大 气 科 学 2002
Chinese Journal of Atmospheric Sciences
Volume 26, No.2, pp. 114–140

Formulation of the Canonical Ensemble Correlation Prediction for Seasonal Precipitation

Samuel S. P. Shen
(National Reserach Council, Washington, DC 20418, U.S.A. and Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada)
William K. M. Lau
(Climate and Radiation Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771, U.S.A.)
Kyu–Myong Kim
(Science Systems and Applications, Inc., Lanham, MD 20706, U.S.A.)
Guilong Li
(Atmospheric Science Division, Meteorological Service of Canada, 4905 Dufferin Street, Downsview, ON M3H 5T4, Canada)
Alan Basist
(NOAA National Climatic Data Center, Asheville, NC 28801, U.S.A.)

Manuscript received April 18, 2002.

This paper describes the mathematical formulation of the canonical ensemble correlation (CEC) forecasting model proposed by Lau et al. The model was developed mainly for forecasting the seasonal precipitation over the United States but also can be used to make long term forecasts of other climate parameters. In CEC, each individual forecast is made through a multivariate regression using canonical correlation analysis (CCA) in the spectral spaces whose bases are empirical orthogonal functions (EOF). Optimal weights are used to make ensemble forecasting and crucially depend on the mean square error of each individual forecast. An estimate of the mean square error of a CCA prediction is also made by using the spectral method. The error is decomposed onto EOFs of the predictand and decreases linearly according to the correlation between the predictor and predictand. Since the new CCA scheme is derived for continuous fields of predictor and predictand, an area–factor is automatically included. Thus our model is an improvement of the spectral CCA scheme of Barnett and Preisendorfer. The improvements include (i) the estimation of prediction error, and (ii) the optimal ensemble of multiple forecasts. The CEC model is applied to the seasonal forecasting of the United States (US) precipitation field. The predictor is the sea surface temperature (SST). The US Climate Prediction Center's reconstructed SST (1951–2000) is used as the predictor's historical data. The US National Center for Environmental Prediction's optimally interpolated precipitation (1951–2000) is used as the predictand's historical data. CEC optimally utilizes intrinsic sea surface temperature (SST) variability in major ocean basins. While the tropical Pacific, i.e., El Niño, contributes to the largest share of potential predictability in the southern tier States during boreal winter, intrinsic SST variability in the North Pacific and the North Atlantic is responsible for enhanced predictability in the northern Great Plains, Midwest and the southwest US during boreal summer. Most importantly, CEC significantly reduces the spring–summer predictability barrier over the conterminous United States, thereby raising the skill bar for dynamic predictions.

Key words: seasonal prediction; canonical ensemble correlation method; prediction error; the United States precipitation.

1. INTRODUCTION

Lau et al. (2002) introduced the CEC (canonical ensemble correlation) method for seasonal climate prediction. The method optimally utilizes intrinsic sea surface temperature (SST) variability in major ocean basins and yields a remarkable (10%–20%) increase in baseline prediction skills for seasonal precipitation over the United States for all seasons. This increase significantly reduces the spring–summer predictability barrier over the conterminous United States (US). The purpose of this paper is to describe the mathematical formulation of the CEC method.

CEC is used to maximize the predictability from multiple predictors while the prediction from each individual predictor is made by a linear regression using canonical correlation analysis (CCA). For example, to maximally extract precursory signals from global SST, the world ocean is partitioned into non–overlapping sectors: Tropical Pacific (TPAC), North Pacific (NPAC), Tropical Atlantic (TATL), North Atlantic (NATL), and Indian Ocean (IND). Each ocean basin can be counted as a predictor for making a forecast. The CEC forecast is then obtained at each grid box as a weighted average of the individual forecasts. In the making of each individual forecast, some improvements have been made in the spectral CCA scheme of Barnett and Preisendorf (1987), called BPCCA: an area–factor is included to reduce the noise of the data over the high latitude areas, and an error estimate is made to give a statistical assessment of the forecast quality.

CEC is a statistical model for long term forecasting. Why is another statistical model needed? Despite the great potential of the coupled GCM, its long–term forecasting ability is still limited because of numerous unknown mechanisms, including land–surface processes and cloud structure. The chaotic nature of nonlinear dynamics requires both spatially and temporally dense observations to limit the computing drift when solving a stack of nonlinear differential equations in a GCM. It is not conclusive yet whether the statistical or dynamic model is the best for forecasting. At the US Climate Prediction Center (CPC), it was stated in 1998 that " neither the dynamic nor the statistical models, as a group, perform significantly better than the other" (Barnston and He, 1998). CCA, OCN (optimal climate normals), and CA (constructive analogue) are all offcial statistical forecasting models used in CPC (Barnston and Smith, 1996). These statistical results are compared with the coupled GCM model forecasts. A subjective determination is made when announcing the NOAA's offcial long term forecasting. The classical CCA algorithm requires an inversion of cross–covariance and auto–covariance matrices. In the climate forecasting practice, all these matrices are not full rank because of the short history of climate observations. The BPCCA scheme in the EOF space avoided the inversion of matrices of non–full rank. BPCCA was a milestone work and made the use of CCA possible as an operational forecasting tool. According to Kim and North (1998; 1999), CCA is an accurate method compared with other statistical methods, such as the methods of POP (principal oscillation pattern) (von Storch and Zwiers, 1999), Markov (Xue et al., 2000), EOF extrapolation (Kim and North, 1998), and MCA (maximum covariance analysis) (Lau and Wu, 1999). Because of our error assessment formula, the mean forecasting quality at a given location is assessed simultaneously as the forecast is issued.

These error estimates are crucial to optimally combine the forecasts from different models and data to form an optimal ensemble forecast.

US seasonal precipitation is a difficult predictand because of its large variance. However, the clear impact of El Niño / Southern Oscillation (ENSO) on the US recitation indicates that SST does have potential predictability for the US precipitation, but the prediction skill in traditional methods drops dramatically in the spring, reaches a minimum in the warm season, and rises steadily from fall to winter. The dramatic reduction in forecast skill from winter to summer through the spring season is known as the "spring predictability barrier," which has been endemic in both statistical and dynamic forecasts of El Niño (Chen et al., 1997). Recently, Wang et al. (1999) found significant predictability from tropical and extratropical Pacific SST on warm season precipitation over the upper Great Plains and Atlantic States of the US during El Niño summers. The increased predictability is a quantitative validation of earlier findings on the relation between the US precipitation and the tropical and North Pacific SST (Ting and Wang, 1997; Ropelewski and Halpert, 1986). However, the forecasting skill was still relatively low in summer, even during the time of a strong SST signal in the tropical Pacific.

It has been suggested that the reduced precipitation predictability in the summer over the US stems from the weaker and more poleward position of the upper level westerly flow in the northern hemisphere, making it difficult for tropical SST influence to be transmitted to the US continent (Lau and Peng, 1992). As a result, the influence of tropical Pacific SST on the US summertime precipitation diminishes significantly. However, SST variability in other ocean basins, especially the extratropics, may begin to have an impact on US precipitation in summer (Higgins et al., 2000). In addition to SST, other factors such as soil moisture, snow cover and vegetation may influence US precipitation predictability for both summer and winter. The CEC of Lau et al. (2002) provides a scheme that can systematically explore potential predictability from various factors and be applied to climate predictions.

The numerical results of this paper are generated using the the US Climate Prediction Center's reconstructed SST (1951–2000) as the predictor's historical data. The US National Center for Environmental Prediction's optimally interpolated precipitation (1951–2000) is used as the predictand's historical data. Our forecast experiments show that the new ensemble canonical correlation scheme renders a reasonable forecasting skill.

The content of this paper is arranged as follows. The basic concept of CEC is described in Section 2. The prediction as a multi–variate regression of the canonical variables is presented in Section 3. Section 4 studies the prediction error estimation. Section 5 shows canonical ensemble forecasts. Numerical results are displayed in Section 6. Section 7 provides summary and conclusions.

2. BASIC CONCEPT OF CEC

Fig. 1 shows the basic concept of CEC. Given an initial climate state, denoted by Ω_i, the real climate system evolves into a climate state Ω_f at some future time $\Delta t > 0$. Because of changing boundary conditions and the chaotic nature of climate evolution from the initial state, Ω_i or states very close to Ω_i, a range of possible future states, within the "event cone", R, may be possible. A given predictor, generating a set of canonical variables, may produce an event cone P_1, which maximizes the overlap (or minimizes the distance metric) with R. If P_1 and R have a large overlap, then P_1 is considered a good predictor that can yield useful forecasts for Ω_f. Hence, P_1 here signifies the predictor data, prediction model and prediction results. Other predictors such as P_2 may produce a lesser overlap with R, but may capture different climatic states spanned by the event cone R, not covered by P_1. This situation can

be applied to other predictors P_n, which span yet other portions of the event cone R. The objective of the CEC is to construct, based on the totality of individual canonical correlation predictions, an "ensemble event cone" that maximizes the overlap with R, taking into account all possible outcomes. In CEC, the number of predictors, represented by P_n, is unlimited, and the predictors are not subject to orthogonality constraints.

Fig. 2 shows the procedures of CEC by using five ocean basins' SST as predictors for the US seasonal precipitation: Tropical Pacific (TPAC), North Pacific (NPAC), Tropical Atlantic

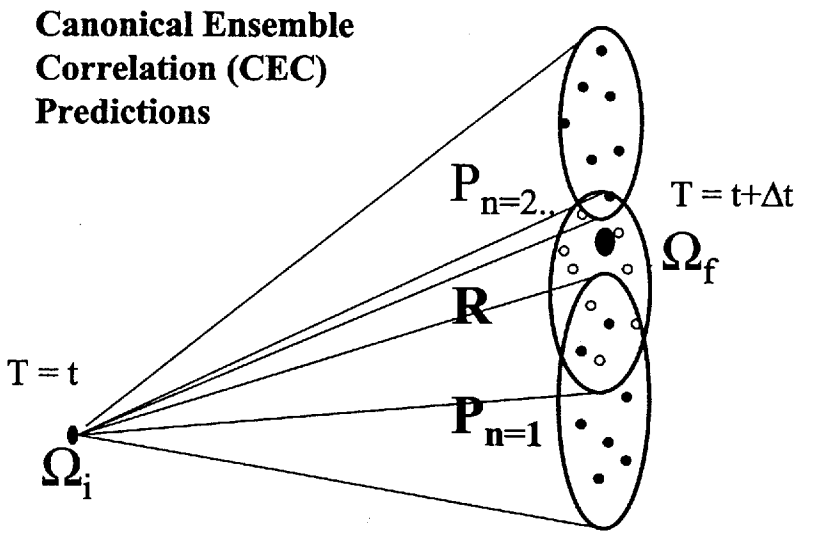

Fig. 1. Schematic diagram illustrating the basic concept of canonical ensemble prediction.

