sea in relation to the spatial distributions of SST, which are represented by a set of empirical orthogonal functions (EOFs). The principal components (PCs) associated with these EOFs are then correlated statistically with

CCC0899-8418/95/091011-09

the annual TC activity in each of the regions using a linear stepwise regression procedure. Those PCs found to be significant statistically are used as predictors of the annual TC activity in each of the two regions. The prediction equations are derived using the entire sample as well as an 'independent' sample generated with the jack-knife method (see e.g. Miller, 1974).

The data used and the methodology adopted in this study are described in section 2. The 'fidelity' of the EOFs derived from the data sample is also discussed. Correlations between the PCs of SST and the annual TC activity in each of the regions are presented in section 3. The predictions derived from the significant parameters and their verifications are also described. A summary of the relevant results is given in section 4.

2. DATA AND THE EOFS

2.1. *Data*

Tropical cyclones over the WNP between 1970 and 1988 form the basic data set for this study. The starting year is chosen as 1970 because SST data before this year are not available to the author. Best tracks of these cyclones are obtained from the Royal Observatory, Hong Kong. Tracks for years prior to 1984 were analysed only in the region west of 160°E. Therefore, this study is limited only to cyclones west of this longitude. This limitation, however, should not restrict the usefulness of the study for two reasons. Most of the TCs over the WNP form west of this longitude (Xue and Neumann, 1984) so that those east of 160°E probably do not have too significant a contribution to the overall distribution of TC activity. Furthermore, Dong and Holland (1994) suggested that most of the variability in TC activity in this ocean basin occurs west of this longitude.

In this study, the TC activity (defined as the number of TCs that have occurred) over the entire ocean basin as well as the South China Sea (SCS) are analysed. The area of the former is defined by 0–40°N, 100–160°E. The latter region is chosen because TC activity here directly affects Hong Kong. The area chosen is bounded by 15–25°N and 110–120°E.

The SST data are provided by the US Climate Analysis Center and consists of grid-point, monthly-mean values over the entire WNP for the years 1970–1989. The horizontal resolution is 2° longitude × 2° latitude. Only SST data for the months of May through October are included because most of the TCs occur during these months (Xue and Neumann, 1984).

2.2. *Representation of the SST fields*

The spatial distribution of the (SST) is projected on to a set of empirical orthogonal functions using the procedure described by Shaffer and Elsberry (1982). The size of the data matrix is 698 × 114, where the rows are the number of grid-points and the columns the number of cases in the data sample.

North *et al.* (1982) proposed a method to evaluate the sampling errors associated with the derivation of the EOFs. They derived the sampling error of a particular eigenvalue λ as

$$\delta\lambda = \lambda\sqrt{\frac{2}{N}} \tag{1}$$

where N is the size of the data sample. If the spacing between two neighbouring eigenvalues is comparable to or larger than $\delta\lambda$, then the sampling errors for the EOF associated with λ will be comparable to the size of the neighbouring EOF. In this case, the sample eigenvector may be a random mixture of the true eigenvectors.

To test whether the EOFs derived from the available SST data can be used to represent the actual fields, equation (1) is applied. The results (Table I) suggest that the first five EOFs have eigenvalues much farther apart from one another than the theoretical sampling error $\delta\lambda$. Therefore, the derived eigenvectors should be representative of the actual ones.

It also can be seen from Table I that the first three EOFs together explain nearly three-quarters of the total variance. To ascertain the validity of these results, the SST field for a month chosen at random (July 1970) was

Table I. Eigenvalues, theoretical spacing ($\delta\lambda$) between neighbouring eigenvalues, and cumulative percentage of variance explained by each of the first five EOFs for the SST data used in the present study. See text for the method used to calculate $\delta\lambda$

Eigenvector number	Eigenvalue	$\delta\lambda$	Cumulative percentage variance explained
1	355·11	47·04	51
2	95·17	12·61	65
3	67·29	8·91	74
4	36·30	4·81	79
5	21·26	2·82	82

reconstructed using only the first EOF and a combination of the first three EOFs. The reconstructed SST patterns are then compared with the actual SST pattern (Figure 1). It can be seen that even with only one EOF, the high SST over the tropics east of the Philippines was largely reproduced. Refinement of this pattern using the second and three EOFs give even closer agreement with the actual field. The EOFs so derived should therefore be truly representative of the actual ones.

