

## A Conceptual Framework for Predictability Studies

TAPIO SCHNEIDER

*Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey*

STEPHEN M. GRIFFIES

*NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey*

(Manuscript received 24 September 1998, in final form 19 January 1999)

### ABSTRACT

A conceptual framework is presented for a unified treatment of issues arising in a variety of predictability studies. The predictive power (PP), a predictability measure based on information-theoretical principles, lies at the center of this framework. The PP is invariant under linear coordinate transformations and applies to multivariate predictions irrespective of assumptions about the probability distribution of prediction errors. For univariate Gaussian predictions, the PP reduces to conventional predictability measures that are based upon the ratio of the rms error of a model prediction over the rms error of the climatological mean prediction.

Since climatic variability on intraseasonal to interdecadal timescales follows an approximately Gaussian distribution, the emphasis of this paper is on multivariate Gaussian random variables. Predictable and unpredictable components of multivariate Gaussian systems can be distinguished by predictable component analysis, a procedure derived from discriminant analysis: seeking components with large PP leads to an eigenvalue problem, whose solution yields uncorrelated components that are ordered by PP from largest to smallest.

In a discussion of the application of the PP and the predictable component analysis in different types of predictability studies, studies are considered that use either ensemble integrations of numerical models or autoregressive models fitted to observed or simulated data.

An investigation of simulated multidecadal variability of the North Atlantic illustrates the proposed methodology. Reanalyzing an ensemble of integrations of the Geophysical Fluid Dynamics Laboratory coupled general circulation model confirms and refines earlier findings. With an autoregressive model fitted to a single integration of the same model, it is demonstrated that similar conclusions can be reached without resorting to computationally costly ensemble integrations.

### 1. Introduction

Since Lorenz (1963) realized that chaotic dynamics may set bounds on the predictability of weather and climate, assessing the predictability of various processes in the atmosphere-ocean system has been the objective of numerous studies. These studies are of two kinds (Lorenz 1975). *Predictability studies of the first kind* address how the uncertainties in an initial state of the climate system affect the prediction of a later state. Initial uncertainties amplify as the prediction lead time increases, thus limiting predictability of the first kind. For example, in weather forecasting, the uncertainty in the predicted state reaches, at a lead time of a few weeks, the climatological uncertainty, the uncertainty as to which atmospheric state may be realized when only the climatological mean is available as a prediction. Day-

to-day weather variations are not predictable beyond this lead time.

*Predictability studies of the second kind* address the predictability of the response of the climate system to changes in boundary conditions. The fact that the state of the climate system is not completely determined by the boundary conditions limits predictability of the second kind. For example, the internal variability of the atmosphere renders a multitude of atmospheric states consistent with a configuration of sea surface temperatures (SSTs). It is uncertain which atmospheric state will be realized at a given time, even if the SST configuration at that time is known. A deviation of the SST from its climatological mean results in a predictable atmospheric response only if it reduces the uncertainty as to which atmospheric state may be realized to less than the climatological uncertainty.

These two types of predictability studies have a number of common features. Each, of course, requires a model that provides predictions of the process under consideration. Hence, predictability is always to be understood as predictability within a given model frame-

---

*Corresponding author address:* Tapio Schneider, AOS Program, Princeton University, Princeton, NJ 08544-0710.  
E-mail: tapio@splash.princeton.edu

work. Each type of study also requires a quantitative measure of predictability. Suggestions for such measures abound. Shukla (1981, 1985), Hayashi (1986), Murphy (1988), and Griffies and Bryan (1997b), for example, offer quantitative definitions of the term *predictability* itself, and Stern and Miyakoda (1995) define the concept of *reproducibility*. All of the above measures are based, in one way or another, on comparing the root-mean-square (rms) error of a univariate model prediction with the rms error of the prediction that consists of the climatological mean. The examined process is considered predictable if the rms error of the model prediction is significantly smaller than the rms error of the climatological mean prediction. Such predictability measures have made possible the definition of local predictability indexes and the study of regional variations in the predictability of geophysical fields [see Shukla (1985) for a review].

Difficulties arise, however, when one tries to generalize these predictability measures for univariate variables to the multivariate case, as one does, for example, when interested not in estimating the predictability of a single scalar variable grid point by grid point, but in estimating the overall predictability of several geophysical fields in some larger region. The initialization of ensemble integrations for numerical weather predictions (see, e.g., Palmer 1996, Palmer et al. 1998, and references therein) is an example of an inherently multivariate problem. Difficulties for multivariate predictions arise because the rms prediction error depends on the basis in which the fields are represented. This means that, although there is not always a natural choice of a metric to measure the prediction error, the outcome of the analysis depends on which metric is chosen.

Another shortcoming of error-based predictability indexes is that they assume the error distributions to be approximately Gaussian. This may be too restrictive an assumption in many cases. The *potential predictive utility* of Anderson and Stern (1996) partially overcomes this drawback of more traditional predictability measures. Anderson and Stern do not merely compare the rms error of a model-derived prediction with that of the climatological mean prediction—that is, the standard deviations of the corresponding error distributions—but they compare the entire error distributions, without making assumptions about their shape. If the error distributions differ significantly, potential predictive utility exists; otherwise, it does not. However, in contrast to the ratio of the rms errors, for example, the potential predictive utility does not give a measure of a prediction's "degree of uncertainty" but only makes statements about whether or not a given model prediction is better than the climatological mean prediction.

In addition to these drawbacks, many predictability measures have been defined only for specific study designs. Even in recent studies, authors have found it necessary to introduce predictability measures of their own. This circumstance highlights the lack of an overarching

conceptual framework that is sufficiently general to encompass currently used study designs. Still, whether one examines predictability of the first kind or predictability of the second kind, whether one employs comprehensive general circulation models (GCMs) to generate ensemble integrations or simpler empirical models fitted to observations—all predictability studies have some essential features in common.

Focusing on the fundamental structure that all predictability studies share, we will here develop a unified conceptual framework. In section 2, we first reduce the stochastic problems arising in predictability studies to their basic structure by stripping them of application-specific details; at the same time, we introduce the terminology and notation used throughout the remainder of this paper. In this general context, we then turn to address issues that frequently arise in predictability studies.

The key constituent of the methodology to be presented is a predictability index that uses concepts from information theory to measure the uncertainty of a prediction. [Shannon and Weaver (1949), Brillouin (1956), and Papoulis (1991, chapter 15) provide surveys on information theory.] The information-theoretical predictability index, the *predictive power* (PP), is defined in section 3. The PP applies to univariate as well as to multivariate predictions. In contrast to measures based on rms errors, the PP is invariant under arbitrary linear coordinate transformations; thus, the difficulties arising from the arbitrariness of an error metric are circumvented. Moreover, in its most general form, the PP does not rely on specific assumptions about either the distributions of the random variables involved or the modeling framework. In the special case of univariate and normally distributed predictions, the PP reduces to the ratio of the rms prediction error over the rms error of the climatological mean prediction (or, according to our conventions, to one minus this ratio). The PP can therefore be understood as a generalization of the above-cited predictability indexes.

Since empirical evidence (e.g., Toth 1991) suggests that aggregated climatic variables, such as space or time averages of geophysical fields, follow an approximately Gaussian distribution, the bulk of this paper focuses on multivariate Gaussian random variables. For Gaussian systems, questions such as, "what are the most predictable features of the system?" can be answered in a systematic manner. When the PP is used as the measure of predictive information in multivariate predictions, then the most predictable linear combination of state space variables, or the most predictable *component*, is the one that maximizes the PP. In section 4, we adapt a procedure from discriminant analysis (see, e.g., McLachlan 1992) to extract predictable components of a system: seeking components with large PP leads to an eigenvalue problem, whose solution yields uncorrelated components that are ordered by PP from largest to smallest. This way of extracting a system's predict-

able components, the *predictable component analysis*, is then compared with principal component analysis and with recently proposed approaches for determining predictable components (e.g., Hasselmann 1993).

Sections 5 and 6 give details for the application of the general methodology to specific types of predictability studies. Section 5 deals with studies that use ensemble integrations to estimate predictability; predictability studies of both the first and the second kind are considered. Section 6 discusses how autoregressive models, a paradigmatic class of empirical models, can be employed to assess the predictability of processes in the climate system.

In section 7, we illustrate the PP concept and the predictable component analysis by investigating the predictability of multidecadal North Atlantic variability (Halliwell 1997, 1998; Griffies and Bryan 1997a,b). Two approaches are taken; first, ensemble integrations of a coupled atmosphere–ocean GCM, as performed by Griffies and Bryan (1997b), are reanalyzed with the new methods; this will confirm and refine the earlier findings of Griffies and Bryan. Second, it will be demonstrated that with an autoregressive model fitted to a single integration of the same GCM, similar conclusions can be reached without performing computationally costly ensemble integrations.

In section 8, we summarize our conclusions and comment on their relevance for future research. The appendix contains computational details of procedures laid out in the body of this paper.

## 2. The basic structure of predictability studies

Suppose the state of a time-evolving system at time instant  $\nu$  is represented by an  $m$ -dimensional *state vector*  $\mathbf{X}_\nu$ . Since we are concerned with the evolution of distributions of states, rather than the evolution of a single state, we take a stochastic perspective on the dynamics of the system: the state is viewed as a random vector and as such is characterized by a probability distribution function whose domain is the *state space*, the set of all values the state vector can possibly attain. Given, for example, the time evolution of a geophysical field, the state space may be the  $m$ -dimensional vector space of a representation of the field by  $m$  linearly independent grid point values or spectral coefficients. The probability distribution associated with the state  $\mathbf{X}_\nu$  is the *climatological distribution* of the geophysical field and reflects the uncertainty in the system's state when the climatological mean is the only available predictive information. In this stochastic framework, a particular observation  $\mathbf{x}_\nu$  is called a *realization* of the random state vector  $\mathbf{X}_\nu$ . (To avoid ambiguities, we make the distinction between a random variable and one of its realizations explicit by using capital letters for the random variable and lowercase for the realization.)

Consider now the prediction of the state  $\mathbf{x}_\nu$ . An individual *prediction*  $\hat{\mathbf{x}}_\nu$  is usually a function of the states

at previous instants  $\nu - 1$ ,  $\nu - 2$ , etc. The prediction might, for example, be obtained as the mean of an ensemble of GCM integrations. In predictability studies of the first kind, each member of the ensemble corresponds to a different initial condition drawn from an initial state whose probability distribution reflects observational uncertainties. In predictability studies of the second kind, the ensemble members form a sample of the distribution of those states that are consistent with a given boundary condition—with a particular configuration of SST, for example. As an alternative to ensemble integrations, the prediction may be based on an empirical model fitted to observed or simulated data.

The index  $\nu$  labels the time for which a prediction is made. In predictability studies of the first kind, the index  $\nu$  designates the forecast lead time. Since the climatological distribution associated with the state  $\mathbf{X}_\nu$  does not vary much over typical forecast lead times, it is usually assumed to be stationary and hence independent of  $\nu$ . In predictability studies of the second kind, the index  $\nu$  usually designates the time of year for which a prediction is made—a particular season, for example—and the climatological distribution associated with the state  $\mathbf{X}_\nu$  depends on  $\nu$ . We will discuss the analysis of a prediction for a single instant  $\nu$ , but to make the dependence on the prediction time explicit, we still index prediction-time dependent variables by  $\nu$ .

Because of the system's stochastic nature, the prediction  $\hat{\mathbf{x}}_\nu$  does not necessarily coincide with the actual realization  $\mathbf{x}_\nu$ , but is afflicted with a random *prediction error*  $\mathbf{e}_\nu \equiv \mathbf{x}_\nu - \hat{\mathbf{x}}_\nu$ . The probability distribution of the corresponding random variable  $\mathbf{E}_\nu$  reflects the uncertainty that remains in the state after a prediction has become available. If the prediction is obtained as the mean of an ensemble, the differences between the individual ensemble members and their mean form a sample of the prediction error distribution. If the prediction is obtained from an empirical model, the distribution of prediction errors must be derived from the assumptions intrinsic to the empirical modeling framework.

Since the prediction error  $\mathbf{e}_\nu$  is the difference between the actually realized state  $\mathbf{x}_\nu$  and the prediction  $\hat{\mathbf{x}}_\nu$ , the realization  $\mathbf{x}_\nu$  of the system's state can be written as the sum  $\mathbf{x}_\nu = \hat{\mathbf{x}}_\nu + \mathbf{e}_\nu$ . Expressed in terms of the corresponding random variables, this statement reads

$$\mathbf{X}_\nu = \hat{\mathbf{X}}_\nu + \mathbf{E}_\nu,$$

where the *predictor*  $\hat{\mathbf{X}}_\nu$  is the random function of which the prediction  $\hat{\mathbf{x}}_\nu$  is a realization. Fundamental to the following line of reasoning is the interpretation of the associated probability distributions: the distribution of the state  $\mathbf{X}_\nu$  is the climatological distribution, which reflects the prior uncertainty as to which state may be realized before any predictive information besides the climatological mean is available; the distribution of the prediction error  $\mathbf{E}_\nu$  reflects the posterior uncertainty that remains in the state after a prediction has become available.

Here we are exclusively concerned with how the random nature of the prediction error affects the predictability of state vector realizations. We assume that the prediction error has no systematic component, which would show as a nonzero error mean; that is, we assume that the predictor is unbiased.<sup>1</sup> The condition of unbiasedness is automatically satisfied if the prediction is obtained as the mean of an ensemble (see section 5). Note, however, that unbiasedness in our context does not necessarily mean that the model provides unbiased predictions of the actual empirical system being modeled; we stay exclusively within the framework set by the model that provides the predictions and merely require that, within this framework, the prediction error have zero mean.

Within the given modeling framework, we now ask the questions, how much information about state realizations does the predictor provide? and, if the system has any predictable components, which of those are the most predictable? More precisely, we want an appropriate measure of predictability and a decomposition of the state space into subspaces that are ordered from most predictable to least predictable.

### 3. The predictive power

#### a. Derivation of the general form

If no more specific prediction for  $\mathbf{x}_v$  is available than the climatological mean, then the uncertainty in the system's state is the climatological or prior uncertainty associated with the climatological probability density that characterizes the state vector  $\mathbf{X}_v$ . The effect of the predictor is to provide predictive information on the system's state, thus reducing the state's prior uncertainty to the posterior uncertainty that remains after a specific prediction has become available. A state is not predictable when its posterior uncertainty is as large as its prior uncertainty—that is, when the prediction does not contain predictive information in excess of the climatological mean—and its predictability increases with increasing predictive information.

Rendering this intuitive notion of predictability quantitative requires a precise definition of the degree of uncertainty associated with the probability density  $p_{\mathbf{X}}(\mathbf{x})$  of a random variable  $\mathbf{X}$ . Such a definition, which is at the heart of information theory (Brillouin 1956; Shannon and Weaver 1949), was introduced by Shannon (1948), who showed that the *entropy*

$$S_{\mathbf{X}} \equiv -k \int d\mathbf{x} p_{\mathbf{X}}(\mathbf{x}) \log p_{\mathbf{X}}(\mathbf{x}) \quad (1)$$

<sup>1</sup> The assumption that the prediction error has no systematic component does not imply a “perfect model” assumption. Sections 6 and 8 contain examples of how the proposed framework applies to “non-perfect model” contexts, namely, to modeling with autoregressive models and to the performance evaluation of forecasting models.

is a natural measure of the uncertainty associated with a random variable  $\mathbf{X}$ . (The quantity  $S_{\mathbf{X}}$  is sometimes called the *information* of the random variable  $\mathbf{X}$ , meaning that on average the additional information  $S_{\mathbf{X}}$  is needed to specify completely a realization of  $\mathbf{X}$ .) Shannon derived the entropy functional from a set of heuristic requirements that any measure of uncertainty should fulfill and showed that the entropy is, up to the constant factor  $k$ , the unique measure fulfilling these requirements. The value of the constant  $k$  determines the units in which the entropy is measured. For thermodynamic systems,  $k$  is Boltzmann's constant. For discrete random variables, the integration in (1) must be replaced by a sum, and  $k = 1/\log 2$  is chosen so that the entropy  $S_{\mathbf{X}}$  becomes the expected number of binary digits, or bits, needed to specify a particular realization of  $\mathbf{X}$ . We set  $k = 1/m$ , where  $m$  is the dimension of the state space, so that  $S_{\mathbf{X}}$  becomes the mean entropy per state vector component. Defining the entropy relative to the state space dimension makes it possible to compare the entropies of random vectors of different dimensions.