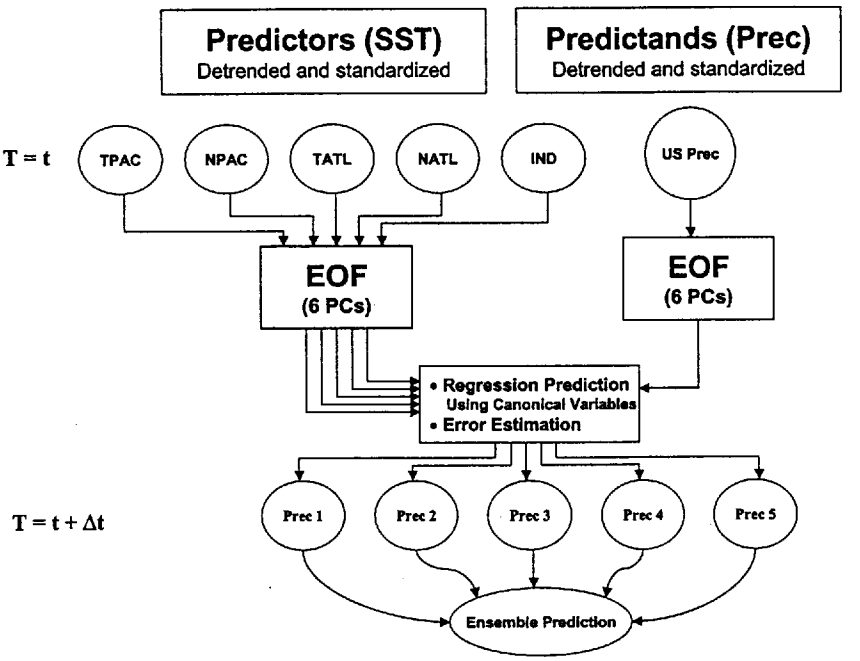

Fig. 2. The flow chart of canonical ensemble correlation (CEC) algorithm for forecasting.

(TATL), North Atlantic (NATL), and Indian Ocean (IND). Separate forecasts each using up to 6 canonical variables are made based on different ocean sectors. This step is crucial, because predictions from individual sectors allow the natural SST variability in that sector to be fully utilized in prediction. The CEC forecast is then obtained at each grid box as a weighted average of the individual forecasts. Various kinds of weighting schemes can be used, such as uniform weighting, maximum skill weighting, super–ensemble weighting, and error weighting. The maximum skill weighting uses only the forecast from the predictor that has the maximum skill for a given grid box. The super–ensemble weights are determined by a linear regression. The error weights are determined by the error of each predictor by using the principle of minimum mean square error. Extrapolation weighting can be used as well. In this weighting scheme, some weights are negative, corresponding to the forecasts that serve the role of correction.

The CEC prediction model is based on linear regressions that maximize the correlation between the weighted integrals of SST and precipitation fields. The discretization of the canonical correlation in EOF spectral space for the spatially continuous predictor and predictand fields leads to an equal area–factor correction. The regression error is estimated for each EOF mode. The matrix for the correlation eigen–problem is solved in the EOF space to obtain the maximum correlation between the canonical correlation variables of SST and precipitation. The predicted precipitation field is then expressed in terms of the canonical variables of predictors.

3. MATHEMATICAL FORMULATION OF CEC

3.1 *CCA for Two Continuous Fields*

The CCA forecasting scheme is a regression between two canonical variables $U(t)$ and $V(t')$, which are weighted integrals of two corresponding variables, predictor and predictand. Here t and $t'= t+ \Delta t$ signify time, and Δt is the lead time of forecast for the predictand. The higher the absolute value of the correlation between the two random variables, the more significant the regression, and consequently, the stronger is the linear relationship between the two random variables. To make the regression statistically significant, the absolute value of the correlation between the two canonical variables is maximized by choosing optimal weight functions.

Let random variables $X(x, t)$ and $Y(y, t)$ represent the predictor and predictand fields, respectively. Here x is the position vector defined in the predictor domain Ω_X, y is the position vector in the predictand domain Ω_Y, and t signifies time. The corresponding weight functions are $u(x)$ for X and $v(y)$ for Y, which are called canonical correlation patterns and are to be determined by an optimization. Two weighted integrals are defined by

$$U(t)= \int_{\Omega_X} X(x, t)u(x)\mathrm{d}\Omega_X, \tag{1}$$

$$V(t)= \int_{\Omega_Y} Y(y, t)v(y)\mathrm{d}\Omega_Y, \tag{2}$$

The unit of the weight function $u(x)$ is $[X\text{–unit}]^{-1}$ $[\text{area}]^{-1}$ and that of $v(y)$ is $[Y\text{–unit}]^{-1}$ $[\text{area}]^{-1}$. Thus, the canonical variables $U(t)$ and $V(t)$ are dimensionless. Since the theory to be developed is still a stationary theory, the processed data of X and Y are often the standardized and detrended anomalies. The area is standardized by R^2, where R is the radius of Earth.

Weight functions often imply certain physical meanings of $U(t)$ and $V(t)$. For instance, if

$$u(\boldsymbol{x})=\frac{1}{\text{Area of } \Omega_X},$$

and if X is dimensionless, such as the standardized anomaly, then $U(t)$ is the spatial average of the X field over the domain Ω_X. If X is not standardized, a quantity, such as the average of the spatial standard deviation, should be included in the denominator of the above formula in order to make $U(t)$ dimensionless.

The CCA scheme searches for the optimal weight functions $u(\boldsymbol{x})$ and $v(\boldsymbol{y})$ so that the correlation between U and V reaches extrema. The extrema can be either positive or negative. As will be seen later, the canonical patterns will be determined by an eigenvalue problem. The orthonormal eigenvectors are unique up to a positive or negative sign. Therefore, one can select the eigenvectors so that the correlations are always non–negative, and the correlation's extrema can be regarded as maxima. The details of the eigenvector selection are given later.

The correlation between U and V is denoted by ρ. Hence, our intention is to find the optimal weight functions $u(\boldsymbol{x})$ and $v(\boldsymbol{y})$ such that

$$\rho=\max \text{Corr}(U, V). \tag{3}$$

Since the weight functions are to be determined, one can always adjust them so that U and V have unit variance. The above correlation optimization problem is then equivalent to a covariance optimization problem

$$\rho=\max \text{Cov}(U, V) \tag{4}$$

under a constraint

$$\text{Var}(U)=\text{Var}(V)=1. \tag{5}$$

The above continuous theory can become computationally possible only when the covariance functions and cross–covariance of the predictor and predictand are known. However, in almost all the climatological applications, they are unknown. The detailed mathematics of the CCA for two continuous fields can be found in Shen et al. (2001). A discrete version of CCA must be considered. Here we derive the discrete CCA from integrals; hence, the area–factor is automatically included. The inclusion of the area–factor is important when the data are defined over the uniform latitude–longitude grid boxes and when the domains Ω_X and / or Ω_Y contain high latitude regions, since the data over the low latitude boxes represent larger areas and are, thus, weighted higher. The details of the covariance matrices and EOFs with area–factor are described in North et al. (1982) and Shen et al. (1998).

Since CCA for the CEC is a spectral model like that of BPCCA, we first revisit the computation of the discrete EOF by following the derivation of North et al. (1982). The exact eigenvalue problem is

$$\int_{\Omega_X} \rho(\boldsymbol{x}, \boldsymbol{x}')\psi_k(\boldsymbol{x}')\mathrm{d}\Omega'=\lambda_k\psi_k(\boldsymbol{x}).$$

Here $\rho(\boldsymbol{x}, \boldsymbol{x}')$ is the covariance function of the X field, $\psi_k(\boldsymbol{x})$ is the kth EOF (or mode) and λ_k is the variance (eigenvalue) of $X(\boldsymbol{x})$ on the kth mode ($k=1, 2, \cdots$). The approximate eigenvalues of the above continuum eigen–problem can be estimated by a discretization procedure given by

$$\sum_j \hat{\rho}_{ij}\hat{v}_j^{(k)}=\hat{\lambda}_k\hat{v}_i^{(k)}, \tag{6}$$

where

$$[\hat{\rho}_{ij}]=[\sqrt{A_i}\,\rho(x_i, x_j)\sqrt{A_j}] \tag{7}$$

is the covariance matrix with area–factor, and

$$\hat{v}_j^{(k)}=v_k(x_j)\sqrt{A_j} \tag{8}$$

is the eigenvectors with area–factor, and A_j is the area associated with the grid x_j. For uniform latitude–longitude grid boxes, one has

$$A_j=\cos\varphi_j\,\Delta\theta\Delta\varphi,$$

where φ_j is the latitude of x_j, and the $\Delta\theta$ and $\Delta\varphi$ are the zonal and meridional box dimensions, respectively, which are measured in radians. The linear spatial unit (i.e., the length unit) is in the radius of earth: $R=6376$ km.

With the above preparation, if the area of the region associated with x_i is denoted by A_i and that associated with y_j by B_j, the covariance matrices with the area–factor are

$$\sum_{XX}^{A}=[\sqrt{A_i}\langle X_iX_j\rangle\sqrt{A_j}],$$
$$\sum_{YY}^{A}=[\sqrt{B_i}\langle Y_iY_j\rangle\sqrt{B_j}],$$
$$\sum_{XY}^{A}=[\sqrt{A_i}\langle X_iY_j\rangle\sqrt{B_j}],$$

The discrete EOFs are e_n^X and e_m^Y. They are related to the continuous EOFs by

$$e_n^X(i)=\psi_n(x_i)\sqrt{A_i},\quad e_m^Y(j)=\varphi_m(y_j)\sqrt{B_i}. \tag{9}$$

Here $e_n^X(i)$ is the ith element of the vector e_n^X. The magnitude of e_n^X is equal to one, since

$$\sum_{i=1}^{J_X}[e_n^X(i)]^2=\sum_{i=1}^{J_X}[\psi_n(x_i)\sqrt{A_i}]^2=\int_{\Omega_X}\psi_n^2(x)\mathrm{d}\Omega_X=1,$$

where J_X is the total number of grid points in domain Ω_X.