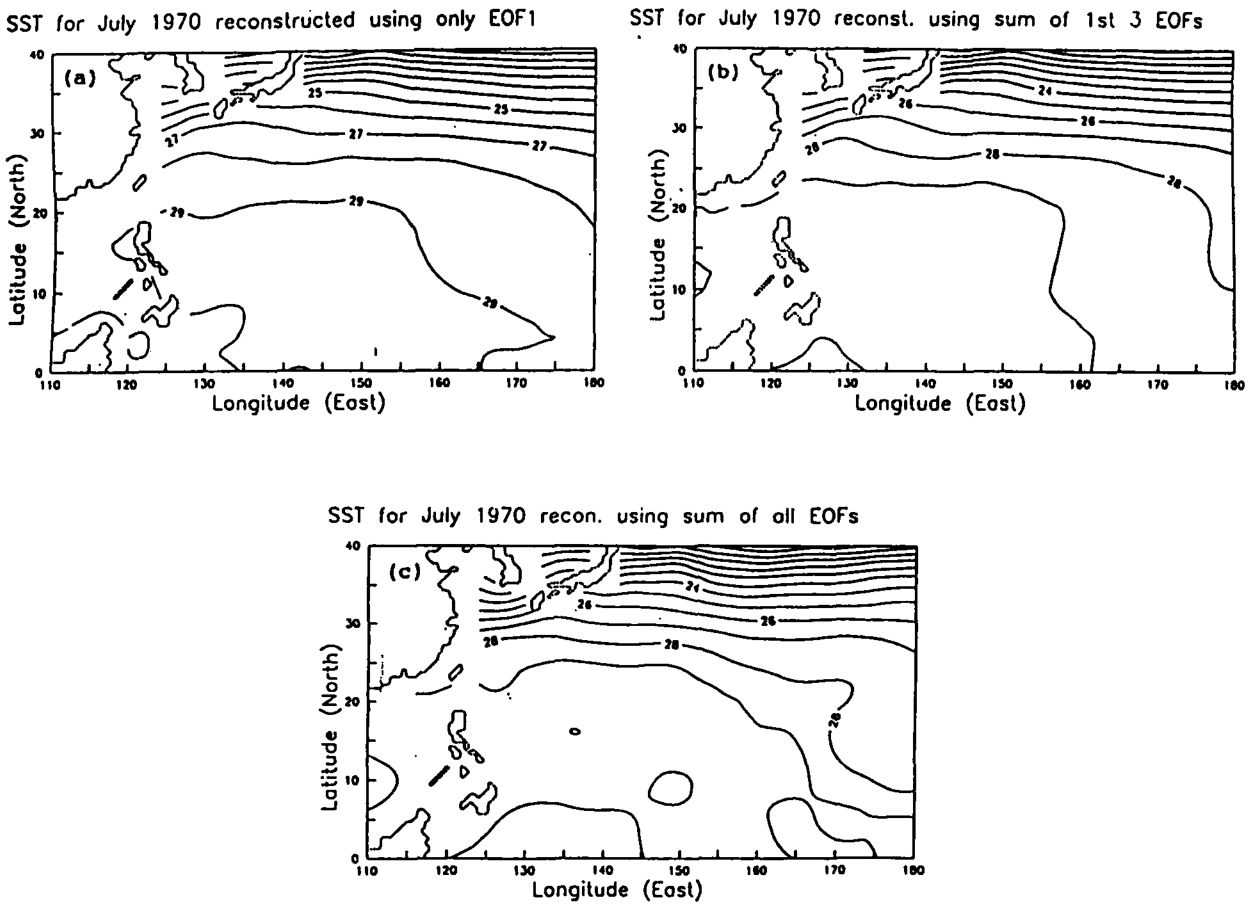

Figure 1. Distribution of SST (°C) for July 1970 reconstructed using (a) only EOF1, (b) the sum of the first three EOFs and (c) the sum of all the EOFs (i.e. the actual)

