Vol. 7 No. 2 **Advances in Atmospheric Sciences** May 1990

Correlations between Sea Surface Temperature in Eastern Equatorial Pacific and Rain Days over China in Summer

Huang Jiayou (黄嘉佑)
Department of Geophysics, Peking University, Beijing 100871

Received March 18, 1989

ABSTRACT

Sea-surface temperature (SST) in the eastern, equatorial Pacific and rain days over China in summer are analysed using correlation moments that is proposed by author and principal component analysis(PCA). Occurrences of the strong rain-day anomalies over China are associated with extreme SSTs in some years. Areas significantly affected by the phenomena include North and Northeast China.

I. INTRODUCTION

A number of studies have shown that climatic extremes tend to be discrete rather than continuous in nature. Examples are the serious North China drought of 1972 and the summer flood of 1964 in Southeast China. These years were associated with El Nino and non-El Nino events (defined by Angell,1981). It may be possible to explain such phenomena by the air-sea interaction.

Coupling the ocean and the atmosphere has concerned many meteorologists. Walsh et al. (1981), for example, used principal component analysis (PCA) to reveal an association between surface temperature over the United States and the North Pacific Ocean. Nitta (1986) studied the response of the atmosphere to the sea using cloud data. Chung (1982) used stratified correlations to show relationships between SSTs in Atlantic and rainfall in northeastern Brazil. Similarly, this study is concerned with the response of surface meteorological elements to an anomalously warm or cold ocean. More specifically, this paper examines rain days over China during summer. Rain days may be a better index than total precipitation for the study of the climatic change (cf. Englehart,1985), since it reflects the frequency of rain-producing synoptic regimes.

To analyse the relationships between meteorological fields, one often employs correlation, particularly pair-wise correlation between time series of the two fields at corresponding grid points. Significantly correlated regions are those in which the correlation coefficients surpass some threshold. In order to explain sea-air interaction, for example, Kawamuka (1986) correlated 500hPa geopotential height fluctuations and time coefficients of the first mode of SST in Pacific. In studies of atmospheric teleconnectivity, a one-point correlation map is frequently employed (e.g. Wallace et al., 1981; and Nitta,1986). Correlation between two meteorological elements, however, can only reveal a limited number of linear relationships contained in the sample. It is impossible to describe nonlinear relationships as well as the variation with time of the correlations between the fields. The correlation moment analysis proposed in this paper can overcome several of the drawbacks in quantifying relationships between SST and rain days over China.

Principal component analysis or empirical orthogonal functions (EOFs) have been used extensively by meteorologists to evaluate meteorological fields (Richman,1986). A promising one, called rotated PCA (cf. Horel, 1984), has been used repeatedly by meteorologists to obtain the principal components (PCs) of a single field. In this study, however, the PCA is based on correlation moments between a meteorological element (i.e. SST) and another meteorological element (i.e. rain days).

II. THE CORRELATION MOMENT AND DATA

A correlation moment between variables x and y_k, where y_k is from a variable field containing p variables, is

$$c_k = xy_k \quad (k = 1,2,\cdots,p) \tag{1}$$

Each c_k is based upon n observations or cases, and x, in this paper, is the seasonal SST in the eastern, equatorial Pacific Ocean (180°E–90°W, 0°–10°S) in summer (June–August). El Nino has taken place frequently in this region according to Angell (1981). The variables y_k $(k=1,2,\cdots,p)$ represent summer rain days at 46 stations (Fig.1) in China. The stations are located in the main agricultural regions of China. These data were drawn from Beijing Meteorological Centre (BMC) for the period 1951–1980.

A correlation moment (c_k) describes a systematic relationship between x and y_k. If x and y_k are deviation variables, the average of the new variables c_k is the covariance between them. When they are standardized variables, as used in this paper, the average of variables c_k is the correlation coefficient (r_k) between them.

