Application of Principal Component Analysis in Long-Range Forecast of Mei-Yu and Typhoon

by Ernest C. Kung
Department of Atmospheric Science
University of Missouri-Columbia
Columbia, Missouri 65211 USA

Abstract:

A systematic investigation of principal components of the circulation in eastern Asia has begun for the development of a long-range forecast scheme of Mei-Yu and typhoon in Taiwan. The data include twice-daily 1957-86 NMC Northern Hemisphere octagonal grid data, monthly sea surface temperature from 40°S to 60°N for 1957-86, and twice-daily FGGE 1.875° x 1.875° grid data.

With the monthly and daily means of surface and upper air fields, major principal components are being identified for a region of 0°-50°N and 60°-160°E prior to, during, and after Mei-Yu. Within this region a fine mesh area of 10°-35°N and 100°-150°E is established for the diagnosis of principal components with twice-daily data. With both the large outer and small inner areas, principal components in the daily synoptic series during Mei-Yu will also be obtained for the FGGE year.

Individual principal components are expected to describe different dynamic and thermodynamic responses of the circulation to various forcing functions at different time and space scales. Thus, through identification of major principal components and their global teleconnections, we may examine the predictability of Mei-Yu related circulation for independent sources of variability. This identification is the key in selecting the proper predictands and predictors, particularly the former. After the selection of proper predictands, the multiple regression forecast scheme will be constructed for seasonal ranges.

The construction of the typhoon long-range forecast scheme will proceed in a similar manner. However, the principal components of the circulation will be used as predictors only.

1. Introduction:

Before the eventual goal of the long-range forecasting capability of the mathematical-dynamical models is reached, it is necessary to develop empirical, regional forecasting schemes at the seasonal range. Development of such schemes would not only benefit economic and societal activities, but would also contribute to the physical understanding of the behavior of the atmospheric general circulation through knowledge acquired empirically. As suggested by the NOAA work group on global climate change (1987), identifying the spatial and temporal variations at various scales is an important mission-oriented research goal, on which the development of worldwide prediction capabilities of natural variations is based.

One of the possible approaches to the long-range forecast is the application of principal component analysis. One of its specific advantages is that through identification of major principal components we may examine the predictability of the circulation system for independent sources of variability. At the University of Missouri-Columbia, comprehensive research is in progress on principal components and the global teleconnection of the Asian summer and winter monsoons As part of the Asian monsoon study, we are initiating principal component analysis and forecast model construction for Mei-Yu in Taiwan. In a separate effort, we are also attempting the typhoon forecast in the long-ranges.

This paper reports the scientific and technological background of the Mei-Yu and typhoon research and outlines the systematic research plan that has just begun. In applying the principal component analysis in the forecast schemes, Mei-Yu and typhoons pose different problems and thus necessitate different approaches. This report concentrates on the Mei-Yu scheme and makes pertinent references to typhoons. Because the research is still in its initial stage, the results have not been compiled at the time of the writing of this paper. The initial results thus will be shown at the time of the oral report in the symposium. However, this paper by itself is complete in describing the systematic research plan.

2. Background and approach:

As revealed in recent years, during the concentrated study and monitoring of the El Niño/Southern Oscillation (ENSO) phenomena, the regional manifestations of the ocean-atmosphere system are mostly associated with the anomalous development of global atmospheric circulation patterns, often at distant locations. Since most major perturbations of the coupled ocean-atmosphere system are consequences of nonlinear interactions of many thermodynamic and dynamic processes, the regional conditions of the global atmosphere exhibit considerable interannual variability with certain antecedent conditions of the general circulation. Hastenrath (1986) reviewed recent progress in tropical long-range prediction and attributed the apparent success of Kung and Sharif (1982) and Wu and Hastenrath (1986) in the multiple regression forecast of the Indian summer monsoon rainfall and onset date to the comprehensive diagnostic approach of climate and circulation anomalies. Shukla and Mooley (1987), with their selection of proper predictors, also indicate noticeable predictability of the Indian summer monsoon rainfall. Kung and Tanaka (1985) extended the multiple regression approach to the monthly temperature and rainfall in various middle-latitude locations, and confirmed similar predictability of these parameters.

Despite clearly existent predictability, the development and research of long-range forecasting schemes are often hampered by complex variations of the general circulation patterns. To clarify the often confusing cross-correlation patterns, one remedy is to use a very long time series as the basis of the statistical analysis. However, since the mode and prevailing patterns of the general circulation oscillate over the period of one decade or two, the use of long-term data beyond 20 years does not substantially contribute to the improvement of the forecast (see Kung and Sharif, 1980, 1982). Since the meteorological parameters used as predictands and predictors are complex manifestations of various scales of variability, it is preferable to resolve patterns of meteorological fields into a manageable number of components and explore the predictability of such components.