Assuming state space vectors to be determined only up to some fixed accuracy allows one to think of the state space as discrete. The *prior entropy*  $S_{\mathbf{X}_v}$  is then the average number of state vector bits that are undetermined when only the climatological mean is known. Similarly, the *posterior entropy*  $S_{\mathbf{E}_v}$ , the conditional entropy of the state given a prediction, is the average number of state vector bits that remain undetermined after a prediction has become available. The difference between these entropies is the predictive information

$$R_v \equiv S_{\mathbf{X}_v} - S_{\mathbf{E}_v}, \quad (2)$$

the average information about the state contained in a prediction. For a discrete system, the predictive information is the average number of state vector bits that a prediction determines.<sup>2</sup>

The *predictive power* is defined as

$$\alpha_v \equiv 1 - e^{-R_v}. \quad (3)$$

Since no predictor should increase the uncertainty in a state to above the prior uncertainty, implying  $S_{\mathbf{X}_v} \geq S_{\mathbf{E}_v}$ , the predictive information  $R_v$  is a positive semi-definite quantity. Hence, the PP exhibits proper limiting behavior: it is an index  $0 \leq \alpha_v \leq 1$  that is zero if the predictive information vanishes and that monotonically increases with increasing predictive information, eventually approaching unity in the limit of infinite predictive information.

The PP can be interpreted geometrically. If, as is common practice, the entropy  $S_{\mathbf{X}}$  is evaluated with  $k = 1$  in definition (1), then the exponential  $\exp S_{\mathbf{X}}$  is the state

<sup>2</sup> Readers familiar with information theory will recognize the close analogy between the predictive information and the rate of transmission in a noisy channel as considered by Shannon (1948). The posterior entropy  $S_{\mathbf{E}_v}$  corresponds to Shannon's *equivocation*.

space volume enclosing “reasonably probable” or “typical” realizations of the random vector  $\mathbf{X}$  (Shannon and Weaver 1949, p. 59; Papoulis 1991). We evaluate the entropy  $S_{\mathbf{X}}$  with  $k = 1/m$  and call the exponential  $\exp S_{\mathbf{X}}$  the *typical range*. The typical range is the  $m$ th root of the volume of typical realizations and measures the mean size of the set of values typically taken by a random vector component. Thus, the term  $\exp(-R_{\nu}) = \exp S_{\mathbf{E}_{\nu}} / \exp S_{\mathbf{X}}$ , in the predictive power (3) is a ratio of typical ranges: it is the “typical range of a state vector component given a prediction” over the “typical range of a climatological state vector component.” That is to say, the term  $\exp(-R_{\nu})$  is the fraction of the climatological typical range that lies within a prediction’s “range of uncertainty.” The complement  $1 - \exp(-R_{\nu})$ , the PP, is the typical range fraction that the predictor eliminates from the climatological typical range. Therefore, the PP indicates the efficacy of the predictor in narrowing the typical range of a state vector component.

Besides exhibiting proper limiting behavior and having an intuitive interpretation, any adequate predictability index should also be independent of the basis in which state vectors are represented. If, for example, the state is a compound of several geophysical fields, its predictability index should not depend on the units in which these fields are measured. Changing the dimensional scaling of some components of a state vector amounts to transforming state space vectors  $\mathbf{x}$  to the rescaled vectors  $\mathbf{x}'$  by multiplication with a diagonal matrix. Such a transformation should leave a predictability measure unchanged. More generally, we require the predictability measure to be invariant under linear coordinate transformations that transform state space vectors  $\mathbf{x}$  to  $\mathbf{x}' = \mathbf{U}\mathbf{x}$  with arbitrary nonsingular matrices  $\mathbf{U}$ . To check if the PP is invariant under such transformations, note that the probability density functions  $p_{\mathbf{x}}$  in the original coordinates and  $p_{\mathbf{x}'}$  in the transformed coordinates are related by  $p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = p_{\mathbf{x}'}(\mathbf{x}') d\mathbf{x}'$ , from which it follows that  $p_{\mathbf{x}}(\mathbf{x}) = |\det \mathbf{U}| p_{\mathbf{x}'}(\mathbf{x}')$ . Using these relations, one finds that the entropy (1) of the transformed variable  $\mathbf{X}'$  compared to that of the original variable  $\mathbf{X}$  is changed only by the additive constant  $k \log |\det \mathbf{U}|$ , which involves the determinant  $\det \mathbf{U}$  of the transformation matrix  $\mathbf{U}$  (Shannon and Weaver 1949, p. 59). In the predictive information—the difference (2) between the prior and the posterior entropies—the constant terms  $k \log |\det \mathbf{U}|$  cancel. Thus, the PP is indeed invariant under arbitrary linear transformations of state space coordinates.

The PP hence has desirable properties and is defined under general circumstances; neither assumptions about the modeling framework nor assumptions about the dimension or distributions of the relevant random variables were needed for the derivation. For univariate and possibly for low-dimensional state vectors, the entropy can be estimated using standard procedures, which involve estimation of the probability density (see, e.g., Silverman 1986; Scott 1992) and of the entropy as a

functional thereof (see, e.g., Prakasa Rao 1983; Dmitriev and Tarasenko 1973; Ahmad and Lin 1976; Joe 1989; Hall and Morton 1993). Thus, it may be possible to obtain a predictability measure for, say, local precipitation, a field for which neither the climatological distribution nor the prediction error distribution is Gaussian and for which a predictability index based on rms errors may be inappropriate.

Whereas the PP in its most general form is applicable to low-dimensional predictions, for high-dimensional states estimation of the entropy from (1) may not be feasible when the available dataset is small. Our emphasis, however, is on intraseasonal to interannual climate predictability, as opposed to the predictability of shorter-term weather processes. That the former kind of variability follows an approximately Gaussian distribution (see, e.g., Toth 1991) considerably simplifies the discussion.

*b. Simplifications for Gaussian random variables*

For an  $m$ -dimensional Gaussian random vector  $\mathbf{X}$ , the probability density takes the form

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{(2\pi)^{-m/2}}{(\det \Sigma)^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \langle \mathbf{X} \rangle)^T \Sigma^{-1} (\mathbf{x} - \langle \mathbf{X} \rangle) \right],$$

where  $\langle \mathbf{X} \rangle$  is the mean of  $\mathbf{X}$ , the superscript  $(\cdot)^T$  indicates the transpose of  $(\cdot)$ , and  $\Sigma^{-1}$  is the inverse of the nonsingular covariance matrix  $\Sigma$ . The entropy integral (1) of the Gaussian density  $p_{\mathbf{X}}$  is readily carried out and yields the entropy

$$S_{\mathbf{X}} = \frac{k}{2} (m + m \log 2\pi + \log \det \Sigma) \tag{4}$$

as a function of the covariance matrix determinant. Denoting the covariance matrix of the state, the *climatological covariance* matrix, by

$$\Sigma_{\nu} = \text{Cov}(\mathbf{X}_{\nu})$$

and the covariance matrix of the prediction error by

$$\mathbf{C}_{\nu} = \text{Cov}(\mathbf{E}_{\nu}),$$

one finds from (2) the predictive information

$$R_{\nu} = -\frac{k}{2} \log \left( \frac{\det \mathbf{C}_{\nu}}{\det \Sigma_{\nu}} \right).$$

Using the product theorem for determinants and substituting  $k = 1/m$  leads to the PP

$$\alpha_{\nu} = 1 - (\det \Gamma_{\nu})^{1/(2m)}, \tag{5}$$

where

$$\Gamma_{\nu} \equiv \mathbf{C}_{\nu} \Sigma_{\nu}^{-1} \tag{6}$$

is called the *predictive information matrix*. The predictive information matrix is well-defined, provided that the climatological covariance matrix  $\Sigma_{\nu}$  is positive def-

inite so that its inverse exists and is likewise symmetric and positive definite. Positive definiteness of the climatological covariance matrix is assumed in the following theoretical developments. The complications arising in practice from singular covariance matrix estimates will be dealt with in section 4c.

The interpretation of the PP as a ratio of typical ranges can now be made more concrete. From the entropy (4) of an  $m$ -dimensional Gaussian random vector  $\mathbf{X}$  with covariance matrix  $\Sigma$  follows, again taking  $k = 1/m$ , the typical range,

$$\exp S_{\mathbf{X}} = \sqrt{2\pi e} (\det \Sigma)^{1/(2m)}. \tag{7}$$

For a univariate random variable with  $m = 1$ , the covariance matrix  $\Sigma$  is a scalar variance, and the square root of this variance is the standard deviation  $\sigma$ . Therefore, the typical range of a univariate Gaussian random variable,  $\exp S_{\mathbf{X}} = \sqrt{2\pi e} \sigma \approx 4.13\sigma$ , is proportional to the standard deviation. For an  $m$ -dimensional Gaussian random vector  $\mathbf{X}$ , the ellipsoid  $\mathcal{E}_p(\mathbf{X})$  that is centered on the mean of  $\mathbf{X}$  and encloses some fraction  $0 < p < 1$  of the cumulative probability distribution has a volume proportional to  $(\det \Sigma)^{1/2}$  (Anderson 1984, 263). Since the volume of an ellipsoid is proportional to the product of the lengths of its semi-axes, the factor  $(\det \Sigma)^{1/(2m)}$  in the typical range (7) is proportional to the geometric mean of the semi-axis lengths of the ellipsoid  $\mathcal{E}_p(\mathbf{X})$ . Hence, the term  $(\det \Gamma_{\nu})^{1/(2m)} = (\det \mathbf{C}_{\nu})^{1/(2m)} (\det \Sigma_{\nu})^{-1/(2m)}$  in the PP is the ratio of the geometric mean of the semi-axis lengths of the prediction error ellipsoid  $\mathcal{E}_p(\mathbf{E}_{\nu})$  over the geometric mean of the semi-axis lengths of the climatological ellipsoid  $\mathcal{E}_p(\mathbf{X}_{\nu})$ . This interpretation of the PP as a ratio of geometric means of semi-axis lengths specializes the above general interpretation of the PP to Gaussian random variables.

For univariate state vectors, the covariance matrices in (6) are scalar variances; the predictive information matrix is the ratio of these variances; and the square root of these variances, the standard deviations, are the rms errors. Therefore, the predictive power (5) reduces to one minus the ratio of the rms error of a prediction over the rms error of the climatological mean prediction. Similar predictability measures have been employed by several authors, for example, Hayashi (1986), Murphy (1988), and Stern and Miyakoda (1995). Thus, the PP can be understood as a generalization of the univariate error-ratio predictability measures to multivariate states with arbitrary probability distributions.

When the distribution of states is multivariate Gaussian, one might think that arguments based on a comparison of prediction errors also lead to an adequate predictability measure. The mean-squared prediction error corresponds to the sum of the diagonal elements, or the trace  $\text{tr} \mathbf{C}_{\nu}$ , of the prediction error covariance matrix  $\mathbf{C}_{\nu}$ . Analogously, the trace  $\text{tr} \Sigma_{\nu}$  of the climatological covariance matrix  $\Sigma_{\nu}$  gives the mean-squared error of the climatological mean prediction. Taking one minus

the ratio of the rms errors as a predictability index, one obtains

$$1 - \left( \frac{\text{tr} \mathbf{C}_{\nu}}{\text{tr} \Sigma_{\nu}} \right)^{1/2}. \tag{8}$$

Traces, however, are only invariant under orthogonal transformations, a subclass of the general linear transformations considered above. A scaling transformation, for example, generally changes the predictability index (8). The expression (5) for the PP, on the other hand, involves a ratio of determinants that remains invariant under arbitrary linear coordinate transformations, including scaling transformations. The invariance under linear coordinate transformations is a principal advantage of information theory arguments over those based on considerations of prediction errors.

#### 4. Predictable component analysis

Adapting a procedure from discriminant analysis, we will now show that, for Gaussian random variables, knowledge of the predictive information matrix  $\Gamma_{\nu}$  allows us to derive a decomposition of the state space into subspaces that are ordered according to decreasing PP.

##### a. State space decomposition

The state vector  $\mathbf{X}_{\nu}$  consists of  $m$  components  $X_{\nu}^1, \dots, X_{\nu}^m$ , which are univariate random variables. If the state vector is a grid representation of a geophysical field, for example, the component  $X_{\nu}^k$  is the random variable associated with the geophysical field at grid point  $k$ . These univariate random variables are generally correlated and are not ordered by PP. From the  $m$  components  $X_{\nu}^k$ , we want to construct  $m$  linear combinations  $Y_{\nu}^k = (\mathbf{u}_{\nu}^k)^T \mathbf{X}_{\nu}$ , such that the first component  $Y_{\nu}^1$  has the largest PP attainable by any linear combination of state vector components, and subsequent components  $Y_{\nu}^2, \dots, Y_{\nu}^m$  are mutually uncorrelated and ordered according to decreasing PP.

The transformed state vector  $\mathbf{Y}_{\nu}$  with components  $Y_{\nu}^k$  is related to the original state vector  $\mathbf{X}_{\nu}$  with components  $X_{\nu}^k$  by  $\mathbf{Y}_{\nu} = \mathbf{U}_{\nu}^T \mathbf{X}_{\nu}$ , where the weight vectors  $\mathbf{u}_{\nu}^k$  form the columns of the matrix  $\mathbf{U}_{\nu}$ . We restrict ourselves to nonsingular transformations  $\mathbf{U}_{\nu} \in \mathfrak{R}^{m \times m}$ , for which the original state vectors can be reconstructed from the transformed ones via  $\mathbf{X}_{\nu} = \mathbf{V}_{\nu} \mathbf{Y}_{\nu}$  with

$$\mathbf{U}_{\nu} \mathbf{V}_{\nu}^T = \mathbf{U}_{\nu}^T \mathbf{V}_{\nu} = \mathbf{I}. \tag{9}$$

Written componentwise, the inverse transformation  $\mathbf{X}_{\nu} = \mathbf{V}_{\nu} \mathbf{Y}_{\nu}$  reads  $X_{\nu}^k = \sum_{l=1}^m Y_{\nu}^l v_{\nu}^{kl}$ , where  $\mathbf{v}_{\nu}^k$  is the  $k$ th column of the matrix  $\mathbf{V}_{\nu}$ . The random variables  $Y_{\nu}^k$  can thus be viewed as the components of the state vector  $\mathbf{X}_{\nu}$  when  $\mathbf{X}_{\nu}$  is expanded in the state space basis  $\mathbf{v}_{\nu}^1, \dots, \mathbf{v}_{\nu}^m$ . The basis vectors  $\mathbf{v}_{\nu}^k$  and the weight vectors, or dual basis vectors,  $\mathbf{u}_{\nu}^k$  are related by the completeness and biorth-

ogonality conditions (9). For orthogonal transformations, basis vectors  $\mathbf{v}_\nu^k$  and their duals  $\mathbf{u}_\nu^k$ , and hence the matrices  $\mathbf{V}_\nu$  and  $\mathbf{U}_\nu$ , are identical. However, as the PP is invariant under arbitrary linear coordinate transformations, the transformation  $\mathbf{U}_\nu$  need not be orthogonal, and (9) holds with matrices  $\mathbf{V}_\nu$  and  $\mathbf{U}_\nu$  that are generally not identical.

