EOF-Based Linear Prediction Algorithm: Theory

KWANG-Y. KIM AND GERALD R. NORTH

Climate System Research Program, Texas A&M University, College Station, Texas

(Manuscript received 1 August 1997, in final form 16 December 1997)

ABSTRACT

This study considers the theory of a general three-dimensional (space and time) statistical prediction/extrapolation algorithm. The predictor is in the form of a linear data filter. The prediction kernel is based on the minimization of prediction error and its construction requires the covariance statistics of a predictand field. The algorithm is formulated in terms of the spatiotemporal EOFs of the predictand field. This EOF representation facilitates the selection of useful physical modes for prediction. Limited tests have been conducted concerning the sensitivity of the prediction algorithm with respect to its construction parameters and the record length of available data for constructing a covariance matrix. Tests reveal that the performance of the predictor is fairly insensitive to a wide range of the construction parameters. The accuracy of the filter, however, depends strongly on the accuracy of the covariance matrix, which critically depends on the length of available data. This inaccuracy implies suboptimal performance of the prediction filter. Simple examples demonstrate the utility of the new algorithm.

1. Introduction

An important class of linear estimation problems is the extrapolation (prediction) of a field outside the data domain. Needs for reliable predictions arise from many different areas of climate studies, El Niño being a prototypical example. The importance of reliable prediction cannot be overemphasized because any significant changes in climatic variables typically have large impacts on climatic, environmental, and economic variables. Considered here is a general three-dimensional statistical predictor.

Prediction schemes may be classified into one of the three categories—statistical, dynamical, and statistical–dynamical. In El Niño predictions, examples of statistical schemes include Barnett et al. (1988) and Xu and von Storch (1990). Examples of dynamical schemes include Cane et al. (1986) and Latif and Flügel (1991). Inoue and O'Brien (1984) used a combination of the two called the statistical–dynamical scheme. This study presents a statistical prediction algorithm based on space–time empirical orthogonal functions (EOFs) of the predictand field.

The present method, in essence, is similar to purely statistical predictors such as autoregressive (AR) predictors discussed in many time series textbooks (e.g., Newton 1988). In the latter, a covariance matrix is used to construct the best unbiased predictor for a given time series. The so-called prediction normal equation is solved to make predictions. The predictor is ''best'' in the sense of minimum error variance. Also, the predictor field is always a subset of the predictand field. In this study, a similar predictor is developed in terms of EOFs.

EOFs naturally come into play because the covariance function (or matrix) is decomposed into a series of EOFs. One important motivation for this EOF representation is that it facilitates a space–time generalization of the method. Spatial information is decomposed into ''independent'' modes that can be predicted separately. Further, the algorithm can easilty be carried over to other basis functions including complex EOFs or cyclostationary EOFs (Kim et al. 1996; Kim and North 1997). A certain set of basis functions may have better representation of true physical modes and using it can improve the performance of predictors. This is an important aspect of the present prediction algorithm. Finally, some EOF modes may be more beneficial for prediction than others. The EOF representation allows the selection of particular (physical) modes for prediction thereby improving the predictability.

Section 2 of this article describes the linear prediction algorithm and the prediction error variance. The algorithm is constructed so that the prediction error variance is minimized. Section 3 shows how to compute finite-domain EOFs, which are an essential ingredient in the formulation of the prediction algorithm. The section discusses the actual implementation of the predictor from a continuum representation in the previous section. In

Corresponding author address: Dr. Kwang-Yul Kim, Dept. of Meteorology/Climate System Research Prog., College of Geosciences and Maritime Studies, Texas A&M University, College Station, TX 77843-3150.
E-mail: kykim@csrp.tamu.edu

section 4, the performance and properties of the predictor are illustrated in terms of two artificial Markov processes. Actual applications of the predictor require extensive cross validations and have been postponed to a later study. It is followed by sensitivity tests in section 5. In this section sensitivity of the predictor is examined in terms of its construction parameters and the accuracy of covariance statistics. Concluding remarks follow in section 6.

2. Method

a. Formulation of optimal filter

Because of the simplicity, derivation here assumes continuum forms. In the next section, actual implementation of the predictor using discrete variables in space and time will be discussed.

We have a space–time domain $\mathcal{R}$ in which a field $T(x)$, $x = (\hat{\mathbf{r}}, t)$, is defined. Throughout we will deal with the anomaly field such that $\langle T(x)\rangle = 0$, where $\langle \cdot \rangle$ denotes ensemble average. We wish to predict $T(x)$ for a particular realization at a point $x \in \mathcal{R}$ (Fig. 1). The prediction is to be based upon measurements of $T(x)$ taken from that same realization in a subset $\mathcal{D}$ of $\mathcal{R}$, namely, $\mathcal{D} \subset \mathcal{R}$. We will call $\mathcal{D}$ data domain and $\mathcal{R}$ prediction domain. The prediction is to be optimal in the least mean-square-error (mse) sense. Let $\hat{T}(x)$ be the prediction value at point x. Then, our predictor can be defined as

$$\hat{T}(x) = \int_{\mathcal{D}} \Gamma(x, x')T(x')\, dx' \tag{1}$$

$$= \int_{\mathcal{R}} \chi(x')\Gamma(x, x')T(x')\, dx', \tag{2}$$

where

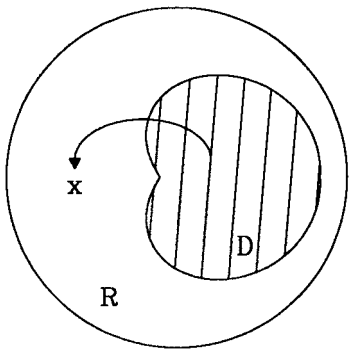

FIG. 1. Schematic drawing of the two domains $\mathcal{D}$ (data domain) and $\mathcal{R}$ (prediction domain). Prediction is from the data domain to a point x in the prediction domain.

$$\chi(x) = \begin{cases} 1, & x \in \mathcal{D} \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

and $\Gamma(x, x')$ is a filter which we tailor to our needs. The filter provides a weighting over the data that leads to the prediction.

The error made in predicting $T(x)$, $x \in \mathcal{R}$, for a particular realization is

$$\epsilon(x) = T(x) - \hat{T}(x), \tag{4}$$

so that the mse is given by

$$\langle \epsilon^2(x)\rangle = \langle (T(x) - \hat{T}(x))^2\rangle. \tag{5}$$

Our goal is to minimize $\langle \epsilon^2(x)\rangle$ by adjusting the shape of the filter $\Gamma(x, x')$.