The standard deviation of variables c_k describes the average deviation of the variable from its mean. It is denoted by s_k. If it is small the representative of the correlation coefficient between them will be good. So, it can be used to illustrate the close–connected degrees of linear relationship between the two variables in sample if the absolute value of the corresponding average is considered simultaneously.

Correlation analysis between x and y_k can also be completed using the averages of their correlation moments. Only a few stations (Fig.1) surpass the significant test at 0.05 level (i.e. the absolute values of the correlation coefficients were >0.36). It can be seen from Table 1 that the averages of c or the correlation coefficient (denoted by r) at five stations are nearly significant. Their standard deviations, however, are so large that it can bear comparison with the variational range of correlation coefficient. It means that they are not good for describing the representative of their correlation. They are not particularly reliable.

On the other hand, although some stations exhibit significant correlation coefficients in some areas in Fig. 1, the correlation field is not significant. According to the significant test for a correlation field suggested by Livezey et al.(1983), the number of the stations with the significant correlation coefficients should not be less than 8 (significant level 5%) in a field that consists of 46 stations.

Ananthakrishnan et al. (1984) additionally told us that the effective number of degrees of freedom (N_{eff}) is

$$N_{eff} = n / (1 + 2r_1 r'_1 + 2r_2 r'_2 + \cdots) \tag{2}$$

where r_i and r'_i $(i=1,2,\cdots)$ are the autocorrelation coefficient of series x and y with lag i, respectively. The numbers of N_{eff} in five stations are calculated and also showed in Table 1. If the significants made with N_{eff} for the correlation coefficients at stations are done, none is significant at the 0.05 level.

Therefore, it is necessary to study the variations of their correlation moments with time.

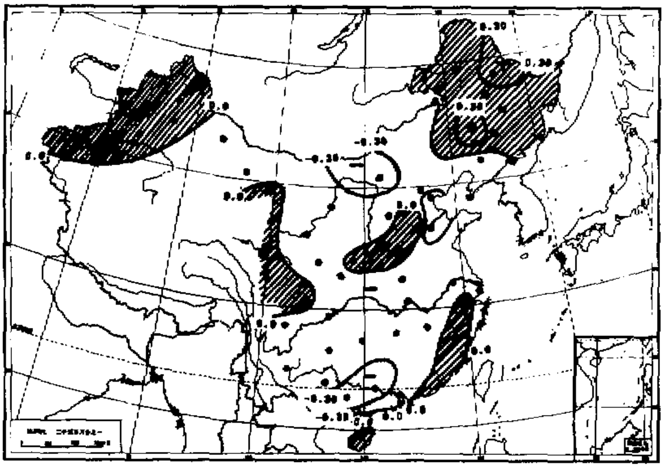

Fig.1. The distributions of the correlation coefficients between the SST and rain days.

Table 1. The Correlation Moments in Five Stations

Station No.	12	11	18	33	41
r_k	0.33	−0.35	−0.33	−0.40	−0.39
s_k	0.90	0.96	0.93	1.09	0.88
N_{eff}	12	11	17	13	17

N_{eff} means effective numbers of degrees of freedom, No.12: Nenjiang (45.4°N, 124.6°E), No.11: Tianjin (39.1°N, 117.2°E), No.18: Hohhot (40.8°N, 111.7°E), No.33: Wuzhou (23.4°N, 111.3°E), No.41: Nanning (22.8°N, 108.3°E)

II. THE RESPONSES OF RAIN DAYS TO ABNORMAL SST

Linear relationships between rain days and SST are obscure due to weak signals of SST and existing nonlinearities. Using an absolute value of the standardized SST of 1.0, the sample can be divided into two groups. One is called the abnormal group (values > 1.0) and it consists of 13 years while the other,called the normal group, has sample size of 17. Correlation moments (variable c) are correspondingly divided into two subsamples under the two SST conditions. Table 2 shows the variations of c in the years under the abnormal SSTs.