The utility of principal component analysis in describing the important modes of variance has been demonstrated by Kutzbach (1967, 1970), Sellers (1968), Kidson (1975), Trenberth (1975), Weare (1977), Heddinghaus and Kung (1980), Barnett (1983), Kawamura (1984, 1986), Lau and Chan (1985), and others. Kutzbach (1967), Stidd (1967), and Davis (1976) describe the procedures for computational analysis of orthogonal pattern vectors and their coefficients. Through principal component analysis, we may reduce a large number of variables into a set of components while retaining the maximum variance of the original variables. In this way we may concisely describe the interrelationship among observed variables. If a large number of observed variables are highly correlated to one another, a principal component may effectively represent the variation of a fundamental mode of observed variables which are highly loaded on the component. The predictability of such a group of variables can be evaluated by examining the correlation between coefficients of the principal components and antecedent variables (or their principal components). By using the principal components as predictands, we can investigate the predictability of regional circulation as a linear combination of independent sources of variability.

In our pilot study (Park and Kung, 1987), we examined the first three principal components of the North American summer temperature field with regard to the antecedent oceanic and atmospheric conditions of the Northern Hemisphere, and showed them to have considerable prognostic value (Fig.1). The cross-correlation patterns and their seasonal variations indicate that the anomalous sea surface temperature (SST) fields of the preceding seasons in the North Pacific are closely related to principal components of the temperature field in North America of the following summer (Figs. 2 and 3). The signal of anomalous SST fields appears early in the preceding fall, reaches the maximum in winter, and diminishes toward summer. As shown in Fig. 1, the summer temperature in the eastern and midwestern United States is determined by the first principal component, whereas in the southern and upper midwestern United States it is determined by the second component, and in the western United States by the third component. In this preliminary study an ENSO mode is detected for the North American summer temperature, indicating the impact of ENSO on the general circulation beyond the winter half of the year. The ongoing phase of our preliminary study is concerned with the North American summer precipitation, and the useful predictability of its principal components is indicated as in the case of summer temperature.

In the ongoing research the pilot study, as described above, is expanded and systematically applied to the Asian summer and winter monsoons. Among major systems of the global circulation, the Asian summer and winter monsoons are ideal for investigating principal components and their seasonal-range predictability. The summer monsoon is a prominent, persistent feature of the Northern Hemisphere, which involve wide ranges of spatial and temporal variations with distinct characteristics from planetary to local scales (see GARP Publication Series, 1976; Lighthill and Pearce, 1981). These characteristic scales of variations exist in response to various scales of forcing functions, such as plateau heating, movement of Hadley cells, land-sea temperature contrast, heating over the ocean, expansion of the southwestern Pacific high pressure, and others. Thus it is reasonable and desirable to resolve the field of monsoon circulations into principal components to be examined with regard to various forcing functions. We may then study the global teleconnection patterns of such principal components for the predictability associated with each component.

In view of the successful evaluation of the predictability associated with principal components of the North American summer temperature and precipitation patterns, which are for a transient eddy region of the middle latitudes, we may reasonably expect that the basic approach can be readily applied to the summer and winter monsoons, which are persistent standing eddy phenomena. The long-range forecast studies of the summer monsoon by Kung and Sharif (1980, 1982), Shukla and Mooley (1987), and Wu and Hastenrath (1986) indicate considerable predictability for monsoons yet to be explored with proper resolutions of predictand and predictor patterns. In major government forecast centers in Asia, empirical schemes are used with mixed results for the long-range outlook of monsoon patterns (e.g., Chu, 1978; Japan Meteorological Agency, 1981; Tseng, 1982; Thapliyal, 1982). Examination of their records indicates that monsoons are better forecasted when a clearly distinguishable antecedent pattern develops, again signifying the existence of considerable predictability and the need for adequate resolution of flow patterns into components.

The large population of the Far East depends on the monsoons for its agricultural production and water supply, directly and indirectly affecting global economic and political stability. Basic research for the seasonal-range empirical forecasting schemes of summer and winter monsoons is thus crucially important for worldwide agricultural and economic planning. For the development of meteorological science and forecast operation, the systematic diagnostic study planned in this proposal will provide important basic information, not only for the formulation of long-range forecasting schemes, but also for comprehensive understanding of the global circulation in reference to monsoon systems. As recognized in previous studies (e.g., Krishnamurti, 1971, 1985; Kung and Chan, 1981), monsoons represent a major energy source of the general circulation. The large amplitude features of monsoon circulations contribute decisively to much of the variance of the general circulation field. Thus monsoon features must realistically be incorporated into the numerical models of the general circulation (e.g., Manabe et al., 1974). The information obtained will serve not only to verify, but also to help formulate general circulation models.

With the background and general approach as discussed above, specific objectives of our ongoing research may be summarized as follows:

1. Systematic identification of characteristic spatial and temporal variations of the summer and winter monsoons in terms of principal components.
2. Systematic diagnosis of global teleconnections among principal components of selected monsoon predictands and predictors.
3. Examination of ENSO mode in the global teleconnection patterns.
4. Linear combination of useful principal components for independent sources of variability in the regional monsoon prediction.