Inserting (2) into (4), we obtain

$$\langle \epsilon^2(x)\rangle = \langle T^2(x)\rangle - 2\int_{\mathcal{R}} \chi(x')\Gamma(x, x')\langle T(x)T(x')\rangle\, dx' + \int_{\mathcal{R}}\int_{\mathcal{R}} \chi(x')\chi(x'')\Gamma(x, x')\Gamma(x, x'')\langle T(x')T(x'')\rangle\, dx'\, dx''. \tag{6}$$

To find the optimum filter we take the variation of $\langle \epsilon^2\rangle$ with respect to the filter function and set it to zero. Then, we find

$$\int_{\mathcal{R}}\int_{\mathcal{R}} \chi(x')\chi(x'')\delta\Gamma(x, x')\Gamma(x, x'')\langle T(x')T(x'')\rangle\, dx'\, dx''$$

$$= \int_{\mathcal{R}} \chi(x')\delta\Gamma(x, x')\langle T(x)T(x')\rangle\, dx'. \tag{7}$$

As usual, if (7) is to be true for arbitrary $\delta\Gamma$, we should have

$$\chi(x')\int_{\mathcal{R}} \chi(x'')\Gamma(x, x'')C(x', x'')\, dx'' = \chi(x')C(x, x'), \tag{8}$$

where

$$C(x, x') = \langle T(x)T(x')\rangle \tag{9}$$

is the space–time lagged covariance of the random field to be predicted. Substituting the above integral equation into (6), we obtain the minimum mse:

$$\langle \epsilon^2(x)\rangle = C(x, x) - \int_{\mathcal{R}} \chi(x')\Gamma_{\text{opt}}(x, x')C(x, x')\, dx'. \tag{10}$$

The major problem remaining in practical applications of this procedure is to solve the appropriate integral equation (8) using a known covariance function. Note that difficulties arise due to the presence of $\chi(x)$ under the integral sign. The expression (8) constitutes an integral equation for $\Gamma(x, x')$, which is the desired filter. To solve the integral equation, we introduce a set of basis functions $\Psi_n(x)$ defined on $\mathcal{R}$; that is,

$$\int_{\mathcal{R}} C(x, x')\Psi_n(x')\, dx' = R_n\Psi_n(x), \tag{11}$$

where R_n are eigenvalues. They are orthonormal since $C(x, x')$ is symmetric:

$$\int_{\mathcal{R}} \Psi_n(x)\Psi_m(x)\, dx \equiv (\Psi_n, \Psi_m)_{\mathcal{R}} \tag{12}$$

$$= \delta_{nm}, \tag{13}$$

where we have introduced an inner product notation in (12) with subscript denoting the domain of integration and Kronecker delta in (13). The basis set is also complete:

$$\sum_n \Psi_n(x)\Psi_n(x') = \delta(x - x'), \qquad x, x' \in \mathcal{R}. \tag{14}$$

Because of the completeness, the covariance function and the filter can be expressed in terms of basis functions as

$$C(x, x') = \sum_n R_n\Psi_n(x)\Psi_n(x'), \qquad x, x' \in \mathcal{R}, \tag{15}$$

$$\Gamma(x, x') = \sum_n \Gamma_n^{\mathcal{R}}(x)\Psi_n(x'), \qquad x' \in \mathcal{R}, \tag{16}$$

where $\Gamma_n^{\mathcal{R}}(x)$ are the expansion coefficients of the filter kernel. The superscript signifies that the expansion is with respect to the $\mathcal{R}$-domain EOFs. The function $\Psi_n(x)$ yields the EOFs defined on the space–time domain $\mathcal{R}$. Note that these eigenfunctions are defined on the whole domain $\mathcal{R}$, but the integral in (8) runs only over the subdomain $\mathcal{D}$ because of the factor $\chi(x)$.

Realizations of the random field $T(x)$ can be expressed as

$$T(x) = \sum_n T_n^{\mathcal{R}}\Psi_n(x), \qquad x \in \mathcal{R}, \tag{17}$$

where the coefficients $T_n^{\mathcal{R}}$ are random numbers, which have zero mean and are uncorrelated with each other:

$$\langle T_n^{\mathcal{R}}\rangle = 0, \tag{18}$$

$$\langle T_n^{\mathcal{R}}T_m^{\mathcal{R}}\rangle = R_n\delta_{nm}. \tag{19}$$

In addition it is convenient to introduce the orthogonal basis set, $\{\Phi_n(x)\}$ defined on the subregion $\mathcal{D}$; that is,

$$\int_{\mathcal{D}} C(x, x')\Phi_n(x')\, dx' = D_n\Phi_n(x), \qquad x, x' \in \mathcal{D}, \tag{20}$$

where D_n are eigenvalues. The basis functions are again orthonormal

$$(\Phi_n, \Phi_m)_{\mathcal{D}} = \delta_{nm}, \tag{21}$$

and complete

$$\sum_n \Phi_n(x)\Phi_n(x') = \delta(x - x'), \qquad x, x' \in \mathcal{D}. \tag{22}$$

Then, it can be shown that

$$T(x) = \sum_n T_n^{\mathcal{D}}\Phi_n(x), \qquad x \in \mathcal{D}, \tag{23}$$

with

$$\langle T_n^{\mathcal{D}}\rangle = 0, \tag{24}$$

$$\langle T_n^{\mathcal{D}}T_m^{\mathcal{D}}\rangle = D_n\delta_{nm}. \tag{25}$$

Further, the covariance function is given by

$$C(x, x') = \sum_n D_n\Phi_n(x)\Phi_n(x'), \qquad x, x' \in \mathcal{D}. \tag{26}$$

Then, (8) can be rewritten as

$$\chi(x')\sum_n \Gamma_n^{\mathcal{D}}(x)D_n\Phi_n(x') = \chi(x')C(x, x'),$$

$$x \in \mathcal{R}, \qquad x' \in \mathcal{D}, \tag{27}$$

where $\Gamma_n^{\mathcal{D}}(x)$ are the expansion coefficients of the filter kernel with respect to the $\mathcal{D}$-domain EOFs $\Phi_n(x')$ [see also (16)]. Next, we project the nth component of (27) [multiply through by $\Phi_n(x')$ and integrate over $\mathcal{D}$] to obtain

$$D_n\Gamma_n^{\mathcal{D}}(x) = \int_{\mathcal{D}} C(x, x')\Phi_n(x')\, dx'. \tag{28}$$