In order to find the linear combination of state vector components that has the largest PP, we must determine the predictive power  $\alpha_\nu^k$  of an arbitrary linear combination  $Y_\nu^k = (\mathbf{u}_\nu^k)^T \mathbf{X}_\nu$  and then maximize this PP with respect to the weight vector  $\mathbf{u}_\nu^k$ . The predictive information matrix of the univariate component  $Y_\nu^k$  is a ratio of scalar variances. These scalar variances are the diagonal elements of the covariance matrices

$$\Sigma'_\nu = \mathbf{U}_\nu^T \Sigma_\nu \mathbf{U}_\nu \quad \text{and} \quad \mathbf{C}'_\nu = \mathbf{U}_\nu^T \mathbf{C}_\nu \mathbf{U}_\nu \quad (10)$$

of the transformed state vector  $\mathbf{Y}_\nu = \mathbf{U}_\nu^T \mathbf{X}_\nu$  and the transformed prediction error  $\mathbf{U}_\nu^T \mathbf{E}_\nu$ . The predictive information matrix (6) of the  $k$ th component thus reduces to the ratio of the  $k$ th diagonal elements,

$$\gamma_\nu^k = \frac{(\mathbf{u}_\nu^k)^T \mathbf{C}_\nu \mathbf{u}_\nu^k}{(\mathbf{u}_\nu^k)^T \Sigma_\nu \mathbf{u}_\nu^k} \quad (11)$$

The scalar  $\gamma_\nu^k$  is called the *Rayleigh quotient* of the weight vector  $\mathbf{u}_\nu^k$  (see, e.g., Golub and van Loan 1993, chapter 8). Substituting the Rayleigh quotient for the predictive information matrix in (5) gives the predictive power of the  $k$ th component

$$\alpha_\nu^k = 1 - (\gamma_\nu^k)^{1/2}. \quad (12)$$

Maximizing the predictive power  $\alpha_\nu^k$  is thus equivalent to minimizing the Rayleigh quotient  $\gamma_\nu^k$ .

The Rayleigh quotient  $\gamma_\nu^k$  is minimized by taking its gradient with respect to the weight vector  $\mathbf{u}_\nu^k$  and equating to zero. This procedure leads to the generalized eigenvalue problem

$$(\mathbf{u}_\nu^k)^T \mathbf{C}_\nu = \gamma_\nu^k (\mathbf{u}_\nu^k)^T \Sigma_\nu,$$

which, still assuming that the climatological covariance matrix  $\Sigma_\nu$  is nonsingular, can be recast into the conventional eigenvalue problem

$$(\mathbf{u}_\nu^k)^T \mathbf{C}_\nu \Sigma_\nu^{-1} = \gamma_\nu^k (\mathbf{u}_\nu^k)^T.$$

This eigenvalue problem determines the weight vector  $\mathbf{u}_\nu^k$  as a *left* eigenvector of the predictive information matrix  $\Gamma_\nu = \mathbf{C}_\nu \Sigma_\nu^{-1}$ . It follows that the minimum value  $\gamma_\nu^1$  is the smallest eigenvalue of the predictive information matrix  $\Gamma_\nu$ . For a nonsymmetric matrix such as  $\Gamma_\nu$ , the completeness and biorthogonality conditions (9) relate left and right eigenvectors. Therefore, the basis vector  $\mathbf{v}_\nu^1$  whose component  $Y_\nu^1$  has the smallest Rayleigh quotient  $\gamma_\nu^1$ , and hence the largest PP, is the *right* eigenvector belonging to the smallest eigenvalue  $\gamma_\nu^1$ ; that is, the basis vector  $\mathbf{v}_\nu^1$  with largest PP satisfies  $\Gamma_\nu \mathbf{v}_\nu^1 = \gamma_\nu^1 \mathbf{v}_\nu^1$ . We call the basis vector  $\mathbf{v}_\nu^1$  the *first predictable pattern*.

We will now argue that an analysis of the remaining

eigenvectors of the predictive information matrix leads to a decomposition of the state space into uncorrelated subspaces that are ordered according to decreasing PP. In making this point, we need some properties of the eigendecomposition of the predictive information matrix.

The predictive information matrix  $\Gamma_\nu$  is a product of the two symmetric matrices  $\mathbf{C}_\nu$  and  $\Sigma_\nu^{-1}$  but is not necessarily symmetric itself. Therefore, the left and right eigenvectors of the predictive information matrix  $\Gamma_\nu$  generally differ and do not form sets of mutually orthogonal vectors, as they would if  $\Gamma_\nu$  were symmetric. However, a generalized orthogonality condition for the eigenvectors follows from a linear algebra theorem on the simultaneous diagonalization of two symmetric matrices (see, e.g., Fukunaga 1990, chapter 2): if the columns of the matrices  $\mathbf{U}_\nu$  and  $\mathbf{V}_\nu$  consist, respectively, of the left and right eigenvectors of the predictive information matrix, then the transformed covariance matrices (10) are both diagonal. The left eigenvectors  $\mathbf{u}_\nu^k$  can be normalized such that the transformed climatological covariance matrix, the covariance matrix of the components  $Y_\nu^k$ , becomes the identity matrix,

$$\Sigma'_\nu = \mathbf{U}_\nu^T \Sigma_\nu \mathbf{U}_\nu = \mathbf{I}. \quad (13)$$

This normalization ensures that the left eigenvectors  $\mathbf{u}_\nu^k$  are orthonormal with respect to the climatological covariance matrix  $\Sigma_\nu$ . Equivalently, this normalization means that the components  $Y_\nu^k$  are mutually uncorrelated and have unit variance. Moreover, one finds from the Rayleigh quotient (11) that, in the transformed coordinates, the prediction error covariance matrix is identical to the diagonalized predictive information matrix,

$$\mathbf{C}'_\nu = \mathbf{U}_\nu^T \mathbf{C}_\nu \mathbf{U}_\nu = \text{Diag}(\gamma_\nu^k); \quad (14)$$

here,  $\text{Diag}(\gamma_\nu^k)$  denotes the diagonal matrix with the eigenvalues  $\gamma_\nu^k$  of the predictive information matrix  $\Gamma_\nu$  as diagonal elements.

An orthogonality condition can be derived for the right eigenvectors  $\mathbf{v}_\nu^k$  as well. Combining the biorthogonality condition (9) with the generalized orthogonality condition (13) for the left eigenvectors  $\mathbf{u}_\nu^k$  yields the relation

$$\Sigma_\nu \mathbf{U}_\nu = \mathbf{V}_\nu \quad (15)$$

between left and right eigenvectors of the predictive information matrix. Solving for  $\mathbf{U}_\nu$  and substituting into (13) leads to

$$\mathbf{V}_\nu^T \Sigma_\nu^{-1} \mathbf{V}_\nu = \mathbf{I}. \quad (16)$$

Therefore, the right eigenvectors  $\mathbf{v}_\nu^k$  are orthonormal with respect to the inverse climatological covariance matrix  $\Sigma_\nu^{-1}$ .

As detailed in the appendix, the eigenvector matrices  $\mathbf{U}_\nu$  and  $\mathbf{V}_\nu$  can be obtained from a sequence of real transformations and are thus real themselves. This means that, despite the fact that the predictive information matrix is not necessarily symmetric, its eigenvalues and

eigenvectors are real. Moreover, as the predictive information matrix is a product of positive semidefinite matrices, the eigenvalues  $\gamma_\nu^k$  are greater than or equal to zero. Since no predictor of a linear combination of state vector components should have a prediction error variance that exceeds that of the climatological mean prediction, the eigenvalues  $\gamma_\nu^k$  should also be less than or equal to one; thus,  $0 \leq \gamma_\nu^k \leq 1$ .

For the remainder of this paper, we adopt the convention that the  $m$  eigenvalues  $\gamma_\nu^k$  of the predictive information matrix  $\Gamma_\nu$  are ordered from smallest to largest, so that the corresponding PPs  $\alpha_\nu^k = 1 - (\gamma_\nu^k)^{1/2}$  are ordered from largest to smallest,

$$1 \geq \alpha_\nu^1 \geq \dots \geq \alpha_\nu^m \geq 0.$$

This ordering implies that the vector  $\mathbf{v}_\nu^1$  is the basis vector whose component  $Y_\nu^1$  is most predictable with predictive power  $\alpha_\nu^1$ . The next eigenvector  $\mathbf{v}_\nu^2$  is the basis vector whose component  $Y_\nu^2$  has the next largest predictive power  $\alpha_\nu^2$ , subject to the constraint that the components  $Y_\nu^1$  and  $Y_\nu^2$  be uncorrelated. Iterating this argument, we arrive at a decomposition of the state space into mutually uncorrelated subspaces that are ordered according to decreasing PP. We call the components  $Y_\nu^1, \dots, Y_\nu^m$  the *predictable components* and the basis vectors  $\mathbf{v}_\nu^1, \dots, \mathbf{v}_\nu^m$  the *predictable patterns*. The expansion of state vectors in terms of predictable patterns is called *predictable component analysis*.

Expressing the predictive information matrix in the predictable pattern basis makes manifest properties of both the PP and the predictable component analysis. Since the determinant of a matrix is the product of its eigenvalues, the determinant of the predictive information matrix can be written as  $\det \Gamma_\nu = \prod_{k=1}^m \gamma_\nu^k$ , whence we infer, from (5), the overall PP

$$\alpha_\nu = 1 - \left( \prod_{k=1}^m \gamma_\nu^k \right)^{1/(2m)}. \quad (17)$$

Written in this form, it is evident that the overall PP is unity if one or more eigenvalues of the predictive information matrix vanish, that is, if the prediction error variance vanishes for at least one state vector component. At the other extreme, the PP is zero if all eigenvalues of the predictive information matrix are unity. The PP is nonzero if at least one eigenvalue of the predictive information matrix is smaller than unity, that is, if there is at least one predictable component with nonzero PP. Conversely, if the overall PP is nonzero, the predictable component analysis discriminates the state vector components with large, nonzero PP from those with small, possibly vanishing PP.

More generally, if the overall PP is nonzero, the predictable component analysis discriminates a more predictable “signal” from an uncorrelated background of less predictable “noise.” The overall PP in the subspace spanned by the first  $r \leq m$  predictable patterns is  $1 - (\prod_{k=1}^r \gamma_\nu^k)^{1/(2r)}$ , which is greater than or equal to the over-

all PP in any subspace of dimension  $r' > r$ . This dimension dependence of the PP particularly implies that the PP  $\alpha_\nu^1 = 1 - (\gamma_\nu^1)^{1/2}$  in the subspace of the first predictable pattern is always greater than or equal to the overall PP in any other subspace, regardless of its dimension. We also conclude that the first  $r < m$  predictable patterns span the  $r$ -dimensional state space portion with the largest PP, the signal, which is uncorrelated with the  $(m - r)$ -dimensional complement, the noise.

### b. Relation to principal component analysis

The transformation  $\mathbf{Y}_\nu = \mathbf{U}_\nu^T \mathbf{X}_\nu$  simultaneously diagonalizes the climatological covariance matrix  $\Sigma_\nu$ , the prediction error covariance matrix  $\mathbf{C}_\nu$ , and the predictive information matrix  $\Gamma_\nu$ . That is to say, when the states and the prediction error are expressed relative to the predictable pattern basis, their components at any fixed instant  $\nu$  are uncorrelated; nevertheless, predictable components at different instants  $\nu$  may be correlated. If we again think of the state vector as a representation of a geophysical field on a spatial grid, the predictable component analysis yields components that are uncorrelated spatially but that may be correlated temporally.

This feature of the predictable component analysis is reminiscent of the principal component analysis, which is the expansion of state vectors in terms of empirical orthogonal functions (EOFs). The principal component analysis of any of the covariance matrices also yields components that are uncorrelated at fixed  $\nu$ . Consider, for example, the principal component analysis of the climatological covariance matrix  $\Sigma_\nu$ . If the EOFs, the mutually orthogonal eigenvectors of  $\Sigma_\nu$ , form the columns of the matrix  $\mathbf{W}_\nu$ , then the matrix  $\Lambda_\nu = \mathbf{W}_\nu^T \Sigma_\nu \mathbf{W}_\nu$  is diagonal with eigenvalues of  $\Sigma_\nu$  as diagonal elements. Rescaling state vectors to unit variance by dividing the principal components  $\mathbf{W}_\nu^T \mathbf{X}_\nu$  by the square root of the eigenvalues transforms the climatological covariance matrix into the identity matrix,

$$\Lambda_\nu^{-1/2} \mathbf{W}_\nu^T \Sigma_\nu \mathbf{W}_\nu \Lambda_\nu^{-1/2} = \mathbf{I}. \quad (18)$$

This transformation is usually called a whitening transformation (see, e.g., Fukunaga 1990, chapter 2). For the variables thus transformed, the predictive information matrix (6) reduces to the transformed covariance matrix of the prediction error,

$$\Lambda_\nu^{-1/2} \mathbf{W}_\nu^T \mathbf{C}_\nu \mathbf{W}_\nu \Lambda_\nu^{-1/2} = \mathbf{K}_\nu, \quad (19)$$

which is generally not diagonal but can be diagonalized by another principal component analysis (cf. appendix). This further orthogonal transformation leaves the transformed climatological covariance matrix, the identity matrix, unchanged. Thus, the predictable component analysis is equivalent to a principal component analysis of  $\mathbf{K}_\nu$ , the prediction error covariance matrix for whitened state vectors.

Principal component analysis and predictable component analysis pursue different goals and optimize dif-



ferent criteria (cf. Fukunaga 1990, chapter 10.1). Expanding state vectors in terms of EOFs and truncating the expansion at some  $r < m$  gives the  $r$ -dimensional subspace that is uncorrelated with the neglected  $(m - r)$ -dimensional subspace and has minimum rms truncation error (e.g., Jolliffe 1986, chapter 3.2). The principal component analysis thus yields an optimal *representation* of states in a reduced basis. As the rms truncation error is invariant solely under orthogonal transformations but is not invariant under, for example, scaling transformations, the EOFs are invariant under orthogonal transformations only; a dimensional rescaling of variables generally changes the outcome of the principal component analysis.

By way of contrast, expanding state vectors in terms of predictable patterns and truncating at some  $r < m$  gives the  $r$ -dimensional subspace that is uncorrelated with the neglected  $(m - r)$ -dimensional subspace and has maximum PP. The predictable component analysis thus yields an optimal *discrimination* between more predictable components and less predictable components. As the predictive power is invariant under arbitrary linear coordinate transformations, so the predictable component analysis is invariant under arbitrary linear transformations of state vectors; in particular, the predictable component analysis does not depend on the dimensional scaling of variables.

### c. Rank-deficient covariance matrices

The expressions for the PP of Gaussian predictions and the predictable component analysis were derived under the assumption that the climatological covariance matrix  $\Sigma_v$  be nonsingular. Yet when the climatological covariance matrix is estimated from data, restrictions in sample size may lead to a sample covariance matrix that is singular. For a sample of size  $N$ , the sample covariance matrix is singular if  $N - 1$ , the number of degrees of freedom in the covariance matrix estimate, is smaller than the state space dimension  $m$ . The sample covariance matrix has at most rank  $N - 1$  or  $m$ , whichever is smaller. In typical studies of climatic predictability, the number  $N$  of independent data points is much smaller than the dimension  $m$  of the full state space of, say, a general circulation model; hence, sample covariance matrices usually do not have full rank.

The correspondence between predictable component analysis and the principal component analysis of the prediction error for whitened state vectors suggests a heuristic for dealing with rank deficiency of sample covariance matrices. Instead of applying the whitening transformation to the full  $m$ -dimensional state vectors, one retains and whitens only those principal components of the climatological covariance matrix that correspond to eigenvalues significantly different from zero. The predictable component analysis is then computed in this truncated state space.