Mei-Yu is a phenomenon in the summer monsoon system, but the study of its long-range forecast in Taiwan is complicated by additional small-scale influences in the region, due to the fact that Taiwan is an isolated island at the edge of the entire monsoon system. Though not yet identified, we may expect these small-scale variations to be isolated in some high components of the principal component analysis. Those small-scale variations are expected to possess different predictability of their own scales from those of the major Mei-Yu system which dominate all of eastern Asia. The design of Mei-Yu study in Taiwan thus must be handled separately from the general study of the Asian summer monsoon.

3. Data

The basic data in this study include the following sets of observations:

1. National Meteorological Center (NMC) Northern Hemisphere analysis data (1955-1986).
2. Australian National Meteorological Research Center (ANMRC)

Southern Hemisphere analysis data (1973-1980).

3. European Centre for Medium-Range Weather Forecast (ECMWF) global analysis data (1980-1986).

4. First GARP (Global Atmospheric Research Program) Global Experiment (FGGE) Level IIIb data sets.

5. Monthly mean SST data (1950-1986).

6. Seasonal SST anomalies over the eastern tropical Pacific (1868-1986).

7. Rainfall time series data over India (1871-1984).

8. Monthly rainfall for 300 stations over China (1951-1980).

9. Monthly sea level pressure for Darwin (1882-1986).

The special regional data sets of Taiwan are:

10. Taiwan station records of precipitation (1941-1987).

11. Records of Mei-Yu onset, recess, and special observations (1941-1987).

12. Taiwan area mesoscale experiment (TAMEX) data (1987).

13. Typhoon records in Taiwan (1945-1987).

This research project has all the basic data sets listed above from 1 to 9. The monthly mean SST data are collected from various archives. In the earlier years they cover from 40°S to 40°N, and from 1970 they cover from 40°S to 60°N. Rainfall data over India and China, Darwin sea level pressure data, and eastern tropical Pacific SST anomaly data were obtained from the University of Maryland, courtesy of Professor J. Shukla.

The regional data in Taiwan will be compiled from the data archive of the Central Weather Bureau. It is likely that additional local data of Taiwan will be needed during the course of the study.

4. Research design

During the next three year period, the research of the Mei-Yu forecast scheme will be conducted in the following four steps. Steps 1 and 2 investigate the principal components of the circulation fields during Mei-Yu. Steps 3 and 4 attempt to use these components with the global predictors in multiple regression schemes of long-range forecast.

(1) Principal components of Mei-Yu circulation

Two grid systems for principal component analysis are established as shown in Fig. 4. The fine-mesh area is superimposed on the 4° x 5° general area so that the smaller-scale variations can be detected around the Taiwan region. The fine mesh of the inner grid area is taken as 2.5° x 2.5° in Fig. 4. The bounds of the large general area are 2°S-70°N and 60°-165°N; those of the inner area are 7.5°-60°N and 90°-150°E.

Orthogonal pattern vectors and their coefficients will be computed for principal components of the geopotential height and temperature fields in the lower, mid, and upper troposphere with monthly means of observations over the large grid area for each month of the year with NMC analyses. For the inner fine-mesh area, the computation will be done with the daily FGGE IIIb analyses during the monsoon months. Additionally the principal components of the monthly and daily precipitation patterns will be computed respectively over the large and fine-mesh areas of analysis.

Obtaining the principal components of height, temperature, and precipitation fields is the basic computational procedure in this research. There are many characteristic local manifestations in the broad planetary monsoon and Mei-Yu circulations. It is possible that important local modifications of planetary monsoons due to orographic barriers and various types of instability fail to distinguish themselves as major principal components. Thus principal component analysis in the restricted local area with the daily fine mesh data will bear as equal importance as the large area analysis. The bounds of large and small areas of analysis will be flexible and properly adjusted as necessitated by the study's progress and computational requirement.

(2) **Examination of Principal Components of Mei-Yu Circulation**

The principal components of Mei-Yu circulation obtained in the preceding step will be examined with regard to the following:

1. Change of monthly patterns in relation to the monsoon development.
2. Association of principal component patterns with specific Mei-Yu phenomena, including onset, recess, dry spell, active and weak Mei-Yu, etc.
3. Spatial and temporal scales of variations, including patterns of interannual variations.
4. Identifiable physical mechanisms at various scales.

After the examination, important principal components are selected for proper description of Mei-Yu processes. These components should either contribute an appreciable percentage of the total variance, or be associated with distinctive monsoon phenomena. These components are also best identifiable with physical forcings of comparable scales. They will be subject to further selection through global cross-correlation analysis to identify predictands in the regression forecast scheme.

(3) **Global Teleconnections**

Both the Northern and Southern Hemisphere analyses by NMC and ANMRC and SST data will be examined to select predictors for principal components identified in the preceding step. For this purpose two sets of variables (predictor variables) will be prepared for global observations and SST respectively to examine the teleconnections of principal components of monsoon circulations. One set is by averaging monthly mean grid values of observation-analysis for some neighboring grids to represent a larger scale than the original grid. The size of the new grid will be determined during the course of study. Another set is major principal components of the global observations and SST which will be obtained separately for the Northern and Southern Hemispheres.