Dividing through by D_n, multiplying by $\Phi_n(x)$, and summing

$$\Gamma_{\text{opt}}(x, x') = \sum_n \frac{\Phi_n(x')}{D_n}\int_{\mathcal{D}} C(x, x'')\Phi_n(x'')\, dx'',$$

$$x \in \mathcal{R} - \mathcal{D}. \tag{29}$$

Finally, it is useful to write the $C(x, x')$ in terms of the $\Psi_n(x)$. Using (15),

$$\Gamma_{\text{opt}}(x, x') = \sum_{n,m} \frac{R_m}{D_n}\Psi_m(x)\Phi_n(x')(\Phi_n, \Psi_m)_{\mathcal{D}},$$

$$x \in \mathcal{R} - \mathcal{D}. \tag{30}$$

Note that $\Psi_m(x)$ is defined for values of x outside of $\mathcal{D}$, which allows $\Gamma_{\text{opt}}(x, x')$ to be defined for x throughout $\mathcal{R}$. The above suggests that the optimal linear filter is easily found if the functions $\Psi_m(x)$, $\Phi_n(x)$ and the eigenvalues R_m, D_n, $n, m = 0, 1, 2, \ldots$ are found in advance. The final formula for the predictor is

$$\hat{T}(x) = \sum_{n,m} \frac{R_m}{D_n}\Psi_m(x)(T, \Phi_n)_{\mathcal{D}}(\Phi_n, \Psi_m)_{\mathcal{D}}. \tag{31}$$

b. Prediction error

The error $\epsilon(x)$ for a particular realization from (1) and (4) is

$$\epsilon(x) = \sum_m T_m^{\mathcal{R}} \Psi_m(x) - \sum_n T_n^{\mathcal{D}} \int_{\mathcal{D}} \Gamma(x, x')\Phi_n(x')\, dx'. \tag{32}$$

From (28)

$$\Gamma_n^{\mathcal{D}}(x) = \sum_j \frac{R_j}{D_n}\Psi_j(x)(\Phi_n, \Psi_j)_{\mathcal{D}} \tag{33}$$

and from (17) and (23),

$$T_n^{\mathcal{D}}(x) = \sum_i T_i^{\mathcal{R}}(\Phi_n, \Psi_i)_{\mathcal{D}}. \tag{34}$$

Then, we can derive

$$\epsilon(x) = \sum_m \left\{\Psi_m(x) - \sum_{n,j} \frac{R_j}{D_n}(\Phi_n, \Psi_m)_{\mathcal{D}}(\Phi_n, \Psi_j)_{\mathcal{D}}\Psi_j(x)\right\} T_m^{\mathcal{R}}. \tag{35}$$

Since $\langle T_m^{\mathcal{R}}\rangle = 0$, the estimate is unbiased.

If x happens to lie inside $\mathcal{D}$, we can make replacements:

$$\begin{aligned}\sum_j R_j(\Phi_n, \Psi_j)_{\mathcal{D}}\Psi_j(x) &= \int_{\mathcal{D}} \Phi_n(x') \sum_j \Psi_j(x')R_j\Psi_j(x)\, dx' \\ &= \int_{\mathcal{D}} \Phi_n(x')C(x, x')\, dx' \\ &= \int_{\mathcal{D}} \Phi_n(x') \sum_j \Phi_j(x')D_j\Phi_j(x)\, dx' \\ &= D_n\Phi_n(x).\end{aligned} \tag{36}$$

Finally,

$$\begin{aligned}\epsilon(x \in \mathcal{D}) &= \sum_m T_m^{\mathcal{R}}(x)\left\{\Psi_m(x) - \sum_n (\Phi_n, \Psi_m)\Phi_n(x)\right\} \\ &= \sum_m T_m^{\mathcal{R}}(x)\{\Psi_m(x) - \Psi_m(x)\} = 0,\end{aligned} \tag{37}$$

where we used the completeness of $\Phi_n(x)$, $x \in \mathcal{D}$, in the last step. Hence, we can say that the error vanishes identically for $x \in \mathcal{D}$, realization by realization and term by term in the sum over EOF index. By a similar reasoning it can be shown that if $x \in \mathcal{D}$, $\Gamma(x, x') = \delta(x - x')$.

It is also useful to think of (35) as a random number that consists of a weighted sum of uncorrelated random numbers, $T_m^{\mathcal{R}}$. The ensemble average of the square of $\epsilon(x)$ can now be conveniently written as

$$\langle\epsilon^2(x)\rangle = \sum_m \left[\Psi_m(x) - \sum_{n,j} \frac{R_j}{D_n}(\Phi_n, \Psi_m)_{\mathcal{D}}(\Phi_n, \Psi_j)_{\mathcal{D}}\Psi_j(x)\right]^2 R_m. \tag{38}$$

Note that if $x \in \mathcal{D}$, each factor in curly braces vanishes for every value of m in the sum. Hence, $\epsilon(x)$ as well as $\epsilon^2(x)$ vanish at any truncation level so long as $x \in \mathcal{D}$. This is true only in a continuum sense.

3. Finite-domain EOFs

In the actual implementation of (30), (31), and (38), it should be interpreted that $\mathcal{D}$ and $\mathcal{R}$ in the derived equations are subsets of the data and prediction domains, respectively. Namely, you do not have to use all the available data for prediction. In practice, $\mathcal{D}$ may well be much smaller than the full data domain for computational efficiency. As will be seen in the sensitivity tests, the performance of the predictor is fairly insensitive to the domain sizes as long as the domain is large enough to include important periodicities in the data.

It is also stressed that covariance function depends only on lag under the stationarity assumption. This implies that $\mathcal{R}$-domain EOFs can be estimated based on data elsewhere. Without the assumption of stationarity or cyclostationarity it is impossible to estimate $\mathcal{R}$-domain EOFs reasonably. Thus, for a nonstationary process the present algorithm cannot be implemented. To obtain reasonable EOFs, it is important to compute covariance function as accurately as possible. The accuracy of covariance function depends strongly on the length of data. For short record, covariance function may be contaminated by sampling errors. For this reason, one may utilize all the available data unless there is a compelling reason to exclude any part of the data.

Preparatory to a presentation of some elementary examples of the predictor algorithm, we need a few results pertaining to stationary time series restricted to a finite interval. We seek the eigenfunction $\Psi_n(x)$, such that

$$\int_{\mathcal{R}} C(x, x')\Psi_n(x')\, dx' = R_n\Psi_n(x), \tag{39}$$

where $C(x, x')$ is the covariance kernel for a stationary process and R_n is the eigenvalue corresponding to $\Psi_n(x)$. For stationary processes, we can conveniently factor the space and time dependence of the eigenfunctions:

$$\Psi_n(x) = \Theta_n(\hat{\mathbf{r}})\psi_n(t). \tag{40}$$

This is an approximation and is true only when the domain size approaches infinity.

