

## Average Predictability Time. Part I: Theory

TIMOTHY DELSOLE

*George Mason University, Fairfax, Virginia, and Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland*

MICHAEL K. TIPPETT

*International Research Institute for Climate and Society, Palisades, New York*

(Manuscript received 27 June 2008, in final form 14 October 2008)

### ABSTRACT

This paper introduces the average predictability time (APT) for characterizing the overall predictability of a system. APT is the integral of a predictability measure over all lead times. The underlying predictability measure is based on the Mahalanobis metric, which is invariant to linear transformation of the prediction variables and hence gives results that are independent of the (arbitrary) basis set used to represent the state. The APT is superior to some integral time scales used to characterize the time scale of a random process because the latter vanishes in situations when it should not, whereas the APT converges to reasonable values. The APT also can be written in terms of the power spectrum, thereby clarifying the connection between predictability and the power spectrum. In essence, predictability is related to the width of spectral peaks, with strong, narrow peaks associated with high predictability and nearly flat spectra associated with low predictability. Closed form expressions for the APT for linear stochastic models are derived. For a given dynamical operator, the stochastic forcing that minimizes APT is one that allows transformation of the original stochastic model into a set of uncoupled, independent stochastic models. Loosely speaking, coupling enhances predictability. A rigorous upper bound on the predictability of linear stochastic models is derived, which clarifies the connection between predictability at short and long lead times, as well as the choice of norm for measuring error growth. Surprisingly, APT can itself be interpreted as the “total variance” of an alternative stochastic model, which means that generalized stability theory and dynamical systems theory can be used to understand APT. The APT can be decomposed into an uncorrelated set of components that maximize predictability time, analogous to the way principle component analysis decomposes variance. Part II of this paper develops a practical method for performing this decomposition and applies it to meteorological data.

### 1. Introduction

This paper proposes a new measure of the overall predictability of a system, independent of lead time. Such a measure allows one system to be characterized as “more predictable” than another. One motivation for this new measure is that traditional measures of the overall predictability of a system often have no obvious generalization to multivariate systems. For instance, the limit of predictability, as defined by Lorenz (1969), is the time beyond which the mean square error exceeds a predefined threshold. However, mean square error de-

pends on the coordinate system used to represent the state and is problematic if different variables with different units and natural variances are mixed. Alternatively, an integral time scale, defined as the integral of some moment of the autocorrelation function with respect to lag, is used often in turbulence studies. However, the autocorrelation function is even more restrictive than mean square error in that it is defined only for a single time series. The predictability time scale of a process is sometimes suggested to be related to the peak of its power spectrum. As an example, ENSO has a spectral peak around 4 yr (Julian and Chervin 1978), which under this assumption implies predictability for about 4 yr. However, real-time forecasts of ENSO demonstrate little skill beyond 1 yr, suggesting an inconsistency. Moreover, even if a reasonable relation between predictability and power spectra could be ascertained for

---

*Corresponding author address:* Timothy DelSole, Center for Ocean–Land–Atmosphere Studies, 4041 Powder Mill Rd., Suite 302, Calverton, MD 20705.  
E-mail: delsole@cola.iges.org

the univariate case, the generalization to multivariate systems would not necessarily be straightforward.

Our proposal for overcoming these limitations has two key elements. The first is adoption of a measure of predictability that is applicable to multivariate systems and invariant to linear transformation of the variables—measures satisfying these properties have been proposed by Leung and North (1990), Schneider and Griffies (1999), and Kleeman (2002) (see DelSole and Tippett 2007 for a review)—and the second is the integration of this measure of predictability with respect to all lead times. The resulting measure has several attractive properties. First, the measure is consistent, in the sense that applying it to the same system represented using a different basis set gives the same result. Second, if the predictability decay follows the same form in different systems, then systems that are predictable at longer lead times will have a larger measure. We restrict our attention to predictability averaged over all initial conditions and call the resulting integral the average predictability time (APT). The APT may not be interpretable as a time scale if the decay of predictability is pathological, but nonetheless the integral still gives useful information about overall predictability.

When only predictability averaged over initial conditions is considered, two natural measures satisfying the above properties are mutual information and the Mahalanobis metric. Unfortunately, mutual information depends on the complete forecast and climatological distributions and is therefore difficult to estimate from finite time series. Even in the case of normally distributed variables, mutual information is difficult to integrate with respect to lead time. In contrast, the Mahalanobis metric depends only on second-order moments and is easier to integrate. Furthermore, it turns out that choosing the Mahalanobis metric as a basis for APT has several other attractive properties. Specifically, the resulting APT can be expressed in terms of power spectra, thus clarifying the connection between predictability and power spectra. This point is demonstrated in the present paper. In addition, the resulting APT can be decomposed into independent components that optimize it, analogous to the way that principal component analysis decomposes variance. This decomposition clarifies the usefulness of APT even if the system is characterized by a wide range of time scales—the decomposition separates different components according to their APT, allowing the full spectrum of APTs to be diagnosed. This decomposition will be derived and illustrated with meteorological data in Delsole and Tippett (2009, hereafter Part II).

The present paper may be summarized briefly as follows: In section 2, we review the measure on which

the APT will be based, namely the Mahalanobis metric. We then show in section 3 that the resulting APT gives sensible results for one-dimensional stochastic models; that is, the APT is inversely proportional to the damping rate and comparable to traditional integral time scales in appropriate cases. On the other hand, we show that the APT gives clearly superior results compared to a widely used integral time scale. Next we discuss the relation between predictability and power spectra. Specifically, we show that the APT equals the integral of the squared modulus of the power spectrum and explain how APT depends on the shape of the power spectrum. In section 6 we evaluate the APT for autonomous, linear stochastic models excited by Gaussian white noise. In addition, we show how the multivariate expressions parallel the univariate results if the variables are transformed appropriately. In section 7, we show that the APT for a set of independent, uncoupled stochastic models equals the average APT of the individual systems. In section 8 we derive the relation between APT and multivariate power spectra and show that the minimum APT occurs when the system is a set of independent, uncoupled white noise processes, consistent with intuition. Bounds on the APT of linear stochastic models are derived in section 9 and a surprising interpretation of APT is discussed in section 10. The bounds are illustrated in section 11 with a simple stochastic model. We conclude with a summary and discussion of results.

## 2. Definition of average predictability time

Consider a system whose state vector is the  $K$ -dimensional vector  $\mathbf{x}$ . Let the forecast distribution at a fixed lead time  $\tau$  have covariance matrix  $\Sigma_\tau$ . This covariance matrix quantifies the uncertainty in the forecast as a function of lead time. If the system is stationary and the forecast is independent of the initial condition at asymptotically long lead times, then the climatological distribution can be identified with the forecast distribution in the limit of large lead time (DelSole and Tippett 2007), namely  $\Sigma_\infty$ .

As discussed in the introduction, the first key to defining APT is to choose a measure of predictability that is invariant to nonsingular linear transformation. Although several such measures exist, one turns out to have certain advantages compared to the others. Specifically, we adopt a measure of predictability called the Mahalanobis signal (DelSole and Tippett 2007), defined as

$$S_\tau = \frac{1}{K} \text{tr}[(\Sigma_\infty - \Sigma_\tau)\Sigma_\infty^{-1}], \quad (1)$$

where “tr” denotes matrix trace. The Mahalanobis signal can be interpreted as 1 minus the mean square error, but with the error measured in a space in which the climatological covariance matrix is the identity matrix [DelSole and Tippett (2007) and Schneider and Griffies (1999) effectively show this too]. The Mahalanobis signal equals 1 for a perfect forecast and vanishes when  $\Sigma_\tau = \Sigma_\infty$ ; that is, it vanishes for a forecast that is no better than a randomly drawn state of the system. The Mahalanobis signal is appropriate only if forecast uncertainty is well characterized by second-moment statistics. As we shall see later, the factor  $1/K$  is used to ensure that a collection of independent, identical systems has the same APT as any subsystem. Schneider and Griffies (1999) note that this factor also allows the predictability of random processes with different dimensions to be compared. Given the above metric, we define APT by

$$S = 2 \int_0^\infty S_\tau \, d\tau. \tag{2}$$

The factor of 2 is introduced to ensure that APT corresponds to an  $e$ -folding time if the autocorrelation is exponential (see section 3). For discrete time, we define APT as

$$S = 2 \sum_{\tau=1}^\infty S_\tau, \tag{3}$$

where the summation starts at time step  $\tau = 1$ .