To examine the predictability of principal components of monsoon circulations, the correlations of each component with antecedent predictor variables are computed and compared. Correlations will be computed with each predictor variable for each month of the three seasons preceding the monsoons. Through a comprehensive analysis of cross-correlation patterns, we will be able to assess the predictability of principal components of monsoon circulations and specific monsoon phenomena. Special attention will be paid to identify possible ENSO modes in these teleconnection patterns.

(4) **Multiple Regression Forecasting Scheme**

After the examination of global teleconnections, an appropriate number of principal components of Mei-Yu fields will be selected as predictands. The number of predictands is to be decided following the preceding step of the study. The predictands selected should be adequate to describe the pattern and development of Mei-Yu, and each of the selected predictands should possess sufficient predictability with identifiable antecedent circulation and SST patterns. By linearly combining these selected principal components as sets of predictands, the predictability of summer and winter monsoons, respectively, may now be examined with independent sources of variability. With identified predictands, a comprehensive description of antecedent elements and following patterns of Mei-Yu will be offered.

For each of the selected predictands for Mei-Yu, a multiple regression equation will be formulated with five predictors. The linear combinations of such regression equations will then be the long-range forecasting scheme of summer and winter monsoons. Formulation of a multiple regression equation is detailed in Kung and Sharif (1980, 1982) and Kung and Tanaka (1985). The forecast experiments with multiple regression equations are to be conducted in two ways. First the regression coefficients are evaluated with the entire time series data, but excluding the data of each forecast year (see Kung and Sharif, 1982; Kung and Tanaka, 1985). Data for the forecast year are used for the real time predictors. Second the regression coefficients are evaluated with the data of earlier years in the time series, and then forecasts are done for successive years following such an evaluation.

5. Typhoon

The description of Mei-Yu long-range forecast research is equally applicable to typhoon. However, except for the precipitation, the principal components of the circulation, including those of SST, are used only as predictors. The statistics of typhoons, such as frequency, strength, path, etc., preclude the use of principal components in their description. Utilities of principal components as predictors in typhoons may be seen from the study of Kawamura (1984, 1986). The first step in typhoon forecast must be the parameterization of typhoon statistics, which will be in progress this year.

6. Concluding remarks

Scientific and technical basis of principal component analysis for the purpose of long-range forecast of Mei-Yu and typhoons is discussed, and the procedures of model construction have been described. This is the first attempt in using principal components in an operational long-range forecast. The principal component analysis for the Mei-Yu circulation has begun, and the study of typhoons is in the stage of parameterizing the typhoon statistics.

7. References

Barnett, T. P., 1983: Interaction of the monsoon and Pacific trade wind system at interannual time scales. Part I: The equatorial zone. Mon. Wea. Rev., **111**, 756-773.

Chu, E. W. K., 1978: A method for forecasting the arrival of cold surges in Hong Kong. Tech. Note No.43, Royal Observatory, Hong Kong, 32 pp.

Davis, R. E., 1978: Predictability of sea level pressure anomalies over the North Pacific Ocean. J. Phys. Oceanogr., **8**, 233-246.

Global Atmospheric Research Programme, 1976: The monsoon experiment. Garp Publication Series No. 18, 123 pp.

Hastenrath, S., 1986: On climatic prediction in the tropics. Bull. Amer. Meteor. Soc., **67**, 696-702.

Heddinghaus, T. K., and E. C. Kung, 1980: An analysis of climatological patterns of the Northern Hemispheric circulation. Mon. Wea. Rev., **108**, 1-17.

Japan Meteorological Agency, 1981: Manual of one-month weather forecasting in Japan. Long-Range Prediction Tech. Note No. 23, 301 pp. (In Japanese).

Kawamura, R., 1984: Relation between atmospheric circulation and dominant sea surface temperature anomaly patterns in the North Pacific during the northern winter. J. Meteor. Soc. Japan, **62**, 910-916.

-----, 1986: Seasonal dependency of atmosphere-ocean interaction over the North Pacific. J. Meteor. Soc. Japan, **64**, 363-371.

Kidson, J. W., 1975: Eigenvector analysis of monthly mean surface data. Mon. Wea. Rev., **103**, 177-186.

Krishnamurti, T. N., 1971: Tropical east-west circulations during northern summer. J. Atmos. Sci., **28**, 1342-1347.

-----, 1985: Summer monsoon experiment--a review. Mon. Wea. Rev., **113**, 1590-1625.

Kung, E. C., and P. H. Chan, 1981: Energetics characteristics of the Asian winter monsoon in the source region. Mon. Wea. Rev., **109**, 854-870.

-----, and T. A. Sharif, 1980: Regression forecasting of the onset of the Indian summer monsoon with antecedent upper air conditions. J. Appl. Meteor., **19**, 370-380.

-----, and -----, 1982: Long-range forecasting of the Indian summer monsoon onset and rainfall with upper air parameters and sea surface temperature. J. Meteor. Soc. Japan, **60**, 672-681.