In practice, the discrete EOFs are computed from the eigenvalue problems of the covariance matrices with area factors.

$$\sum_{XX}^{A}e_n^X=\lambda_n^Xe_n^X,$$
$$\sum_{YY}^{A}e_n^Y=\lambda_n^Ye_n^Y,$$

The correlation ρ defined by (3) has a discrete version

$$\rho=\mathrm{Corr}(X_A\cdot u_A, Y_B\cdot v_B), \tag{10}$$

where the vectors are defined by

$$X_A(t)=[\sqrt{A_i}\,X(x_i, t)],$$
$$Y_B(t)=[\sqrt{B_i}\,Y(y_i, t)],$$
$$u_A=[\sqrt{A_i}\,u(x_i)],$$
$$v_B=[\sqrt{B_i}\,v(y_i)],$$

The maximization of the correlation ρ can then be realized by maximizing the covariance

$$\rho=u'_A\sum_{XY}^{A}v_B, \tag{11}$$

under the constraints

$$u'_A \sum_{XX}^{A} u_A = 1, \tag{12}$$

$$v'_B \sum_{YY}^{B} v_B = 1, \tag{13}$$

In order to find the maximal correlation ρ by solving a single eigenvalue problem, both the fields and weight functions are expressed in terms of discrete EOF. The EOF expansions are

$$X_A(t) \approx \sum_{n=1}^{p} X_n(t) e_n^X \sqrt{\lambda_n^X},$$

$$Y_B(t) \approx \sum_{m=1}^{q} Y_m(t) e_m^Y \sqrt{\lambda_m^Y},$$

$$u_A \approx \sum_{n=1}^{p} u_n e_n^X / \sqrt{\lambda_n^X},$$

$$v_B \approx \sum_{m=1}^{q} v_m e_m^Y / \sqrt{\lambda_m^Y}.$$

The requirement

$$\frac{1}{K_X} \sum_{i=1}^{K_X} X_m(t) X_n(t) = \delta_{mn}, \qquad \frac{1}{K_Y} \sum_{i=1}^{K_Y} Y_m(t) Y_n(t) = \delta_{mn}$$

forces the lengths of the data streams for predictor and predicatand to be the same, i.e., $K_X = K_Y = K$. The covariance matrices between the data point i and the data point j with area–factor are approximately computed by

$$\sum_{XX}^{A}(i,j) = \frac{1}{K} \sum_{i=1}^{K} X_A(x_i, t) X_A(x_j, t),$$

$$\sum_{YY}^{A}(i,j) = \frac{1}{K} \sum_{i=1}^{K} Y_B(y_i, t) Y_B(y_j, t),$$

$$\sum_{XY}^{A}(i,j) = \frac{1}{K} \sum_{i=1}^{K} X_A(x_i, t) Y_B(y_j, t).$$

These matrices allow one to compute the principal components $X_n(t)$ and $Y_m(t)$, such as

$$X_n(t) = \frac{1}{\sqrt{\lambda_n^X}} (e_n^X)' X_A = \sum_{i=1}^{J_X} \frac{1}{\sqrt{\lambda_n^X}} e_n^X(i) X(x_i, t) \sqrt{A_i} = \sum_{i=1}^{J_X} \frac{1}{\sqrt{\lambda_n^X}} \psi_n(x_i) X(x_i, t) A_i. \tag{14}$$

Using what expression in the above depends on the available types of EOFs. The quantity $Y_m(t)$ can be computed similarly.

The cross–covariance matrix in the EOF space is approximated by

$$\tilde{\sum}_{XY} \approx \frac{1}{K} \sum_{i=1}^{K} X_n(t) Y_m(t), \tag{15}$$

Its transpose is

$$\tilde{\sum}'_{XY} = \tilde{\sum}_{YX}. \tag{16}$$

The maximization of the correlation ρ leads to the following eigenvalue problem for $\tilde{u}$:

$$\tilde{\sum}_{XY}\tilde{\sum}_{YX}\tilde{\boldsymbol{u}}=\rho^2\tilde{\boldsymbol{u}}. \tag{17}$$

Similarly, the eigenvalue problem for $\tilde{v}$ can be derived as well:

$$\tilde{\sum}_{YX}\tilde{\sum}_{XY}\tilde{\boldsymbol{v}}=\rho^2\tilde{\boldsymbol{v}}. \tag{18}$$

The CCA patterns are thus the solutions of these eigenvalue problems.

3.2 *Multivariate Regression for Anomaly Prediction*

Our prediction method is a linear regression based upon the data of the predictor and predictand. The predictand is a linear combination of the predictor. The predictand and predictor can be defined either on points or regions, although they are often defined on grid points. Of course, the regression is meaningful only when an approximate linear relationship exists between the predictor and predictand. The regression is statistically significant only when the scatter diagram of the data is not too far away from the regression line (or the regression hyper–plane in the case of multivariate analysis). This condition is equivalent to that the correlation between the two fields is large.

In the case of a single–variate, the regression line

$$y=bx+x_0$$

is statistically significant only when the data points (x_i, y_i) are close to the line. The slope is $b=\rho\sigma_y/\sigma_x$, where ρ is the sample correlation, and σ_x and σ_y are the sample standard deviations of the x–data and y–data, respectively. Distance between the points and the regression line is measured by the sample correlation, ρ, computed from the data. The closer the absolute value of ρ is to one, the closer are the points to a line, and consequently, the more significant the regression. For example, with 15 data points, the 5% significance level requires that the correlation be not less than 0.514.

Our climate prediction uses multivariate regression, which is a linear approximation to a nonlinear relationship. The correlation between the two fields is defined in the previous sections and computed by the CCA algorithms. Our linear approximation is optimal in the sense of the maximization of the correlation between the two fields.

Our linear prediction in the physical space can be written as

$$Y(\boldsymbol{y}_i, t')=\sum_j B_{ij}X(\boldsymbol{x}_j, t). \tag{19}$$

Here, the predictand is at time $t'=t+\Delta t$, the predictor is at time t, the "slope" matrix B is to be found, and the "intercept" vector is set to be zero since both Y and X are anomalies. We will use the canonical variables and canonical correlation patterns, together with EOFs, to express the fields and then find the "slope" matrix B. The correlation between the X–field and Y–field is the key parameter in the regression.

If the truncation order $p>q$, then the correlation between $U^{(k)}(t)$ and $V^{(k)}(t)$ is zero when $k>q$. The same is true for the case $q>p$. Thus, we always make the EOF truncation order for X and Y be the same: $p=q$. The CCA eigenvalue problem has p non–zero eigenvalues. When $s=2p-K+1>0$ (where K is the length of the data stream), the first s eigenvalues are equal to one, which implies a perfect linear relationship that is certainly impossible. Thus, choosing a large p may result in an over–fit. In this paper, we always choose $p<(K-1)/2$, and none of the eigenvalues is equal to one.

The p eigenvectors are mutually orthogonal. The canonical variables are

$$U^{(k)}(t)= \sum_{n=1}^{p} u_n^{(k)} X_n(t), \tag{20}$$

$$V^{(k)}(t')= \sum_{n=1}^{q} v_n^{(k)} Y_n(t'), \quad k= 1, 2, \cdots, p. \tag{21}$$

Now, the sign of $\tilde{\boldsymbol{u}}^{(k)}$ needs to be determined to guarantee the positivity of ρ_k. One can calculate ρ by using

$$\gamma_k = \frac{1}{K} \sum_{t=1}^{K} U^{(k)}(t) V^{(k)}(t+ \Delta t).$$

If γ_k is positive, then $\rho_k = \gamma_k$ and $\tilde{\boldsymbol{u}}^{(k)}$ does not change signs. Otherwise, $\gamma_k < 0$, and then $\rho_k = -\gamma_k$ and $\tilde{\boldsymbol{u}}^{(k)}$ is changed to $-\tilde{\boldsymbol{u}}^{(k)}$. After this operation, the correlation at each mode (k) is non–negative.

The operations above are among the dimensionless variables. Because the matrices $[u_n^{(k)}]_{p\times p}$ and $[v_m^{(k)}]_{p\times p}$ are orthogonal matrices, their inverses are equal to their transposes. Such an inverse operation for a matrix is physically meaningful only when the matrix is dimensionless.

The inverse of the matrix equations (20) and (21) leads to

$$X_n(t)= \sum_{k=1}^{p} u_n^{(k)} U^{(k)}(t), \tag{22}$$

$$Y_n(t')= \sum_{k=1}^{p} v_n^{(k)} V^{(k)}(t'), \tag{23}$$

The field X in the spectral space is expressed by the sum of the products of the canonical variable $U^{(k)}$ and eigenvector $\tilde{\boldsymbol{u}}^{(k)}$. The eigenvector $\tilde{\boldsymbol{u}}^{(k)}$ is the canonical pattern in spectral space. Its physical space correspondence is $\tilde{\boldsymbol{u}}_A^{(k)}$ given by formula (14).

For the purpose of one–tier prediction, the predictand time t' is larger than the predictor time t by Δt, and, thus, the cross–covariance matrix $\tilde{\sum}_{XY}$ in this section is a lagged cross–covariance. The lag time is $t'- t= \Delta t$.

The vectors $(U^{(1)}, U^{(2)}, \cdots, U^{(p)})'$ and $(V^{(1)}, V^{(2)}, \cdots, V^{(p)})'$ are all in p–dimensional vector space. A unique transform matrix must exist that transforms one to the other. Let this transformation be written as

$$V^{(k)}= \sum_{l=1}^{p} F_{kl} U^{(l)} + b_k, \quad k= 1, 2, \cdots, p.$$

The formulas

$$\langle V^{(k)} U^{(l)} \rangle = \rho_k \delta_{kl},$$
$$\langle U^{(k)} \rangle = 0, \quad \langle (U^{(k)})^2 \rangle = 1,$$
$$\langle V^{(k)} \rangle = 0, \quad \langle (V^{(k)})^2 \rangle = 1,$$

imply that

$$V^{(k)}(t')= \rho_k U^{(k)}(t), \quad k= 1, 2, \cdots, p. \tag{24}$$

The prediction is therefore made for $t'= t+ \Delta t$ when the predictor at t is known:

$$X(\boldsymbol{x}, t) \rightarrow X_n(t) \rightarrow U^{(k)}(t) \rightarrow V^{(k)}(t') \rightarrow Y_n(t') \rightarrow Y(y, t').$$

The equation (24) is thus the main prediction equation. The predictand at t' is

$$\hat{Y}(y, t') = \sum_{n=1}^{p} \left(\sum_{k=1}^{p} v_n^{(k)} \rho_k U^{(k)}(t) \right) \sqrt{\lambda_n^Y}\, \varphi_n(y). \tag{25}$$

This is the one–tier prediction scheme: knowing the predictor at time t and predicting the predictand at $t' = t + \Delta t$. The CCA scheme can also be applied to the so called two–tier forecasting: forecasting the predictor from time t to time $t' = t + \Delta t$ by one method and predicting the predictand at time t' from the predictor also at time t' by another method. CCA can also be used as the latter method. The CCA scheme is exactly the same, except that the lagged cross covariance matrix should be replaced by a cross–covariance matrix without time lag, i.e., $\Delta t = 0$.

4. ESTIMATE OF THE PREDICTION ERROR

This section estimates the mean square error (MSE) of prediction. The importance of this error estimate is three–fold. First, it gives the mean property of the forecasting quality. The smaller the error, the more accurate the forecast. Second, the MSE is directly related to the correlation skill of the prediction which is often used to assess the quality of a prediction. Third, the MSE errors are needed to compute the optimal weights to form a canonical ensemble forecast.

The CCA scheme requires the fields to be stationary. The computing algorithm requires the fields to be ergodic; i.e., the ensemble mean can be replaced by time mean for a long time. Our estimated error, as an ensemble value, cannot reflect the fluctuation of the fields according to time. However, if the data used in the prediction are changed, then the error should reflect this change. One may regard an observational network as a filter. When a filter is changed, the filtered results should change accordingly. Thus, the time dependence of the error is on the observational data and the data processing method, not on the intrinsic fluctuations of the fields. Thus, for a given observational network and a given data–processing method, our forecasting error is time independent.

The error field is

$$\varepsilon^2(y) = \langle (Y(y, t') - \hat{Y}(y, t'))^2 \rangle, \tag{26}$$

where Y is the unknown true field to be predicted, and $\hat{Y}$ is the prediction that approximates Y. The prediction skill of $\hat{Y}$ is often measured by its correlation with the original field Y, defined by

$$r = \frac{\langle Y\hat{Y} \rangle}{\sqrt{\langle Y^2 \rangle \langle \hat{Y}^2 \rangle}} \tag{27}$$

From the above two equations, one can derive

$$r = \frac{1}{2} \left[\sqrt{\frac{\langle Y^2 \rangle}{\langle \hat{Y}^2 \rangle}} + \sqrt{\frac{\langle \hat{Y}^2 \rangle}{\langle Y^2 \rangle}} - \frac{\varepsilon^2}{\langle Y^2 \rangle \langle \hat{Y}^2 \rangle} \right]. \tag{28}$$

Several special cases are interesting. Case 1 is the perfect prediction: $\varepsilon = 0$ implies $r = 1$. Case 2 is the worst prediction: $\varepsilon^2 = \langle Y^2 \rangle + \langle \hat{Y}^2 \rangle$ implies $r = 0$. Case 3 is the random prediction: $\varepsilon^2 = \langle Y^2 \rangle$ implies $r = (1/2)\sqrt{\langle \hat{Y}^2 \rangle / \langle Y^2 \rangle}$.