### 3. One-dimensional case

To illustrate the reasonableness of the above definitions, consider the one-dimensional stochastic model

$$\frac{dx}{dt} = ax + w, \tag{4}$$

where  $a$  is a negative number and  $w$  is a Gaussian white noise process with zero mean and time-lagged covariance given by

$$\langle w(t + \tau)w(t) \rangle = \sigma_w^2 \delta(\tau), \tag{5}$$

where the angle brackets denote an ensemble average and  $\delta$  denotes the Dirac delta function. The solution to (4) at time  $\tau$  for a particular realization of the forcing can be solved by elementary methods as

$$x_\tau = e^{a\tau} x_0 + e^{a\tau} \int_0^\tau e^{-as} w(s) ds. \tag{6}$$

Taking the mean of an ensemble of solutions with mean initial condition mean  $\mu_0$  gives

$$\langle x_\tau \rangle = e^{a\tau} \mu_0. \tag{7}$$

It follows that the mean of the ensemble decays exponentially with a decay rate determined by  $a$ . The total variance of the ensemble is

$$\begin{aligned} \sigma_\tau^2 &= \langle (x_\tau - e^{a\tau} \mu_0)^2 \rangle \\ &= e^{2a\tau} \langle x_0 - \mu_0 \rangle^2 + e^{2a\tau} \int_0^\tau \int_0^\tau e^{-as} \langle w(s)w(s') \rangle e^{-as'} ds ds' \\ &= e^{2a\tau} \sigma_0^2 + e^{2a\tau} \int_0^\tau e^{-2as} ds \\ &= e^{2a\tau} \sigma_0^2 - \frac{1 - e^{2a\tau}}{2a} \sigma_w^2, \end{aligned} \tag{8}$$

where the initial ensemble and stochastic forcing are assumed to be independent. For simplicity, we assume the initial variance vanishes; that is,  $\sigma_0^2 = 0$ . In this case, the only source of uncertainty is the stochastic forcing, and the variance of the forecast ensemble given by (8) is

$$\sigma_\tau^2 = (1 - e^{2a\tau}) \sigma_\infty^2, \tag{9}$$

where we have used  $\sigma_\infty^2 = -\sigma_w^2/(2a)$  from (8). If the initial condition is not perfect (i.e.,  $\sigma_0 \neq 0$ ), then the predictability is reduced relative to the perfect initial condition case.

The predictability of the system is computed by substituting (9) into the predictability measure (1), which gives

$$S_\tau = (\sigma_\infty^2 - \sigma_\tau^2)(\sigma_\infty^2)^{-1} = e^{2a\tau}. \tag{10}$$

This result shows that the predictability decays monotonically with lead time  $\tau$ . The APT is found by integrating this expression with respect to lead time and multiplying by 2. The result is

$$S = 2 \int_0^\infty e^{2at} dt = -\frac{1}{a}. \tag{11}$$

This result shows that APT is inversely related to the decay rate, consistent with physical intuition—systems with strong damping have less “memory” and thus less predictability.

### 4. Connection between APT and autocorrelation

Intuitively, processes with large autocorrelations are more predictable than processes with small autocorrelations. To formalize this intuition, the relation between forecast uncertainty and time-lagged covariance needs to be defined. For linear prediction models of stationary processes, this relation can be computed readily. Consider the linear prediction model

$$\hat{x}_\tau = L_\tau x_0 + b. \tag{12}$$

Standard regression yields the parameters that minimize the mean square prediction error as

$$L_\tau = c_\tau/c_0, \quad b = \mu - L_\tau \mu \quad \tau > 0, \tag{13}$$

where the time-lagged covariance  $c_\tau$  is defined as

$$c_\tau = \langle (x_\tau - \mu)(x_0 - \mu) \rangle \tag{14}$$

and  $\mu$  is the stationary mean. The forecast error variance of this model is thus

$$\begin{aligned} \sigma_\tau^2 &= \langle (x_\tau - L_\tau x_0 - b)^2 \rangle \\ &= \langle (x_\tau - \mu)^2 \rangle - 2L_\tau \langle (x_\tau - \mu)(x_0 - \mu) \rangle + L_\tau^2 \langle (x_0 - \mu)^2 \rangle \\ &= c_0 - 2L_\tau c_\tau + L_\tau^2 c_0 = c_0 - c_\tau^2/c_0. \end{aligned} \tag{15}$$

For stationary processes, the autocorrelation is  $\rho_\tau = c_\tau/c_0$  and  $c_0 = \sigma_\infty^2$ , in which case the autocorrelation can be written in terms of the forecast covariance as

$$\sigma_\tau^2 = \sigma_\infty^2 (1 - \rho_\tau^2). \tag{16}$$

Substituting this expression into the expressions for APT gives

$$S = 2 \int_0^\infty S_\tau d\tau = 2 \int_0^\infty (1 - \sigma_\tau^2/\sigma_\infty^2) d\tau = 2 \int_0^\infty \rho_\tau^2 d\tau. \tag{17}$$

We see that the APT of a one-dimensional regression model is twice the integral of the *square* of the autocorrelation function. This function was proposed by DelSole (2001) as a measure of the time scale of a process; it emerges here as a measure of overall predictability. As a check, note that the time-lagged covariance for the stochastic model (4) is

$$c_\tau = \langle (x_\tau - e^{a\tau} \mu_0)(x_0 - \mu_0) \rangle = e^{a\tau} \sigma_\infty^2, \quad \tau > 0, \tag{18}$$

which implies  $\rho_\tau = \exp(a\tau)$ . Substituting this into (17) gives  $S = -1/a$ , consistent with (10).

Having related the APT of a linear regression model to the autocorrelation function of the random process, we may now choose any permissible autocorrelation function for demonstration purposes. Accordingly, consider the damped oscillation function

$$\rho_\tau = e^{a\tau} \cos(\omega_0 \tau), \quad \tau > 0, \tag{19}$$

which is a permissible autocorrelation function for a stationary process (Gelb 1974, p. 82). The APT of a regression model for this process is

$$S = 2 \int_0^\infty \rho_\tau^2 d\tau = \frac{1}{2a} \left( 1 + \frac{a^2}{a^2 + \omega_0^2} \right). \tag{20}$$

The limit  $\omega_0 \rightarrow 0$  recovers the one-dimensional case  $S = -1/a$ . However, the limit  $\omega_0 \rightarrow \infty$  gives  $S = -1/2a$ , or half the APT of the case with no oscillations. It is instructive to compare the APT with the familiar integral time scale

$$T_1 = \int_0^\infty \rho_\tau d\tau. \tag{21}$$

For the damped oscillation (19), this integral time scale is

$$T_1 = -\frac{a}{a^2 + \omega_0^2}. \tag{22}$$

If  $\omega_0 = 0$ , then the integral time scale  $T_1$  coincides with the APT given in (11). This shows that the APT is consistent with the integral time scale  $T_1$  for random processes with nonoscillating autocorrelation functions. However, if  $\omega_0 \neq 0$ , then the integral time scale is less than the APT, and in fact  $T_1 = 0$  in the limit  $\omega_0 \rightarrow \infty$  while  $S = -1/2a$ . Thus, the APT and integral time scale  $T_1$  have dramatically different dependence on the oscillation frequency. To illustrate this fact, consider the autocorrelations for select values of  $a$  shown on the right-hand side of Fig. 1. As  $\omega_0$  increases, the frequency of oscillation increases, but the bounding envelope remains the same. In the case  $a = -1$  and  $\omega_0 = 4$ , the integral time scale  $T_1$  evaluated from (22) is  $1/17$ , but strong correlations persist well beyond this value. It is evident, then, that the integral time scale defined in (21) is not appropriate for oscillatory correlation functions because the oscillations lead to cancellations in the integral. In contrast, the APT gives a more appropriate estimate of the time scale of a process.