-----, and H. Tanaka, 1985: Long-range forecasting of temperature and precipitation with upper air parameters and sea surface temperature in a multiple regression approach. J. Meteor. Soc. Japan, **63**, 619-631.

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

-----, 1970: Large-scale features of monthly mean Northern Hemisphere anomaly maps of sea-level pressure. Mon. Wea. Rev., **98**, 708-716.

Lau, K.-M., and P. H. Chan, 1985: Aspects of the 40-50 day oscillation during the northern winter as inferred from outgoing longwave radiation. Mon. Wea. Rev., **113**, 1889-1909.

Lighthill, J. and R. P. Pearce (Eds.), 1981: Monsoon Dynamics, Cambridge University Press, 735 pp.

NOAA Office of Oceanic and Atmospheric Research/Office of Climatic and Atmospheric Research, 1987: Work Group Report--NOAA Global Climatic Change Workshop, 187 pp.

Manabe, S., D. G. Hahn, and J. L. Holloway, 1974: The seasonal variation of the tropical circulation as simulated by a global model of the atmosphere. J. Atmos. Sci., **31**, 43-83.

Park, C. K., and E. C. Kung, 1987: Principal components of the North American summer temperature field and the antecedent oceanic and atmospheric circulations. (Submitted for publication.)

Sellers, W. D., 1968: Climatology of monthly precipitation patterns in the western United States, 1931-1966. Mon. Wea. Rev., **96**, 585-595.

Shukla, J. and D. A. Mooley, 1987: Empirical prediction of the summer monsoon rainfall over India. Mon. Wea. Rev., **117**, 695-703.

Stidd, C. K., 1967: The use of eigenvectors for climatic estimates. J. Appl. Meteor., **6**, 255-264.

Thapliyal, V., 1982: Stochastic dynamic model for long-range prediction of monsoon rainfall in penisular India. Mausan, **33**, 399-404.

Trenberth, K. E., 1975: A quasi-biennial standing wave in the Southern Hemisphere and interrelations with sea surface temperature. Quart. J. Roy. Meteor. Soc., **101**, 55-74.

Tseng, C. F., 1982: Analysis and prediction of Taiwan Mei-Yu. Central Weather Bureau Research Paper, Taipei, Taiwan, China, 78 pp.

Weare, B. C., 1977: Empirical orthogonal analysis of Atlantic Ocean surface temperature. Quart. J. Roy. Meteor. Soc., **103**, 553-564.

Wu, M.-C. and S. Hastenrath, 1986: On the interannual variability of the Indian monsoon and southern oscillation. Arch. Met. Geoph. Biocl., Ser. B, **36**, 239-261.