The MSE can be estimated by using EOFs. The error can be written in two parts:

$$\varepsilon^2(y)= \varepsilon_q^2(y)+ \varepsilon_R^2(y), \tag{29}$$

where

$$\varepsilon_q^2(y)= \sum_{m=1}^{q} \varepsilon_m^2 \varphi_m^2(y) \tag{30}$$

is the MSE computed from the first q modes with

$$\varepsilon_m^2 = \langle (Y_m(t')- \hat{Y}_m(t'))^2 \rangle, \tag{31}$$

and the residual $\varepsilon_R^2(y)$ is the contribution of higher modes to the MSE:

$$\varepsilon_R^2(y)= \sum_{m=q+1}^{\infty} \lambda_m \varphi_m^2(y). \tag{32}$$

Apparently, this residue error depends on the truncation order, which, in turn, depends on the resolution of data and the spatial scale of the field. The higher the data resolution, the larger the q value. In practice, one uses discrete data and obtains only a finite number of modes, say, q'. The prediction uses the first q modes among the $q' > q$ modes. Then, the residual error may formally written as

$$\sum_{m=q+1}^{q'} \lambda_m^Y \varphi_m^2(y).$$

This q' is usually chosen to be K, the length of the data stream. However, when the data streams are not sufficiently long, they cannot resolve the EOFs of very high order. Because the higher order eigenvalues are close to each other, the computed EOFs can be very much distorted from the true ones (See North's rule–of–thumb, North et al., 1982). Thus, the above estimate of residual error can be unreliable if the data streams are not sufficiently long.

Usually, the q modes of the predictand should resolve at least 50% total variance. The residual error should be smaller than or comparable to the prediction error of the first q modes. With the omission of the residual error in the computation, the error $\varepsilon_q^2(y, t')$ computed based on the first q modes can be regarded as the lower bound of the MSE.

The average MSE of the error field is the integral of $\varepsilon^2(y, t')$

$$\varepsilon_A^2 = \frac{1}{A}\int_{\Omega_Y} \varepsilon^2(y, t')\mathrm{d}\Omega_Y, \tag{33}$$

where A is the total area of the domain Ω_Y. When using the first q modes, the average MSE error is

$$\varepsilon_A^2 = \frac{1}{A}\sum_{m=1}^{q} \varepsilon_m^2. \tag{34}$$

This MSE's unit is still [Y unit]^2, as was expected.

The error for the mth mode is computed as follows:

$$\begin{aligned}\varepsilon_m^2 &= \lambda_m^Y \langle (Y_m(t')- \hat{Y}_m(t'))^2 \rangle = \lambda_m^Y \left\langle \left(\sum_{k=1}^{q} v_m^{(k)} V^{(k)}(t')- \sum_{k=1}^{q} v_m^{(k)} \rho_k U^{(k)}(t') \right)^2 \right\rangle \\ &= \lambda_m^Y \left\langle \left[\sum_{k=1}^{q} (V^{(k)} - \rho_k U^{(k)}) v_m^{(k)} \right]^2 \right\rangle = \lambda_m^Y \sum_{k=1}^{q} (1- \rho_k^2)(v_m^{(k)})^2. \end{aligned} \tag{35}$$

The error is scaled down for higher EOF modes of the Y–field by its eigenvalues λ_m^Y since $\lambda_m^Y \to 0$ as $m \to \infty$, but scaled up by the CCA eigenvalues ρ_k^2 because $1-\rho_k^2 \to 1$ as $k \to \infty$. If the correlations ρ_k^2 are large, say, close to one, then $\varepsilon_m \approx 0$. If the predictability is low, then the correlation is small. If the largest correlation is even less than 0.4, the factor $1-\rho_1^2$ is larger than 0.84, which is relatively large. If both the predictor and predictand are spatial white noise, then no correlation exists between the two fields, and $\rho_k^2 = 0$ and

$$\varepsilon_m^2 = \lambda_m^Y,$$

and

$$\varepsilon_A^2 = \sum_m \lambda_m^Y = \left\langle \int_{\Omega_Y} Y^2(y, t)\mathrm{d}\Omega_Y \right\rangle.$$

Hence, the prediction error is equal to the total variance of the field and the prediction skill is zero.

5. OPTIMAL ENSEMBLE OF MANY FORECASTS

Different data sets of predictor and predictand may yield many forecasts, such as the forecasts of precipitation by using SST, soil moisture, SLP, SLP, or SST from different ocean basins, etc. An optimal weight needs to be assigned to each forecasting result at every grid point in order to form a canonical ensemble forecast. This method is similar to the superensemble idea proposed by Krishnamurti et al. (2000).

This section will show the method of finding the optimal weights and the resulting prediction error. The spectral method is again used in our computing. The use of the spectral method requires a stationarity assumption. Because of the requirement, which may be true in a short period, the forecasting naturally deteriorates rapidly as the lead time of the prediction increases.

The spectral representation of the predictand is

$$Y(y, t) = \sum_{n=1}^{\infty} \sqrt{\lambda_n^Y}\, Y_n(t)\varphi_n(y).$$

Its estimator is

$$\hat{Y}(y, t) = \sum_{n=1}^{q} \sqrt{\lambda_n^Y}\, \hat{Y}_n(t)\varphi_n(y).$$

Suppose that there are H forecasts. The ensemble forecast for the principal component $Y_n(t)$ is

$$\hat{Y}_n(t) = \sum_{h=1}^{H} w_n^{(h)} \hat{Y}_n^{(h)}(t), \tag{36}$$

where $w_n^{(h)}$, to be determined, is the weight for model h and mode n. The weights satisfy a constraint

$$\sum_{h=1}^{K} w_n^{(h)} = 1, \tag{37}$$

but they do not have to be all positive.

The truncated MSE forecasting error is a function of y and is written as

$$E^2(y)=\langle(Y(y,t)-\hat{Y}(y,t))^2\rangle=\left\langle\left[\sum_{n=1}^{q}\left(Y_n(t)-\sum_{h=1}^{H}w_n^{(h)}\hat{Y}_n^{(h)}(t)\right)\sqrt{\lambda_n^Y}\,\varphi_n(y)\right]^2\right\rangle.$$

The principal components of different modes are independent. The errors $Y_n-\hat{Y}_n$ $(n=1, 2, 3, \cdots)$ are assumed to be independent. Then, the error function becomes

$$E^2(y)=\sum_n E_n^2\varphi_n^2(y), \tag{38}$$

where

$$E_n^2=\lambda_n^Y\langle(Y_n(t)-\sum_{h=1}^{H}w_n^{(h)}\hat{Y}_n^{(h)}(t))^2\rangle \tag{39}$$

is the MSE error for forecasting the nth principal component $Y_n(t)$.

The forecasting errors are caused by noisy data or inaccuracy of the forecasting model. For each EOF coeffcient, the forecasting errors for different prediction models are assumed to be uncorrelated. With this assumption, one then has the following:

$$E_n^2=\lambda_n^Y\langle(\sum_{h=1}^{H}w_n^{(h)}(Y_n(t)-\hat{Y}_n^{(h)}(t)))^2\rangle=\lambda_n^Y\sum_{h=1}^{H}(w_n^{(h)})^2(\varepsilon_n^{(h)})^2, \tag{40}$$

where

$$(\varepsilon_n^{(h)})^2=\langle(Y_n(t)-\hat{Y}_n^{(k)}(t))^2\rangle. \tag{41}$$

The minimization of ε_n^2 expressed as above under the constraint (37) leads to the optimal weights. The weight for a model is inversely proportional to the model's forecasting error. Thus,

$$w_n^{(h)}=\frac{1/(\varepsilon_n^{(h)})^2}{\sum_{l=1}^{H}1/(\varepsilon_n^{(l)})^2}. \tag{42}$$

Then the resulting MSE error for the mode n is

$$E_n^2=(\sum_{h=1}^{H}1/(\varepsilon_n^{(h)})^2)^{-1}. \tag{43}$$

That the errors from different prediction models are independent is an important assumption. The different models usually mean different predictors, but the same prediction scheme. Thus, to ensure the independence of the errors, one should choose the predictors which are not strongly correlated. This practice is intuitively correct, since the information from the similar sources is redundant.

A special case occurs when all the models have the same amount of errors, $(\varepsilon_n^{(0)})^2$. In this case, the optimal weights are all equal to $1/H$, and the resulting error is

$$\varepsilon_n^2=\frac{(\varepsilon_n^{(0)})^2}{H}. \tag{44}$$

This is the conventional error–formula of homogeneous statistics.

The method of finding optimal weights here is slightly different from that of Krishnamurti et al. (2000), although the purposes of the two methods are the same: optimal

combination of multiple forecasts. Krishnamurti et al. (2000) used the multivariate regression to determine the weights, and they did not require the sum of the weight to be one. Since their regression is for the best fit of the linear superensemble model to data, the forecasted expected values are approximately the same as the observed expected values and hence the constraint of the sum of all the weights equal to one is not needed. In our case, however, the constraint is needed to guarantee that the forecasted expected values are approximately the same as the observed expected values. Otherwise, the weights might be too large, and, consequently, the forecasted expected value will be out of the range of possible climate and hence, the forecast is not valid. Despite the difference between the present method and that of Krishnamurti et al. (2000), the results are actually close. We applied the formula (43) to the Y–component of their Fig. 5b. Formula (43) provides an RMSE 1.7, which is about the same as that of Krishnamurti et al. (2000).

6. DATA SETS AND THE RESULTS OF THE US PRECIPITATION FORECASTING

6.1 *Data Sets*

Our predictor's data set is the historical monthly SST reconstructed by Smith et al. (1996). The spatial coverage is from 45°S to 69°N. The spatial resolution is $2° \times 2°$. The temporal coverage is from January 1951 to December 2000. The temporal resolution is a month. The data are available via ftp at: ftp:/ / ftp.ncdc.noaa.gov/ pub / data / ocean / rsst / . Our linear regression scheme cannot resolve the nonlinear process imbedded in small scales. Thus, the $2° \times 2°$ SST data are further averaged into a data set of 4–degree latitude by 6–degree longitude. This average removes the noise of smaller scales.

The predictand's data set is based upon the latest version of the Xie–Arkin global land monthly precipitation data set as reported in Chen et al. (2001). The temporal coverage of the precipitation data is from January 1948 to December 2000, but only the part overlapped with the reconstructed SST period, January 1951–December 2000, is used in our forecasting calculation. The spatial resolution is $2.5° \times 2.5°$. The gridded data result from the interpolation of over 17,000 stations collected in the Global Historical Climatology Network (GHCN) Version 2 and the Climate Anomaly Monitoring System (CAMS). The data are available via ftp at ftp.ncep.noaa.gov / pub / precip / 50–yr.