Complications similar to those with the climatological

covariance matrix  $\Sigma_v$  may arise with the prediction error covariance matrix  $\mathbf{C}_v$ . If the number of degrees of freedom  $n$  available for the estimation of  $\mathbf{C}_v$  is smaller than the state space dimension  $m$ , the estimated prediction error covariance matrix is singular. A singular error covariance matrix leads to vanishing eigenvalues of the predictive information matrix  $\Gamma_v$ . Vanishing eigenvalues of the predictive information matrix correspond to states that have zero prediction error variance for at least one state vector component, but if  $n < m$ , at least  $m - n$  of the vanishing eigenvalues may be spurious: they correspond to state space directions in which the prediction error variance is zero because of sparse sampling but could become nonzero if the sample were larger. As above, a way to circumvent these difficulties is to perform a principal component analysis of the climatological covariance matrix, retaining at most  $n$  components for further analysis.

If the state vectors consist of variables with different dimensions, the principal component analysis depends on the dimensional scaling of the variables. For state vectors that are, for example, compounds of different geophysical fields, it is therefore advisable to compute the principal components of each field separately and assemble the state vectors for the predictable component analysis from selected principal components of each field. The principal components should be selected in such a way that the resulting state space dimension is small enough to ensure adequate sampling and nonsingular covariance matrix estimates. Section 7a contains an example that illustrates how principal components may be selected for a predictable component analysis.

Estimating predictable components from sparse data is an *ill-posed problem*, a problem in which the number of parameters to be estimated exceeds the sample size. Methods for solving ill-posed problems are known as *regularization techniques* (see, e.g., Tikhonov and Arsenin 1977; Engl et al. 1996; Hansen 1997; Neumaier 1998). We refer to the above approach as *regularization by truncated principal component analysis*. The computational algorithm in the appendix shows that regularization by truncated principal component analysis amounts to replacing the ill-defined inverse of the estimated climatological covariance matrix by a Moore–Penrose pseudoinverse (see, e.g., Golub and van Loan 1993, chapter 5). Since the principal component analysis and the pseudoinverse can be computed via a singular value decomposition of a data matrix (see, e.g., Jolliffe 1986, chapter 3.5), regularization by truncated principal component analysis is equivalent to regularization by truncated singular value decomposition, a method extensively discussed in the regularization literature (e.g., in Hansen 1997, chapter 3). More sophisticated regularization techniques (e.g., McLachlan 1992, chapter 5; Friedman 1989; Cheng et al. 1992; Krzanowski et al. 1995) may yield better estimates of the predictive information matrix; however, these techniques are less

transparent than regularization by truncated principal component analysis.

*d. Related work in the statistics and climatic predictability literature*

Predictable component analysis is a variant of a method known in multivariate statistics as discriminant analysis. [For introductory surveys, see Ripley (1996, chapter 3) and Fukunaga (1990, chapter 10).] Discriminant analysis seeks those linear combinations of state variables that optimize a criterion called the discriminant function. Discriminant functions are usually ratios of determinants or of traces of covariance matrices and thus resemble the PP.

In discriminant analysis, one considers only the weight vectors  $\mathbf{u}_\nu^k$  and the associated components  $Y_\nu^k$ , which are commonly referred to as canonical variates. Our additional interest in the predictable patterns  $\mathbf{v}_\nu^k$  led us to the above generalizations of standard results from discriminant analysis. Instead of focusing solely on the left eigenvectors  $\mathbf{u}_\nu^k$  of the predictive information matrix (or on the right eigenvectors of its transpose), we have considered both the right and the left eigenvectors as well as their interdependence. In this respect, the above derivations extend those in the literature on discriminant analysis.

In climate research, a number of authors have used some of the above methods, particularly in the detection of climate change [e.g., Bell (1982, 1986); Hasselmann (1993); see Hegerl and North (1997) for a review]. Hasselmann (1993), for example, takes a *climate change signal*  $\mathbf{v}_\nu^1$  as given and determines the linear combination  $Y_\nu^1 = (\mathbf{u}_\nu^1)^T \mathbf{X}_\nu$  of climatic variables that best discriminates between the climate change signal and a background noise of natural variability. He obtains from the signal  $\mathbf{v}_\nu^1$  the *optimal fingerprint*  $\mathbf{u}_\nu^1$  via the relation  $\mathbf{u}_\nu^1 = \Sigma_\nu^{-1} \mathbf{v}_\nu^1$ , which is a special case of the relation (15) between predictable patterns  $\mathbf{v}_\nu^k$  and weight vectors  $\mathbf{u}_\nu^k$ .

Another example of a method that is used in climate research and resembles discriminant analysis is the state space decomposition discussed by Thacker (1996). Thacker's state space decomposition formally parallels the predictable component analysis but derives from a different motivation, namely, seeking dominant modes of variability in datasets in which the data are affected by uncertainties.

The predictable component analysis unifies these approaches. Grounding the analysis in the literature on multivariate statistics should make a host of further methods accessible to climate research.

**5. Ensemble integrations**

Corresponding to the distinction between predictability studies of the first kind and predictability studies of the second kind, ensemble studies are divided into two kinds. Since analyzing these two kinds of studies

requires differing techniques, we will consider the two cases separately.

*a. Predictability studies of the first kind*

Studies of the first kind address the evolution of uncertainties in the initial condition for a prediction. In studies using ensemble integrations of a numerical model,  $M$  initial model states  $\mathbf{x}_0^1, \dots, \mathbf{x}_0^M$  are chosen such as to sample a probability distribution that represents uncertainties in the initial condition. Each initial state  $\mathbf{x}_0^i$  is then integrated forward in time, evolving into the state  $\mathbf{x}_\nu^i$  at instant  $\nu$ . Just as the initial states  $\mathbf{x}_0^1, \dots, \mathbf{x}_0^M$  form a sample of a distribution that represents uncertainties in the initial condition, the states  $\mathbf{x}_\nu^1, \dots, \mathbf{x}_\nu^M$  form a sample of a distribution that represents uncertainties in the prediction for lead time  $\nu$ .

The predictive information matrix is the product of the prediction error covariance matrix and the inverse of the climatological covariance matrix, and these covariance matrices are estimated as sample covariance matrices from the ensemble of model integrations. Since the climatological statistics are often almost stationary over typical forecast lead times, the climatological covariance matrix  $\Sigma = \Sigma_\nu$  is usually assumed to be independent of the lead time  $\nu$ . The climatological covariance matrix depends only, for example, on the month or the season for which a forecast is made. If, in addition to the ensemble integration, a longer control integration of the model is available, the climatological covariance matrix can be estimated from this control integration as the sample covariance matrix

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{\nu=1}^N (\mathbf{x}_\nu - \bar{\mathbf{x}})(\mathbf{x}_\nu - \bar{\mathbf{x}})^T. \tag{20}$$

The sample mean

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{\nu=1}^N \mathbf{x}_\nu \tag{21}$$

is an estimate of the climatological mean, and the index  $\nu$  runs over those  $N$  instants of the control integration that have the same climatological statistics as the instant for which the forecast is made. The sample covariance matrix  $\hat{\Sigma}$  is an estimate of the unknown climatological covariance matrix  $\Sigma$ .

The mean of the  $M$ -member ensemble

$$\hat{\mathbf{x}}_\nu \equiv \frac{1}{M} \sum_{i=1}^M \mathbf{x}_\nu^i \tag{22}$$

is a prediction of the model state  $\mathbf{x}_\nu$  at lead time  $\nu$  that evolved from some initial state  $\mathbf{x}_0$  drawn from the distribution representing initial uncertainties. The ensemble mean prediction is unbiased because the residuals

$$\mathbf{e}_\nu^i = \mathbf{x}_\nu^i - \hat{\mathbf{x}}_\nu,$$

which form a sample of the prediction error distribution,

have zero mean. The sample covariance matrix of the residuals

$$\begin{aligned} \hat{\mathbf{C}}_\nu &= \frac{1}{M-1} \sum_{i=1}^M \mathbf{e}_\nu^i (\mathbf{e}_\nu^i)^T \\ &= \frac{1}{M-1} \sum_{i=1}^M (\mathbf{x}_\nu^i - \hat{\mathbf{x}}_\nu)(\mathbf{x}_\nu^i - \hat{\mathbf{x}}_\nu)^T \end{aligned} \quad (23)$$

is an estimate of the prediction error covariance matrix.

The predictive information matrix is estimated from the sample covariance matrices as  $\hat{\mathbf{\Gamma}}_\nu = \hat{\mathbf{C}}_\nu \hat{\mathbf{\Sigma}}_\nu^{-1}$ , and the estimate  $\hat{\mathbf{\Gamma}}_\nu$  is substituted for the actual predictive information matrix  $\mathbf{\Gamma}_\nu$  in all of the above analyses. Thus, one can estimate predictive information matrices for a sequence of forecast lead times  $\nu$  and obtain the overall PP at each  $\nu$  from (5). Examining the PP as a function of lead time  $\nu$  will reveal typical timescales over which the predictability varies. As illustrated in section 7b, one can test by Monte Carlo simulation at which lead times  $\nu$ , if at any, the PP estimate is significantly greater than zero. At those lead times  $\nu$  at which the PP is significantly greater than zero, there exist predictable state vector components, and these can be identified by a predictable component analysis. The sequence of predictable patterns with a PP significantly greater than zero will disclose the system's predictable features as functions of forecast lead time  $\nu$ . The first predictable pattern is that pattern whose component is predictable with the smallest rms error relative to the rms error of the climatological mean prediction.

Since the estimate  $\hat{\mathbf{C}}_\nu$  of the prediction error covariance matrix and the estimate  $\hat{\mathbf{\Sigma}}_\nu$  of the climatological covariance matrix are computed from different datasets, finite sample effects may cause their difference  $\hat{\mathbf{\Sigma}}_\nu - \hat{\mathbf{C}}_\nu$  not to be positive semidefinite; that is, for some components, the prediction error variance may exceed the climatological variance. If the difference  $\hat{\mathbf{\Sigma}}_\nu - \hat{\mathbf{C}}_\nu$  is not positive semidefinite, the estimate of the predictive information matrix  $\hat{\mathbf{\Gamma}}_\nu$  has eigenvalues  $\hat{\gamma}_\nu^k$  that are greater than one, and such eigenvalues may lead to negative PPs. Negative PPs can be avoided by setting all estimated eigenvalues  $\hat{\gamma}_\nu^k > 1$  to  $\hat{\gamma}_\nu^k = 1$ . The predictable patterns and weight vectors that belong to eigenvalues greater than one are not reliably estimated; however, since they correspond to state space portions with small PP, they are of little interest.

*b. Predictability studies of the second kind*

Studies of the second kind address the predictability of the response of a system to changes in boundary conditions. Internal variability of the system renders a multitude of states consistent with a particular boundary condition, but the distributions of possible state realizations may differ from one boundary condition to another. Predictability of the second kind rests on the separability of the distributions of possible realizations: the more separable the distributions are according to different boundary conditions and the more the distribu-

tions are localized in state space, the more a prediction, based on knowledge of a particular boundary condition, reduces the uncertainty of which state may be realized.

In ensemble studies, each member  $i = 1, \dots, M$  of the ensemble is a model state that is consistent with a given boundary condition. The scatter of the  $M$  ensemble members around their mean reflects the internal variability. The climatic variability, reflected by the scatter of states around the climatological mean, is composed of the internal variability plus the variability of states induced by variability in the boundary conditions. In ensemble studies, variability in the boundary conditions is accounted for by determining the model's response to  $J$  different boundary conditions  $j = 1, \dots, J$ , which are chosen so as to sample the climatological distribution of boundary conditions. Thus, the simulated data consist of model states  $x_\nu^{ij}$ , where the indices  $i$  and  $j$  label the ensemble member and the boundary condition, respectively, and  $\nu$  designates the time for which predictability characteristics are being examined. For example, in a study that aims to assess the predictability of the response of the atmosphere to changes in SST,  $\nu$  may label the season and  $j$  a particular configuration of SST drawn from the climatological distribution of SST in season  $\nu$ . To perform such a study in practice, time-varying SST observations of various years may be prescribed as a boundary condition in a GCM. For each season  $\nu$ , the SST configurations in the years  $j = 1, \dots, J$  form a sample of the climatological distribution of SST. Each of the model states  $x_\nu^{ij}$  is one possible atmospheric state consistent with the SST configuration  $j$  in season  $\nu$ .

The analysis of such ensemble integrations uses techniques from the multivariate analysis of variance (MANOVA). [See, e.g., Johnson and Wichern (1982, chapter 6) for an introduction to MANOVA; among others, Harzallah and Sadourny (1995), Stern and Miyakoda (1995), and Zwiers (1996) have used univariate analysis of variance techniques in predictability studies of the second kind.] MANOVA tests whether  $J$  groups of multivariate random variables are separable. Similarly, predictability studies of the second kind are concerned with the separability of state distributions according to  $J$  different conditions on the system's boundary.

The climatological covariance matrix is estimated as the sample covariance matrix

$$\hat{\mathbf{\Sigma}}_\nu = \frac{1}{N-1} \sum_{j=1}^J \sum_{i=1}^M (\mathbf{x}_\nu^{ij} - \bar{\mathbf{x}}_\nu)(\mathbf{x}_\nu^{ij} - \bar{\mathbf{x}}_\nu)^T, \quad (24)$$

which measures, at time  $\nu$ , the scatter of the  $N = JM$  sample vectors  $\mathbf{x}_\nu^{ij}$  around the sample mean

$$\bar{\mathbf{x}}_\nu = \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^M \mathbf{x}_\nu^{ij}. \quad (25)$$

The sample mean is an estimate of the climatological mean.

Given a boundary condition  $j$  at time  $\nu$ , the ensemble mean

$$\hat{\mathbf{x}}_\nu^j = \frac{1}{M} \sum_{i=1}^M \mathbf{x}_\nu^{ij} \quad (26)$$

provides a prediction of the model state. As above, this prediction is unbiased because the residuals

$$\mathbf{e}_\nu^{ij} = \mathbf{x}_\nu^{ij} - \hat{\mathbf{x}}_\nu^j,$$

which form a sample of the prediction error distribution, have zero mean. The sample covariance matrix of the residuals

$$\begin{aligned} \hat{\mathbf{C}}_\nu^j &= \frac{1}{M-1} \sum_{i=1}^M \mathbf{e}_\nu^{ij} (\mathbf{e}_\nu^{ij})^T \\ &= \frac{1}{M-1} \sum_{i=1}^M (\mathbf{x}_\nu^{ij} - \hat{\mathbf{x}}_\nu^j) (\mathbf{x}_\nu^{ij} - \hat{\mathbf{x}}_\nu^j)^T \end{aligned}$$

is an estimate of the prediction error covariance matrix. From the estimate  $\hat{\mathbf{C}}_\nu^j$  of the prediction error covariance matrix and the estimate  $\hat{\mathbf{\Sigma}}_\nu$  of the climatological covariance matrix, one could compute the predictive information matrix and hence the PP and the predictable component analysis for each individual boundary condition  $j$  at time  $\nu$ . However, attention is seldom focused on predictability characteristics associated with individual boundary conditions but is often focused on predictability characteristics averaged over all boundary conditions that typically occur at time  $\nu$ . For example, atmospheric predictability characteristics associated with a particular SST configuration are often of less interest than average atmospheric predictability characteristics associated with SST configurations that typically occur in season  $\nu$ . For this reason, the estimated covariance matrices of the prediction error are often combined to an average covariance matrix

$$\begin{aligned} \hat{\mathbf{C}}_\nu &= \frac{1}{J} \sum_{j=1}^J \hat{\mathbf{C}}_\nu^j \\ &= \frac{1}{N-J} \sum_{j=1}^J \sum_{i=1}^M (\mathbf{x}_\nu^{ij} - \hat{\mathbf{x}}_\nu^j) (\mathbf{x}_\nu^{ij} - \hat{\mathbf{x}}_\nu^j)^T, \quad (27) \end{aligned}$$

where  $N - J = J(M - 1)$  is the number of degrees of freedom in the averaged estimate. Taking the average covariance matrix  $\hat{\mathbf{C}}_\nu$  in place of the individual covariance matrices  $\hat{\mathbf{C}}_\nu^j$  has the advantage of increasing the number of degrees of freedom in the estimate of the prediction error covariance matrix from  $M - 1$  to  $J(M - 1)$ . Averaging thus regularizes the estimate of the prediction error covariance matrix (Friedman 1989).