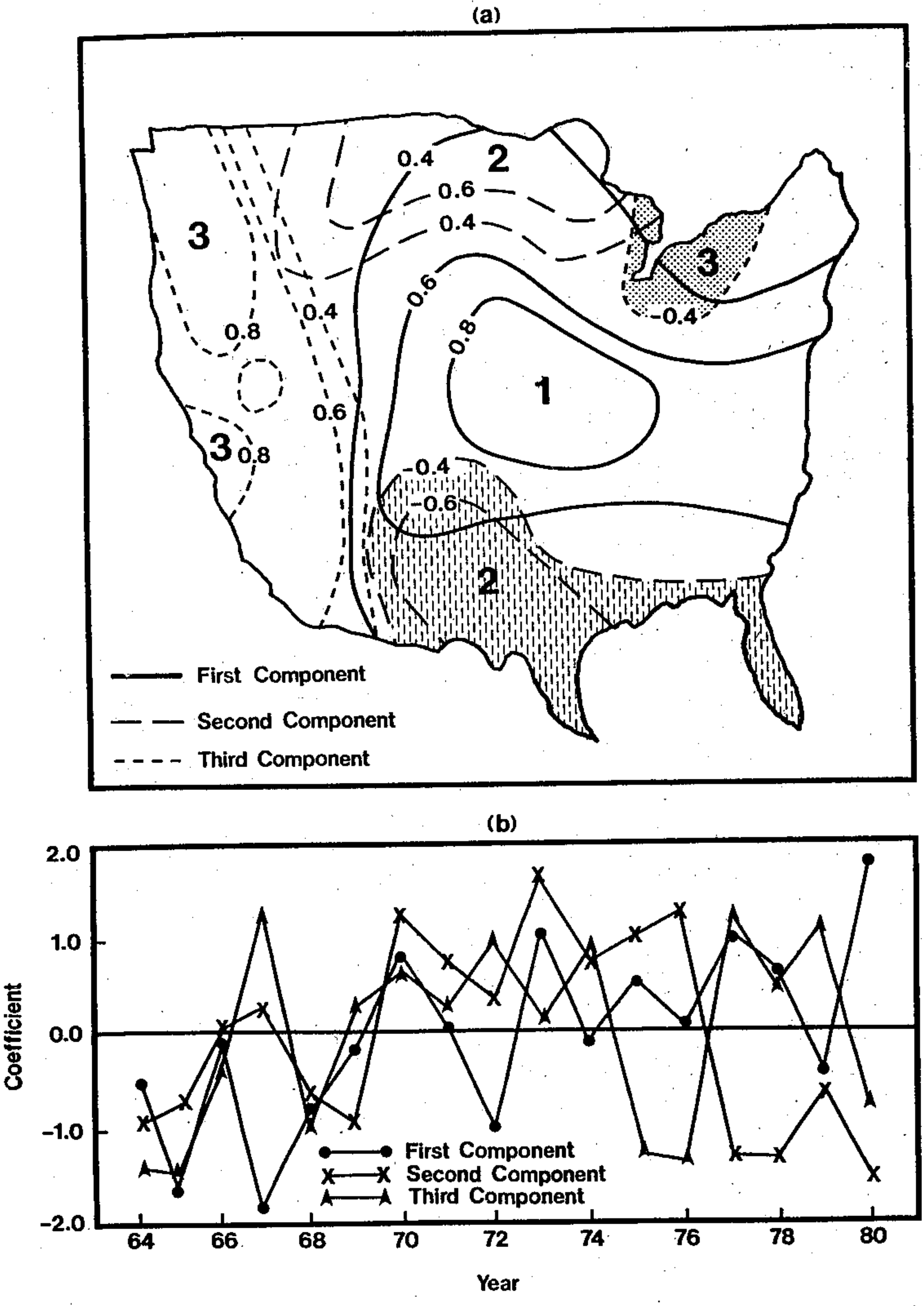

Fig. 1 (a) Characteristic patterns for the first three components of the summer surface temperature of North America and (b) the coefficients of the corresponding principal components.

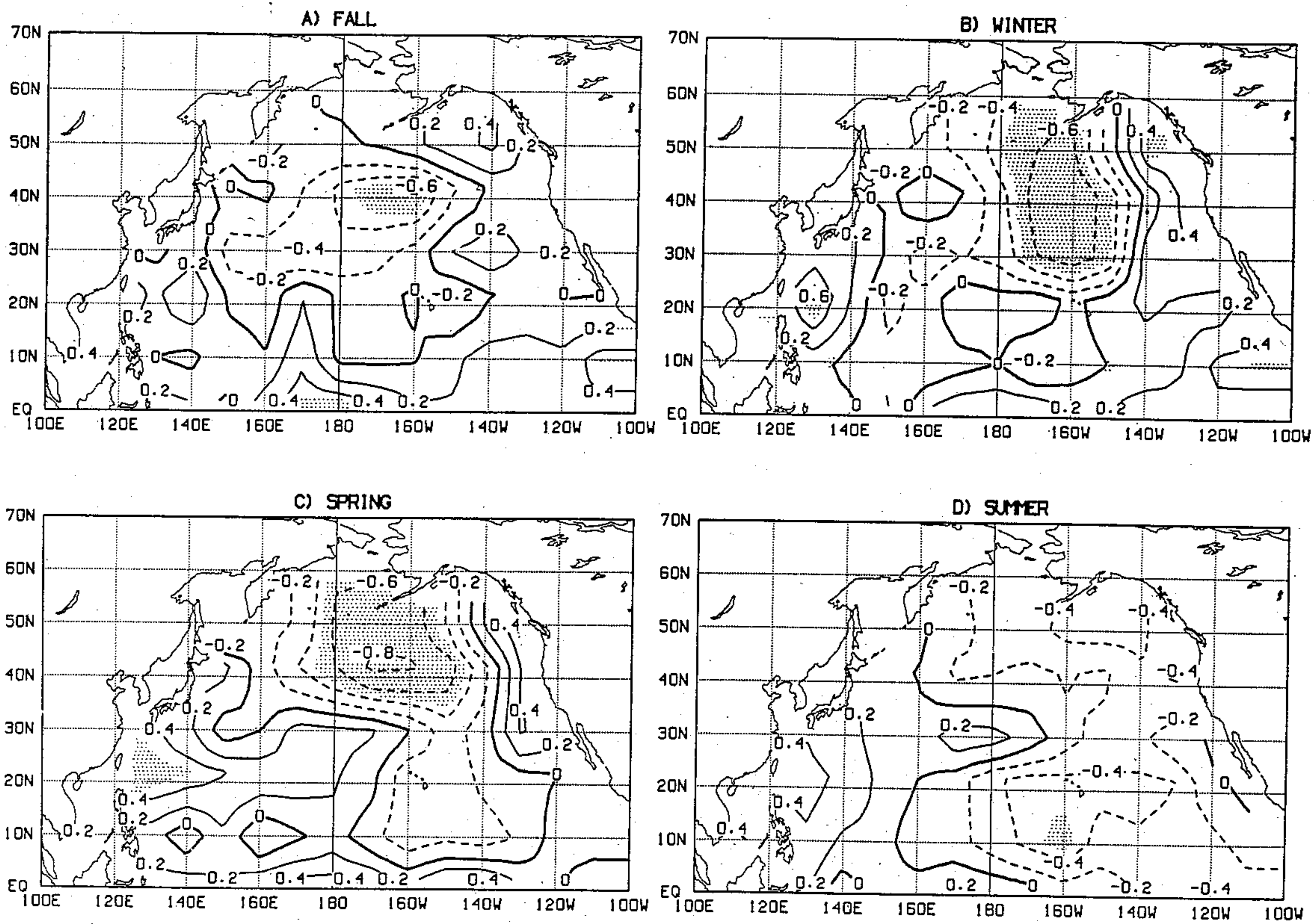

Fig. 2 Cross-correlation between the seasonal SST and the first component of the North American summer temperature. Shaded areas indicate correlations exceeding the 5% level of significance.