### 5. Connection between APT and power spectrum

We now derive the relation between APT and power spectra. Certain basic points about the relation between predictability and the power spectrum can be inferred by considering two extreme cases. First, white noise can be considered to be the least predictable process because its value at one time is completely independent of its value at any other time. The power spectrum of white noise is constant, or “flat.” Conversely, consider a perfect sine wave. A sine wave is perfectly predictable because the value at a finite number of times can be used to specify the value at all other times. The power spectrum of a sine wave is a delta function. These considerations suggest that strongly peaked power spectra correspond to highly predictable time series, whereas

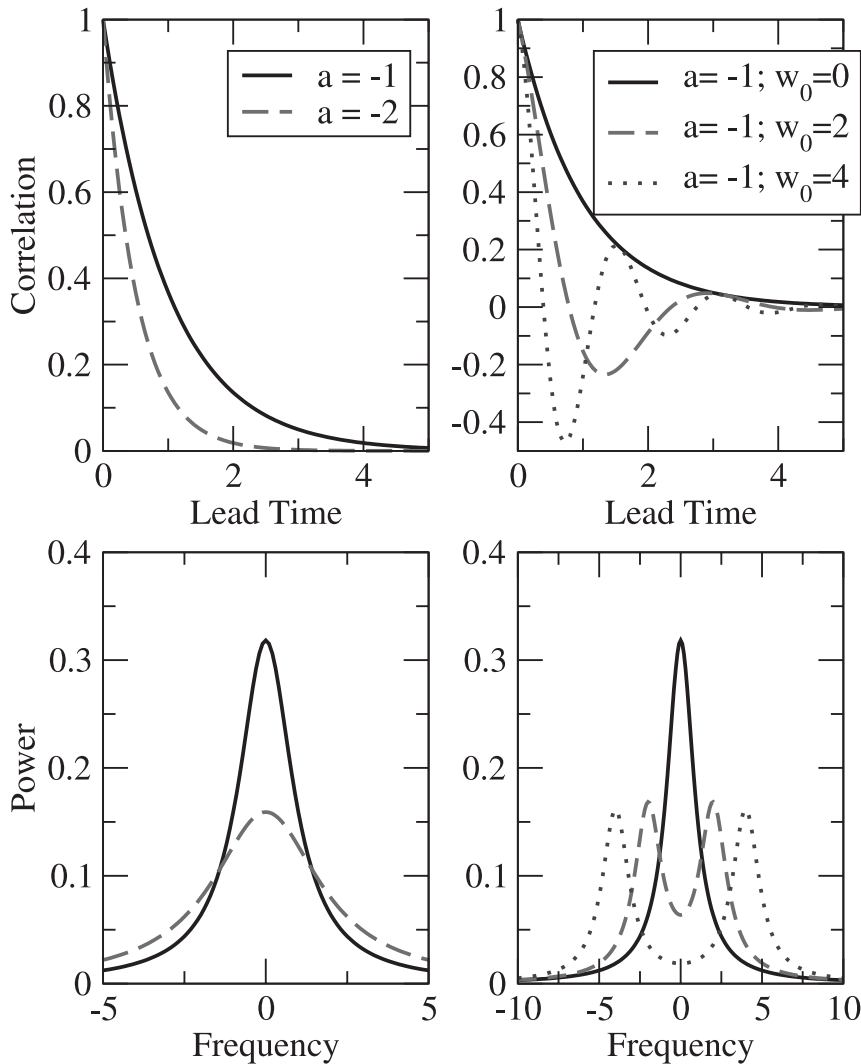


FIG. 1. (top) Autocorrelation function and (bottom) corresponding power spectrum of the damped oscillation function (19), for the values of the parameters indicated in the figures.

flat or near-flat power spectra correspond to weakly predictable time series.

The above arguments suggest that predictability is related to the degree of flatness or “peakiness” of power spectra, but they are not precise enough to suggest a quantitative relation. We show in this section what properties of power spectra influence predictability as measured by APT. As is well known, the power spectrum  $p_\omega$  of a stationary process equals the Fourier transform of the time-lagged covariance  $c_\tau$ ; specifically,

$$c_\tau = \int_{-\infty}^{\infty} p_\omega e^{i\omega\tau} d\omega \quad \text{or} \quad p_\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} c_\tau e^{-i\omega\tau} d\tau. \quad (23)$$

Substituting the latter relation into the APT (17) gives

$$\begin{aligned} S &= 2 \int_0^\infty \rho_\tau^2 d\tau = \frac{1}{\sigma_\infty^4} \int_{-\infty}^\infty c_\tau c_\tau^* d\tau \\ &= \frac{1}{\sigma_\infty^4} \int_{-\infty}^\infty \left( \int_{-\infty}^\infty p_\omega e^{i\omega\tau} d\omega \right) \left( \int_{-\infty}^\infty p_{\omega'}^* e^{-i\omega'\tau} d\omega' \right) d\tau \\ &= \frac{1}{\sigma_\infty^4} \int_{-\infty}^\infty \int_{-\infty}^\infty \left( \int_{-\infty}^\infty e^{i\tau(\omega-\omega')} d\tau \right) p_\omega p_{\omega'}^* d\omega d\omega' \\ &= 2\pi \int_{-\infty}^\infty \frac{p_\omega p_\omega^*}{\sigma_\infty^2 \sigma_\infty^2} d\omega, \end{aligned} \quad (24)$$

where we have used the relation

$$\int_{-\infty}^\infty e^{i\tau(\omega-\omega')} d\tau = 2\pi\delta(\omega-\omega'). \quad (25)$$



We call  $p_w/\sigma_\infty^2$  the whitened power spectrum. By definition, the whitened power spectrum has a unit integral. The above result shows that the APT for a linear regression model is proportional to the integral of the squared modulus of the power spectrum. To see how this relation relates to the shape of the power spectrum, consider the identity

$$\overline{p^2} = \overline{(p - \bar{p})^2} + (\bar{p})^2, \tag{26}$$

where overbars denote the mean value over any chosen set of frequencies. The first term on the right-hand side measures the deviations about the mean value. The closer the spectrum is to its mean value (on average), the smaller the first term. A spectrum can be “flattened,” and hence rendered less predictable, by setting the spectrum equal to its mean value over any set of bands; that is, by setting  $p = \bar{p}$ . This flattening process preserves the mean spectrum over the band and therefore preserves the total integral over all frequencies; it also produces positive power spectra because the mean spectrum can never be negative. These considerations suggest that the integral of the squared whitened spectrum is a measure of the degree to which the spectrum differs from a constant—that is, APT is a measure of the peakiness of a power spectrum. The closer the spectrum is to its mean value, in the sense that the first term in (26) is small, and the “flatter” it is, the more closely it approximates white noise and the less predictable the corresponding random process. Conversely, the more the spectrum deviates from the mean, in an average square sense, the stronger the peakiness and the more predictable the corresponding process.

It is perhaps worth mentioning that the above definition of peakiness is not the only one possible. For instance, the Burg entropy is defined as the integral of the log of the power spectrum, and it is known (Priestley 1981, p. 604) that its minimum value, for all power spectra with the same total power, is achieved when the spectrum is a constant (i.e., flat). We propose (24) as a measure of flatness because it has certain practical and theoretical advantages compared to other definitions; that is, it can be estimated from only second-order moments and can be decomposed into components ordered by their APT (the latter point will be demonstrated in Part II).

As an illustration of the above results, consider again the autocorrelation (19) for a damped oscillation. The corresponding power spectrum is simply the Fourier transform of (19), which is

$$p_\omega = -\frac{1}{4\pi} \left[ \frac{a}{(\omega - \omega_0)^2 + a^2} + \frac{a}{(\omega + \omega_0)^2 + a^2} \right] \tag{27}$$

(we have used the fact that for any real stationary process  $c_{-\tau} = c_\tau$ ). The power spectrum and autocorrelation function for various choices of  $a$  and  $\omega_0$  are illustrated in Fig. 1. As expected, the autocorrelation function generally increases as  $a$  approaches zero (for fixed  $\tau$ ) and the peak of the power spectrum becomes sharper (i.e., less flat) as  $a$  approaches zero. Results for different values of  $\omega_0$  are shown in the right panel, but inferences of APT given the power spectrum are not as straightforward (e.g., they require visually integrating the square of the power).

These considerations clarify the error in relating predictability to the *location* of spectral peaks. Returning to the ENSO example mentioned in the introduction, we note that the Niño-3 index is tolerably fit by the autocorrelation function (19) with  $a^{-1} = 16$  months and  $\omega_0^{-1} = 48/(2\pi)$  months, as shown in Fig. 2 (We do not recommend this fitting procedure in general; it is used here primarily for illustration purposes.). These parameter values give  $S = 9.5$  months, which differs considerably from the period of the spectral peak, which is  $2\pi/\omega_0 = 48$  months. In the limit of large  $\omega_0$ , there is only a modest effect on  $S$ ,  $S \rightarrow -1/(2a) = 8$ . Thus, the location of the peak is almost irrelevant to the predictability time scale. The reason is that predictability depends on the peakiness of the power spectrum, which is more strongly controlled by the damping time  $-1/a$ , rather than on the location of the peak itself ( $\omega_0$ ).