Ideally, if two variables have close relation (positively or negatively) their correlation moments in sample should have the same signs as the one of their correlation coefficient. One can see that, from Table 2, although most signs of variable c at each station are consistent with the ones of the corresponding correlation coefficient, some of them are contrary. For example, the correlation coefficient at station 12 is positive with quite high value but the negative correlation moments with high values occurred in 1975, 1965 and 1976. The linear correlation between SST and rain days at station 34 is very weak (the correlation coefficient equals −0.06), but the strong values of correlation moment between them occurred in 1955, 1970, 1965 and 1972. The cases can be found in the other stations too. The nonlinear relationship is also found in the strong variations of the averages of c under two conditions of SST (see Table 3). The slope of the regression straightlines in the two subsamples has strong changes. The difference of the averages can be tested using Student test (cf. Harnack et al., 1984). The statistic t is calculated by

$$t = \frac{\bar{c}_a - \bar{c}_n}{(n_1 s_a^2 + n_2 s_n^2)^{1/2}} \left(\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2} \right)^{1/2} \tag{3}$$

where n_1 and n_2 denote the sample size in the two subsamples under abnormal and normal SSTs respectively. Unfortunately, the differences at the five stations are not all significant at 5% level (see Table 3). However, It is worth noting that, from Table 3, the responses of the amplitudes of c (without regard to their signs) in the years to abnormal SSTs are obvious. Therefore, similar inter–comparisons between the absolute values of c under the two conditions of SST had been done (see the bottom block in Table 3). It can be found that their differences are significant. Some of them can reach quite high significance (<1% level). The same tests have been done on whole c field. The results show that there are 30 stations being significant at the 0.05 level in the field. Obviously, the difference between the two fields of c under the different conditions is significant. It follows that the responses of rain days to SST in the extreme group are stronger than the other group. However, the responses are not linear. It will conclude that the nonlinear relationships between SST and rain days represent strongly in the sample, especially in abnormal SST.

Table 2. The Values of c at Five Stations during Abnormal SSTs

Year	SST	12	11	34	33	41
1955	−1.37	1.63	0.63	−1.02	−2.55	−1.71
1956	−1.16	−0.11	−1.44	−0.50	0.05	−1.28
1964	−1.37	1.24	−0.98	0.28	−0.34	1.23
1970	−1.02	1.07	−0.47	1.02	0.34	−0.10
1974	−1.02	1.80	0.07	−0.11	−0.55	−1.42
1975	−1.63	−1.31	0.11	0.08	−0.17	0.54
1957	1.85	1.22	−0.36	0.21	0.73	−0.87
1958	1.03	−0.49	−0.60	0.61	−1.40	−0.63
1963	1.03	0.53	−1.01	0.11	−0.05	−2.11
1965	1.64	−1.49	−2.09	3.06	−1.52	−2.43
1969	1.17	1.44	0.54	−0.24	−1.26	−1.23
1972	2.25	2.14	−3.39	−4.42	−4.74	−1.72
1976	1.23	−1.30	1.05	−0.65	0.49	0.65

Station 34: Chongqing (29.5°N, 106.5°E) the others are the same as Table 1.

Table 3. The Inter–comparisons between the Averages of c_a and c_n under Different Conditions of SST

Condition *	12	11	34	33	41
$\bar{c}_a$	0.49	−0.67	−0.12	−0.85	−0.85
$\bar{c}_n$	0.21	−0.10	−0.02	−0.07	−0.03
t	0.5	−0.8	−0.1	−1.1	−1.8
$\lvert\bar{c}_a\rvert$	1.21	1.05	0.95	1.09	1.23
$\lvert\bar{c}_n\rvert$	0.40	0.39	0.34	0.38	0.30
t	4.3	1.8	1.4	1.7	5.0

(*): The subscript a and n mean the states under the conditions of abnormal and normal SSTs, respectively. t denotes the calculated value from Eq.(3).