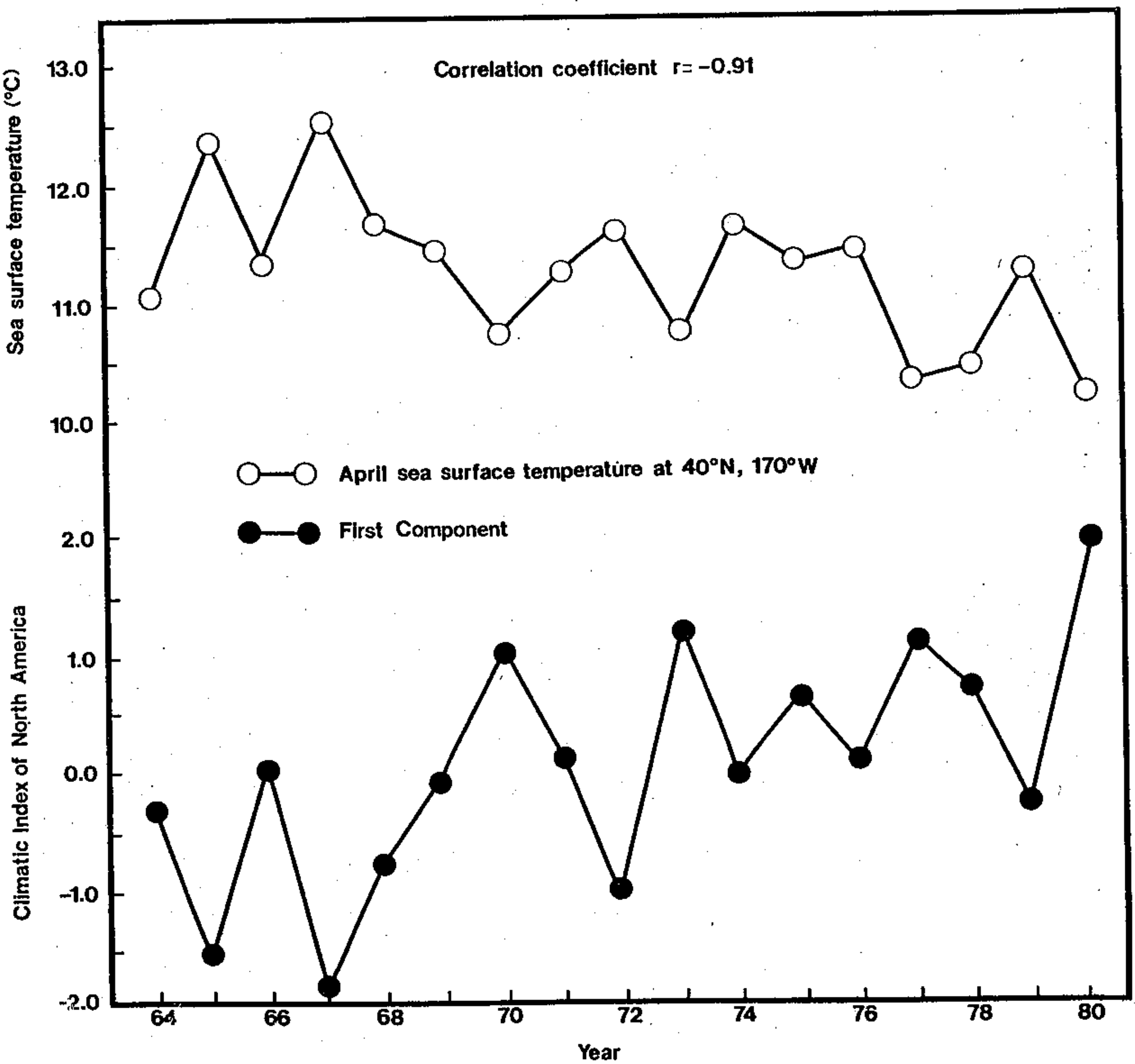

Fig. 3 Interannual variations of the central North Pacific SST (°C) in April and the coefficients of the first component of the North American summer temperature.