The predictive information matrix is estimated from the sample covariance matrices as  $\hat{\mathbf{\Gamma}}_\nu = \hat{\mathbf{C}}_\nu \hat{\mathbf{\Sigma}}_\nu^{-1}$ . With the estimates  $\hat{\mathbf{\Gamma}}_\nu$  of predictive information matrices, the overall PP at a sequence of  $\nu$  can be obtained from (5). If the index  $\nu$  labels seasons, for example, examining the PP as a function of  $\nu$  will reveal how the system's average predictability varies seasonally. By Monte Carlo

simulation or with Wilks' lambda statistic (see, e.g., Anderson 1984; chapter 8.4), it can be tested at which times  $\nu$ , if at any, the estimated PP is significantly greater than zero. At times  $\nu$  when the PP is significantly greater than zero, the predictable component analysis will yield the predictable patterns. The first predictable pattern is that pattern whose component varies most strongly, relative to its climatological variability, from one boundary condition to another.

In studies of the second kind, the climatological covariance matrix and the covariance matrix of the prediction error are estimated from a single dataset. Therefore, one would expect that the predictive information matrix can be estimated consistently in that all estimated eigenvalues  $\hat{\gamma}_\nu^k$  lie between zero and one. When the estimated covariance matrices  $\hat{\mathbf{\Sigma}}_\nu$  and  $\hat{\mathbf{C}}_\nu$  are computed from (24) and (27), respectively, it can be verified that the eigenvalues  $\hat{\gamma}_\nu^k$  of the predictive information matrix estimate  $\hat{\mathbf{\Gamma}}_\nu$  are greater than zero and are bounded above by  $(N - 1)/(N - J)$ . In the limit of large sample sizes  $N$ , the upper bound approaches unity from above, but for finite  $N$ , the eigenvalues  $\hat{\gamma}_\nu^k$  are not guaranteed to be less than or equal to unity. For the sake of consistent estimation, one may use biased covariance matrix estimates in which both the factor  $1/(N - 1)$  in the climatological sample covariance matrix (24) and the factor  $1/(N - J)$  in the sample covariance matrix (27) of the prediction error are replaced by  $1/N$ . These replacements ensure that the predictive information matrix estimate  $\hat{\mathbf{\Gamma}}_\nu$  has eigenvalues between zero and one so that the PP always lies between zero and one as well. However, the resulting PP estimate is biased toward larger values.

## 6. AR models as a class of empirical models

The complexity of comprehensive GCMs makes the direct computation of the probability distributions of model states and prediction errors impossible. Ensembles of states and predictions are simulated to infer the model statistics indirectly from samples. If, however, the process whose predictability is to be assessed can be modeled by a simpler empirical model, predictability characteristics can often be derived without the computational expense of ensemble integrations. For linear stochastic models, for example, statistics of states and predictions can be computed directly from the model parameters and the assumptions intrinsic to the model. Observational climate data or data simulated by a GCM are required only to estimate the adjustable parameters in a linear stochastic model. Whereas the GCMs used in ensemble integration studies are deterministic models, in which the underdetermination of an initial condition or the underdetermination of the state given boundary conditions limits the predictability of state vector realizations, in stochastic models it is the stochastic nature of the model itself that limits the predictability of model states.

Given a sample of a time series and a set of initial states, the predictability of future states of the time series can be investigated with multivariate autoregressive (AR) models. An *autoregressive model of order p* [AR( $p$ ) model] is a model of the form

$$\mathbf{x}_\nu = \mathbf{z} + \sum_{l=1}^p \mathbf{A}_l \mathbf{x}_{\nu-l} + \boldsymbol{\epsilon}_\nu, \quad \nu = 1, \dots, N, \quad (28)$$

for a stationary time series of  $m$ -dimensional state vectors  $\mathbf{x}_\nu$ . The  $p$  matrices  $\mathbf{A}_l \in \mathfrak{R}^{m \times m}$  ( $l = 1, \dots, p$ ) are called *coefficient matrices*, and the vectors  $\boldsymbol{\epsilon}_\nu = \text{noise}(\mathbf{S})$  are uncorrelated  $m$ -dimensional random vectors with zero mean and covariance matrix  $\mathbf{S}$ . The  $m$ -dimensional parameter vector of intercept terms  $\mathbf{z}$  allows for a non-zero mean

$$\langle \mathbf{X}_\nu \rangle = (\mathbf{I} - \mathbf{A}_1 - \dots - \mathbf{A}_p)^{-1} \mathbf{z}$$

of the time series (Lütkepohl 1993, chapter 2). The mean exists if the AR model is stable. Stability of the AR model will be assumed in what follows.

A sample of size  $N$  and  $p$  presample values of the state vectors  $\mathbf{x}_\nu$  ( $\nu = 1 - p, \dots, N$ ) are assumed to be available. The appropriate model order  $p$ , the coefficient matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$ , the intercept vector  $\mathbf{z}$ , and the noise covariance matrix  $\mathbf{S}$  must be estimated from the sample of state vectors. Methods for the identification of a model that is adequate to represent given time series data are well known to time series analysts. But since they appear to be largely unknown in the climate research community, we will summarize some model identification techniques before describing how predictions and the predictive information matrix can be obtained from a fitted AR model.<sup>3</sup>

#### a. Model identification

The model identification process comprises three phases (Tiao and Box 1981): (i) selecting the model order  $p$ ; (ii) estimating the coefficient matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$ , the intercept vector  $\mathbf{z}$ , and the noise covariance matrix  $\mathbf{S}$ ; and (iii) diagnostic checking of the fitted model's adequacy to represent the given time series.

#### 1) ORDER SELECTION

The number of adjustable parameters in an AR( $p$ ) model increases with the order  $p$  of the model. As the model order increases, one gains the flexibility to model a larger class of time series so that one can fit a model more closely to the given data. However, overfitting, that is, fitting a model too closely to the given time series realization, results in a fitted model with poor predictive capabilities. Selecting the model order means

finding an optimum between gaining flexibility by increasing the model order and avoiding the deterioration of predictions caused by overfitting.

The model order is commonly chosen as the minimizer of an order selection criterion that measures the goodness of an AR model fit. [For a discussion of order selection criteria, see Lütkepohl (1993, chapter 4).] Asymptotic properties in the limit of large sample sizes furnish the theoretical foundation of order selection criteria. Since statements valid in the limit of large sample sizes may be of only limited validity for the sample sizes available in practice, and since small-sample properties of order selection criteria are difficult to derive analytically, Lütkepohl (1985) compared the small-sample performance of various order selection criteria in a simulation study. Among all tested criteria in Lütkepohl's study, the Schwarz Bayesian criterion (SBC; see Schwarz 1978) chose the correct model order most often and also led, on the average, to the smallest mean-squared prediction error of the fitted AR models. Neumaier and Schneider (1997) proposed a modified Schwarz criterion (MSC) that, on small samples, estimates the model order yet more reliably than the original SBC.

In studies of climatic predictability, prior information about the model order is not usually available, so the model order must be estimated from the given data. Based on the above-cited studies, we recommend using SBC or MSC as criteria to select the AR model order.

#### 2) PARAMETER ESTIMATION

Under weak conditions on the distribution of the noise vectors  $\boldsymbol{\epsilon}_\nu$  in the AR model, it can be shown that the least squares (LS) estimators of the coefficient matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$ , of the intercept vector  $\mathbf{z}$ , and of the noise covariance matrix  $\mathbf{S}$  are consistent and asymptotically normal (Lütkepohl 1993, chapter 3). The LS estimators thus have desirable asymptotic properties. Since, beyond that, they also perform well on small samples and can be computed efficiently (Neumaier and Schneider 1997), LS estimation is our recommended method for estimating parameters in AR models, unless the estimated AR model is unstable or nearly unstable, in which case the computationally more expensive exact maximum likelihood method of Ansley and Kohn (1983, 1986) may be preferable.

#### 3) CHECKING MODEL ADEQUACY

After one has obtained the AR model that fits the given data best, it is necessary to check whether the model is adequate to represent the data. Adequacy of a fitted model is necessary for analyses of its predictability characteristics or of its dynamical structure to be meaningful. A variety of tests of model adequacy are described, for example, in Lütkepohl (1993, chapter 4),

<sup>3</sup> For an introduction to modeling multivariate time series with AR models, see Lütkepohl (1993).

Brockwell and Davis (1991, chapter 9.4), and Tiao and Box (1981).

As one approach to testing model adequacy, one can test whether the fitted model and the data are consistent with the assumptions intrinsic to AR models. A principal assumption in AR models is that the noise vectors  $\epsilon_\nu$  be uncorrelated. Uncorrelatedness of the noise vectors is, for example, invoked in the derivation of LS estimates and will be implicit in the discussion of predictions with AR models in section 6b. To test if the fitted model and the data are consistent with this assumption, the uncorrelatedness of the residuals

$$\hat{\epsilon}_\nu = \mathbf{x}_\nu - \hat{\mathbf{z}} - \sum_{l=1}^p \hat{\mathbf{A}}_l \mathbf{x}_{\nu-l}, \quad \nu = 1, \dots, N \quad (29)$$

can be tested. The superscript ( $\hat{\cdot}$ ) refers, as above, to estimated quantities, here to the LS estimates of the AR model parameters. Uncorrelatedness of the residuals can be tested by examining their autocorrelation function (Brockwell and Davis 1991, chapter 9.4) or by performing statistical tests such as the multivariate portmanteau test of Li and McLeod (1981).

*b. Predictions with an estimated model*

After an AR( $p$ ) model has been identified that is adequate to represent a given time series of state vectors, future state vectors can be predicted with the estimated model. Suppose that  $p$  initial states  $\mathbf{x}_0, \mathbf{x}_{-1}, \dots, \mathbf{x}_{1-p}$  are given independently of the sample from which the AR model was estimated and that the state  $\mathbf{x}_\nu$ ,  $\nu$  steps ahead of  $\mathbf{x}_0$ , is to be predicted. The  $\nu$ -step prediction

$$\hat{\mathbf{x}}_\nu = \hat{\mathbf{z}} + \sum_{l=1}^p \hat{\mathbf{A}}_l \hat{\mathbf{x}}_{\nu-l}, \quad (30)$$

with  $\hat{\mathbf{x}}_j = \mathbf{x}_j$  for  $j \leq 0$ , predicts the state  $\mathbf{x}_\nu$  optimally in that it is the linear prediction with minimum rms prediction error (Lütkepohl 1993, chapters 2.2 and 3.5).

We take into account two contributions to the error in predictions with estimated AR models. The first contribution to the prediction error arises because AR models are stochastic models whose predictions are always subject to uncertainty, even when the model parameters are known. The second contribution to the prediction error arises because the AR parameters are estimated, as opposed to being known, and are thus afflicted with sampling error. The uncertainty in the estimated parameters adds to the uncertainty in the predictions. A third contribution to the prediction error arises from uncertainty about the correct model order and uncertainty about the adequacy of an AR model to represent the given data. This third contribution, which results from uncertainty about the model structure, will be ignored in what follows. Draper (1995) discusses how the uncertainty about a model structure affects predictions.

For the  $\nu$ -step prediction (30), the first two contributions to the prediction error are uncorrelated (Lüt-

kepohl 1993, chapter 3.5.1). Therefore, the covariance matrix of the  $\nu$ -step prediction error is the sum

$$\mathbf{C}_\nu = \mathbf{C}_\nu^{\text{mod}} + \mathbf{C}_\nu^{\text{spl}} \quad (31)$$

of the prediction error covariance matrix  $\mathbf{C}_\nu^{\text{mod}}$  of an AR model with known parameters and the covariance matrix  $\mathbf{C}_\nu^{\text{spl}}$  of the sampling error in the prediction. Estimates of the error covariance matrices  $\hat{\mathbf{C}}_\nu^{\text{mod}}$  and  $\hat{\mathbf{C}}_\nu^{\text{spl}}$  for arbitrary forecast lead times  $\nu$  are given in Lütkepohl (1993, chapter 3.5).

As the forecast lead time  $\nu$  approaches infinity, the  $\nu$ -step prediction (30) approaches the estimate of the mean

$$\hat{\boldsymbol{\mu}} = (\mathbf{I} - \hat{\mathbf{A}}_1 - \dots - \hat{\mathbf{A}}_p)^{-1} \hat{\mathbf{z}} \quad (32)$$

of the AR process. The mean of the AR process is the optimal long-range prediction of the model state, and it is the optimal prediction when no initial states are available (cf. Lütkepohl 1993, chapter 2.2).

The climatological covariance matrix of the fitted AR model—the error covariance matrix of the prediction consisting of the estimated mean  $\hat{\boldsymbol{\mu}}$ —is the sum  $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}^{\text{mod}} + \hat{\mathbf{M}}$  of the covariance matrix  $\boldsymbol{\Sigma}^{\text{mod}}$  of AR model states and the covariance matrix  $\mathbf{M}$  of the sampling error in the estimate  $\hat{\boldsymbol{\mu}}$  of the mean. The prediction error covariance matrices  $\mathbf{C}_\nu^{\text{mod}}$  approach the covariance matrix  $\boldsymbol{\Sigma}^{\text{mod}}$  of the states as  $\nu$  approaches infinity. An estimate  $\hat{\boldsymbol{\Sigma}}^{\text{mod}}$  of the covariance matrix of model states can be computed from the estimated AR parameters by the method in Lütkepohl (1993, chapter 2.1).<sup>4</sup> An estimate  $\hat{\mathbf{M}}$  of the covariance matrix of the sampling error in the mean  $\hat{\boldsymbol{\mu}}$  follows by substituting estimated parameters for the exact parameters in the asymptotic expression for the matrix  $\mathbf{M}$  given in Lütkepohl (1993, chapter 3.4). Finally, the sum

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}^{\text{mod}} + \hat{\mathbf{M}} \quad (33)$$

is an estimate of the climatological covariance matrix.

The predictive information matrix of an AR model is estimated from the prediction error covariance matrix  $\hat{\mathbf{C}}_\nu$  and the climatological covariance matrix  $\hat{\boldsymbol{\Sigma}}$  as  $\hat{\boldsymbol{\Gamma}}_\nu = \hat{\mathbf{C}}_\nu \hat{\boldsymbol{\Sigma}}^{-1}$ . The PP and the predictable component analysis can then be computed, just as in ensemble studies of the first kind, for a sequence of forecast lead times  $\nu$ .<sup>5</sup>

As mentioned above, in ensemble studies of the first kind, finite sample effects may cause the estimated PPs to be negative. The same remark applies to AR models because the referenced estimators of the sampling error covariance matrices are asymptotic estimators that be-

<sup>4</sup> To ensure that the estimated covariance matrices are compatible with each other, the estimate  $\hat{\boldsymbol{\Sigma}}^{\text{mod}}$  of the state covariance matrix should be computed directly from the fitted AR parameters. The state covariance matrix should not be estimated as the sample covariance matrix of the states.

<sup>5</sup> Box and Tiao (1977) offer an analysis of AR models that resembles the predictable component analysis but neglects sampling errors in the prediction.

come exact only in the limit of infinite sample sizes. In finite samples, the difference  $\hat{\Sigma} - \hat{C}_v$  between the estimated climatological covariance matrix and the estimated prediction error covariance matrix is not necessarily positive semidefinite. Negative PPs can again be avoided by setting to unity all those eigenvalues of the estimated predictive information matrix that are greater than unity.

**7. Example: North Atlantic multidecadal variability**

To assess the predictability of multidecadal variability in the North Atlantic, Griffies and Bryan (1997a,b) have performed an ensemble study of the first kind with the Geophysical Fluid Dynamics Laboratory coupled climate model (Delworth et al. 1993, 1997). With univariate methods, they have investigated the predictability of individual principal components of various oceanic fields. Here we reexamine a part of Griffies and Bryan’s dataset with the multivariate methods proposed.