For simplicity, let us consider a univariate time series problem—that is, $x = t$. Equation (39) is reduced to

$$\int_0^L C(t, t')\psi_n(t')\, dt' = R_n\psi_n(t). \tag{41}$$

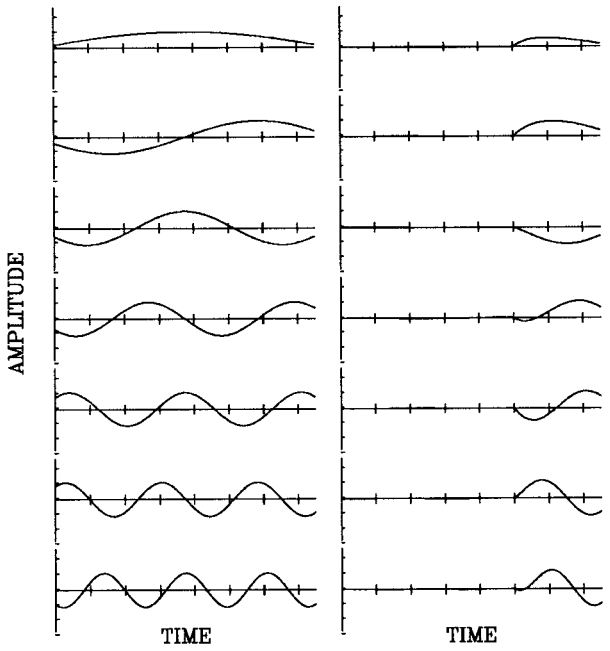

FIG. 2. The first seven $\mathcal{R}$-domain EOFs (left) and the associated errors (right) of the first-order Markov process ($\tau_0 = 10$ and $\sigma_F = 1$). Error occurs in representing $\mathcal{R}$-domain EOFs in terms of $\mathcal{D}$-domain EOFs. The length of $\mathcal{R}$ domain and $\mathcal{D}$ domain are 150 and 100 units, respectively.

We may represent $C(\tau)$, $\tau = t - t'$, by its Fourier integral in terms of the spectral density, $S(f)$:

$$C(\tau) = \int_{-\infty}^{\infty} S(f) e^{2\pi i f \tau}\, df. \tag{42}$$

Since the spectral density is symmetric about $f = 0$, $C(\tau)$ is a real function. For a discrete time series, the eigenfunctions $\psi_n(t)$ can be approximated on the interval $(0, L)$ as solutions of a matrix equation

$$\sum_{m=0}^{L} C_{mm'} \psi_n(m') = R_n \psi_n(m), \tag{43}$$

where

$$C_{mm'} = C(t = m;\, t' = m'), \qquad m, m' = 0, 1, \ldots, L. \tag{44}$$

Note that this defines a general problem of finding empirical orthogonal vectors for a stationary process restricted to a finite segment of length L with a known discrete covariance kernel. The eigenvectors converge to exact eigenfunctions (sine and cosine functions) as $L \to \infty$. The spectrum of eigenvalues R_n also approaches $S(f = n/L)$ for an infinitely long time series.

For a three-dimensional discrete dataset spatial EOFs may be computed as eigenvectors of spatial covariance matrix $\{\mathbf{C}_{jk}\}$ where j and k are station indices. Then, temporal EOFs are computed as eigenvectors of each principal component (PC) time series by invoking the factorization (40). Note that the continuum equations such as (30), (31), and (38) can readily be generalized for discrete variables by replacing summations for integrals with the understanding that T, Φ_n, and Ψ_n are vectors. For instance, (12) represents a dot product of two vectors for discrete variables.

4. One-dimensional prediction examples

a. A first-order Markov process

Consider a prediction problem for a first-order Markov process. The process is governed by a Langevin equation:

$$\frac{dT(t)}{dt} + \frac{T(t)}{\tau_0} = F(t), \tag{45}$$

where $F(t)$ is assumed to be a white noise process and satisfies

$$\langle F(t) \rangle = 0, \tag{46}$$

$$\langle F(t) F(t') \rangle = \sigma_F^2 \delta(t - t'). \tag{47}$$

Complete solution of (45) is obtained in the form

$$T(t) = e^{-t/\tau_0}\left(A + \int^{t} e^{p\tau_0} F(p)\, dp\right)$$

$$= e^{-t/\tau_0} \int_{-\infty}^{t} e^{p\tau_0} F(p)\, dp, \tag{48}$$

where we let $T(t)$ be zero as $t \to -\infty$.

Now, the covariance function is given by

$$C(t, t') = \frac{\tau_0 \sigma_F^2}{2} e^{-|t'-t|/\tau_0}. \tag{49}$$

The spectral density function is an inverse Fourier transform of the covariance function. From (49), we can show that

$$S(f) = \int_0^{\infty} \tau_0 e^{-\tau/\tau_0} \cos(2\pi f \tau)\, d\tau$$

$$= \frac{\tau_0^2 \sigma_F^2}{1 + (2\pi f \tau_0)^2}. \tag{50}$$

Eigenvectors and eigenvalues were, respectively, computed from (49) and (50) as suggested in the previous section.

As should be expected, $\mathcal{D}$-domain and $\mathcal{R}$-domain EOFs resemble sine and cosine functions for the latters are exact eigenfunctions of (49). Discrepancy occurs mainly near the end points and originates with the finiteness of the domain size. To gain further insight into the prediction error let us examine (35) carefully. For each mode number m, (35) without the last multiplicative term, $T_m^{\mathcal{R}}$, represents error in predicting the amplitude of each EOF of the $\mathcal{R}$ domain in terms of $\mathcal{D}$-domain EOFs. Figure 2 shows the first seven $\mathcal{R}$-domain

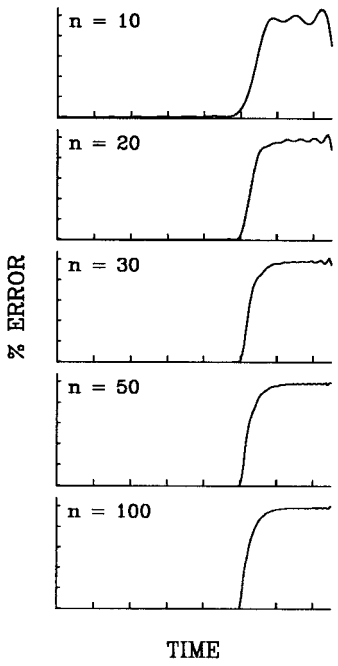

FIG. 3. Plot of prediction error squared as a function of the number of retained modes for the first-order Markov process ($\tau_0 = 10$ and $\sigma_F = 1$). Each curve approaches the plateau representing no predictability (100% prediction error).

EOFs and the associated errors. Error occurs because the $\mathcal{D}$-domain EOFs are not complete in the $\mathcal{R}$ domain. As proved in (37), error should be zero when $x \in \mathcal{D}$, which indeed is the case in Fig. 2.

Now, Fig. 3 shows the prediction error squared, (38), as a function of the number of modes retained. Because of discretization error, variance is not exactly zero in $\mathcal{D}$ domain for the first panel. We expect that the representation (35) of the $\mathcal{R}$-domain eigenfunctions in terms of the $\mathcal{D}$-domain eigenfunctions will be improved if we retain more modes. Indeed, the representation of prediction error improves as the number of retained modes is increased. The improvement is shown in the figure by the error curves approaching a theoretical one $1 - (1 - 1/\tau_0)^h$, h being the prediction horizon (Newton 1988), with the diminution of ripples as more modes are retained. By retaining more modes, however, prediction error may not be reduced beyond a certain limit because it is the property of the particular process involved not of the number of modes retained. The prediction error quickly reaches the plateau implying that there is no prediction beyond that point (100% prediction error variance = total variance).