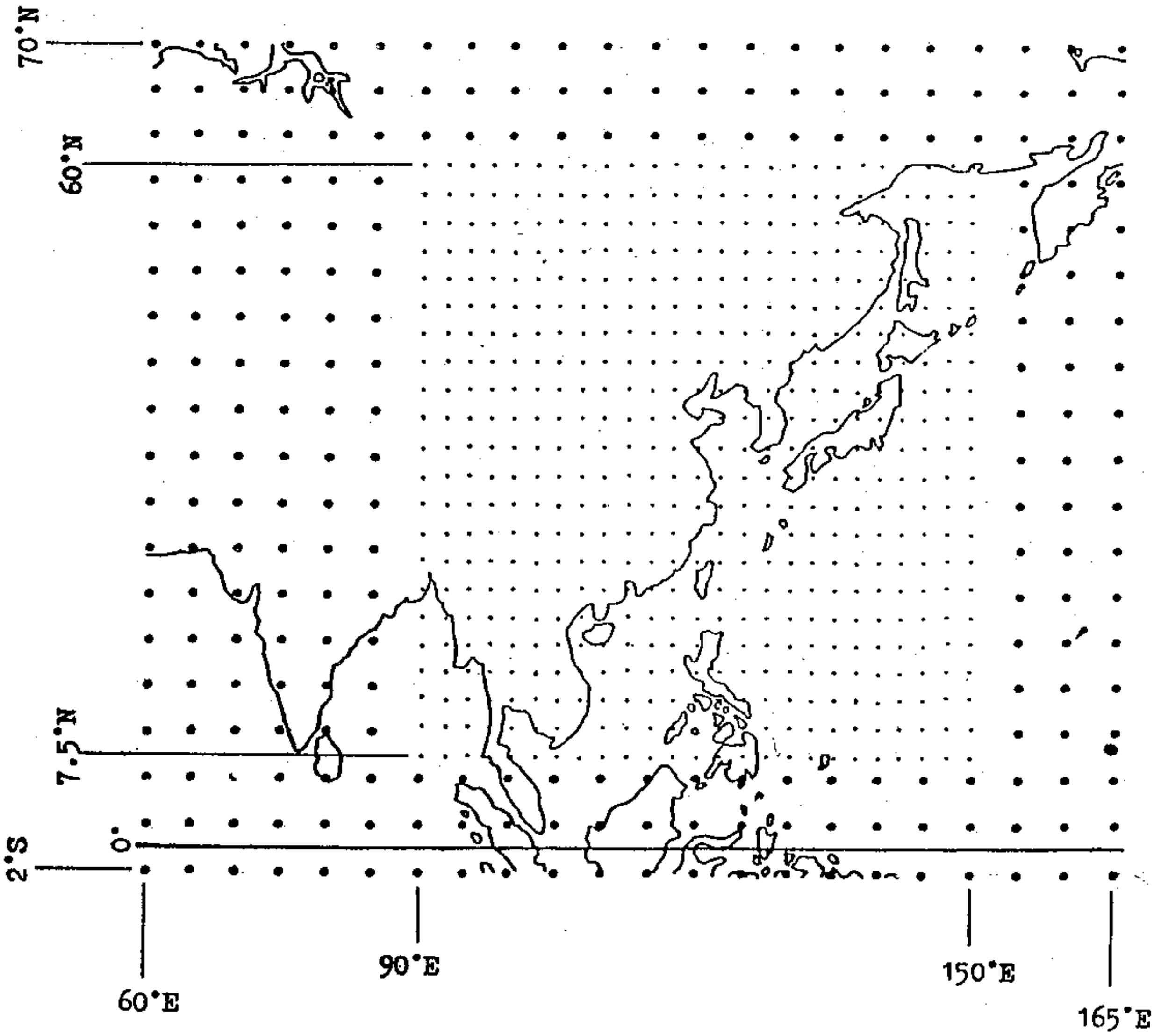

Fig. 4 Outer and inner grid areas for the principal component analysis.

主成份分析在梅雨及颱風長期預報上之應用

孔 震 村

密蘇里大學可倫比亞校區大氣科學系

摘 要

台灣區梅雨及颱風之長期預報發展上，對東亞環流型式已開始做有系統的主成份分析。所使用資料包含 1957 － 1986 年每日兩次之 NMC北半球八角網格資料，1957 － 1986 年40°S至60°N之逐月海平面月平均溫度資料，以及 FGGE觀測期間每日兩次之1.875° ×1.875° 網格分析資料。

用上述逐月及逐日平均地表與各層大氣場資料，對梅雨前、期間、過後之主要環流成份，進行綜合檢討，主成份計算的範圍有大小兩網格，大網格是 0° －50°N，60° － 160°E。小網格是10° －35°N，100° － 150°E。用小網格的目的是爲觀察梅雨期間主要成份的逐日中規模變化。

不同的主要主成份，可以描述對不同熱力及力學上的導引函數的大氣反應。而且這些反應表示在不同的時間與空間尺度。因此經過主要的主成份和其全球性遙地相關綜合分析，我們可就獨立個別的變化原因，調查梅雨環流形態的預報度。選擇適當主成份爲預報值與預報因子，乃是季節尺度長期復式廻歸預報方式的關鍵。

對颱風而言，亦可使用同樣方式，建立長期預報模式，但主成份僅可應用爲預報因子。至於預報值，則用各種季候函數，表示颱風長期的逐年觀測。