In this illustrative application of the PP concept and the predictable component analysis, we focus on the annual mean of the dynamic topography. The set of simulated data to be examined [“Ensemble A” of Griffies and Bryan (1997b)] consists of a control integration with 200 yr of data and an ensemble of  $M = 12$  integrations with 30 yr of data each. We restrict our investigation to the North Atlantic sector extending from the equator to 72°N and containing  $m = 247$  model grid points. In this region, the interannual variability of the model’s dynamic topography is dominated by an oscillation with a period of about 40 yr. Griffies and Bryan have found individual principal components of this oscillation to be predictable up to 10–20 yr in advance. Regarding the oscillation as an interaction among several principal components in a multidimensional state space, we will estimate the overall PP of this oscillation and determine its predictable patterns.

For this purpose, predictive information matrices  $\hat{\Gamma}_v = \hat{C}_v \hat{\Sigma}^{-1}$  must be estimated for a sequence of forecast lead times  $v$ . Since the state space dimension  $m = 247$  exceeds the number of degrees of freedom in estimates of both the climatological covariance matrix  $\hat{\Sigma}$  and the prediction error covariance matrices  $\hat{C}_v$ , it is necessary to regularize the estimates. To do so, we perform a principal component analysis, truncate it, and then examine the predictability characteristics of the dynamic topography in the state space of the retained principal components.

*a. Principal component analysis of the dynamic topography*

The EOFs and the variances of the principal components would traditionally be estimated as the eigenvectors and eigenvalues of the sample covariance matrix of the control integration. The eigenvalues of the sample

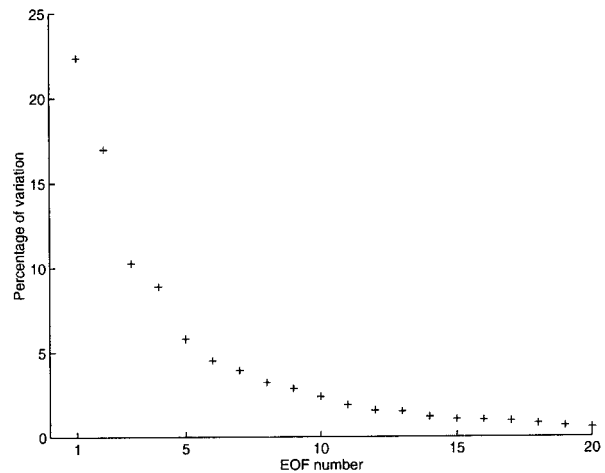


FIG. 1. Percentage of total variation accounted for by each of the first 20 EOFs of North Atlantic dynamic topography.

covariance matrix are, however, biased estimates of the variances (Lawley 1956). The bias of the eigenvalues is positive for the leading principal components and negative for the trailing principal components. The positive bias of the eigenvalues of the leading principal components, the components retained in a truncated principal component analysis, implies a positive bias of the regularized estimate of the climatological covariance matrix. This selection bias would lead to a PP estimate that is biased toward larger values.

We avoid the selection bias of the PP estimate by partitioning the control integration into two parts with  $N = 100$  randomly drawn years each and by selecting principal components from one part of the dataset and estimating the climatological covariance matrix from the other (cf. Miller 1984). We denote the sample covariance matrices of the two parts of the control integration by  $\hat{\Sigma}^{(1)}$  and  $\hat{\Sigma}^{(2)}$  and refer to the eigenvectors  $\hat{w}^k$  ( $k = 1, \dots, m$ ) of the matrix  $\hat{\Sigma}^{(1)}$  as EOFs and to the projections  $(\hat{w}^k)^T \mathbf{x}_v$  of state vectors  $\mathbf{x}_v$  onto the EOFs as principal components. If the  $r$  EOFs retained in a truncated principal component analysis form the columns of the matrix  $\hat{W} \in \mathbb{R}^{m \times r}$ , then the generally nondiagonal estimate  $\hat{W}^T \hat{\Sigma}^{(2)} \hat{W}$  of the covariance matrix of the retained principal components is not affected by selection bias.

Selecting a truncation of the principal component analysis means finding a trade-off between reducing the sampling variability of the estimated covariance matrices by including fewer EOFs and reducing the truncation error by including more EOFs. The fewer EOFs retained, the smaller the sampling error in the estimated covariance matrices. The eigenvalue associated with an EOF gives the mean-squared truncation error that omission of this EOF would entail.

Figure 1 shows the spectrum of the eigenvalues associated with the first 20 EOFs of the dynamic topography. The eigenvalues are normalized such as to in-

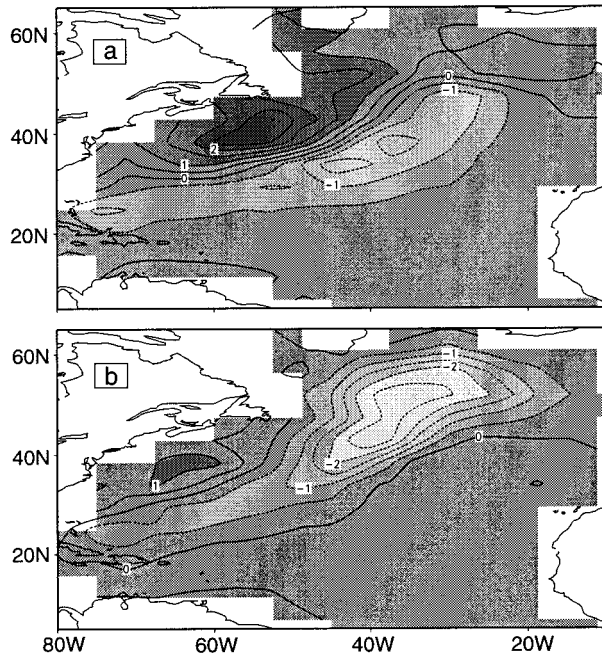


FIG. 2. (a) First EOF and (b) second EOF of North Atlantic dynamic topography [dynamic cm]. The patterns are scaled by the standard deviations of their associated principal components.

indicate the percentage of total sample variation accounted for by each of the EOFs. Figure 1 suggests a truncation after the fourth EOF, at the eigenvalue spectrum's point of greatest curvature. However, the third and fourth EOFs represent centennial trends in the model's deeper ocean (Griffies and Bryan 1997b). Since we are interested in multidecadal oscillations and not in long-term model trends, we focus on the first two EOFs.

Figure 2 shows the first and second EOF patterns. The EOF patterns, normalized to norm unity, are in this plot multiplied by the standard deviation of their principal components, so the amplitude of the patterns indicates the rms variability of the dynamic topography. EOF1 represents variations in the strength of the North Atlantic Current's northeastward drift. The dynamic topography variations have maximum amplitude in the model's western boundary region. EOF2 is more elongated meridionally than EOF1, and it extends farther northward into the sinking region of the model's subpolar North Atlantic. EOF2 represents gyre-shaped variations in the North Atlantic circulation with the strongest current variations directed northeastward and located in the central portion of the basin. The principal components associated with these patterns exhibit irregular oscillations with a dominant period of 40–45 yr. Since the two principal component time series are roughly in quadrature, the EOFs can be viewed as different phases of an oscillatory mode of dynamic topography variability.

Together, the first two EOFs account for 39% of the total sample variation in North Atlantic dynamic to-

pography. In the central and western North Atlantic, where these EOFs have the largest amplitude, the correlation coefficients between the principal component time series and the local variations in dynamic topography exceed 0.8 (Griffies and Bryan 1997b). Hence, predictability of the two EOFs would imply predictability of a large portion of local variability in these dynamically active regions.

#### b. Predictability estimates from the ensemble integration

To obtain regularized estimates of the predictive information matrices, we project the data from the ensemble integration onto the  $r = 2$  EOFs retained in the truncated principal component analysis. For each lead time  $\nu = 1, \dots, 30$  yr, we estimate a prediction error covariance matrix  $\hat{\mathbf{C}}_\nu \in \mathbb{R}^{r \times r}$  from the sample covariance matrix of the residuals (23). The full sample covariance matrix  $\hat{\mathbf{\Sigma}}^{(2)} \in \mathbb{R}^{m \times m}$  of  $N = 100$  yr of the control integration is used in the estimate  $\hat{\mathbf{\Sigma}} = \hat{\mathbf{W}}^T \hat{\mathbf{\Sigma}}^{(2)} \hat{\mathbf{W}} \in \mathbb{R}^{r \times r}$  of the climatological covariance matrix in the truncated EOF basis. With the Cholesky factorization technique in the appendix, PPs and the predictable component analysis are computed from the covariance matrices  $\hat{\mathbf{C}}_\nu$  and  $\hat{\mathbf{\Sigma}}$ .

This approach to estimating the predictive information matrices avoids selection bias. However, the estimation of the climatological covariance matrix draws upon only 100 yr out of the 200 yr of available data. But since for the estimation of the prediction error covariance matrices  $\hat{\mathbf{C}}_\nu$  only 11 degrees of freedom are available, against 99 degrees of freedom for the estimation of the climatological covariance matrix  $\hat{\mathbf{\Sigma}}$ , the sampling variability of the predictive information matrices  $\hat{\mathbf{I}}_\nu = \hat{\mathbf{C}}_\nu \hat{\mathbf{\Sigma}}^{-1}$  is dominated by the sampling variability of the prediction error covariance matrices  $\hat{\mathbf{C}}_\nu$ . Therefore, ignoring one-half of the control integration in the estimation of the climatological covariance matrix has little effect on the accuracy of the predictive information matrix estimates.

Figure 3a shows the overall PP of the first two EOFs as a function of forecast lead time  $\nu$ . At each  $\nu$  we estimate, by Monte Carlo simulation of 1000 samples, a 95% confidence interval for the PP estimate. Each of the 1000 samples consists of  $M = 12$  random vectors drawn from a Gaussian distribution with a covariance matrix equal to the sample covariance matrix  $\hat{\mathbf{C}}_\nu$  of the residuals and  $N = 100$  random vectors drawn from a Gaussian distribution with a covariance matrix equal to the estimated covariance matrix  $\hat{\mathbf{\Sigma}} = \hat{\mathbf{W}}^T \hat{\mathbf{\Sigma}}^{(2)} \hat{\mathbf{W}}$  of the principal components. A PP is computed from each sample. Adding the difference between the 97.5th percentile and the mean of the simulated PPs to the overall PP gives the upper bound of the estimated 95% confidence interval, and subtracting the difference between the mean of the simulated PPs and the 2.5th percentile gives the lower bound. Because the thus estimated 95% con-



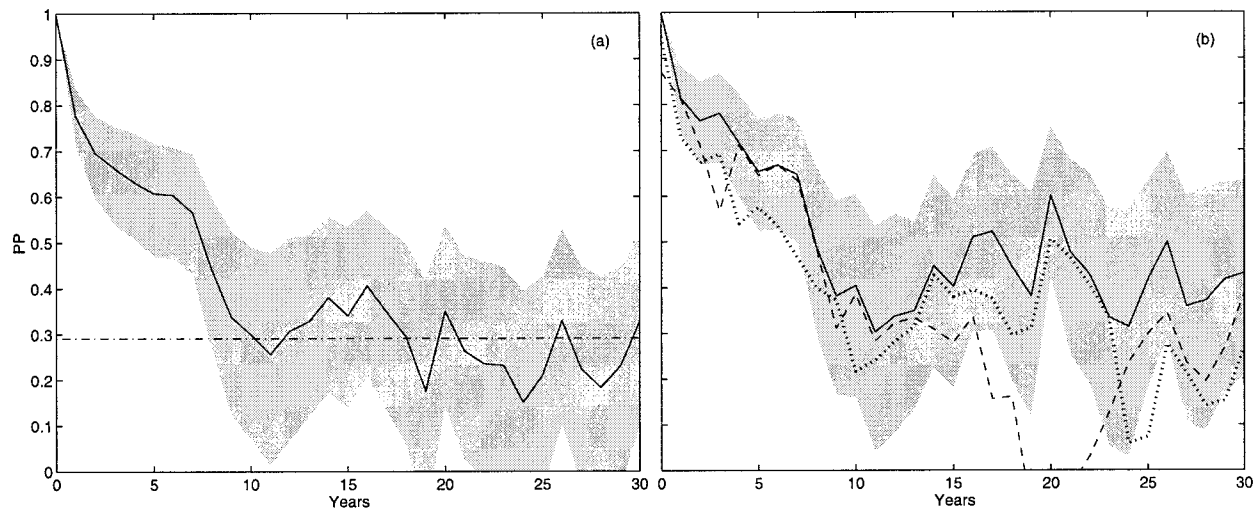


FIG. 3. Predictive power for North Atlantic dynamic topography as a function of forecast lead time. (a) Overall PP (solid line) of the first two EOFs with 95% confidence interval (shaded). PPs above the dash-dotted line are significant at the 5% level. (b) PP of the first predictable pattern (solid line) with 95% confidence interval (shaded). Individual PPs of the first EOF (dashed line) and of the second EOF (dotted line).

confidence interval is centered on the estimated overall PP, it does not account for a bias of the PP estimate.

The difference between the mean of the Monte Carlo simulated PPs and the PP estimate from the GCM ensemble is a measure of the bias of the PP estimate. The mean of the Monte Carlo simulated PPs is always greater than the PP estimate, indicating a bias of the PP estimate toward larger values, but the average PP difference of 0.03 is negligible compared with the sampling error in the PP estimates.

The bias of the PP estimate is small enough that the PP can be considered significantly greater than zero when the 95% confidence interval does not include zero. To justify this heuristic for establishing significance of the PP estimate, we test, also by Monte Carlo simulation, the null hypothesis that the residuals and the state vectors of the control integration are drawn from distributions with equal covariance matrices. If the null hypothesis is true that the climatological covariance matrix  $\Sigma$  and the prediction error covariance matrix  $C_v$  are equal, then there are no predictable components in the state space of the first two EOFs. The null hypothesis is rejected at the 5% significance level for PPs greater than 0.28, the bound marked by the dash-dotted line in Fig. 3a. The lead times at which the 95% confidence interval for the PP estimate does not include zero approximately coincide with the lead times at which the estimated PP is greater than the 0.28 significance bound. The overall PP decays rapidly over the first 10 yr of the forecasting lead time, remains marginally significant up to about year 17, and becomes insignificant beyond year 17.

When the overall PP is significantly greater than zero, there exist predictable linear combinations of state variables, and the predictable component analysis identifies the most predictable of those. Figure 3b shows the pre-

dictive power  $\hat{\alpha}_v^1$  of the first predictable pattern as well as the individual PPs of the two EOFs. The first predictable pattern is the linear combination of EOF1 and EOF2 with the largest PP. As discussed in section 4a, the PP of the first predictable pattern is always greater than both the overall PP of the two EOFs combined and the individual PPs of EOF1 and EOF2.

The confidence interval for the estimated PP of the first predictable pattern is obtained from the same Monte Carlo simulation as the confidence interval for the estimated overall PP. Because of the selection bias introduced by selecting the most predictable linear combination of state variables, exclusion of zero from the confidence interval is not sufficient for a PP estimate  $\hat{\alpha}_v^1$  to be considered significantly greater than zero. The fact that, beyond year 17, the 95% confidence interval for the estimated PP of the first predictable pattern does not include zero cannot be taken as evidence of a PP significantly greater than zero.