The anomaly data are computed according to the 1961–1990 climatology. These anomaly data are further detrended and standardized before they go into the EOF computing. After the forecasting on the standardized data, the forecasting results (for dimensionless quantities) are converted back to the anomaly data with dimension. However, the conversion uses only the mean and standard deviation, and the trend is not included. The forecasting results are then compared with the observed detrended anomaly field to assess the forecasting accuracy.

6.2 *Forecasting Results*

The 50 years (1951–2000) SST and precipitation data are used in our forecasting experiments. Hence, 49 years of data are used for computing the EOFs for both SST and precipitation, and these data are used to predict the remaining year's precipitation. Our scheme requires that the lengths of the data streams for both SST and precipitation be the same. The reconstructed SST data are only for the preiod of 1951–2000. The precipitation data before 1951, although available, are not used in our forecasting experiments. The EOFs of the US precipitation are presented for two seasons: DJF (winter, Fig. 3) and JJA (summer, Fig. 4). For both precipitation and SST fields, we show only the first six EOF modes. The number of modes used for actual forecasting is ten or another number.

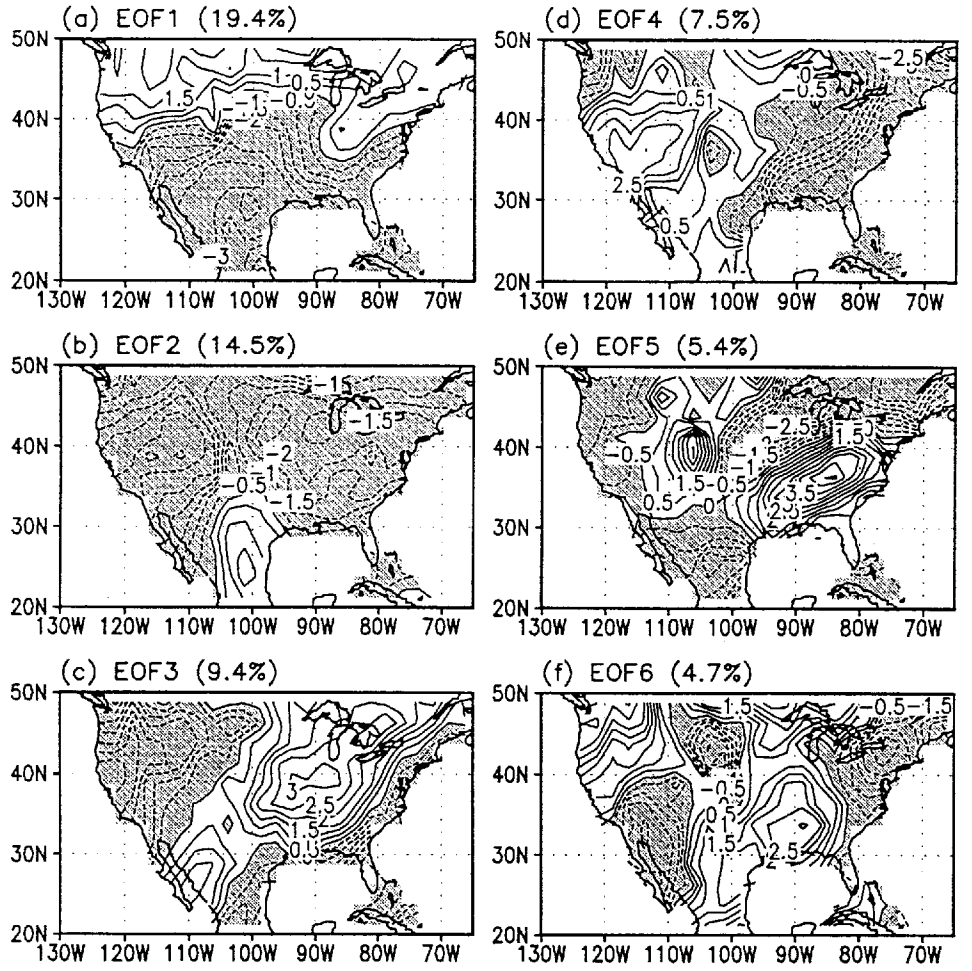

Fig. 3. First six EOFs of US precipitation in DJF. The computation used the three-month mean precipitation data, which had been detrended and standardized.

The characteristics of summer precipitation are very much different from that of winter. The EOFs clearly reflect the difference. For example, the DJF EOF1 demonstrates a strong north–south pattern, while the JJA EOF1 supports a northwest–southeast pattern. The differences are also reflected in higher modes.

The CCA patterns are linear combinations of the EOFs. The first two winter CCA patterns can sometimes be highly mode–selective, preferring only a few from the six retained EOF modes. When performing the SON (SST field)–DJF (precipitation) CCA analysis, the first SST CCA pattern is mainly contributed by 0.75 EOF1 and 0.61 EOF2. Since the first SST EOF is the El Niño mode, the first SST CCA pattern looks like the El Niño mode plus some North Pacific oscillation (see Fig. 5). The contribution from other EOFs is negligible. The first precipitation CCA pattern is mainly contributed to by 0.68 EOF1, 0.38 EOF2, and 0.34 EOF5 (Fig. 5). The contribution from other modes is small. The northwest positive is mainly contributed to by EOF1.

With this pair of CCA patterns, the CCA correlation between the SST and precipitation field (i.e., ρ_1) is 0.87, a relatively large value indicating that a good forecasting result is

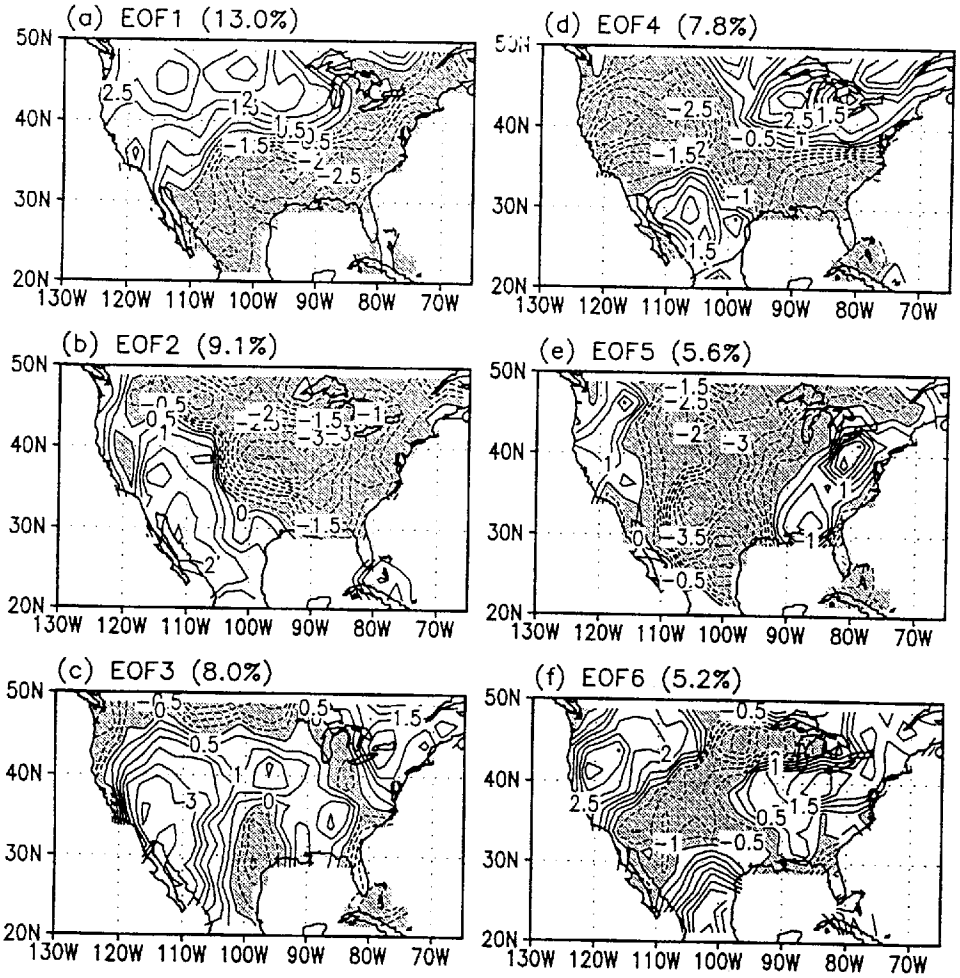

Fig. 4. Same as Fig. 3, except for the summer season JJA.

expected. However, the summer CCA patterns are more uniformly distributed among the ten retained EOFs and do not have apparent dominant modes. For example, the first precipitation CCA pattern selects not only EOF1 and EOF2 , but also EOF6, EOF7, EOF8 and EOF9 (Fig. 6). Comparing Fig. 6a and Fig. 5a, one can see that DJF's first canonical pattern is much more coherent than that of JJA. The same is true for the SST canonical patterns (see Figs. 5b and 6b). This result reflects the diffculty of predicting the summer precipitation.

The 2000 DJF precipitation is predicted by the 1999 SON SST. The observed precipitation anomaly is shown in Fig. 7a and the predicted anomaly in Fig. 7b. The predicted and the observed anomalies agree in most of the southeast areas. The spatial pattern correlation, defined by the formula (45) below, between the predicted and observed is 0.47. The corresponding pattern correlation is higher for El Niño or La Niña years, but the year 2000 was neither an El Niño nor an La Niña year.

The DJF precipitation forecasts are made for every year from 1951 to 2000 by using the previous season's (i.e. SON) SST. The pattern correlation and the Heidke skill score are shown in Fig. 8. Heidke score accounts for only the correct anomaly sign. Pattern correlation accounts for both sign and amplitude. Only the perfect forecast, with both correct sign and