If a sign of the c is positive it is called by a positive response of rain days to abnormal SST. Conversely, if a negative correlation moment in a year means that a negative response of rain days to abnormal SST will occur. In station 12, for example, there were the positive responses of rain days to the extremely cold SST in 1955, 1964, 1970, and 1974. It means that there were much little rain days in summer in those years in station 12. The negative responses of rain days to the extremely warm SST in 1958, 1965, and 1976 occurred at the station with the same states in rain days. However, there were very weakly positive responses of rain days to the normal SST at the station. One can see that, from Table 3, the average value of variable c is 0.21 which is not significant at level 5% using the t test. It means that there is no correlation between SST and rain days under the normal SST.

It seems to say that, in summary, firstly there are probably some nonlinear relationships between SST and rain days, and secondly there were strong responses of rain days to the abnormal SST in the years. They are presented by correlation moment in different regions and years.

III. PRINCIPAL COMPONENTS ON CORRELATION MOMENT FIELD

In order to extract main characteristics of the field of variable c with spatial and temporal variations, it is necessary to perform the PCA for it. The principal component (PC) for vaviables c_k $(k=1,2,\cdots,p)$ can be written as

$$PC = v_1 c_1 + v_2 c_2 + \cdots + v_p c_p \qquad (4)$$

Where v's are the eigenvectors of the correlation matrix associated with the variables of c. There are p principal components if the number of variable $c(p)$ is less than the cases of the sample (n). They are denoted by PC_j $(j=1,2,\cdots,p)$. The explained variances of the first five PCs are listed in Table 4. Because the percentages of the cumulative value of explained variance (CEV) in the five PCs have almost reached 70% they can be representative of the dominant characteristics of c field.

Table 4 shows the explained variances in the PCs.

Table 4. The Explained Variance of PCs

	PC1	PC2	PC3	PC4	PC5	CEV
Explained Variance	0.238	0.161	0.104	0.097	0.073	0.673
Eigenvalue	10.95 *	7.42 *	4.80 *	4.44	3.34	

According to North et al.(1982), the significant PCs can be determined. They showed that the standard error of an eigenvalue λ is roughly $\delta\lambda \approx (2/n)^{1/2}$, where n is the sample size. Thus, unless the difference between two neighboring eigenvalues is greater than $\delta\lambda$, the two associated PCs cannot be properly resolved. Table 4 shows the eigenvalues of the matrix. The star(*) in Table 4 means that the corresponding PC passes the test.

The spatial patterns associated corresponding PCs will illustrate the spatial characteristics of variable c.

The first PC, which accounts for 25 per cent of the total variance roughly, is the dominant component. The $PC1$ pattern (Fig.2a) is the general tendency for almost eastern part over China where the anomalous variations of c have similar phase. The maximums of the loadings in it occur in North of China. It is worth noting that the distribution tendency in

Fig.2a is rather similar to that in Fig.1. It means that the first *PC* is representative of the average feature of their correlation moment field between SST and rain days.

The second *PC* pattern (Fig.2b) strongly represents the variational features of variable *c* in Northeast of China. The distribution tendency runs latitudal mode. It might be associated with the synoptic regime of subtropical high in Pacific Ocean.

The third pattern (Fig.2c) describes the feature about the contrary variation relation between Southeast and Northwest part of China. The fourth and fifth patterns (not shown here) have similar distribution with isolate signs (+—+ / —+—) over China.