### 6. Multivariate linear stochastic models

In this section we derive the APT for linear stochastic models of the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{28}$$

where  $\mathbf{A}$  is a  $K \times K$  matrix, called the dynamical operator, and  $\mathbf{w}$  is a Gaussian white noise process with zero mean and covariance matrix  $\mathbf{Q}$ . The dynamical operator is assumed to be independent of time and stable—that is, it possesses  $K$  distinct eigenvalues with negative real parts. Let  $\lambda_k(\mathbf{A})$  denote the  $k$ th eigenvalue of  $\mathbf{A}$ . Also, let the eigenvector decomposition of  $\mathbf{A}$  be

$$\mathbf{A} = \mathbf{Z}\mathbf{\Lambda}\mathbf{Z}^{-1}, \tag{29}$$

where the columns of  $\mathbf{Z}$  are the eigenvectors of  $\mathbf{A}$ , and  $\mathbf{\Lambda}$  is a diagonal matrix whose diagonal elements are the eigenvalues of  $\mathbf{A}$ . As in the one-dimensional case, assume that the initial ensemble has zero variance. Tippett and Chang (2003) show that the ensemble of solutions to the stochastic model (28) has covariance matrix

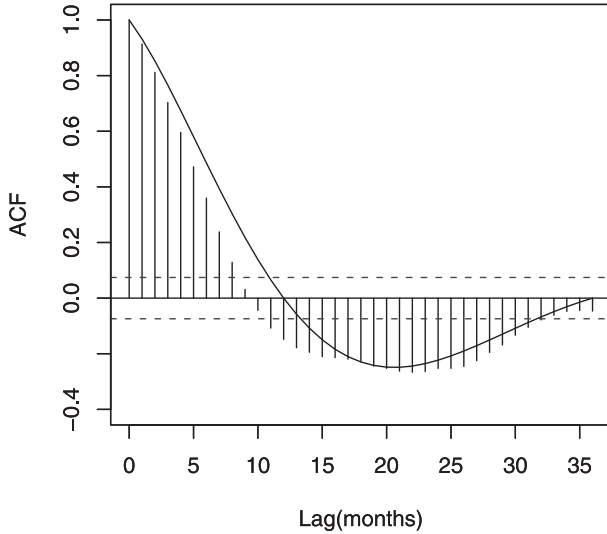


FIG. 2. Autocorrelation function of the Niño-3 index (downloaded from <http://www.cpc.noaa.gov/data/indices/>) during the period 1950–2007 (histogram), and a fit of the autocorrelation to (19) (solid line). The horizontal dashes indicate the 5% significance thresholds of the correlation coefficient.

$$\Sigma_\tau = \mathbf{Z}(\mathbf{Z}^{-1}\mathbf{Q}\mathbf{Z}^{-H} \circ \mathbf{E}_\tau)\mathbf{Z}^H, \quad (30)$$

where superscript  $H$  denotes the conjugate transpose,  $-H$  denotes the inverse of the conjugate transpose,

$$(\mathbf{E}_\tau)_{kl} = \frac{e^{\lambda_k(\mathbf{A})\tau + \lambda_l^*(\mathbf{A})\tau} - 1}{\lambda_k(\mathbf{A}) + \lambda_l^*(\mathbf{A})}, \quad (31)$$

and  $\mathbf{C} \circ \mathbf{D}$  denotes the Hadamard product of  $\mathbf{C}$  and  $\mathbf{D}$ ; that is,  $(\mathbf{C} \circ \mathbf{D})_{ij} = \mathbf{C}_{ij}\mathbf{D}_{ij}$ . [This result also is derived in DelSole and Tippett (2007); the stationary form of this solution can be found in Horn and Johnson (1999, p. 301)]. The solution (30) is the multivariate generalization of (8). A further standard result is that the forecast covariance matrix (30) also satisfies

$$\Sigma_\tau = \Sigma_\infty - e^{\mathbf{A}\tau}\Sigma_\infty e^{\mathbf{A}^H\tau}, \quad (32)$$

which is the multivariate generalization of (9).

Analogous to the one-dimensional case, the climatological covariance matrix is identified with the asymptotic forecast covariance matrix  $\Sigma_\infty$ . Substituting (32) into the predictability measure (1) gives

$$S_\tau = \frac{1}{K} \text{tr}(e^{\mathbf{A}\tau}\Sigma_\infty e^{\mathbf{A}^H\tau}\Sigma_\infty^{-1}). \quad (33)$$

The APT is obtained by integrating this expression over all lead times  $\tau$ . Adopting the convention that repeated indices are summed, we have

$$\begin{aligned} S &= \frac{2}{K} \int_0^\infty \text{tr}(e^{\mathbf{A}\tau}\Sigma_\infty e^{\mathbf{A}^H\tau}\Sigma_\infty^{-1}) d\tau \\ &= \frac{2}{K} \int_0^\infty Z_{ij} e^{\lambda_j(\mathbf{A})\tau} (\mathbf{Z}^{-1}\Sigma_\infty\mathbf{Z}^{-H})_{jk} e^{\lambda_k^*(\mathbf{A})\tau} (\mathbf{Z}^H\Sigma_\infty^{-1})_{ki} d\tau \\ &= \frac{2}{K} (\mathbf{Z}^{-1}\Sigma_\infty\mathbf{Z}^{-H})_{jk} (\mathbf{Z}^H\Sigma_\infty^{-1}\mathbf{Z})_{kj} \int_0^\infty e^{[\lambda_j(\mathbf{A}) + \lambda_k^*(\mathbf{A})]\tau} d\tau \\ &= \frac{2}{K} (\mathbf{Z}^{-1}\Sigma_\infty\mathbf{Z}^{-H})_{jk} (\mathbf{Z}^H\Sigma_\infty^{-1}\mathbf{Z})_{kj} \frac{-1}{\lambda_j(\mathbf{A}) + \lambda_k^*(\mathbf{A})} \\ &= \frac{2}{K} \text{tr}[(\mathbf{E}_\infty \circ \mathbf{Z}^{-1}\Sigma_\infty\mathbf{Z}^{-H})(\mathbf{Z}^H\Sigma_\infty^{-1}\mathbf{Z})]. \end{aligned} \quad (34)$$

Substituting the expression in (30) for  $\Sigma_\infty$  into (34) gives

$$S = \frac{2}{K} \text{tr}[(\mathbf{E}_\infty \circ \mathbf{E}_\infty \circ \mathbf{Z}^{-1}\mathbf{Q}\mathbf{Z}^{-H})(\mathbf{E}_\infty \circ \mathbf{Z}^{-1}\mathbf{Q}\mathbf{Z}^{-H})^{-1}]. \quad (35)$$

The last expression relates APT directly to known parameters of the stochastic model.

An alternative, and particularly revealing, expression for predictability can be obtained by introducing the matrix square root  $\Sigma_\infty^{1/2}$  of the climatological covariance, which satisfies

$$\Sigma_\infty = \Sigma_\infty^{1/2}\Sigma_\infty^{H/2}, \quad (36)$$

where  $H/2$  denotes the Hermitian of the square root matrix. A square root matrix always exists for positive definite  $\Sigma_\infty$ , although it is unique only up to a unitary matrix multiplied on the right-hand side. Substituting (36) into the predictability measure (33) and invoking standard properties of the trace, determinant, and exponential operators yields

$$S = \frac{2}{K} \int_0^\infty \text{tr}(e^{\tilde{\mathbf{A}}\tau} e^{\tilde{\mathbf{A}}^H\tau}) d\tau, \quad \text{where} \quad (37)$$

$$\tilde{\mathbf{A}} = \Sigma_\infty^{-1/2}\mathbf{A}\Sigma_\infty^{H/2}. \quad (38)$$

Expression (37) is the multivariate generalizations of (11). Here,  $\tilde{\mathbf{A}}$  is called the whitened dynamical operator and is the dynamical operator governing the transformed variable

$$\tilde{\mathbf{x}} = \Sigma_\infty^{-1/2}\mathbf{x}. \quad (39)$$

Transformation (39), called the whitening transformation, appears frequently in information theory, predictability theory, and pattern recognition theory (Fukunaga 1990; Schneider and Griffies 1999; Majda et al. 2002; Tippett and Chang 2003). See DelSole and Tippett (2007, 2008) for a review of this transformation and its importance to predictability theory.