3. THE REGRESSION AND PREDICTION PROCEDURES

The annual TC activity over each of the two regions defined in section 2.1 (the WNP and the South China Sea) is correlated with the monthly SST using a forward, stepwise linear regression procedure. Because the increase in the amount of percentage variance of SST explained by the fourth and fifth EOFs is quite small (see Table I), only the principal components (PCs) associated with the first three EOFs are used as potential predictors in the regression procedure. This gives a total of 18 predictors. Because only 19 years of data are available, the regression procedure is performed using PCs associated with each EOF at a time. Those PCs that are found to be significant are then combined and then regressed against the TC activity again. With 19 years of data and anticipating no more than three to four predictors are chosen, a variable will be selected if the *F*-to-enter value is ≥ 4.5 so as to have a confidence limit of ≥ 0.95. The *F*-to-enter value is defined as

$$F\text{-}to\text{-}enter = \frac{SS(residuals) - SS(next)}{SS(next)/(N-p)}$$

where SS(residuals) is the sum of the squares of the residuals at the present step, SS(next) the residual sum of squares after the next independent variable is entered, N the number of cases and p the number of independent variables.

Using the predictors selected, predictions of TC activity are first made from the entire data sample (the dependent sample). A jack-knife procedure is then applied to simulate predictions of an independent sample. A brief description of this procedure is given in the Appendix. Further details and variants of this technique can be found in Miller (1974).

3.1 Entire WNP

The only principal component of SST selected when correlated with the annual number of TCs over the WNP is May PC3, which explains 28 per cent of the variance in TC activity. However, if only TCs with maximum sustained wind speeds of ≥ 17 m s^{-1} (i.e. tropical storms and typhoons) are included, the variance explained by this PC increased to 40 per cent (Table II). Three other PCs are also chosen by the regression procedure, explaining a total of 89 per cent of the variance. This empirical result is quite unexpected in that the annual variability in tropical storm and typhoon activity over the WNP can be explained statistically almost entirely by the variation in the SST distribution.Predictions using the four PCs listed in Table II for both the dependent and simulated independent samples are shown in Table III. Those for the dependent sample are very close to the observed numbers, with an average absolute error of only ≈ 1. Notice also that years with much above (e.g. 1971) or much below (e.g. 1975) normal tropical storm and typhoon activity are predicted correctly. This conclusion also applies to the independent sample although, as expected, the errors are larger than those of the dependent sample. A linear regression between the observed numbers of tropical storms and typhoons and those predicted by the jack-knife method gives a correlation of 0·89 (Figure 2). The plot of Figure 2 shows the technique tends to underforecast slightly.

Table II. Variables selected in the stepwise linear regression between the first three principal components of SST and anomalous tropical storm and typhoon activity over the WNP

Step	Variable	Cumulative percentage variance
1	May PC3	40
2	June PC2	52
3	September PC2	81
4	September PC3	89

Table III. Prediction of the number of tropical storms and typhoons over the entire WNP using PCs of SST (May PC3, June PC2, September PC2, and September PC3) for both the dependent and the independent samples. Predictions for the independent sample are derived using the 'jack-knife method' described in the text: σ = standard deviation

		Dependent sample		Independent sample	
Year	Observed	Predicted	Error	Predicted	Error
1970	21	21	0	21	0
1971	36	36	0	37	1
1972	28	27	−1	26	−2
1973	20	22	2	22	2
1974	31	31	0	31	0
1975	19	19	0	18	−1
1976	25	24	−1	23	−2
1977	19	21	2	21	2
1978	27	26	−1	24	−3
1979	23	24	1	25	2
1980	24	26	2	26	2
1981	25	27	2	27	2
1982	26	25	−1	25	−1
1983	23	23	0	22	−1
1984	25	26	1	27	2
1985	25	26	1	27	2
1986	28	26	−2	25	−3
1987	21	20	−1	20	−1
1988	28	26	−2	24	−4
Mean	25				
σ	4·3				
Average absolute error			1·1		1·7