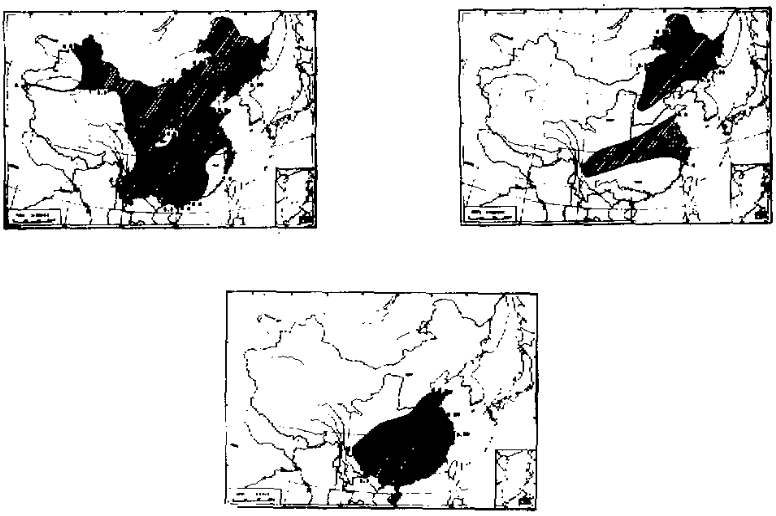

Fig.2. Component loadings (× 100) for the first (a) to third (c) *PC* using the standardized variable *c*.

The *PCs* (scores) which obtain from Eq.(4) can be representative of the characteristics of the temporal variations in the correlation moment field between SST and rain days. They have strong variations with time. Table 5 shows the values of *PCs* in the years mentioned in Table 2 for comparison. It can be found that when the standardized SST occurred with the extreme anomaly in the year, in general the PC also occurred with the strong amplitude, especially in 1972. Because the spatial patterns present the different characteristics in regions, it implies that when the extreme anomalies of SST occurred in the years the strong response with the extreme anomalies of rain days occurred in different regions where the spatial patterns are associated. However, the response of the anomalous values of rain days to SST is not linear. For example, there are negative responses in most years, in general, concerning *PC*1 on the SST. It means that there is the negative relationship between SST and rain days in the positive loading region (e.g., the eastern part of China in Fig.2a) in most years, especially in 1972. However, there were the positive responses occurred in 1975, 1957 and 1976. It can be considered as the main characteristics of the distribution with time in correlation moment field.

There are some comparisons between the standardized values of SST and the five *PCs* in the years. The *AA* is used for denoting the average of the amplitude of the *PCs* which have been standardized for the comparison. It is written as

$$AA_i = (\frac{1}{5}\sum_{i=1}^{5} PCj_i^2)^2 \tag{5}$$

and is the representative of the average feature in the *PCs* with time. The *AAs* in the years are shown in last column in Table 5. It can be found that the *AAs* have the extreme values (that reach or surpass about 1.0) when the values of SST are the extremes in most years. There is stronger amplitude (with average 1.32 and standard deviation 0.57) to describe the variation of the variable *c* in the years. In order to compare the difference of the responses of *PCs* between the two groups of abnormal and normal SST , the difference of the averages in two groups with the absolute values and the values of *PCs* are tested by *t*—test, too. The results (see Table 6) show that the differences of averages of the absolute values of *PCs* and *AA* in the two groups are significant. The nonlinear relationships with close—connected degrees in the responses of rain days to SST are further proved.

Table 5. The Values *PCs* in Years

Year	SST	PC1	PC2	PC3	PC4	PC5	AA
1955	−1.37	0.77	−0.03	−1.39	0.85	2.19	1.27
1956	−1.16	−0.86	−2.81	1.03	2.09	−2.68	2.06
1964	−1.37	−0.50	0.07	1.85	0.71	1.63	1.17
1970	−1.02	0.10	0.15	−0.38	−0.91	−0.72	0.58
1974	−1.02	0.33	0.52	−1.02	−0.93	0.01	0.67
1975	−1.63	1.60	−0.92	−2.44	−0.91	−0.57	1.45
1957	1.85	1.31	3.07	−1.33	2.16	−2.37	2.15
1958	1.03	0.17	−1.37	−0.53	−2.01	−0.04	1.11
1963	1.03	−0.14	0.65	−0.16	−0.65	−0.13	0.43
1965	1.64	−1.42	1.79	1.67	−2.78	−0.87	1.82
1969	1.17	0.27	1.17	1.37	1.79	1.50	1.32
1972	2.25	−4.36	0.21	−2.00	1.02	0.52	2.21
1976	1.23	0.66	−1.85	−0.47	0.15	0.83	0.98