### 7. Normal case

We now consider linear stochastic models of the form (28) with diagonal dynamical operator  $\mathbf{A}$  having diagonal elements  $\lambda_k(\mathbf{A})$  and diagonal noise covariance matrix  $\mathbf{Q}$  having diagonal elements  $Q_{kk}$ . This case corresponds to a set of uncoupled, damped oscillators that are excited independently. In this case, the eigenvector matrix  $\mathbf{Z}$  is the identity matrix, and the forecast covariance matrix computed from (30) is diagonal, with the  $k$ th diagonal element being

$$(\Sigma_\tau)_{kk} = \frac{e^{2\text{Re}[\lambda_k(\mathbf{A})]\tau} - 1}{2\text{Re}[\lambda_k(\mathbf{A})]} Q_{kk}. \tag{40}$$

We assume that the noise variances  $Q_{kk}$  are strictly positive so that all the eigenmodes are excited and that the climatological is invertible. Substituting this expression into the predictability measure (1) gives

$$S_\tau = \frac{1}{K} \sum_{k=1}^K e^{2\text{Re}[\lambda_k(\mathbf{A})]\tau}. \tag{41}$$

Integrating this expression over all lags gives

$$S = -\frac{1}{K} \sum_{k=1}^K \frac{1}{\text{Re}[\lambda_k(\mathbf{A})]}, \tag{42}$$

which is the mean damping time of the eigenmodes. Comparing this expression with its one-dimensional counterpart (11) shows that the APT of a set of independently excited, uncoupled oscillators equals the average APT of the individual oscillators.

Because APT is invariant to linear transformations, the above results hold for a wider class of models than merely those with a diagonal dynamical operator and noise covariance matrix. In fact, if the dynamical operator has the eigenvector decomposition (29), then the above result holds if the noise covariance matrix is of the form

$$\mathbf{Q} = \mathbf{Z}\mathbf{D}\mathbf{Z}^H, \tag{43}$$

where  $\mathbf{D}$  is a diagonal matrix with real positive entries. Thus, a more general result is that if there exists a basis set in which both the dynamical operator and noise covariance matrix are diagonal, then the APT of the system is given by (42). Intuitively, if both the dynamical operator and noise covariance matrix can be rendered diagonal by a suitable linear coordinate transformation, then the dynamical system is fundamentally equivalent to a set of independently excited, uncoupled oscillators. It is interesting to note that this condition is precisely

the condition for the stochastic model to satisfy the principle of detailed balance (Weiss 2003).

The condition (43), where  $\mathbf{Z}$  is the eigenvector matrix of  $\mathbf{A}$ , can be expressed in a more revealing way. Substituting (43) into the forecast covariance matrix (30) gives

$$\Sigma_\tau = \mathbf{Z}(\mathbf{D} \circ \mathbf{E}_\tau)\mathbf{Z}^H. \tag{44}$$

Let us now compute the whitened dynamical operator defined in (38). Because the square root matrix is unique up to a unitary matrix, the most general square root matrix is

$$\Sigma_\tau^{1/2} = \mathbf{Z}(\mathbf{D} \circ \mathbf{E}_\tau)^{1/2}\mathbf{U}^H, \tag{45}$$

where  $\mathbf{U}$  is an arbitrary unitary matrix (this square root is well defined because  $\mathbf{D} \circ \mathbf{E}_\tau$  is a real, positive definite, diagonal matrix). Substituting this expression into the definition of the whitened dynamical operator (38) gives

$$\tilde{\mathbf{A}} = \mathbf{U}(\mathbf{D} \circ \mathbf{E}_\tau)^{-1/2}\mathbf{Z}^{-1}\mathbf{A}\mathbf{Z}(\mathbf{D} \circ \mathbf{E}_\tau)^{1/2}\mathbf{U}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \tag{46}$$

where we have used  $\mathbf{Z}^{-1}\mathbf{A}\mathbf{Z} = \mathbf{\Lambda}$  and the fact that diagonal matrices commute. This result shows that if the noise covariance matrix  $\mathbf{Q}$  is of the form (43), then  $\tilde{\mathbf{A}}$  is normal. We conclude then that if the whitened dynamical operator  $\tilde{\mathbf{A}}$  is normal, then the system is fundamentally a set of independent, uncoupled oscillators and the APT is given by (42), which depends only on the real part of the eigenvalues. It is worth noting that although the question of whether  $\mathbf{A}$  is normal depends on the coordinate system, the normality of  $\tilde{\mathbf{A}}$  is invariant under nonsingular linear coordinate transformations.

### 8. Relation between the power spectrum and predictability

We now derive the relation between APT and power spectrum for multivariate systems, thereby generalizing the results of section 5. We assume the time series is stationary and band limited, in which case the power spectra are limited to finite frequencies and only discrete time steps are needed. According to the Wiener-Khinchin theorem, the time-lagged covariance matrix

$$\mathbf{C}_\tau = \langle x(t + \tau)\mathbf{x}^H(t) \rangle \tag{47}$$

is related to the power spectrum matrix  $\mathbf{P}_\omega$  as

$$\mathbf{C}_\tau = \int_{-\pi}^{\pi} \mathbf{P}_\omega e^{i\tau\omega} d\omega, \quad \mathbf{P}_\omega = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \mathbf{C}_\tau e^{-i\tau\omega}. \tag{48}$$

The time-lagged covariance matrix  $\mathbf{C}_\tau$  can be related to the forecast covariance matrix  $\Sigma_\tau$  only after a forecast



model is invoked explicitly. Here we consider linear forecast models of the form

$$\hat{\mathbf{x}}(t + \tau) = \mathbf{L}_\tau \mathbf{x}(t), \quad \tau \geq 0, \tag{49}$$

where  $\hat{\mathbf{x}}$  is a prediction of  $\mathbf{x}$  and  $\mathbf{L}_\tau$  is a linear operator. Determination of the linear operator that minimizes the sum of squared forecast errors is a standard regression problem with solution

$$\mathbf{L}_\tau = \mathbf{C}_\tau \mathbf{C}_0^{-1}, \quad \tau \geq 0. \tag{50}$$

The forecast error covariance matrix for this model is therefore

$$\begin{aligned} \Sigma_\tau &= \langle [\mathbf{x}(t + \tau) - \mathbf{L}_\tau \mathbf{x}(t)][\mathbf{x}(t + \tau) - \mathbf{L}_\tau \mathbf{x}(t)]^H \rangle \\ &= \mathbf{C}_0 - \mathbf{C}_\tau \mathbf{L}_\tau^H - \mathbf{L}_\tau \mathbf{C}_\tau^H + \mathbf{L}_\tau \mathbf{C}_0 \mathbf{L}_\tau^H \\ &= \mathbf{C}_0 - \mathbf{C}_\tau \mathbf{C}_0^{-1} \mathbf{C}_\tau^H. \end{aligned} \tag{51}$$

This expression is the multivariate generalization of (15). Furthermore, this expression is equivalent to (32) with the identification  $\mathbf{L}_\tau = \exp(\mathbf{A}\tau)$ , reflecting the close relation between linear stochastic models and linear regression models.

Having computed the forecast covariance matrix (51) for the linear prediction model, we can now substitute it into the Mahalanobis signal (1), which gives

$$S_\tau = \frac{1}{K} \text{tr}(\mathbf{C}_\tau \mathbf{C}_0^{-1} \mathbf{C}_\tau^H \mathbf{C}_0^{-1}). \tag{52}$$

A subtlety is that the linear prediction model (50) is valid only for positive time lags  $\tau$ . However, for any stationary process,  $\mathbf{C}_{-\tau} = \mathbf{C}_\tau^H$ . Substituting this identity into the above equation implies  $S_{-\tau} = S_\tau$ ; that is, our expression for  $S_\tau$  is an even function of lag. Therefore, we may sum this expression over positive and negative lags. Substituting the Wiener–Khinchin relation for the power spectrum, while noting that  $S_0 = 1$ , gives

$$S + 1 = \frac{1}{K} \sum_{\tau=-\infty}^{\infty} \text{tr}(\mathbf{C}_\tau \mathbf{C}_0^{-1} \mathbf{C}_\tau^H \mathbf{C}_0^{-1}) \tag{53}$$

$$= \frac{1}{K} \sum_{\tau=-\infty}^{\infty} \text{tr} \left[ \left( \int_{-\pi}^{\pi} \mathbf{P}_\omega e^{-i\tau\omega} d\omega \right) \mathbf{C}_0^{-1} \left( \int_{-\pi}^{\pi} \mathbf{P}_{\omega'} e^{-i\tau\omega'} d\omega' \right)^H \mathbf{C}_0^{-1} \right] \tag{54}$$

$$= \frac{1}{K} \text{tr} \left[ \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathbf{P}_\omega \mathbf{C}_0^{-1} \mathbf{P}_{\omega'}^H \mathbf{C}_0^{-1} \left( \sum_{\tau=-\infty}^{\infty} e^{i\tau(\omega' - \omega)} \right) d\omega d\omega' \right]. \tag{55}$$