The optimal prediction kernel of the first-order Markov process as derived in (30) is shown in Fig. 4. The crest of sharp peaks approaches the Dirac delta function $\delta(t - t')$ (actually Dirac delta sequence). This is the consequence of the fact that for a minimum mse predictor the predicted value should be the same as the

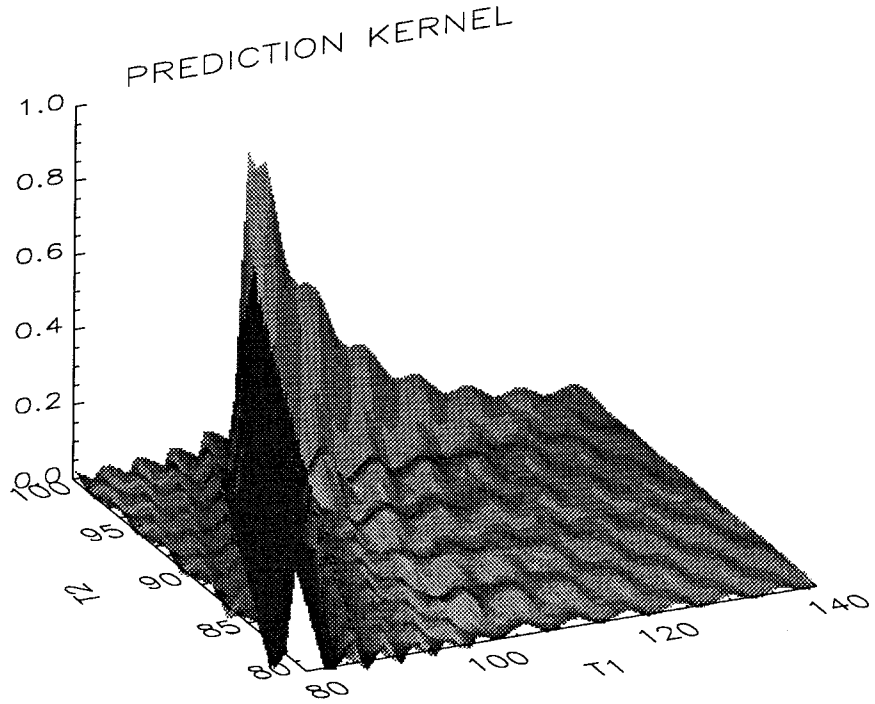

FIG. 4. The optimal prediction kernel, $\Gamma_{opt}(t, t')$, of the first-order Markov process ($\tau_0 = 10$ and $\sigma_F = 1$). Here, (T1, T2) represents (t, t'), where $t \in \mathcal{R}$ and $t' \in \mathcal{D}$. Only the region $(t, t') \in [80, 150] \times [80, 100]$ is plotted.

observed value in the $\mathcal{D}$ domain. Of course, $\hat{T}(x) = T(x)$, $x \in \mathcal{D}$, if $\Gamma(x, x') = \delta(x - x')$ in (1). The largest peak behind the delta function in Fig. 4 decays exponentially both in the postive t and the negative t' directions. This implies that predictability comes from only a small number of observations close to the $\mathcal{D}$ boundary, hardly a surprise given the definition of a Markov process. Also, predictability decreases away from the $\mathcal{D}$-domain boundary into the future in a time equal to the autocorrelation time. The small undulations of the filter function in the foreground indicates the degree of resolution of the covariance function in terms of finite-domain discrete eigenfunctions. The small ripples will disappear if all the EOFs are employed in constructing a prediction filter.

b. A second-order Markov process

As our second example, we consider a second-order Markov sequence representing a damped oscillator forced by noise. The process is governed by a second-order Langevin equation:

$$\frac{d^2T(t)}{dt^2} + \frac{2}{\tau_0}\frac{dT(t)}{dt} + \omega_0^2 T(t) = F(t), \qquad (51)$$

where $F(t)$ is a white noise with constant spectral density σ_F^2. Using the influence function (one-sided Green's function) technique, complete solution of (51) is obtained in the form

$$T(t) = \int_{-\infty}^{t} -\frac{1}{\beta} e^{\alpha(t-p)} \sin\beta(p - t)F(p)\, dp. \qquad (52)$$

After tedious algebraic manipulation, the covariance function is obtained as

$$C(\tau) = \frac{\sigma_F^2}{4\alpha\omega_0^2} e^{-\alpha|\tau|}\left[\cos\beta|\tau| + \frac{\alpha}{\beta}\sin\beta|\tau|\right]. \qquad (53)$$

Further the corresponding spectral density function is given by

$$S(f) = \int_{-\infty}^{\infty} C(\tau)e^{-2\pi i f\tau}\, d\tau$$

$$= \frac{\sigma_F^2}{(\omega^2 - \omega_0^2)^2 + 4\alpha^2\omega^2}, \qquad (54)$$

where $\omega = 2\pi f$.

Figure 5 shows the prediction error squared, (38), as a function of the number of modes retained. As the number of EOF modes is increased, prediction error is better resolved. As in the case of the first-order Markov process, the prediction error quickly reaches the plateau of no predictability (100% prediction error).

The optimal prediction kernel of the second-order Markov process is shown in Fig. 6. The crest of sharp peaks represent the Dirac delta function $\delta(t - t')$. The shape of the peaks of the kernel in the prediction domain $\mathcal{R} - \mathcal{D}$ is much different from that in Fig. 4 for a first-order Markov process. The functions again decay exponentially both in the postive t and the negative t' directions but some undulation is obvious reflecting the oscillatory nature of the second-order Markov process. Information useful for prediction is limited to the region close to the $\mathcal{D}$ boundary. Predictability quickly decays away from the $\mathcal{D}$-domain boundary into the future.