Table 6. The *t*'s Values for Testing the Difference of Averages of the Absolutes and Values of *PCs* in Two Groups under the Abnormal and Normal SST

Component		1	2	3	4	5
Unrotated	*PC*	−0.5	0.2	−0.7	0.1	−0.3
	|*PC*|	2.4	5.3	3.0	9.0	2.5
Rotated	*PC*	0.1	−0.5	−0.4	−0.6	−1.0
	|*PC*|	2.5	2.1	6.1	3.1	6.1

IV. REGIONALIZED CHARACTERISTICS

The regionalized characteristics, furthermore, can be better represented by means of a simple structure rotation method, which is called rotated component analysis (RPCA), as shown by Richman(1986). Five representative subregions are shown in Fig.3 where the number of the subregion is associated with the corresponding rotated *PC*. The absolute values of the loading on *PCs*, which are greater than 0.5, are drawn in Fig.3. In order to emphasize the response of the anomalies of rain days to the abnormal SST in regions, the similar comparisons have done and shown in Table 7 using the rotated *PCs*. One can find, from Table 7, that the strongly negative responses of rain days to the warm SST in subregion 1 (i.e. North of China in Fig.3) occurred in 1965 and 1972. Serious droughts occurred in the subregion in the years. The strongly positive responses on the warm SST in 1957 and 1972 occurred in subregion 2 and 4, respectively. When the strongly negative anomaly of SST occurred in 1956, the strongly negative correlation moment controlled over subregion 2 (i.e. Northeast of China). However, the strongly positive value of variable *c* occurred in 1957 and 1965 in the warm years of SST. Obviously, there is nonlinear relationship between SST

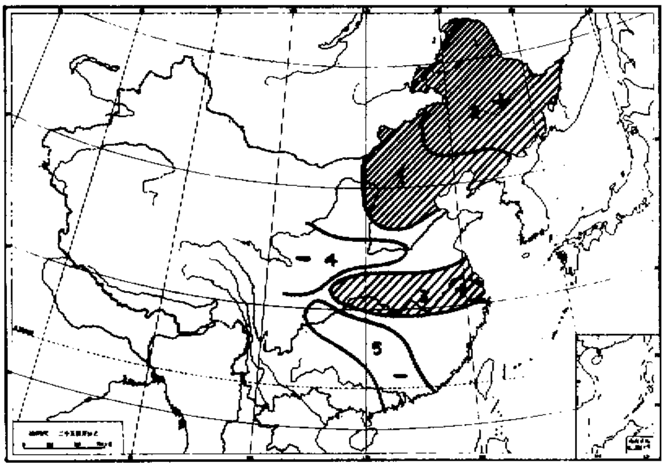

Fig.3. Five homogeneous regions as depicted by maximum rotated component loadings using maximum variance rotated method for variable *c*.

Table 7. The Rotated PCs of Variable c

Year	SST	PC1	PC2	PC3	PC4	PC5
1955	−1.37	1.90	0.35	−0.16	−0.67	1.33
1956	−1.16	−0.62	−3.37	−1.48	2.73	−1.21
1964	−1.37	0.04	−0.92	2.36	−0.12	0.82
1970	−1.02	−0.42	0.57	−0.55	−0.42	−0.52
1974	−1.02	0.18	0.78	−0.49	−0.71	0.16
1975	−1.63	1.29	0.79	−2.59	−1.09	−0.97
1957	1.85	1.16	2.98	−0.88	3.41	−0.18
1958	1.03	−0.43	−0.46	−1.03	−2.04	−0.82
1963	1.03	−0.50	0.75	−0.04	−0.25	−0.06
1965	1.64	−3.56	1.62	0.94	−0.45	−1.47
1969	1.17	0.90	0.12	2.28	1.17	1.80
1972	2.25	−2.12	−0.46	−1.38	0.48	4.21
1976	1.23	1.25	−1.40	−0.47	−1.04	0.07