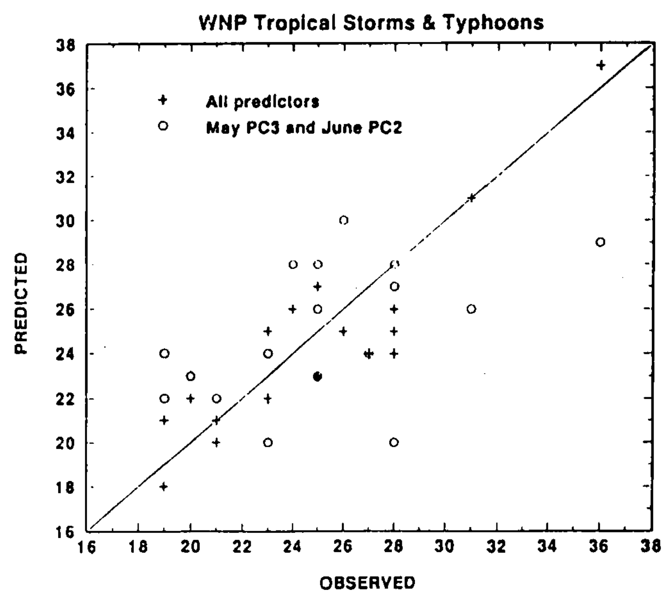

Figure 2. Scatterplot of the annual observed number of tropical storms and typhoons over the western North Pacific versus the number predicted by the jack-knife method using all predictors (+) and only the two predictors May PC3 and June PC2 (o). The straight line indicates the perfect prediction

Nevertheless, even the largest error of −4 is within one standard deviation of the observed number. Therefore, it appears that the four SST EOFs can be used to predict the number of tropical storms and typhoons over the entire WNP with a high degree of certainty.

However, because two of the predictors are from the SST in September, the prediction does not have much practical use except in modifying later in the season any prediction made using the other two PCs. Fortunately, these two PCs, the May PC3 and June PC2, together explain over half of the total variance in tropical storm and typhoon activity (see Table II). Therefore, another set of predictions is made using these two predictors alone. As expected, the predictions are not as good as those using all available predictors (Table IV and Figure 2). The correlation between the observed numbers and those predicted from the independent sample is 0·52. Nevertheless, with the exception of a few cases, they are still within one standard deviation of the observed number. Therefore, in practice, a preliminary prediction using these two predictors can be made by the end of June. The predicted number can then be modified at the end of September by incorporating the two SST PCs of this month.

3.2 South China Sea

The first principal component of SST chosen using the stepwise regression procedure for correlating with the annual TC activity in this region is PC1 in July, which explains 46 per cent of the variance. The second (and only other) variable chosen is PC2 of May, which reduces the variance by another 15 per cent. If tropical depressions (i.e. TCs with maximum sustained winds of ≤ 17 m s^{-1}) are excluded, only the July PC1 is chosen, with 53 per cent of the variance explained.

Table IV. Prediction of the number of tropical storms and typhoons over the entire WNP using May PC3 and June PC2 of SST for the independent sample derived using the 'jack-knife method' described in the text: σ = standard deviation

		Independent sample	
Year	Observed	Predicted	Error
1970	21	22	1
1971	36	29	−7
1972	28	27	−1
1973	20	23	3
1974	31	26	−5
1975	19	22	3
1976	25	28	3
1977	19	24	5
1978	27	24	−3
1979	23	24	1
1980	24	28	4
1981	25	26	1
1982	26	30	4
1983	23	20	−3
1984	25	23	−2
1985	25	26	1
1986	28	28	0
1987	21	22	1
1988	28	20	−8
Mean	25		
σ	4·3		
Average absolute error			2·7

Table V. Prediction of the number of tropical cyclones over the South China Sea using July PC1 and May PC2 of SST for both the dependent and the independent samples. Predictions for the independent sample are derived using the 'jack-knife method' described in the text: σ = standard deviation

		Dependent sample		Independent sample	
Year	Observed	Predicted	Error	Predicted	Error
1970	14	12	−2	11	−3
1971	15	15	0	15	0
1972	9	9	0	9	0
1973	13	9	−4	9	−4
1974	14	13	−1	13	−1
1975	9	10	1	10	1
1976	6	8	2	10	4
1977	10	11	1	11	1
1978	11	10	−1	10	−1
1979	12	10	−2	10	−2
1980	14	15	1	16	2
1981	9	8	−1	8	−1
1982	8	9	1	10	2
1983	9	9	0	10	1
1984	9	11	2	11	2
1985	10	10	0	10	0
1986	7	8	1	9	2
1987	6	6	0	7	1
1988	9	10	1	10	1
Mean	10				
σ	2·3				
Average absolute error			1·1		1·5