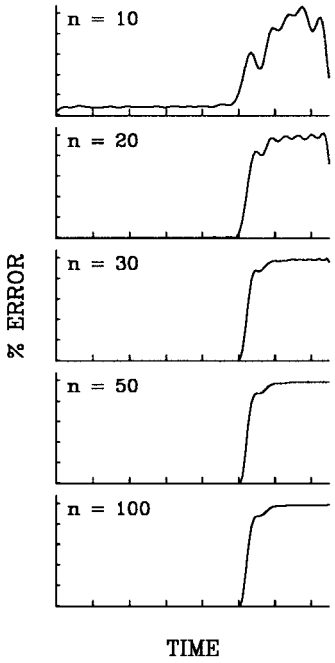

FIG. 5. Plot of prediction error squared as a function of the number of retained modes of the second-order Markov process ($\tau_0 = 10$, $\omega_0 = 0.3$, and $\sigma_F = 1$). Each curve approaches the plateau representing no predictability (100% prediction error).

c. Deterministic plus AR-1 process

Let us now consider a time series consisting of a deterministic signal and a stochastic noise (see Fig. 7). The time series of length 150 was constructed by adding a cosine function and an AR-1 (order-1 autoregressive) process given by

$$T(t + 1) - \alpha T(t) = \epsilon(t), \qquad (55)$$

where $\alpha = 0.78$ and $\epsilon(t)$ is a white noise process with variance 0.11. In statistical predictions it is often advised that known deterministic signals be removed from data prior to prediction. This is because some statistical predictors such as AR cannot handle deterministic components correctly. The process of removing deterministic components, however, is not necessarily easy in practice. Since deterministic processes have unrestricted predictability it is important for a predictor to handle them accurately. As shown in Fig. 7, the developed predictor does a good job in predicting the deterministic component outside the domain $\mathcal{D}$. This is because the

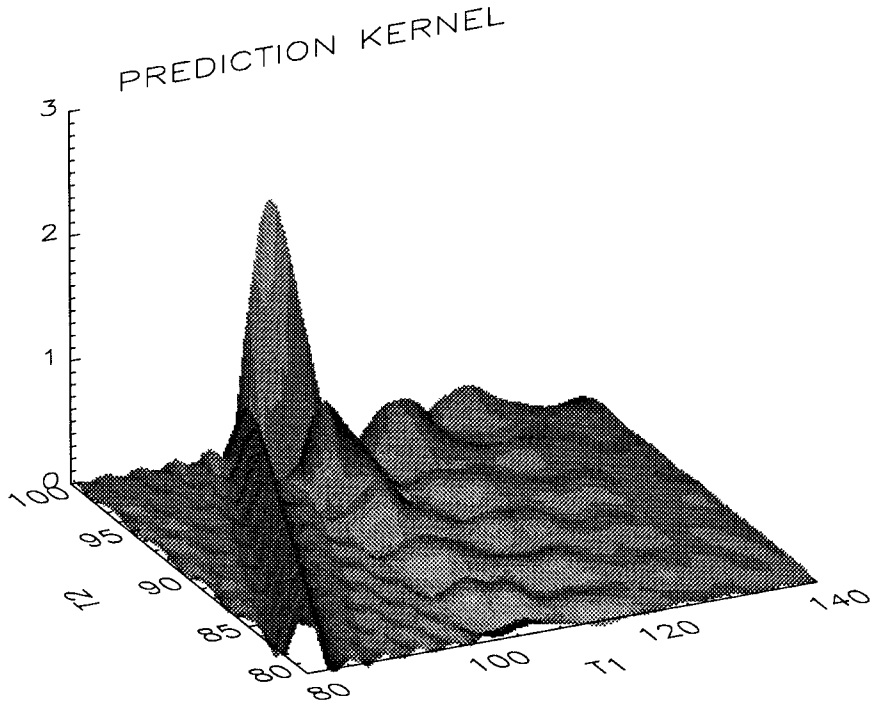

FIG. 6. The optimal prediction kernel, $\Gamma_{opt}(t, t')$, of the second-order Markov process ($\tau_0 = 10$, $\omega_0 = 0.3$, and $\sigma_F = 1$). Here, (T1, T2) represents (t, t') where $t \in \mathcal{R}$ and $t' \in \mathcal{D}$. Only the region $(t, t') \in [80, 150] \times [80, 100]$ is plotted.

data domain is large enough to capture a significant portion of the deterministic component and as a consequence it was well resolved in EOFs. As expected, the AR-1 predictor cannot properly handle the deterministic part. One has to either remove the deterministic component (using differencing) or use a higher-order AR model to resolve the deterministic signal.

Although the deterministic part was predicted accurately, the stochastic part was predicted reasonably for only a few points. The latter part, of course, sets an inherent prediction limit. The prediction error variance is almost comparable to that of an AR-1 prediction for the stochastic part only (see Fig. 8). This is reasonable because both are best unbiased linear predictors. The error variance of an AR-1 predictor is described in Newton (1988). Similarly, for processes other than AR processes one can find best linear unbiased predictors that would perform comparably to the present predictor. Of course, an important motivation for the present study is to develop a predictor that is convenient and easy to understand and that can easily be generalized into three dimensions.

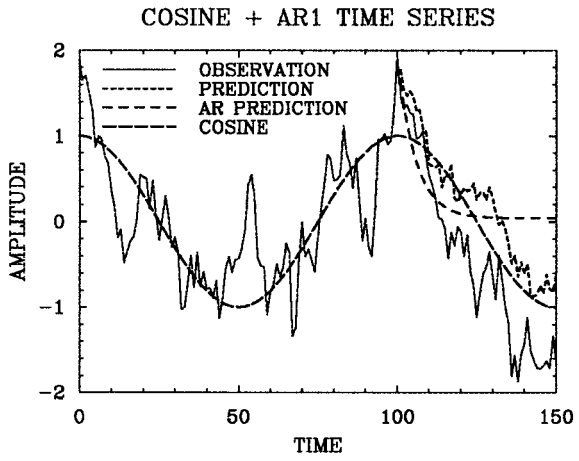

FIG. 7. Plot of cosine plus AR-1 time series (solid line) and 50 predicted values (dotted line) based on the first 100 points of the time series in comparison with the AR-1 prediction (short-dashed line). The long-dashed line is the deterministic component (cosine function) of the time series.

5. Sensitivity tests

There are four parameters regarding the construction of a prediction filter. They are the size of data domain $\mathcal{D}$ and prediction domain $\mathcal{R}$, and EOF truncation mode numbers n and m in the two domains [see (30)]. Another important practical consideration is the accuracy of the covariance matrix which is essentially governed by the total record length. A general and exhaustive sensitivity test employing many different types of datasets is difficult to conduct and is not included here. Instead, sensitivity tests here are limited to an AR-1 process that we discussed earlier in (55).