and rain days in that subregion. In fact, serious floods occurred in Northeast of China in the two years. The nonlinear relationships between them strongly presented in subregion 3 in which the positive response occurred in warm (SST) years (such as in 1964 and 1969) but the negative response occurred in cold (SST) year (e.g. in 1975) generally. The result using *t*–test shows that the difference of the averages of absolute *PC*3 under the conditions of abnormal and normal SST is very significant (see Table 6). Because the first three *PCs* have the most percentages of explained variance the corresponding subregions (i.e. North and Northeast of China) can be considered strongly presentation of the nonlinear correlations in the response area.

V. CONCLUSIONS

The relationships between rain days in 46 stations over China and SST in the Pacific of the equator region in summer have been analysed in this paper using correlation coefficient between them. The results show that there are weak negative correlations between them. However, the significant test shows that their correlation fields are not significant, and the linear relationships are obscure.

The correlation moment between SST and rain days can reveal their correlations in some years when the extremes of SST occurred. Furthermore, the spatial and temporal patterns can be extracted by PCA and RPCA based on the field of the correlation moment between them.

There are some nonlinear relationships between rain days over China and SST by means of the methods mentioned above. When the extreme anomalies of SST occurred in the years, the amplitudes of correlation moment between them also occurred with the extreme values. The years are consistent with El Nino and non–El Nino ones. It implies that when the background of the atmospheric circulation has suddenly changed the surface meteorological elements have the strong response in variations in the years.

The results show that the correlation moment analysis proposed in this paper is more powerful than the traditional correlation analysis for researching the relationship between two variable fields.

REFERENCES

Ananthakrishnan, R. and B. Parthasarathy (1984), Indian rainfall in relation to the sunspot cycle, *J. Climatology*, 4: 149–169.

Angell, J.K. (1981), Comparison of variation in atmospheric quantities with sea surface temperature variations in equatorial Eastern Pacific, *Mon. Wea. Rev.*, **109**: 230–243.

Chung, J.C., (1982), Correlations between the tropical Atlantic trade winds and precipitation in Northeastern Brazil, *J. Climatology*, **2**: 35–46.

Englehart, P.J. (1985), A statistical analysis in the conterminous United States, including comparisons with precipitation totals, *J. Climate and Appl. Meteor.*, **24**: 350–362.

Harnack, J. and R.P. Harnack (1984), A Southern Hemisphere sea level pressure data set for use in climatic studies *J. Climatology*, **4**: 187–204.

Horel, J.D. (1984), A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500hPa height field, *Mon. Wea. Rev.*, **109**: 2080–2092.

Hsu Huang–Hsiung and J.M.Wallace (1985), Vertical structure of winter– time teleconnection patterns, *J. Atoms. Sci.*, **42**: 1693–1710.

Kawamuka, R. (1984), Relation between atmospheric circulation and dominant sea surface temperature anomaly patterns in the North Pacific during the Northern winter, *J. Met. Soc. Japan*, **62**: 910–916.

Livezey, R.E. and W.Y. Chen (1983), Statistical field significance and its determination by Monte Carlo techniques, *Mon. Wea. Rev.*, **111**: 46–59.

Nitta, T. (1986), Long–term variations of cloud amount in the Western Pacific Region, *J. Met. Soc. Japan*, **64**: 373–390.

North, G.R., T.L. Bell and R.F. Cahalan (1982), Sampling errors in the estimation of empirical orthogonal functions, *Mon. Wea. Rev.*, **110**: 699–706.

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

Richman, M.B. (1986), Review article, rotation of principal components, *J. Climatolgy*, **6**: 293–355.

Wallace, J.M. and D.S. Gutzler (1981), Teleconnections in the geopotential height field during the Northern Hemisphere winter, *Mon. Wea. Rev.*, **109**: 784–812.

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