Questions as to which combination of state variables contributes most to the predictability of state realizations are only meaningful when the overall PP is greater than zero. However, in a statistical test of whether the overall PP is consistent with zero, it is possible that the null hypothesis of zero overall PP is accepted because of a lack of power of the test, not because it is in fact true. With subset selection techniques that exclude from the state space components of small PP (cf. McLachlan 1992, chapter 12), it might then be possible to identify a lower-dimensional subspace in which the test has greater power and yields a significant overall PP. For an example of a similar phenomenon in a different multivariate test, see Johnson and Wichern (1982, chapter 5B). In our two-dimensional example, however, Fig. 3b shows that, at most lead times, both EOFs contribute to the overall PP, so analyzing them jointly seems appro-

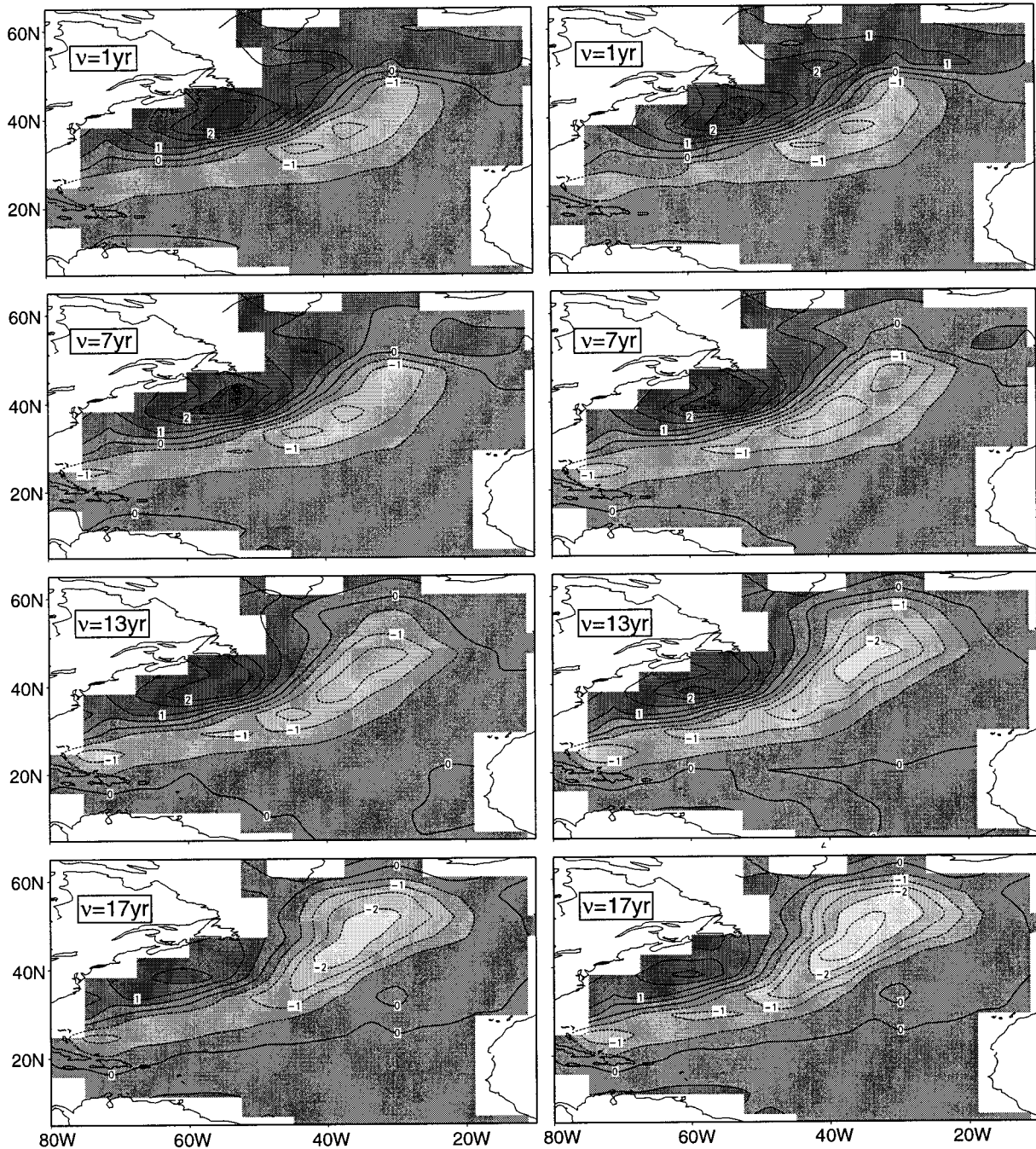


FIG. 4. First predictable patterns of dynamic topography [dynamic cm] for lead times  $\nu = 1, 7, 13,$  and  $17$  yr. Left column: first predictable pattern of ensemble study. Right column: first predictable pattern of AR model fitted to 100 yr of GCM data.

appropriate. From the above Monte Carlo simulations we therefore conclude that the overall PP is insignificant beyond year 17, and we only consider the first predictable patterns and their PPs up to this lead time.

Figure 3b suggests that during most of the first 13 yr, EOF1 has a greater PP than EOF2; conversely, between years 14 and 17, EOF2 has a greater PP than EOF1. The succession of first predictable patterns  $\hat{\nu}_\nu^1$  at lead

times  $\nu = 1, 7, 13,$  and  $17$  yr, displayed in the left column of Fig. 4, also reflects the relative magnitudes of the individual PPs of the two EOFs. EOF1 dominates the first predictable pattern at lead times 1 and 7 yr. At year 13, EOF2 starts to contribute significantly to the first predictable pattern. At year 17, EOF2 dominates the first predictable pattern.

Both the normalization and the sign of the predictable

patterns are matters of convention. According to the normalization convention (16), the predictable patterns are orthogonal with respect to the inverse climatological covariance matrix  $\hat{\Sigma}^{-1}$ . Therefore, amplitudes of the patterns in Fig. 4 indicate the rms variability of the dynamic topography. The sign of the first predictable pattern  $\hat{\mathbf{v}}_\nu^1$  at lead time  $\nu$  is chosen so as to minimize the squared Mahalanobis distance  $(\hat{\mathbf{v}}_\nu^1 - \hat{\mathbf{v}}_{\nu-1}^1)^T \hat{\Sigma}^{-1} (\hat{\mathbf{v}}_\nu^1 - \hat{\mathbf{v}}_{\nu-1}^1)$  to the pattern  $\hat{\mathbf{v}}_{\nu-1}^1$  at lead time  $\nu - 1$ . The first predictable pattern  $\hat{\mathbf{v}}_0^1$  at  $\nu = 0$  is taken to be the normalized model state at the initialization of the ensemble integration. This sign convention ensures that the first predictable patterns evolve smoothly with forecast lead time.

In the initial conditions for the ensemble integration, the Atlantic overturning circulation is anomalously weak. Figure 4 shows that what is most predictable 1 and 7 yr in advance is the state vector component associated with a weak-drift anomaly in the North Atlantic Current. Most predictable 13 yr in advance is the state vector component associated with a spreading of the weak-drift anomaly into subpolar regions and with the beginning formation of a gyre-shaped current anomaly in the central North Atlantic. Most predictable 17 yr in advance is the state vector component associated with a decrease in amplitude of the weak-drift anomaly in the North Atlantic Current and with an increase in amplitude and a northward spreading of the gyre-shaped anomaly in the central North Atlantic. The first predictable patterns represent those features of the dynamic topography whose components are predictable with the smallest rms error relative to the rms error of the climatological mean prediction.

The examination of the PP, of the significance of the PP, and of the first predictable patterns confirms and complements the analyses of Griffies and Bryan (1997b). The results presented above were found to be robust: they do not depend on the particular random partition of the control integration chosen to estimate the EOFs and the climatological covariance matrix.

### c. Empirical predictability estimates from an AR model

Predictability characteristics can also be derived empirically from AR models fitted to the same set of GCM data projected onto two EOFs. The AR model identification is performed with the software package ARFIT (Schneider and Neumaier 1997).

Among AR( $p$ ) models of order  $p = 0, \dots, 6$ , the order selection criteria SBC and MSC indicate that an AR(1) model best fits the first 100 yr of the GCM control integration. To check whether the fitted AR(1) model is adequate to represent the GCM data, we test the residuals (29) of the fitted model for uncorrelatedness. For  $N$  realizations of a white noise process, approximately 95% of the sample autocorrelations are expected to lie within the bounds  $\pm 1.96(N)^{-1/2}$  (e.g., Brockwell and Davis 1991, chapter 7.2). For the 99 bivariate residuals

of the fitted AR(1) model, all but 2 of the 40 sample autocorrelations between lags 1 and 20 lie within the bounds  $\pm 1.96(99)^{-1/2}$ . Additionally, the modified portmanteau statistic of Li and McLeod (1981) does not reject, at the 5% significance level, the null hypothesis that the residuals are uncorrelated. Therefore, within the class of AR models, an AR(1) model fits the first 100 yr of the control integration best; the residuals of the fitted model provide no grounds for rejecting the hypothesis that the model is consistent with the GCM data.

The AR model identification procedure was repeated with segments of the GCM data of various lengths. AR(1) models were consistently found to be the best fitting, and diagnostic tests of the fitted models provided no grounds for rejecting the hypothesis of model adequacy. But the fact that an AR(1) model seems to represent adequately the particular set of oscillatory principal component time series considered here is not to be taken as a justification for the indiscriminate use of first-order models. The adequacy of a fitted model must be assessed, and, when linear stochastic models are at all adequate, then higher-order models will often be more appropriate than first-order models. [For an example of how models of inappropriately low order can produce misleading results, see Tiao and Box (1981).] As discussed in section 6, the PP concept and the predictable component analysis are applicable to AR models of any order.

A  $\chi^2$ -test of the skewness and kurtosis of the residuals (29) of the fitted AR(1) model does not reject, at the 5% significance level, the hypothesis that the residuals are a realization of a Gaussian process. The estimators of the prediction error covariance matrices in section 6b are valid for Gaussian processes and can thus be expected to yield reliable estimates of the PP of the AR model (cf. Lütkepohl 1993, chapters 3.5 and 4.5).

Figure 5 shows the PP of the fitted AR(1) model as a function of forecast lead time. Since the prediction error variance of an AR model is a monotonically increasing function of lead time (Lütkepohl 1993, chapter 2.2), the PP decreases monotonically. The overall PP of the AR(1) model fitted to the first 100 yr of the GCM control integration reaches zero at a lead time of about 20 yr. Beyond year 20, the error variances of a model prediction are estimated to be as large as the error variances of the climatological mean prediction.

Included in Fig. 5 is the overall PP of an AR model fitted to only 30 yr of GCM data. Sampling errors contribute to the uncertainty in predictions, and since the sampling errors in models decrease with increasing length of the time series from which the model parameters are estimated, the PP of the model fitted to 30 yr of data is smaller than the PP of the model fitted to 100 yr of data. The PP of the model fitted to 30 yr of data already vanishes at a lead time of 11 yr. However, if sampling errors in the estimated parameters are not taken into account and the PP is estimated only from the generic prediction error matrices  $\hat{\mathbf{C}}_\nu^{\text{mod}}$  and  $\hat{\Sigma}^{\text{mod}}$ , then the

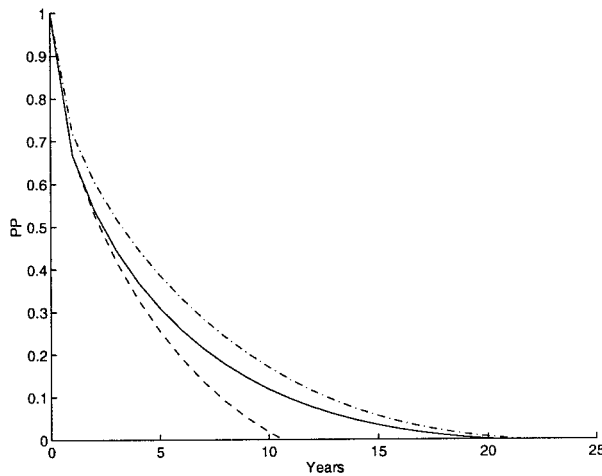


FIG. 5. Here, PP of AR models as a function of forecast lead time: overall PP (solid line) and PP of first predictable pattern (dash-dotted line) of AR model fitted to 100 yr of GCM data; overall PP of AR model fitted to 30 yr of GCM data (dashed line).

PPs of the two AR models are almost indistinguishable and nonzero up to year 40. Therefore, the difference between the two PP curves in Fig. 5 can be attributed entirely to larger sampling errors in the parameters of the model fitted to the shorter time series. Neglecting sampling errors leads to a gross overestimation of the predictive capabilities of fitted AR models.

The PP of the first predictable pattern of the AR(1) model fitted to 100 yr of GCM data is shown in Fig. 5, and the right column of Fig. 4 displays the first predictable pattern itself. The sign and normalization conventions of section 7b are applied. The qualitative features of the first predictable patterns of the AR model are the same as those of the first predictable patterns inferred from the ensemble integration. Since the predictive information matrix of an AR model does not depend on the initial condition for a prediction, the significance of the agreement between the first predictable patterns inferred from the ensemble integration and from the AR model goes beyond indicating that the AR model fits the GCM data well. At lead time 1 year, the first predictable pattern of the AR model is dominated by the anomaly, represented by EOF1, in the strength of the North Atlantic Current's northeastward drift. This pattern suggests that predictions of the AR model are particularly reliable when initialized during either an extremely strong or an extremely weak phase of the oscillation in the strength of this drift. The GCM ensemble was initialized during an extremely weak phase of the North Atlantic Current, which explains the agreement between the predictable patterns.

Thus, the estimation of predictability characteristics from the ensemble integration and from an AR model fitted to a small fraction of the GCM data leads to similar results. For the GCM and the AR model, the lead-time scales over which the overall PP is distinguishable from

zero coincide, and the same features of the dynamic topography field are associated with large PP. Such an agreement of results is possible if, as in our example, the process in question can be modeled as a linear superposition of stochastically forced damped-oscillatory and relaxatory modes, modes that an AR model is able to represent.

## 8. Concluding remarks

We have presented a conceptual framework for the multivariate analysis of predictability studies. The predictability measure in this framework, the PP, indicates by how much a prediction reduces the uncertainty as to which state of the predicted process will occur. The uncertainties in the state before and after a specific prediction is made are quantified by the prior entropy and the posterior entropy. The difference between these two entropies is the predictive information contained in a prediction. The PP, an index between zero and one, is based on an exponential of the predictive information and measures the efficacy of predictions in narrowing the range of values typically taken by state vector components.

To quantify predictability, the information content of predictions must be measured relative to the background information available prior to the issue of a prediction. Since climatological statistics are accessible in the types of predictability studies discussed in this paper, we chose to measure the predictive power of predictions relative to the climatological mean prediction as a baseline. The prior entropy thus became the entropy of the climatological distribution of states, or the entropy of the distribution of errors in the climatological mean prediction. Other choices of a baseline are, however, possible. To evaluate the performance of weather forecasting models, for example, one might choose the persistence forecast as a baseline. The methods presented above can then be applied with the prediction error of the persistence forecast substituted for the prediction error of the climatological mean prediction.

For Gaussian random variables, the PP is a function of the determinant of the predictive information matrix, the product matrix of the prediction error covariance matrix and the inverse of the climatological covariance matrix. Estimating the PP thus reduces to estimating the predictive information matrix from samples of data or from estimated parameters of empirical models. We have discussed how the predictive information matrix is obtained from ensemble integration studies of the first and the second kind and from AR models fitted to observed or simulated data. The application of the PP concept in an ensemble integration study of the predictability of multidecadal North Atlantic variability illustrates how confidence intervals and significance bounds for the PP estimate can be established and how the PP is to be interpreted.

When the estimated PP of a process is significantly

greater than zero, the process has predictable components, and these can be discriminated from the unpredictable components by a predictable component analysis, an eigendecomposition of the predictive information matrix. If state vectors are expanded in terms of predictable patterns—that is, in terms of the right eigenvectors of the predictive information matrix—then their first component is the most predictable, and subsequent components are mutually uncorrelated and ordered by PP from largest to smallest. The examination of North Atlantic variability illustrates the interpretation of the first predictable pattern. The sequence of predictable patterns for forecast lead times between 1 and 17 yr shows the most predictable features of a multi-decadal oscillation in the dynamic topography field. In this example, the analysis of an AR model adequately representing the oscillation in the dynamic topography and the analysis of an ensemble of GCM integrations yield similar lead-time scales of nonzero PP and similar predictable patterns.