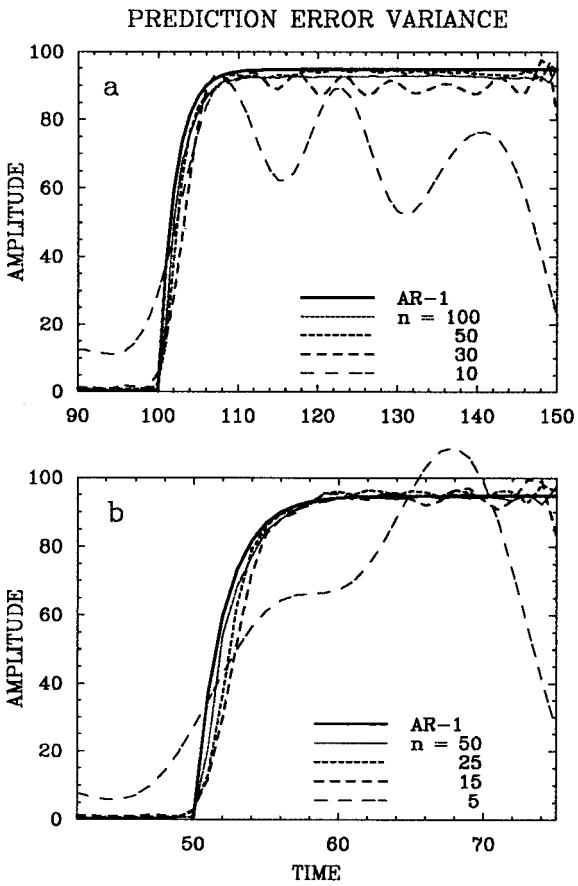

FIG. 8. Plot of prediction error variance vs the number of retained EOF modes: (a) $\mathcal{D} = 100$, $\mathcal{R} = 150$, and (b) $\mathcal{D} = 50$, $\mathcal{R} = 75$. The number of retained EOF modes are 100%, 50%, 30%, and 10% of the $\mathcal{D}$-domain size. The thick solid line represents the error variance of an AR-1 predictor. Error variance is normalized with respect to the total variance of the retained EOF modes.

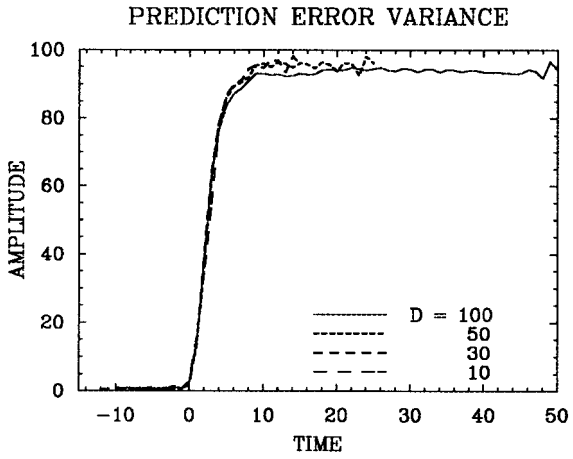

FIG. 9. Plot of normalized prediction error variance vs the $\mathcal{D}$-domain size. Prediction begins at time zero, which is the boundary of the data domain. The selected data domain sizes are 100, 50, 30, and 10 sampling points, respectively. The prediction domain size $\mathcal{R}$ is 1.5 times the data domain. The number of retained EOF modes is half the data domain size.

a. Sensitivity to retained mode numbers

To simplify the problem let us first set $n = m$. In other words, we use an equal number of EOFs in the data and prediction domains in constructing the prediction filter (30). Note that the variance explained by an equal number of EOFs is approximately the same in the data and prediction domains, respectively. Figure 8 shows the prediction error variance versus the number of retained EOF modes. The variance was normalized with respect to the total variance of the retained modes in the $\mathcal{D}$ domain. The error variance of an AR-1 predictor was also plotted for a comparison. As shown in the figure the prediction error variance (38) approaches its theoretical limit as the number of retained modes increases (see also Fig. 3). The prediction error variance asymptotically approaches less than 100% because we choose to set $n = m$. As a result percent variance explained by the retained EOF modes in the prediction domain is slightly less than that in the data domain. This, in turn, causes a slight underestimation of the prediction variance. In the case of $n = m = 10$, prediction error variance is incorrect because the covariability of the data is insufficiently resolved by 10 modes that explain about 60% of the total variability. This test indicates that 30 EOF modes (80% of the total variance) should suffice to design a reasonably accurate filter. This accuracy implies near-optimal performance of the filter. The insensitivity of the performance of the filter to the number of retained modes also implies that the simplification $n = m$ is acceptable for all practical purposes.

b. Sensitivity to domain sizes

Figure 9 shows the error variance versus the size of the data domain $\mathcal{D}$. As shown in the figure, the performance of the predictor is fairly insensitive to the data domain size for this specific test case. This is an expected result since Figs. 4 and 6 show that predictability comes mainly from the data domain boundary. If there is a deterministic component, however, the data domain may preferably be large enough to include a meaningful portion of this deterministic component. A crude test with a cosine function plus AR-1 noise indicates that the data domain should be at least half the cycle of the cosine function for accurate filter function (Fig. 10).

The performance of the predictor is also insensitive to the relative size of the prediction domain $\mathcal{R}$ to the data domain. As shown in Fig. 11, prediction error variance is essentially the same for prediction domains of different size.

c. Sensitivity to the total length of data

The most important ingredient in constructing the prediction filter is a set of EOFs that is derived from a

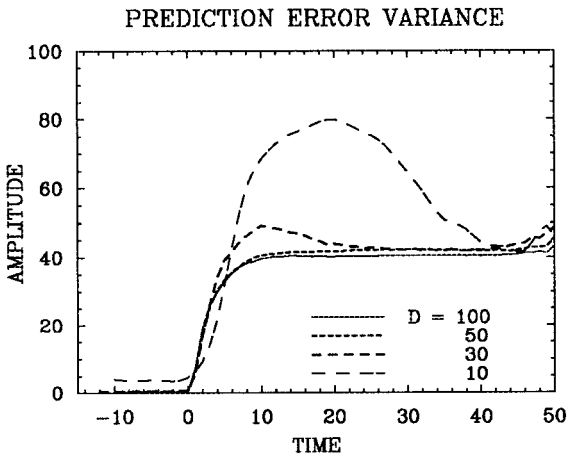

FIG. 10. Same as Fig. 9 but for a time series consisting of a deterministic cosine function with the period of 100 units plus an AR-1 process.

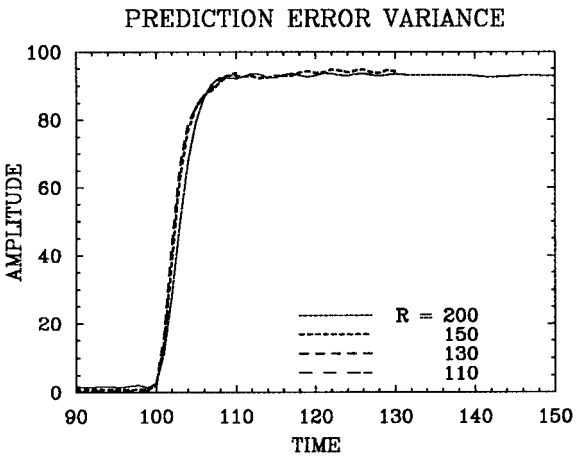

FIG. 11. Plot of normalized prediction error variance vs the $\mathcal{R}$-domain size. The selected prediction domain sizes are 200, 150, 130, and 110, respectively. The $\mathcal{D}$-domain size is 100 sampling points and the prediction filter was represented in terms of 50 EOFs.

sample covariance matrix. The accuracy of covariance matrix determines the accuracy of EOFs, which in turn determines the accuracy of the prediction filter. The length of available data is an important factor governing the accuracy of a covariance matrix. The shorter the record length the larger the effect of sampling error is in computing covariance matrix.