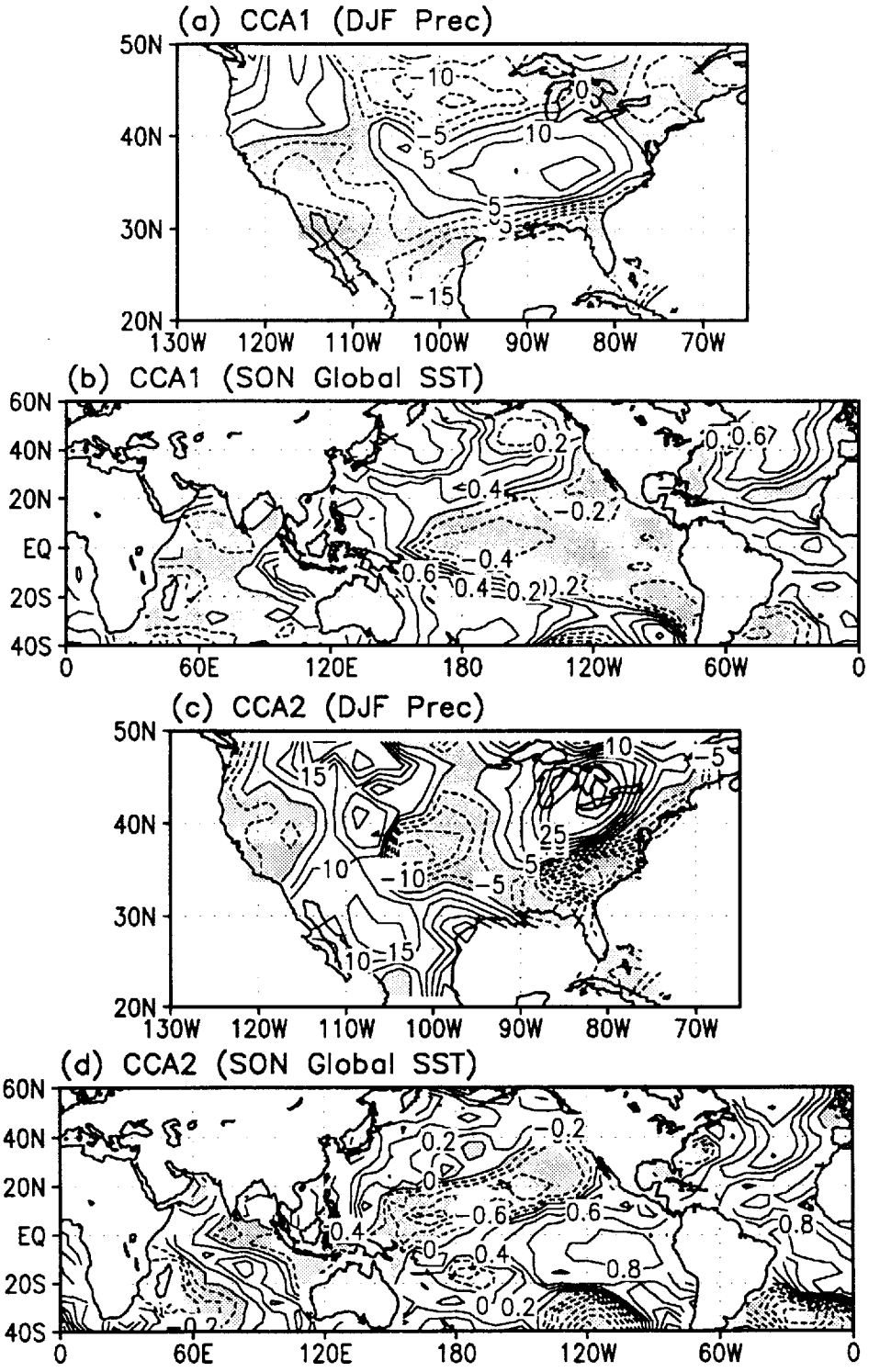

Fig. 5. First two canonical correlation patterns for SON SST and DJF US precipitation.

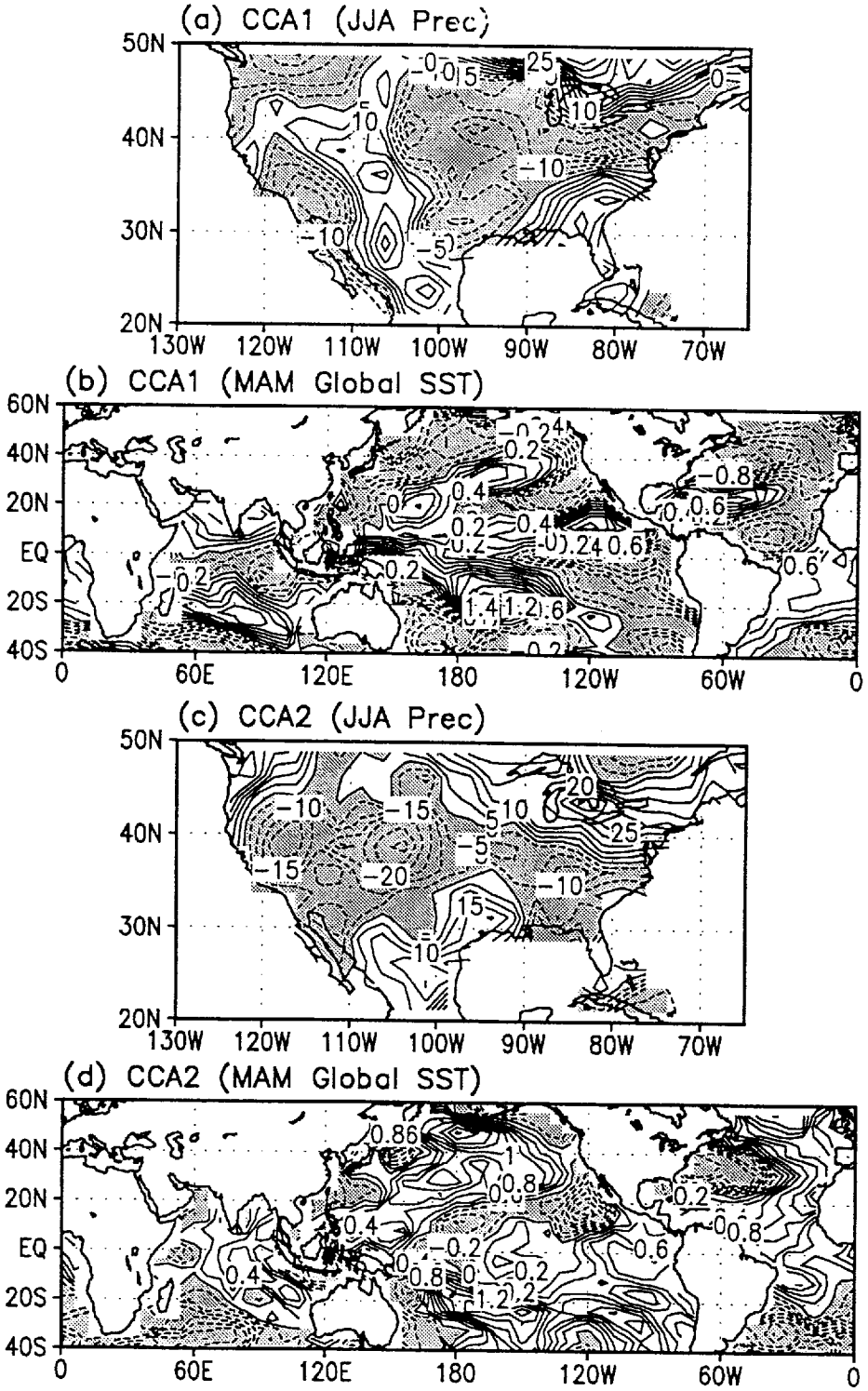

Fig. 6. First two canonical correlation patterns for MAM SST and JJA US precipitation.

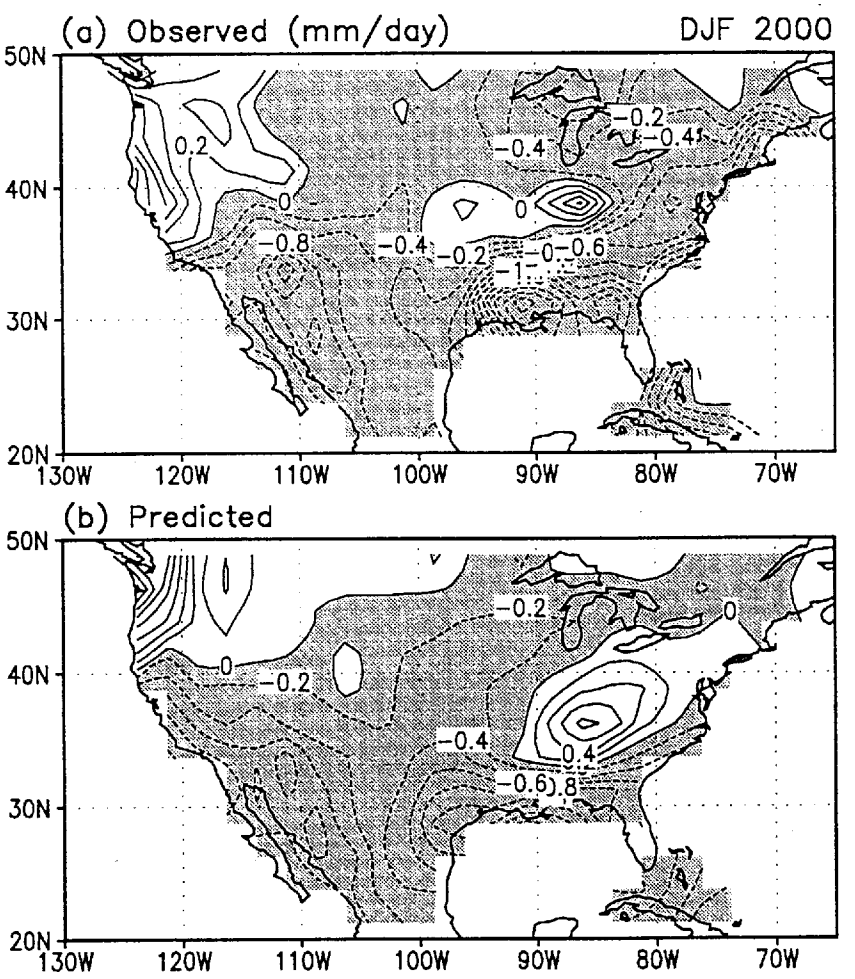

Fig. 7. Observed and predicted precipitation field of DJF 2000. The prediction was made from the previous season SON SST. The time lag is hence one season. The unit is mm d^{-1}.

amplitude, has pattern correlation equal to one. Fig. 8 indicates that our forecasting results are reasonable.

The spatial pattern correlation is the correlation between the forecasted ($\hat{R}_i$) and observed anomalies (R_i) for a given season:

$$\gamma = \frac{\sum_i R_i \hat{R}_i}{\sqrt{\sum_i R_i^2}\sqrt{\sum_i \hat{R}_i^2}}. \tag{45}$$

When the pattern correlation reaches 0.4, the forecasted field has a good similarity to the observed. Among our 50 years of forecasts, 29 years have the pattern correlation equal to or greater than 0.4. Our maximum pattern correlation is 0.78, which indicates an extremely good forecast (see Fig. 8a).

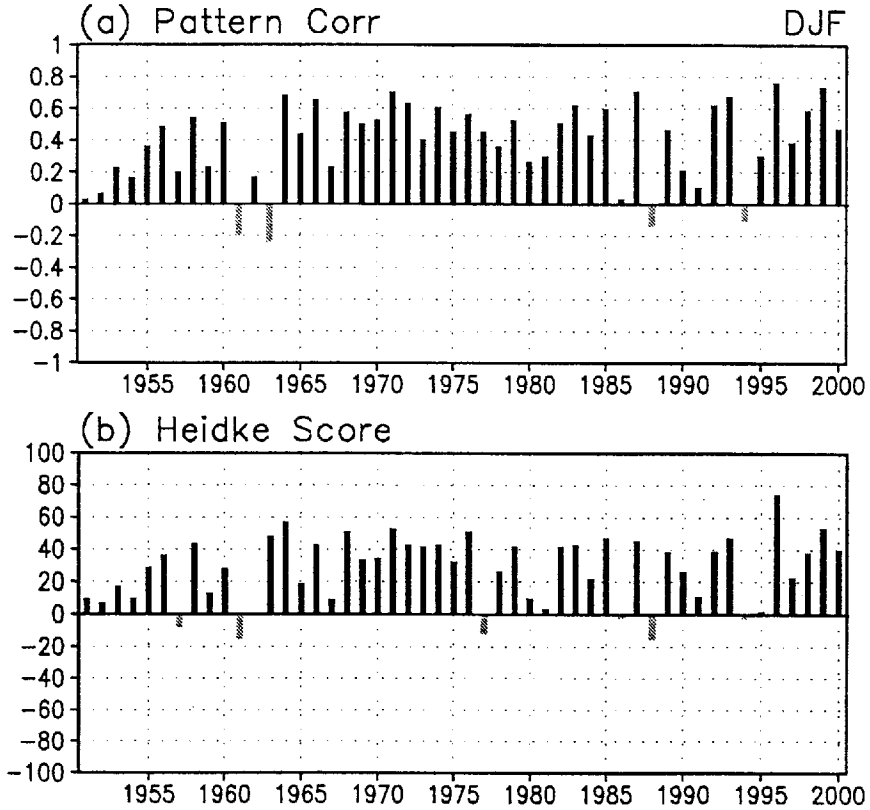

Fig. 8. Forecasting skills of using previous season's SON SST to predict DJF precipitation: (a) pattern correlation (defined by Eq. (45)), and (b) Heidke score (defined by Eq. (46)). The EOFs were computed by using the data from 1951–1999.