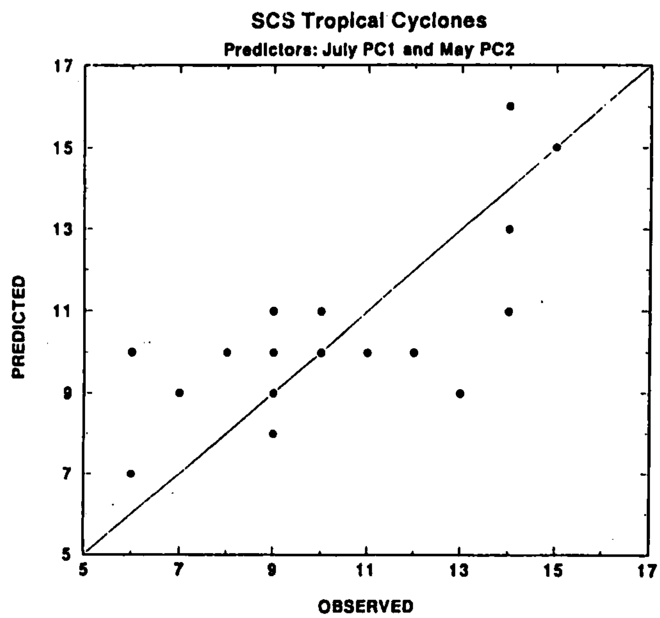

Figure 3. As in Figure 2 except for the annual number of TCs over the SCS using the predictors July PC1 and May PC2

Table VI. Prediction of the number of tropical storms and typhoons over the South China Sea using July SST PC1 for both the dependent and the independent samples. Predictions for the independent sample are derived using the 'jack-knife method' described in the text: σ = standard deviation

		Dependent sample		Independent sample	
Year	Observed	Predicted	Error	Predicted	Error
1970	10	9	−1	9	−1
1971	14	11	−3	10	−4
1972	8	7	−1	7	−1
1973	11	7	−4	7	−4
1974	11	9	−2	9	−2
1975	7	8	1	8	1
1976	6	5	−1	4	−2
1977	8	8	0	8	0
1978	8	8	0	8	0
1979	5	8	3	8	3
1980	11	12	1	12	1
1981	8	8	0	8	0
1982	8	8	0	8	0
1983	7	7	0	6	−1
1984	7	10	3	10	3
1985	7	8	1	8	1
1986	6	8	2	8	2
1987	6	7	1	7	1
1988	8	9	1	9	1
Mean	8				
σ	2·3				
Average absolute error			1·3		1·5

With the exception of 1973, predictions of the annual number of TCs in this area using the SST PCs are within one standard deviation of the actual number for the dependent sample (Table V). In fact, only five out of the 19 years have errors of ≥ 2. The errors for the independent sample are only slight larger, with an average absolute error of $\approx 0{\cdot}5\sigma$. The correlation between the observed and predicted (independent sample) number of TCs is 0·72 (Figure 3). A slight overprediction can be seen from the plot. Because the TC season for this region is between May and October (Xue and Neumann, 1984), this result is of great value because a prediction can be made after July.

The predictions of the annual number of tropical storms and typhoons are slightly worse (Table VI). This might be expected because only 53 per cent of the variance is explained. However, the average absolute errors for both the dependent and independent samples (the latter having a correlation of 0·53 with those observed) are still relatively small ($\approx 0{\cdot}5\sigma$) and only in four out of the 19 cases are the errors $> 1\sigma$. Again, this result has a high operational value.

4. SUMMARY

In the past, internannual variations of tropical cyclone activity have been related to the ENSO phenomenon, expressed in terms of either the Southern Oscillation Index or the SST over the eastern equational Pacific. This paper shows that such variations in the western North Pacific are, in addition, related to the spatial distribution of SST over the same region. Representing the SST distribution by a set of empirical orthogonal functions, the corresponding principal components can be used to produce rather good predictions of the annual number of tropical cyclones occurring over the western North Pacific as well as the South China Sea. Most of these predictions make use of predictors that are available early or in the middle of the tropical cyclone season. Therefore, these results can be used operationally as a predictive tool for the annual tropical cyclone activity over these two regions.