Figure 12 shows the prediction error variance versus the record length. The loss of precision of the prediction filter is obvious with the decreased record length. The degradation of the performance is more severe for larger domains for the same record length as should be expected. The limited test indicates that the record length should be at least 10 times the size of the prediction domain for a reasonably accurate prediction filter. An inaccurate prediction filter may imply that prediction error variance will not be minimal in the ensemble sense. Thus, the predictor will be suboptimal.

6. Summary and concluding remarks

Developed in this study is a general statistical predictor based on EOFs. It is the best linear unbiased predictor and minimizes the prediction error variance. The predictor conceptually is an extrapolator based on the past covariance structures and hence the accuracy of this covariance structure is an important limiting factor for the performance of the predictor. It is obtained in the form of a filter kernel which is applied to a space–time stream of data for prediction. The developed predictor was applied to one-dimensional examples for testing the validity and performance.

As any other statistical predictors the developed predictor cannot perform better than the prediction limit inherent in the data. If the covariance statistics are accurate the performance of the predictor in this study should approach the prediction limit because its formulation is based on the minimum prediction error. All

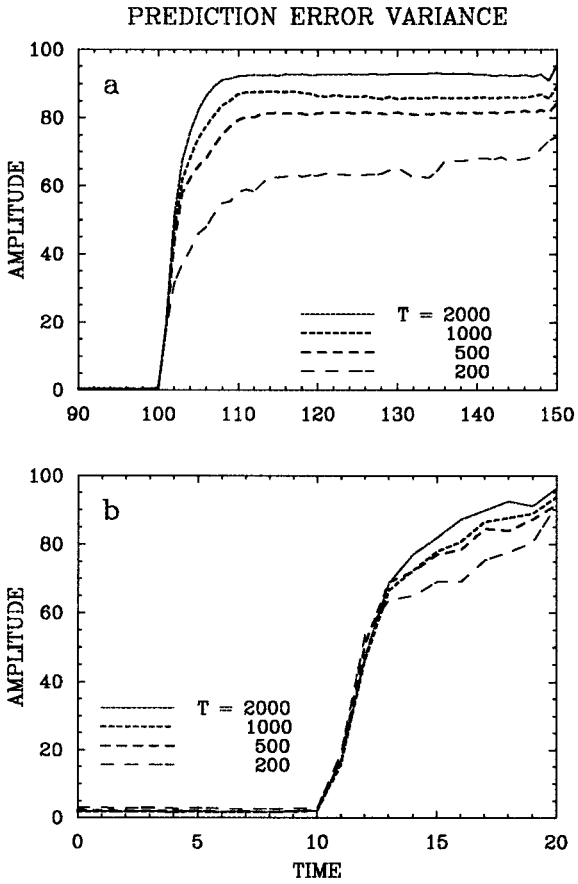

FIG. 12. Plot of normalized prediction error variance vs the record length of data: (a) $\mathcal{D} = 100$, $\mathcal{R} = 150$, and (b) $\mathcal{D} = 10$, $\mathcal{R} = 20$. The lengths of employed data are 2000, 1000, 500, and 200, respectively.

the examples demonstrate that the developed predictor performs as should be expected.

Four parameters concerning the construction of a predictor are the sizes of data domain $\mathcal{D}$ and prediction domain $\mathcal{R}$, and the numbers of EOF modes retained in the data and prediction domains, respectively. The sensitivity tests reveal that the performance of the predictor is fairly insensitive to a wide range of these construction parameters. Another important factor governing the performance of the predictor is the length of available data. It turns out that the accuracy of the predictor depends crucially on the accuracy of covariance function, which in turn depends on the avilable record length. Inaccuracy of the filter implies suboptimal performance of the filter. A limited test indicates that the record length should be at least 10 times the prediction domain size for near-optimal performance of the predictor.

There is also a further improvement to be made. In some prediction studies, it may be beneficial to include the phase information. For one thing, statistics of data may not be stationary. Statistics of surface temperature field, for example, exhibit significant seasonal dependency (Kim et al. 1996; Kim and North 1997). Also some phenomena such as El Niño may be strongly phase locked with the seasonal cycle (Jin et al. 1994; Tziperman et al. 1994; Chang et al. 1994, 1995). This phasic dependency of the predictor can be incorporated easily by using cyclostationary EOFs (Kim and North 1997). The algorithm was so constructed that such improvement can be incorporated easily.

Acknowledgments. We gratefully acknowledge support for this work by the Department of Energy (DEFG03-98ER62610) via a grant to Texas A&M University. The Department of Energy does not necessarily endorse any of the conclusions drawn in the paper.

REFERENCES

Barnett, T., N. Graham, M. A. Cane, S. E. Zebiak, S. Dolan, J. J. O'Brien, and D. Legler, 1988: On the prediction of the El Niño of 1986–1987. *Science,* **241,** 192–196.

Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature,* **321,** 827–832.

Chang, P., B. Wang, T. Li, and L. Ji, 1994: Interactions between the seasonal cycle and the southern oscillation—Frequency entrainment and chaos in a coupled ocean–atmosphere model. *Geophys. Res. Lett.,* **21,** 2817–2820.

——, L. Ji, B. Wang, and T. Li, 1995: Interactions between the seasonal cycle and El Niño–Southern Oscillation in an intermediate coupled ocean–atmosphere model. *J. Atmos. Sci.,* **52,** 2353–2372.

Inoue, M., and J. J. O'Brien, 1984: A forecasting model for the onset of El Niño. *Mon. Wea. Rev.,* **112,** 2326–2337.

Jin, F.-F., J. D. Neelin, and M. Ghil, 1994: El Niño on the devil's staircase: Annual subharmonic steps to chaos. *Science,* **264,** 70–72.

Kim, K.-Y., and G. R. North, 1997: EOFs of harmonizable cyclostationary processes. *J. Atmos. Sci.,* **54,** 2416–2427.

——, ——, and J. Huang, 1996: EOFs of one-dimensional cyclostationary time series: Computations, examples, and stochastic modeling. *J. Atmos. Sci.,* **53,** 1007–1017.

Latif, M., and M. Flügel, 1991: An investigation of short range climate predictability in the tropical Pacific. *J. Geophys. Res.,* **96,** 2661–2673.

Newton, H. J., 1988: *TIMESLAB: A Time Series Analysis Laboratory.* Wadsworth and Brooks, 623 pp.

Tziperman, E., L. Stone, M. A. Cane, and H. Jarosh, 1994: El Niño chaos: Overlapping of resonances between the seasonal cycle and the Pacific ocean–atmosphere oscillator. *Science,* **264,** 72–74.

Xu, J.-S., and H. von Storch, 1990: Predicting the state of the Southern Oscillation using principal oscillation pattern analysis. *J. Climate,* **3,** 1316–1329.