Although the PP and the predictable component analysis have been derived under the assumption that the states and prediction errors follow Gaussian distributions, it is the ellipsoidal symmetry of the distributions that is more important than their detailed shape (Friedman 1989). Hence, the assumption of Gaussian distributions can, in practice, be relaxed to a symmetry assumption.

The framework that has been presented in this paper is applicable to a wider range of studies than that explicitly covered. The above-mentioned performance evaluation of weather forecasting models is but one example of further applications. Grounding our analyses in the literature on multivariate statistics will, we hope, facilitate the extension of the framework to other applications.

*Acknowledgments.* We wish to express our thanks to Jeff Anderson and Arnold Neumaier, who drew our attention to some of the referenced literature on climatic predictability and discriminant analysis, respectively. Jeff Anderson, Ruth Michaels, Thomas Müller, Arnold Neumaier, Heidi Swanson, and Jens Timmer carefully read drafts of this paper. We gratefully acknowledge their comments and criticism, which led to substantial improvements in the final version.

APPENDIX

**Computation of the Predictable Component Analysis**

The predictable component analysis simultaneously diagonalizes the climatological covariance matrix and the prediction error covariance matrix. This fact can be exploited in the practical computation of predictable components and predictable patterns (cf. Fukunaga 1990, chapter 2).

First, we compute the principal component analysis  $\Sigma_v = \mathbf{W}_v \Lambda_v \mathbf{W}_v^T$  of the climatological covariance matrix  $\Sigma_v$ . The orthogonal matrix  $\mathbf{W}_v$ , whose columns are the eigenvectors of  $\Sigma_v$ , and the diagonal eigenvalue matrix  $\Lambda_v = \text{Diag}(\lambda_v^k)$  are then used in a whitening transformation (18), which transforms the climatological covariance matrix into the identity matrix and the prediction error covariance matrix into the matrix  $\mathbf{K}_v$ , given in (19). The matrix  $\mathbf{K}_v$  is symmetric but not necessarily diagonal. It can be diagonalized by a further orthogonal transformation  $\mathbf{T}_v$ , with columns of  $\mathbf{T}_v$  formed by eigenvectors of  $\mathbf{K}_v$ , such that, as in (14),

$$\mathbf{T}_v^T \Lambda_v^{-1/2} \mathbf{W}_v^T \mathbf{C}_v \mathbf{W}_v \Lambda_v^{-1/2} \mathbf{T}_v = \mathbf{T}_v^T \mathbf{K}_v \mathbf{T}_v = \text{Diag}(\gamma_v^k). \tag{A1}$$

The identity matrix (18) remains unchanged under this transformation:

$$\mathbf{T}_v^T \Lambda_v^{-1/2} \mathbf{W}_v^T \Sigma_v \mathbf{W}_v \Lambda_v^{-1/2} \mathbf{T}_v = \mathbf{T}_v^T \mathbf{I} \mathbf{T}_v = \mathbf{I}. \tag{A2}$$

Comparing the expressions (A1) and (A2) for the transformed covariance matrices with the corresponding expressions (14) and (13) in the predictable pattern basis, we obtain for the weight vector matrix

$$\mathbf{U}_v = \mathbf{W}_v \Lambda_v^{-1/2} \mathbf{T}_v. \tag{A3}$$

From the biorthogonality condition (9) follows

$$\mathbf{V}_v = \mathbf{W}_v \Lambda_v^{1/2} \mathbf{T}_v \tag{A4}$$

for the matrix  $\mathbf{V}_v$  with predictable patterns as columns. The implementation of this algorithm can be checked for consistency by verifying that the weight vector matrix  $\mathbf{U}_v$  and the predictable pattern matrix  $\mathbf{V}_v$  satisfy the completeness and biorthogonality conditions (9).

If the climatological covariance matrix  $\Sigma_v$  is singular, one or more of the eigenvalues  $\lambda_v^k$  in  $\Lambda_v = \text{Diag}(\lambda_v^k)$  is zero and  $\Lambda_v^{-1/2} = \text{Diag}[(\lambda_v^k)^{-1/2}]$  does not exist. Regularization by truncated principal component analysis proceeds by setting to zero both the eigenvalues  $\lambda_v^k$  not significantly different from zero and the square root  $(\lambda_v^k)^{-1/2}$  of their inverses. Zeroing these contributions to the inverse climatological covariance matrix  $\Sigma_v^{-1} = \mathbf{W}_v \Lambda_v^{-1} \mathbf{W}_v^T$  amounts to replacing the ill-defined inverse  $\Sigma_v^{-1}$  by a Moore–Penrose pseudoinverse (see, e.g., Golub and van Loan 1993, chapter 5). If, after regularization, only the first  $r$  of the  $m$  eigenvalues  $\lambda_v^k$  are nonzero, the predictable component analysis is computed in the truncated  $r$ -dimensional state space.

Computing the predictable component analysis by a principal component analysis of the climatological covariance matrix  $\Sigma_v$  followed by a principal component analysis of the prediction error covariance matrix  $\mathbf{K}_v$  has several advantages: besides the predictable patterns  $\mathbf{V}_v$ , it produces the EOFs  $\mathbf{W}_v$ ; and the predictable component analysis can be regularized by truncating the principal component analysis of the climatological covariance matrix. However, when no regularization needs to be performed, it is numerically more efficient to replace the eigendecomposition  $\Sigma_v = \mathbf{W}_v \Lambda_v \mathbf{W}_v^T$  of the clima-

tological covariance matrix by a Cholesky decomposition  $\Sigma_\nu = \mathbf{L}_\nu \mathbf{L}_\nu^T$  and to use the Cholesky factor  $\mathbf{L}_\nu$  in place of  $\mathbf{W}_\nu \Lambda_\nu^{1/2}$  and  $\mathbf{L}_\nu^{-T}$  in place of  $\mathbf{W}_\nu \Lambda_\nu^{-1/2}$  (Press et al. 1992, 455).

In predictability studies of the first kind, the climatological covariance matrix  $\Sigma_\nu$  usually does not depend on the forecast lead time  $\nu$ , so that the whitening transformation (18) with the matrix  $\Lambda_\nu^{-1/2} \mathbf{W}_\nu^T$  or with the Cholesky factor  $\mathbf{L}_\nu^{-1}$  need be computed only once. Only the eigendecomposition  $\mathbf{T}_\nu$  of the transformed prediction error covariance matrix  $\mathbf{K}_\nu$  must be computed for each forecast lead time  $\nu$ . In predictability studies of the second kind, all of the above transformations must be computed for each time  $\nu$  for which a prediction is made.

## REFERENCES

- Ahmad, T. A., and P.-E. Lin, 1976: A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Trans. Inf. Theory*, **22**, 372–375.
- Anderson, J. L., and W. F. Stern, 1996: Evaluating the potential predictive utility of ensemble forecasts. *J. Climate*, **9**, 260–269.
- Anderson, T. W., 1984: *An Introduction to Multivariate Statistical Analysis*. 2d ed. Series in Probability and Mathematical Statistics, Wiley, 675 pp.
- Ansley, C. F., and R. Kohn, 1983: Exact likelihood of a vector autoregressive moving average process with missing or aggregated data. *Biometrika*, **70**, 275–278.
- , and —, 1986: A note on reparameterising a vector autoregressive moving average model to enforce stationarity. *J. Stat. Comput. Sim.*, **24**, 99–106.
- Bell, T. L., 1982: Optimal weighting of data to detect climatic change: Application to the carbon dioxide problem. *J. Geophys. Res.*, **87**, 11 161–11 170.
- , 1986: Theory of optimal weighting of data to detect climatic change. *J. Atmos. Sci.*, **43**, 1694–1710.
- Box, G. E. P., and G. C. Tiao, 1977: A canonical analysis of multiple time series. *Biometrika*, **64**, 355–365.
- Brillouin, L., 1956: *Science and Information Theory*. Academic Press, 320 pp.
- Brockwell, P. J., and A. Davis, 1991: *Time Series: Theory and Methods*. 2d ed. Springer, 577 pp.
- Cheng, Y.-Q., Y.-M. Zhuang, and J.-Y. Yang, 1992: Optimal Fisher discriminant analysis using the rank decomposition. *Pattern Recognit.*, **25**, 101–111.
- Delworth, T., S. Manabe, and R. J. Stouffer, 1993: Interdecadal variations of the thermohaline circulation in a coupled ocean–atmosphere model. *J. Climate*, **6**, 1993–2011.
- , —, and —, 1997: Multidecadal climate variability in the Greenland Seas and surrounding regions: A coupled model simulation. *Geophys. Res. Lett.*, **24**, 257–260.
- Dmitriev, Y. G., and F. P. Tarasenko, 1973: On the estimation of functionals of the probability density and its derivatives. *Theory Probab. Appl.*, **18**, 628–633.
- Draper, D., 1995: Assessment and propagation of model uncertainty. *J. Roy. Stat. Soc. B*, **57**, 45–97.
- Engl, H. W., M. Hanke, and A. Neubauer, 1996: *Regularization of Inverse Problems*. Kluwer, 321 pp.
- Friedman, J. H., 1989: Regularized discriminant analysis. *J. Amer. Stat. Assoc.*, **84**, 165–175.
- Fukunaga, K., 1990: *Introduction to Statistical Pattern Recognition*. 2d ed. Academic Press, 591 pp.
- Golub, G. H., and C. F. van Loan, 1993: *Matrix Computations*. 2d ed. Johns Hopkins University Press, 642 pp.
- Griffies, S. M., and K. Bryan, 1997a: Predictability of North Atlantic multidecadal climate variability. *Science*, **275**, 181–184.
- , and —, 1997b: A predictability study of simulated North Atlantic multidecadal variability. *Climate Dyn.*, **8**, 459–488.
- Hall, P., and S. C. Morton, 1993: On the estimation of entropy. *Ann. Inst. Stat. Math.*, **45**, 69–88.
- Halliwel, G. R., 1997: Decadal and multidecadal North Atlantic SST anomalies driven by standing and propagating basin-scale atmospheric anomalies. *J. Climate*, **10**, 2405–2411.
- , 1998: Simulation of North Atlantic decadal/multidecadal winter SST anomalies driven by basin-scale atmospheric circulation anomalies. *J. Phys. Oceanogr.*, **28**, 5–21.
- Hansen, P. C., 1997: *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM Monogr. on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics, 247 pp.
- Harzallah, A., and R. Sadourny, 1995: Internal versus SST-forced atmospheric variability as simulated by an atmospheric circulation model. *J. Climate*, **8**, 474–495.
- Hasselmann, K., 1993: Optimal fingerprints for the detection of time-dependent climate change. *J. Climate*, **6**, 1957–1971.
- Hayashi, Y., 1986: Statistical interpretation of ensemble-time mean predictability. *J. Meteor. Soc. Japan*, **64**, 167–181.
- Hegerl, G. C., and G. R. North, 1997: Comparison of statistically optimal approaches to detecting anthropogenic climate change. *J. Climate*, **10**, 1125–1133.
- Joe, H., 1989: Estimation of entropy and other functionals of a multivariate density. *Ann. Inst. Stat. Math.*, **41**, 683–697.
- Johnson, R. A., and D. W. Wichern, 1982: *Applied Multivariate Statistical Analysis*. Prentice-Hall, 594 pp.
- Jolliffe, I. T., 1986: *Principal Component Analysis*. Springer Series in Statistics, Springer, 271 pp.
- Krzanowski, W. J., P. Jonathan, W. V. McCarthy, and M. R. Thomas, 1995: Discriminant analysis with singular covariance matrices: Methods and applications to spectroscopic data. *Appl. Stat.*, **44**, 101–115.
- Lawley, D. N., 1956: Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128–136.
- Li, W. K., and A. I. McLeod, 1981: Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Stat. Soc. B*, **43**, 231–239.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.
- , 1975: Climatic predictability. *The Physical Basis of Climate and Climate Modelling*, B. Bolin et al., Eds., GARP Publication Series, Vol. 16, World Meteorological Organization, 132–136.
- Lütkepohl, H., 1985: Comparison of criteria for estimating the order of a vector autoregressive process. *J. Time Ser. Anal.*, **6**, 35–52; Correction, **8**, 373.
- , 1993: *Introduction to Multiple Time Series Analysis*. 2d ed. Springer-Verlag, 545 pp.
- McLachlan, G. J., 1992: *Discriminant Analysis and Statistical Pattern Recognition*. Series in Probability and Mathematical Statistics, Wiley, 544 pp.
- Miller, A. J., 1984: Selection of subsets of regression variables. *J. Roy. Stat. Soc. A*, **147**, 389–425.
- Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **114**, 463–493.
- Neumaier, A., 1998: Solving ill-conditioned and singular linear systems: A tutorial on regularization. *SIAM Rev.*, **40**, 636–666.
- , and T. Schneider, cited 1997: Multivariate autoregressive and Ornstein–Uhlenbeck processes: Estimates for order, parameters, spectral information, and confidence regions. [Available online at <http://www.aos.Princeton.EDU/WWWPUBLIC/tapio/arfit/>]
- Palmer, T. N., 1996: Predictability of the atmosphere and oceans: From days to decades. *Decadal Climate Variability: Dynamics and Predictability*, D. L. T. Anderson and J. Willebrand, Eds., NATO ASI Series, Vol. I 44, Springer, 83–155.
- , R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653.
- Papoulis, A., 1991: *Probability, Random Variables, and Stochastic Processes*. 3d ed. McGraw-Hill, 666 pp.

- Prakasa Rao, B. L. S., 1983: *Nonparametric Functional Estimation*. Series in Probability and Mathematical Statistics, Academic Press, 522 pp.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes*. 2d ed. Cambridge University Press, 963 pp.
- Ripley, B. D., 1996: *Pattern Recognition and Neural Networks*. Cambridge University Press, 403 pp.
- Schneider, T., and A. Neumaier, cited 1997: Algorithm: ARfit—A Matlab package for estimation and spectral decomposition of multivariate autoregressive processes. [Available online at <http://www.aos.Princeton.EDU/WWWPUBLIC/tapio/arfit/>.]
- Schwarz, G., 1978: Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464.
- Scott, D. W., 1992: *Multivariate Density Estimation: Theory, Practice, and Visualization*. Series in Probability and Mathematical Statistics, Wiley, 317 pp.
- Shannon, C. E., 1948: A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 370–423, 623–656.
- , and W. Weaver, 1949: *The Mathematical Theory of Communication*. University of Illinois Press, 117 pp.
- Shukla, J., 1981: Dynamical predictability of monthly means. *J. Atmos. Sci.*, **38**, 2547–2572.
- , 1985: Predictability. *Advances in Geophysics*, Vol. 28b, Academic Press, 87–122.
- Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 175 pp.
- Stern, W. F., and K. Miyakoda, 1995: The feasibility of seasonal forecasts inferred from multiple GCM simulations. *J. Climate*, **8**, 1071–1085.
- Thacker, W. C., 1996: Metric-based principal components: Data uncertainties. *Tellus*, **48A**, 584–592.
- Tiao, G. C., and G. E. P. Box, 1981: Modeling multiple time series with applications. *J. Amer. Stat. Assoc.*, **76**, 802–816.
- Tikhonov, A. N., and V. Y. Arsenin, 1977: *Solution of Ill-Posed Problems*. Scripta Series in Mathematics, V. H. Winston and Sons, 258 pp.
- Toth, Z., 1991: Circulation patterns in phase space: A multinormal distribution? *Mon. Wea. Rev.*, **119**, 1501–1511.
- Zwiers, F. W., 1996: Interannual variability and predictability in an ensemble of AMIP climate simulations conducted with the CCC GCM2. *Climate Dyn.*, **12**, 825–848.