The Heidke score (HS) is another commonly used index to indicate the prediction skill (van den Dool et al., 1997). The two–class HS is defined as

$$HS = \frac{H - E}{T - E} \times 100, \tag{46}$$

where T is the total number of forecasts (i.e., the total number of grid points in our case), E is the expected value of correct random forecasts (i.e., $E = T/2$), and H is the actual number of correct forecasts. Since the two–class Heidke score is used, a forecast is defined as correct if the forecasted anomaly has the same sign as the observed anomaly. The maximal value of HS is equal to 100, when every forecast is correct. The HS score for random guessing is zero because $H = E$; hence, the random forecasting model does not have a skill. The HS reaches its minimum value if every forecast has missed and the score is $-100 \times E/(T - E)$, which is equal to −100. Our forecasts show positive skill scores in 45 years of the total of 50 years. The maximum Heidke score is 76 (see Fig. 8b).

The summer precipitation skill is consistently lower than that for winter. Among the 50 years of experiments, there are only 18 years whose pattern correlations are equal to or greater than 0.4. The skill varies strongly with near–zero skill in many years, but the maximum skill can still be high. In our example, the maximum skill is 0.83. The pattern correlation for the 2000 prediction is very low and the forecast is not good. We have chosen to show a fore-

cast of a bit less–than–average skill. The year is 1997. The training period of the predictor and predictand is 1951–1996. The 1997's JJA precipitation is predicted by using the same year's MAM SST. The observed JJA precipitation is shown in Fig. 9a. The predicted result is shown in Fig. 9b. Comparing these two figures, one can see that the forecasting skills are mainly from the Northeast and Southwest. In almost 50% of the US, the forecast has a wrong sign. Both the pattern correlation (0.18) between the forecasted and observed and the Heidke skill score are low (almost zero) (the furtherest bar to the right in Fig. 10). This result is a detailed indication of the diffculty of forecasting the summer precipitation.

Of course, for an actual forecast one does not know the observed precipitation. The reliability can be assessed by the expected value of prediction errors, as was discussed in Section 5. Fig. 11 shows the prediction root mean square errors (RMSE) for two seasons: DJF and JJA. The predictor is the previous season's SST. The errors measure the mean square difference between the actual and forecasted precipitation. In the region where the precipitation variance is large, the error is also large. The variance is proportional to the precipitation

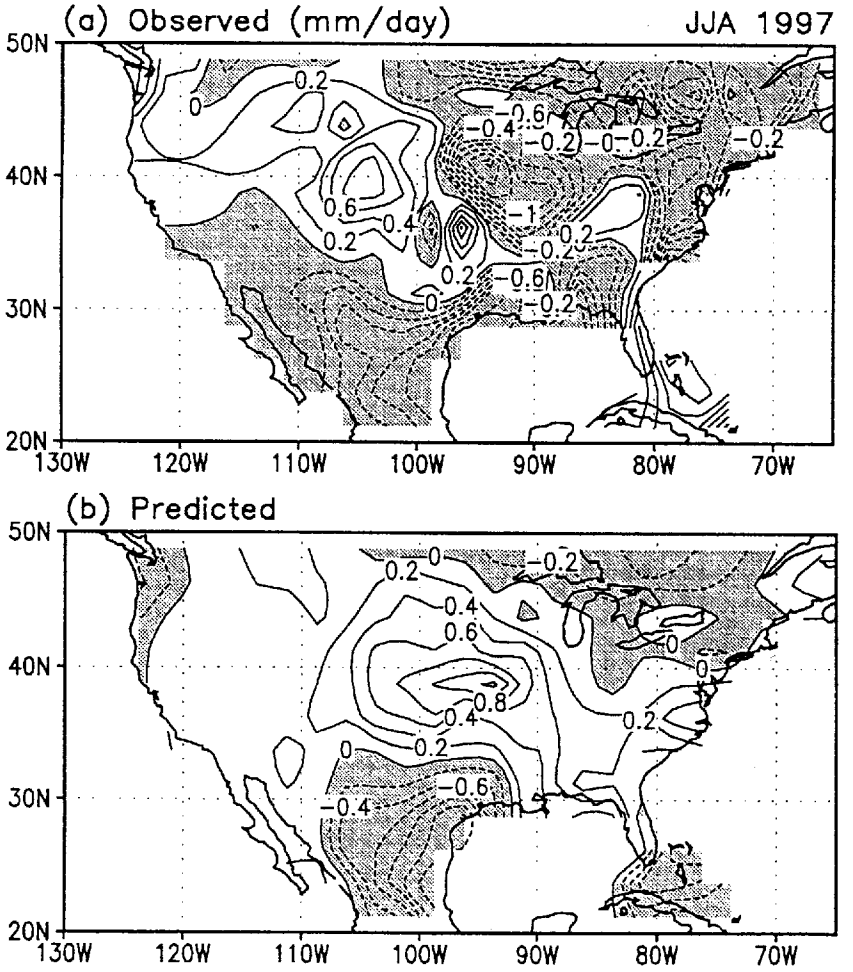

Fig. 9. The observed and the forecasted 1997 JJA precipitation. The predictor was the 1997 MAM SST.

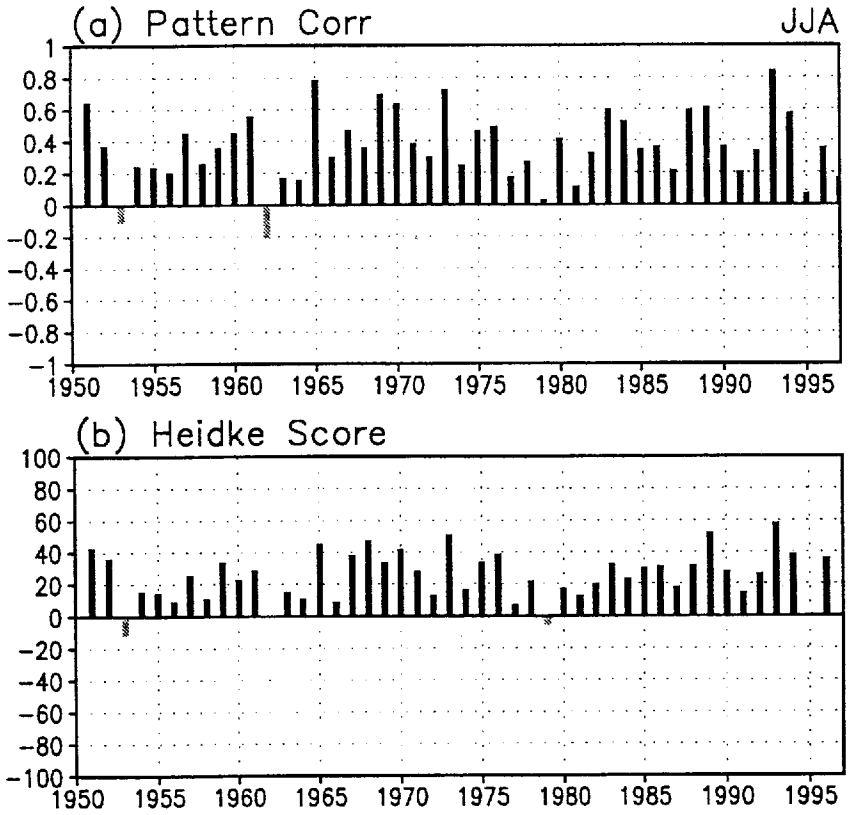

Fig. 10. Forecasting skills of using MAM SST to predict JJA precipitation: (a) pattern correlation (defined by Eq. (45)), and (b) Heidke score (defined by Eq. (46)). The EOFs were computed by using the data from 1951–1996.

climatology. The DJF error map has a Northwest–Southeast pattern (Fig. 11a), and the JJA error error map has a Northeast–Southwest pattern. These patterns are consistent with US precipitation climatology (Higgins et al., 1996).

The RMSE error is the expected value. The regions with large errors are likely to have the wrong prediction, but for an individual forecast, the region with a large error may still have a good forecast.

The optimal combination of two forecasts may enhance the forecasting skill, but one has to be careful when combining them. Since both weights are positive, the optimal combination is an interpolation process, which prefers two good forecasts. Thus, one should combine the two forecasts of reasonably high Heidke scores. Fig. 12c shows the optimal combination of the 2000 DJF precipitation forecasts from the previous season's SST by using the Pacific (Fig. 12a) and Atlantic oceans (Fig. 12b), respectively. The pattern correlation and Heidke score for Pacific forecast are 0.43 and 35, respectively, and those for Atlantic are 0.04 and 31. The optimally combined forecast is shown in Fig. 12c. Its pattern correlation and Heidke scores become 0.47 and 40.

Compared with the observed precipitation in Fig. 7a, the combined forecast has a better skill and is more similar to the observed than the two individual forecasts. If both forecasts are very good, the optimally combined forecast can have a much higher skill.