Using the fact that

$$\sum_{\tau=-\infty}^{\infty} e^{i\tau(\omega' - \omega)} = 2\pi\delta(\omega - \omega') \tag{56}$$

gives

$$S + 1 = \frac{2\pi}{K} \int_{-\pi}^{\pi} \text{tr}(\mathbf{P}_\omega \mathbf{C}_0^{-1} \mathbf{P}_\omega^H \mathbf{C}_0^{-1}) d\omega. \tag{57}$$

Factoring the zero-lag covariance matrix as in (36) and invoking standard properties of the trace operator yields

$$S + 1 = \frac{1}{K} \sum_{\tau=-\infty}^{\infty} \text{tr}(\tilde{\mathbf{C}}_\tau \tilde{\mathbf{C}}_\tau^H) = 2\pi \frac{1}{K} \int_{-\pi}^{\pi} \text{tr}(\tilde{\mathbf{P}}_\omega \tilde{\mathbf{P}}_\omega^H) d\omega, \tag{58}$$

where  $\tilde{\mathbf{C}}_\tau$  and  $\tilde{\mathbf{P}}_\omega$  are the whitened time lag covariance and power spectrum matrices, respectively, defined as

$$\tilde{\mathbf{C}}_\tau = \Sigma_\tau^{-1/2} \mathbf{C}_\tau \Sigma_\tau^{-H/2}, \quad \tilde{\mathbf{P}}_\omega = \Sigma_\omega^{-1/2} \mathbf{P}_\omega \Sigma_\omega^{-H/2}. \tag{59}$$

The relation between APT and power spectrum given in (58) is the multivariate generalization of (24) (but restricted to band-limited spectra).

The restriction to band-limited power spectra allows us to place rigorous bounds on the APT. Specifically, we show in the appendix that

$$2\pi \frac{1}{K} \int_{-\pi}^{\pi} \text{tr}(\tilde{\mathbf{P}}_\omega \tilde{\mathbf{P}}_\omega^H) d\omega \geq 1 \tag{60}$$

with equality if and only if the power spectra is

$$\tilde{\mathbf{P}}_\omega = \frac{1}{2\pi} \mathbf{I}. \tag{61}$$

The later condition implies that minimum predictability occurs when the power spectrum of all variables can be transformed into a set of uncorrelated, white noise

processes, as intuition would predict. We suggest that (58) is a kind of multivariate generalization of spectral peakiness, as defined in section 5, although we emphasize that it is not the only reasonable measure of peakiness. The intuitive result follows that the absolute minimum APT occurs when all processes are independent white noise (i.e., flat). In the next section, we consider the minimum APT subject to the constraint that the dynamical operator is known.

**9. Bounds on APT**

In this section we derive bounds on the APT of linear stochastic models (28). Bounds on predictability at fixed lead time have been derived in Tippett and Chang (2003) and Weiss (2003) and reviewed in DelSole and Tippett (2007). Tippett and Chang (2003) show that a lower bound for the Mahalanobis signal is given by

$$S_\tau \geq \frac{1}{K} \sum_{k=1}^K e^{2\text{Re}[\lambda_k(\mathbf{A})]\tau}. \tag{62}$$

Because the lower bound holds at each lead time individually, a lower bound for APT can be found by integrating both sides over all lead times and multiplying by two, which yields

$$S \geq -\frac{1}{K} \sum_{k=1}^K \frac{1}{\text{Re}[\lambda_k(\mathbf{A})]}. \tag{63}$$

Note that the right-hand side is identical to (42). It follows that the minimum APT occurs when there exists a basis set in which the dynamical operator  $\mathbf{A}$  and noise covariance matrix  $\mathbf{Q}$  are diagonal, which in turn implies that the system is fundamentally a set of uncoupled, damped oscillators that are excited independently. This condition for minimum APT is precisely the condition for minimum predictability at fixed lag, as well as the condition for detailed balance (Tippett and Chang 2003; Weiss 2003). These correspondences are not surprising because predictability is minimized at each individual lead time and hence minimized for the integral over all lead times.

An upper bound on APT can be derived from a majorization property due to Cohen (1988). Specifically, for matrix exponentials, Cohen (1988) proves that

$$\sum_{k=1}^n \log \lambda_k(e^{\mathbf{A}} e^{\mathbf{A}^H}) \leq \sum_{k=1}^n \log \lambda_k(e^{\mathbf{A} + \mathbf{A}^H}), \tag{64}$$

where  $n = 1, 2, \dots, K$ . As is well known, any increasing, convex function preserves majorization (Horn and

Johnson 1999, p. 173). Because  $\exp(2x)$  is an increasing, convex function, it follows that an upper bound on Mahalanobis signal is

$$S_\tau \leq \frac{1}{K} \sum_{k=1}^K e^{2\lambda_k(\mathbf{A}_s)\tau}, \tag{65}$$

where  $\mathbf{A}_s = (\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^H)/2$  is the symmetric part of  $\tilde{\mathbf{A}}$ . These inequalities become equalities if and only if  $\tilde{\mathbf{A}}$  is normal, in which case these upper bounds equal the lower bound (62). Because this upper bound holds at each lead time individually, the upper bound for APT can be found by integrating both sides over all lead times and multiplying by two, which yields

$$S \leq -\frac{1}{K} \sum_{k=1}^K \frac{1}{\lambda_k(\mathbf{A}_s)}. \tag{66}$$

We see that the upper bound (66) and lower bound (63) are proportional to a sum of inverse eigenvalues, but the respective eigenvalues are based on different operators. The operator  $\mathbf{A}_s$  appearing in the upper bound (66) arises in a variety of contexts related to the instantaneous rate of change of variables. For instance, the eigenvectors of  $\mathbf{A}_s$  are called instantaneous optimals and define an orthogonal set of initial states that optimize the instantaneous rate of change of  $\tilde{\mathbf{x}}^H \tilde{\mathbf{x}}$  (see DelSole 2004 for review). In the present situation, however, the initial condition plays no role in the predictability because all unpredictability arises from stochastic forcing (i.e., the initial condition is assumed to be known perfectly). The relevance of the operator  $\tilde{\mathbf{A}}$  in the present situation can be seen as follows: It is straightforward to show from (30) that  $d\tilde{\Sigma}_\tau/d\tau = \tilde{\mathbf{Q}}$  at  $\tau = 0$ . Furthermore, the Lyapunov equation  $\mathbf{A}\Sigma_\infty + \Sigma_\infty\mathbf{A}^H + \mathbf{Q} = 0$  implies  $\tilde{\mathbf{Q}} = -(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^H)$ , where we have used the fact that  $\tilde{\Sigma}_\infty = \mathbf{I}$ , by definition. Combining these results implies that

$$\left(\frac{d\tilde{\Sigma}_\tau}{d\tau}\right)_0 = -(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^H). \tag{67}$$

Thus,  $2\mathbf{A}_s$  is the rate of change of the whitened forecast covariance matrix at  $\tau = 0$ . Interestingly, we would obtain the same rate of change if the original dynamical operator had been  $\mathbf{A}_s$ . It follows that the upper bound (66) can be derived by replacing the original dynamical operator by the normal dynamical operator  $\mathbf{A}_s$ , which gives precisely the same rate of predictability loss at the initial time as the original system, then integrating the predictability of the system over all times.