ACKNOWLEDGEMENTS

Much of the programming work in this study was done by Mr S. H. Lam. The author also wishes to thank the Royal Observatory Hong Kong for providing the best-track data and the Climate Analysis Center of the USA for providing the SST data. The professional help in manuscript preparation of Ms Rolin Ng and Ms Betsy Yu are gratefully acknowledged.

This research is supported by the Indicated Research Grant No. 700039 of the City Polytechnic of Hong Kong and the Earmarked Grant No. 904024 of the Universities and Polytechnics Grants Committee of Hong Kong.

APPENDIX

The jack-knife method

To test the usefulness of a prediction equation derived from a sample of limited size the 'jack-knife method' is often used to simulate an independent sample. The simplest procedure consists of the following:

(i) derive the prediction equation based on all except one data point (say, in this paper, the number of TCs in year 1);
(ii) use the derived equation to predict the number of TCs of year 1 and calculate the error;
(iii) repeat (i) and (ii) but now include year 1 and exclude year 2;
(iv) repeat (iii) until all the years have been excluded once.

With this procedure, each of the predictions can be considered to be independent. If the prediction errors using this sample are reasonable, it may be concluded that the prediction equation using the dependent sample is useful and can be used for future predictions.

REFERENCES

Chan, J. C. L. 1985. 'Tropical cyclone activity in the northwest Pacific in relation to the El Niño/Southern Oscillation phenomenon', *Mon. Wea. Rev.*, **113**, 599–606.
Dong, K. 1988. 'El Niño and tropical cyclone frequency in the Australian region and the northwest Pacific', *Aust. Meteorol. Mag.*, **36**, 219–225.
Gray, W. M. 1984a. 'Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb QBO influences', *Mon. Wea. Rev.*, **112**, 1649–1668.
Gray, W. M. 1984b. 'Atlantic seasonal hurricane frequency. Part II: forecasting its variability', *Mon. Wea. Rev.*, **112**, 1669–1683.
Gray, W M., Landsea, C. W., Mielke, P. W., Jr. and Berry, K. J. 1992. 'Predicting Atlantic seasonal hurricane activity 6–11 months in advance', *Wea. Forecast.*, **7**, 440–455.
Gray, W. M., Landsea, C. W., Mielke, P. W., Jr. and Berry, K. J. 1993. 'Predicting Atlantic basin seasonal tropical cyclone activity by 1 August', *Wea. Forecast.*, **8**, 73–86.
Gray, W. M., Landsea, C. W., Mielke, P. W., Jr. and Berry, K. J. 1994. 'Predicting Atlantic basin seasonal tropical cyclone activity by 1 June', *Wea. Forecast.*, **9**, 103–115.
Miller, R. G. 1974. 'The jackknife—a review', *Biometrika*, **61** 1–15.
Nicholls, N. 1984. 'The Southern Oscillation, sea-surface temperature and interannual fluctuations in Australian tropical cyclone activity', *J. Climatol.*, **4**, 661–670.
Nicholls, N. 1992. 'Recent performance of a method for forecasting Australian seasonal tropical cyclone activity', *Aust. Meteorol. Mag.*, **40**, 105–110.
Nitta, T. 1987. 'Convective activities in the tropical western Pacific and their impact on the Northern Hemisphere summer circulation', *J. Meteorol. Soc. Jpn.*, **65**, 373–390.
North, G. R., Bell, T. L., Cahalan, R. F. and Moeng, F. J. 1982. 'Sampling errors in the estimation of empirical orthogonal functions', *Mon. Wea. Ref.*, **110**, 699–706.
Shaffer, A. R. and Elsberry, R. L. 1982. 'A statistical climatological tropical cyclone track prediction technique using an EOF representation of the synoptic forcing', *Mon. Wea. Ref.*, **110**, 1945–1954.
WMO. 1979. *Operational Techniques for Forecasting Tropical Cyclone Intensity and Movement*, WMO-528, World Meteorological Organization, Geneva, 138 pp.
Xue, Z. and Neumann, C. J. 1984. *Frequency and Motion of Western North Pacific Tropical Cyclones*, NOAA Technical Memorandum NWS NHC 23, National Hurricane Center, Miami, Fl, 89 pp.