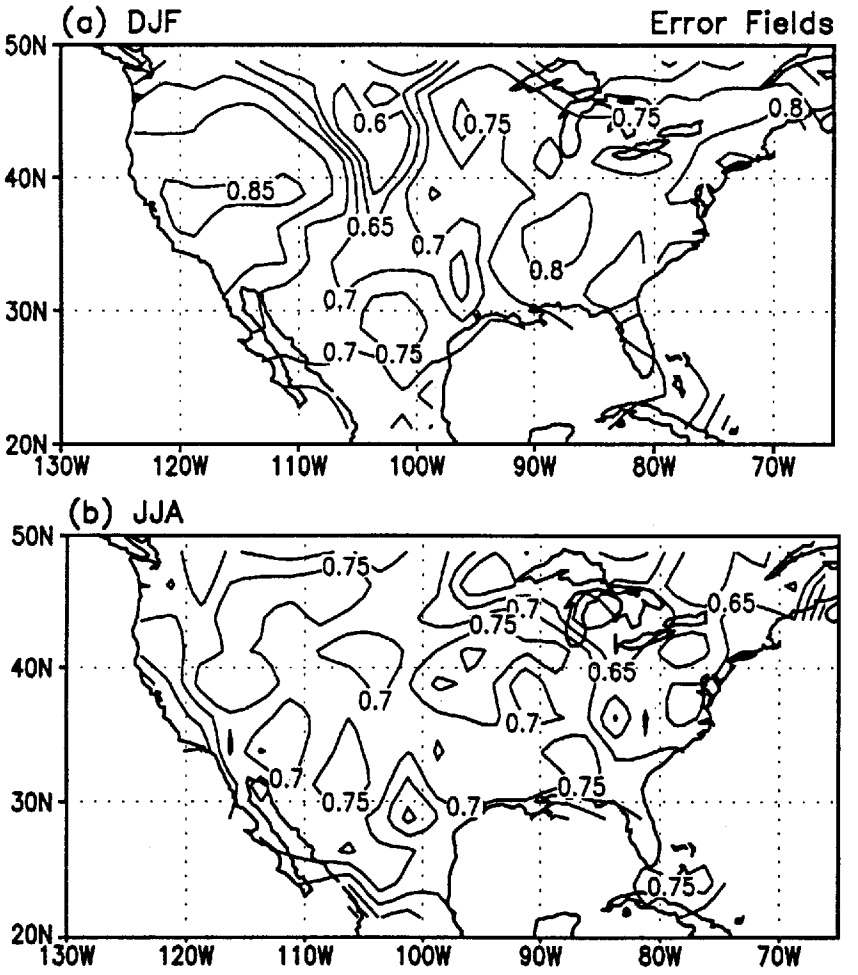

Fig. 11. Expected values of the forecasting error when predicting the (a) DJF precipitation by using the previous season's SON SST, and (b) JJA precipitation by using the same year's MAM SST.

7. SUMMARY AND CONCLUSIONS

This paper has described the mathematical formulation of the canonical ensemble correlation (CEC) forecasting model of Lau et al. (2002). The mean square forecasting error is estimated and gives a quantitative, mean quality assessment of forecasting. The error's spatial distribution reflects the physical interaction mechanisms between the predictor and predictand. The third is the optimal combination of the multiple forecasts. The third is possible only when the error estimates are made in the second.

Our forecast experiments show that the new ensemble canonical correlation scheme renders a reasonable forecasting skill. For example, when using September–October–November SST to predict the December–January–February precipitation of the next year, the spatial pattern correlation between the observed and predicted is positive in 46 of the 50 years of

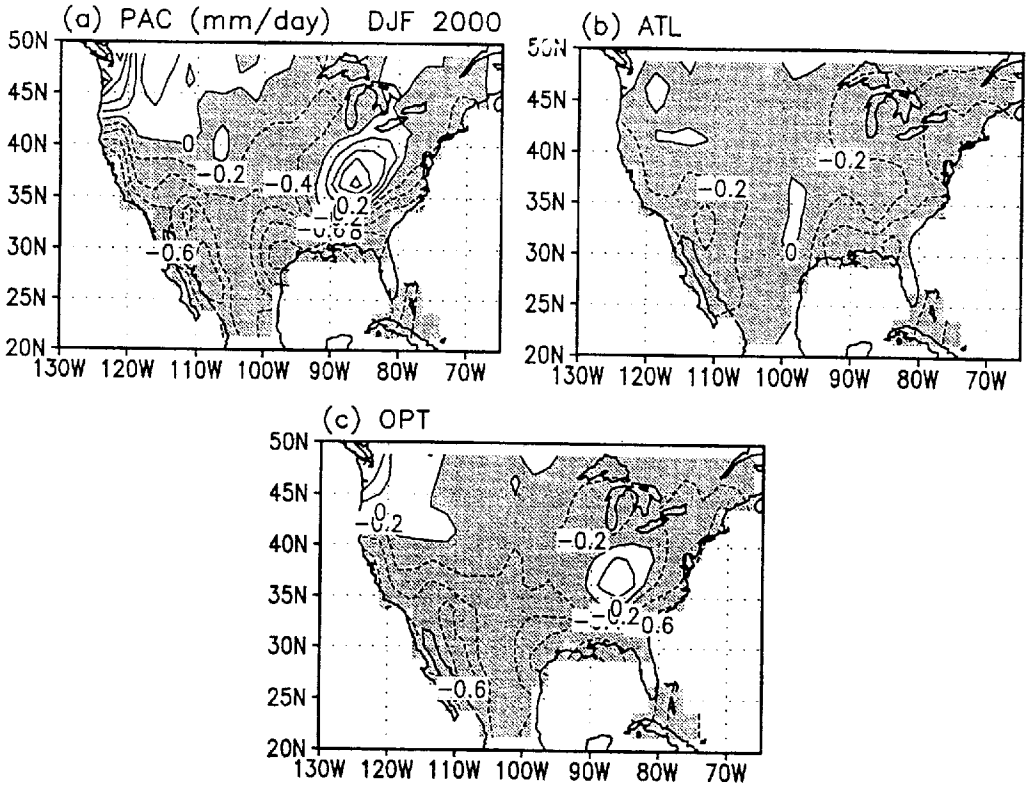

Fig. 12. (a) Predicted 2000 DJF precipitation by using the previous season's SON Pacific SST, (b) Predicted 2000 DJF precipitation by using the previous season's SON Atlantic SST, and (c) the optimal ensemble forecast, which is the optimal combination of the CCA forecasting results from (a) and (b). The spatial pattern correlations between the predicted and the observed (Fig. 7a) are 0.43 for (a), 0.04 for (b) and 0.44 for (c).

experiments. The positive correlations are equal to or greater than 0.4 in 29 years, a result which indicates the excellent performance of the forecasting model. The forecasting skill can be further enhanced when careful selection of ocean basins is made and more predictors are used.

From the results presented in this paper, we may conclude that (i) the CEC model of Lau et al. (2002) is an effective prediction tool, (ii) the error estimate of the new prediction tool is crucial in finding the optimal ensemble forecasting, (iii) other predictors, such as sea level pressure and soil moisture, should be considered, (iv) the GCM simulation data may be used to find the CCA statistical structure, and (v) further research is needed to determine the optimal EOF truncation order.

Of course, precipitation over some areas is not directly related to SST anomalies, but more related to SLP or other predictors. The CEC model has much flexibility and can easily include several predictors. Using multiple predictors can be an effective way to increase the forecasting skill. Therefore, a careful tuning of the model and a careful selection of predictors, such as soil moisture and SLP, can enhance the skill of prediction. Thus, our next project is to include other predictors in our optimal ensemble forecasting model.

The prediction skill can be further enhanced by using an extrapolation type of optimal combination of forecasts. Some of the forecasts with negative temporal correlation signs can also help with enhancing the skill of the optimal ensemble forecasting skill. Some of the weights of the forecasts can be negative. The weights are computed from a linear regression scheme similar to that of Krishnamurti et al. (2000). The assumption of the independence of the forecasting errors of different models is no longer needed. This work is deferred to future studies.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Global Modelling and Data Analysis Program and the Tropical Rainfall Measuring Mission of the NASA Earth Science Enterprise for partial support of this study. This work was carried out while Shen held a National Research Council–(GSFC / NASA) Research Associateship. This work was partially supported by the Atmospheric Environment Service of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Programs of "Overseas Well–known Scholars" and "Overseas Assessors" of the Chinese Academy of Sciences.

REFERENCES

Barnett, T. P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825–1850.

Barnston, A. G., and Y. He, 1998: Skill summary of long–lead prediction of the ENSO conditions from Fall 1996 to Fall 1998, *Proceedings of the Twenty–Third Annual Climate Diagnostics and Prediction Workshop*, Miami, Florid, 86–89.

Barnston, A. G., and T. M. Smith, 1996: Specification and prediction of global surface temperature and precipitation from global SST using CCA, *J. Clim.*, **9**, 2660–2697.

Chen, M., P. Xie, J. E. Janowick, and P. A. Arkin, 2001: Global land precipitation: 1 50–year monthly analysis based on gauge observations, *J. Hydrometeorology*. (submitted for publication)

Chen, D. K., S. E. Zebiak, M. A. Cane, and A. J. Busalacchi, 1997: Initialization and predictability of a coupled ENSO forecast Model, *Mon. Wea. Rev.*, **125**, 773–788.

Higgins, W., J. E. Janowiak, and Y.–P, Yao, 1996: A gridded hourly precipitation data base for the United States (1963–1993), NCEP / Climate Prediction Center ATLAS No. 1, Camp Springs, MD 20746.

Higgins, R. W., A. Leetmaa, Y. Xue, and A. Barnston, 2000: Dominant factors influencing the seasonal predictability of U.S. precipitation and surface air temperature, *J. Climate*, **13**, 3994–4017.

Kim, K.–Y., and G. R. North, 1998: EOF–based linear prediction algorithm: Theory, *J. Clim.*, **11**, 3046–3056.

Kim, K.–Y., and G. R. North, 1999: EOF–based linear prediction algorithm: Examples, *J. Clim.*, **12**, 2076–2092.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Willford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate, *J. Clim.*, **23**, 4196–4216.

Lau, K. M., K. M. Kim, and S. S. P. Shen, 2002: Canonical ensemble correlation prediction scheme for seasonal precipitation, *Geophys. Res. Lett.* (in press).

Lau, K. M., and L. Peng, 1992: Dynamics of atmospheric teleconnections during the northern summer, *J. Climate*, **5**, 140–158.

Lau, K. M., and H.–T. Wu, 1999: Assessment of the impacts of the 1997–1998 El Niño on the Asian–Australia monsoon, *Geophys. Res. Lett.*, **26**, 1747–1750.

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

Ropelewski, C. F. and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with El Niño / Southern Oscillation (ENSO), *Mon. Wea. Rev.*, **114**, 2352–2362.

Shen, S. S. P., K.–M. Lau, K.–M. Kim, and G. Li, 2001: A canonical ensemble correlation prediction model for seasonal precipitation anomaly. NASA / TM–bf 2001–209989, 53pp.

Shen, S. S., T. M. Smith, C. F. Ropelewski and R. E. Livezey, 1998: An optimal regional averaging method with error estimates and a test using tropical Pacific SST Data, *J. Climate*, **11**, 2340–2350.

Smith T. M., R. W. Reynolds, R. E. Livezey, D. C. Stokes, 1996: Reconstruction of historical sea surface temperature using empirical orthogonal functions, *J. Climate*, **9**, 1403–1420.

Ting, M. F. and H. Wang, 1997: Summertime US precipitation variability and its relation to Pacific sea surface temperature, *J. Climate*, **10**, 1853–1873.

van den Dool, H. M., J. Hopinggarner, E. O'Lenic, A. J. Wagner, J. Huang, and R. Churchill, 1997: Second annual review of skill of CPC real time long lead prediction, *Proceedings of the Twenty–Second Annual Climate Diagnostics and Prediction Workshop*, Berkeley, California, pp.10–14.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*, Cambridge University Press, 1999, Chs. 14 and 15.

Wang, H., M. F. Ting, and M. Ji, Prediction of seasonal mean United States precipitation based on El Niño sea surface temperatures. *Geohphys. Res. Lett.*, **26**, 1341–1344.

Xue, Y., A. Leetmaa, and M. Ji, 2000: ENSO prediction with Markov models: the impact of sea level, *J. Climate*, **13**, 849–871.