The upper bound (66) can be written explicitly in terms of model parameters using the fact that

$$\begin{aligned}
 -\frac{1}{\lambda_k(\mathbf{A}_s)} &= -\lambda_k(\mathbf{A}_s^{-1}) = -2\lambda_k[(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^H)^{-1}] \\
 &= 2\lambda_k(\tilde{\mathbf{Q}}^{-1}) = 2\lambda_k(\tilde{\mathbf{Q}}^{-1}\Sigma_\infty) \\
 &= 2\lambda_k[\mathbf{Q}^{-1}\mathbf{Z}(\mathbf{Z}^{-1}\mathbf{Q}\mathbf{Z}^{-H} \circ \mathbf{E}_\infty)\mathbf{Z}^H] \\
 &= 2\lambda_k[(\mathbf{Z}^{-1}\mathbf{Q}\mathbf{Z}^{-H} \circ \mathbf{E}_\infty)(\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Z}^{-H})^{-1}], \quad (68)
 \end{aligned}$$

where we have used (30) and the fact that  $\lambda_k(\mathbf{AB}) = \lambda_k(\mathbf{BA})$  for any two matrices  $\mathbf{A}$  and  $\mathbf{B}$ . Using this identity, the upper bound (66) becomes

$$S \leq -\frac{2}{K} \text{tr}[(\mathbf{E}_\infty \circ \mathbf{Z}^{-1}\mathbf{Q}\mathbf{Z}^{-H})(\mathbf{Z}^{-1}\mathbf{Q}\mathbf{Z}^{-H})^{-1}]. \quad (69)$$

If the noise covariance matrix is of the form (43), then the above upper bound becomes identical to the lower bound (62). It is also apparent that if  $\mathbf{Q}$  is singular, then the upper bound given above is not defined, consistent with the fact that the original upper bound (66) diverges to infinity in this case. Thus, this upper bound is useful only when  $\mathbf{Q}$  is full rank. On the other hand, this upper bound requires knowing the same number of parameters as the full solution, whereas the other bounds require fewer parameters. Nevertheless, the above upper bound is conceptually useful because it relates the APT of linear stochastic models to the predictability at short times.

An alternative upper bound on predictability can be derived from the conjecture of Tippett and Chang (2003) that maximum predictability occurs when the noise covariance matrix is rank 1 and excites all eigenmodes. As Tippett and Chang (2003) show, this conjecture gives the upper bound

$$S_\tau \leq \frac{1}{K} \text{tr}[(\mathbf{E}_\infty - \mathbf{E}_\tau)\mathbf{E}_\infty^{-1}]. \quad (70)$$

Remarkably, the response to rank-1 forcing is independent of detailed structure of the forcing. This independence can be explained as follows: A rank-1 noise covariance matrix implies that each eigenmode is excited in sync with other eigenmodes. The response thus depends only on the amplitude with which each mode is excited, but any set of noise amplitudes can be transformed into any other set by multiplying each eigenmode equation by an appropriate factor, which corresponds to a linear transformation that does not affect predictability. Standard calculus gives

$$\int_0^\infty (\mathbf{E}_\infty - \mathbf{E}_\tau)_{kl} d\tau = \left[ \frac{1}{\lambda_k(\mathbf{A}) + \lambda_l^*(\mathbf{A})} \right]^2 = (\mathbf{E}_\infty \circ \mathbf{E}_\infty)_{kl}, \quad (71)$$

where the definition (31) has been used. Therefore, integrating (70) and multiplying by  $2/K$  gives the conjectural upper bound

$$S = \frac{2}{K} \int_0^\infty S_\tau d\tau \leq \frac{2}{K} \text{tr}[(\mathbf{E}_\infty \circ \mathbf{E}_\infty)\mathbf{E}_\infty^{-1}]. \quad (72)$$

An advantage of this conjectured upper bound is that it is an explicit solution, thus ensuring that it is actually achieved by a specific stochastic model. In contrast, the upper bound (66) is achieved only when the whitened dynamical operator is normal, but then the upper bound equals the lower bound, which is not an interesting limit. Another distinction is that the upper bound (72) depends only on the eigenvalues of the dynamical operator, in contrast to the upper bound (66). Finally, it should be recognized that this upper bound is conjectural. These bounds will be illustrated with a simple example in section 11.

### 10. APT as a solution to an alternative stochastic model

A useful concept in the application of information theory predictability measures to normally distributed variables is that techniques for analysis of predictability are often equivalent to techniques for analysis of variance to applied whitened variables (Schneider and Griffies 1999; DelSole and Tippett 2007). Here we show how APT fits in this framework. Generalized stability analysis relates the variance of linear dynamics forced by homogeneous isotropic stochastic forcing to the stability properties of the linear dynamics (Farrell and Ioannou 1996; Tippett and Marchesin 1999). Here we show that the variance of *whitened* linear dynamics excited by homogeneous isotropic stochastic forcing is related to the predictability of the system as measured by APT.

The asymptotic solution to the stochastic model (28) can be written in the alternative form (Lancaster and Tismensetsky 1985)

$$\Sigma_\infty = \int_0^\infty e^{\mathbf{A}s}\mathbf{Q}e^{\mathbf{A}^Hs} ds. \quad (73)$$

This expression can be transformed into the form of (37) by making the substitutions  $\mathbf{A} \rightarrow \tilde{\mathbf{A}}$  and  $\mathbf{Q} \rightarrow \mathbf{I}$  in (73), then taking the trace and multiplying by  $2/K$ . This equivalence implies that the APT can be interpreted as the “total variance” of a stochastic model with dynamical operator  $\tilde{\mathbf{A}}$  and noise covariance matrix  $\mathbf{I}$ . This result bears repeating: APT can be expressed as

$$S = \frac{2}{K} \text{tr}(\Sigma_y), \quad (74)$$

where  $\Sigma_y$  is the stationary covariance matrix of  $\mathbf{y}$  obtained from the stochastic model

$$\frac{d\mathbf{y}}{dt} = \tilde{\mathbf{A}}\mathbf{y} + \mathbf{v}, \tag{75}$$

where  $\mathbf{v}$  is a Gaussian white noise process with zero mean and covariance matrix  $\mathbf{I}$ . The fact that APT equals the total variance of a stochastic model implies that the intuition developed about total variance produced by stochastic models can be applied directly to APT. As an example, a system with a nonnormal dynamical operator generates more variance than a system with a normal dynamical operator with the same eigenvalues (Ioannou 1995). Applying this theorem to the stochastic model (75) immediately implies that nonnormality of  $\tilde{\mathbf{A}}$  enhances APT, consistent with the result derived in section 9. Another consequence of the above equivalence is that results from dynamical stability theory can be used to understand APT. For instance, it is known that the covariance matrix  $\Sigma_y$  derived from the stochastic model (75) satisfies the Lyapunov equation

$$\tilde{\mathbf{A}}\Sigma_y + \Sigma_y\tilde{\mathbf{A}}^H + \mathbf{I} = 0. \tag{76}$$

Furthermore, there exists a wide variety of bounds on the solutions to Lyapunov equations (Kwon et al. 1996). For instance, bound (106) from Kwon et al. (1996) implies the upper bound (66).

### 11. Example

We illustrate some properties of APT with a simple stochastic model; a more detailed example based on meteorological data will be given in Part II. Consider a two-dimensional linear stochastic model of the form (28) with noise covariance matrix  $\mathbf{Q} = \mathbf{I}$  and dynamical operator

$$\mathbf{A} = \begin{bmatrix} -1/3 & c \\ 0 & -1 \end{bmatrix}, \tag{77}$$

where  $c$  is a parameter measuring the coupling between the two components. Because  $\mathbf{A}$  is upper triangular, its eigenvalues are simply the diagonal elements  $-1/3$  and  $-1$ . The negative eigenvalues imply that the dynamical operator is stable and that the stochastic model gives statistically stationary solutions. The predictability of the system is minimized and achieves the lower bound (63) when there is no coupling between the two components of the system (i.e., when  $c = 0$ ). The time-dependent predictability for the case  $c = 1$  and  $c = 4$  are shown in the top panels of Fig. 3. Also shown are the upper and lower bounds. As expected, as the coupling parameter increases, so too does the predictability, at all

lead times. For  $c = 1$  the APT is  $S = 2.54$ ; for  $c = 4$  the APT is  $S = 3.35$ . For comparison, the conjectured upper bound (72) is 3.5, whereas the lower bound is 2. In general, predictability is an increasing function of coupling parameter  $c$ . The two upper bounds provide tight constraints in different extremes—the bound based on instantaneous optimals performs well at small lead times, whereas the conjectured bound performs well at long lead times. The lower panels show the average predictability time as a function of coupling parameter  $c$ . The upper bound based on instantaneous optimals greatly overestimates the APT at large values of the coupling parameter. A good upper bound is the minimum of the two upper bounds.

### 12. Summary and discussion

This paper introduced average predictability time (APT) for characterizing the overall predictability of a system. The APT is defined as the integral of the Mahalanobis signal with respect to lead time. As such, APT measures an inherent property of a system that is independent of the basis set used to represent the system. The appropriateness of APT was illustrated with a one-dimensional stochastic model in which the APT depends inversely on the damping rate, consistent with the intuition that systems with stronger damping have less memory and hence are less predictable. Furthermore, the APT is comparable to integral time scales in appropriate cases. However, if the autocorrelation is a damped oscillation, then the integral time scale defined in (21) becomes arbitrarily small for large frequencies, which is unrealistic, whereas APT is always within 50% of the  $e$ -folding time. Thus, the APT provides a more suitable measure of time scale than the integral time scale (21). For nonlinear or non-Gaussian processes, the Mahalanobis signal may decay very differently from that of linear stochastic systems (e.g., it may exhibit long tails). Nevertheless, just as the correlation coefficient has meaning even if the variables are not linearly related, the APT still has meaning even if it cannot be interpreted directly as a time scale.

The APT also clarifies the connection between predictability and power spectra. Specifically, if a process is stationary, then the APT of a linear regression forecast is proportional to the integral of the square of the normalized power spectrum. The appropriateness of this relation can be seen from the fact that the latter quantity can be decreased simply by replacing the power in any spectral band by its mean value within the band; that is, the quantity decreases as the spectrum becomes flatter. If the time series is band-limited, then the minimum value occurs when the system comprises a set of

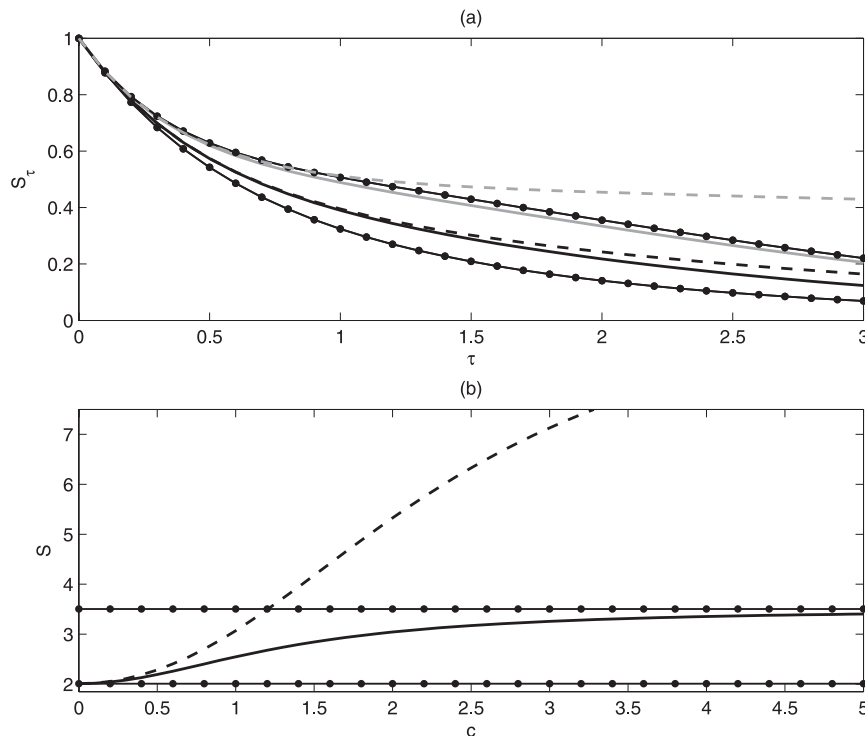


FIG. 3. (a) Mahalanobis signal (solid) and upper bound (66) (dashed) as a function of lead time  $\tau$  for  $c = 1$  (dark) and  $c = 4$  (light) in the  $2 \times 2$  example. The lower bound (63) (lower filled circles) and upper bound conjecture (72) (upper filled circles) are independent of  $c$ . (b) The integrated Mahalanobis signal (solid), lower bound (62) (lower filled circles), upper bound conjecture (72) (upper filled circles) and upper bound (65) (dashed) as a function of the coupling parameter  $c$ .

independent and identically distributed white noise processes, which is obviously the least predictable system. In essence, predictability is related to the width of spectral peaks, with strong narrow peaks associated with high predictability and nearly flat spectra associated with low predictability. As extreme examples, white noise has a constant power spectrum and is minimally predictable, whereas a sine wave has a delta-function power spectrum and is perfectly predictable. Expressing the APT in terms of the power spectra rigorously quantifies this intuitive relation.

Closed form expressions for the APT of linear, multivariate stochastic models were derived. If the dynamical operator and noise covariance matrix are both diagonal, then—remarkably—the APT depends only on the damping rates and equals the average APT of the individual eigenmodes. Because the APT is invariant to linear transformation, this result is true for any system for which there exists a basis set in which both the dynamical operator and noise covariance matrix are diagonal. We further show that this particular system minimizes the APT compared with all linear stochastic models with the same dynamical operator. Thus, the least

predictable system can be transformed into a set of uncoupled, independent stochastic models. It follows that systems that are irreducibly coupled have more predictability than those that are fundamentally uncoupled. Simply put, coupling enhances predictability. APT rigorously justifies this intuitive notion.

As reviewed in DelSole and Tippett (2007), applying the whitening transformation to a system allows familiar concepts from analysis of variance and generalized stability analysis to be applied directly to predictability analysis. We show that this equivalence carries over to APT because, surprisingly, APT itself can be interpreted as the total variance of an alternative stochastic model. The alternative stochastic model is driven by homogeneous white noise and has a dynamical operator equal to the whitened dynamical operator of the original stochastic model. This connection allows one to anticipate that the lower bound for APT occurs when the whitened dynamical operator is normal, which in turn is equivalent to the condition that there exists a basis set in which both the dynamical operator and noise covariance matrix are diagonal—consistent with the lower bound of APT discussed above.



The remarkable equivalence noted above is worth further reflection. Loosely speaking, one way to understand the stability of a dynamical system is to examine its response to white noise forcing. The more variance produced, the less stable the system. This approach is the basis of generalized stability theory, which in turn utilizes concepts from dynamical stability theory. In defining APT, we proposed something apparently different: we proposed that the overall predictability of a system can be quantified by the integral of predictability with respect to lead time, reasoning that more predictable systems have larger integrals. Surprisingly, the integral of predictability is itself the total variance of the system consisting of the whitened dynamical operator forced by white noise. APT measures the stability of the whitened dynamics. Thus, the two methods for understanding predictability turn out to be fundamentally equivalent. An important benefit in this equivalence, however, is that the whitened dynamical operator is unique up to a variance preserving unitary transformation. Thus, the framework proposed in this paper imposes a particular stochastic model for studying predictability, which is essential for applying generalized stability theory. In contrast, the latter theory does not impose the norm or coordinate system; rather, it provides a framework for understanding variance after the norm and coordinate system have been chosen.

An upper bound on the APT of linear stochastic models also was derived. The upper bound is proportional to the sum of inverse eigenvalues of the symmetric part of the whitened dynamical operator. The latter operator arises frequently in the investigation of instantaneous growth rates of variance. It is interesting that the upper bound on APT (or the predictability at any lead time) is related to the predictability at very short time scales in linear stochastic models. The upper bounds imply that as the short time predictability decreases, so too does the maximum possible APT. This intuitive notion is used widely in predictability studies as a justification for drawing conclusions about overall predictability from the characteristics of error growth at short times. Two key points about this relation are worth emphasizing. First, although intuitive, we are aware of no rigorous demonstration of a connection between predictability at short times and overall predictability. Thus, our result provides such a demonstration, at least for linear stochastic models. Second, the connection between short time error growth and APT is most direct when the error growth is measured in whitened space. In other words, our result clarifies the existence of a preferred norm for relating predictability at short and long lead times. Attempts to draw conclusions about long-term predictability from short-term error growth

in other norms may actually be misleading (see DelSole and Tippett 2008).

A conjecture for an upper bound for APT follows from the conjecture of Tippett and Chang (2003), which states that out of all linear stochastic models with the same dynamical eigenvalues, the model with rank-1 noise covariance matrix (with nonzero diagonal elements) has maximum predictability. This conjecture can be motivated by the fact that if minimum predictability occurs when all eigenmodes are forced independently, then perhaps maximum predictability occurs when all eigenmodes are forced in perfect correlation. The resulting upper bound depends only on the eigenmode damping rates; in particular, the upper bound is independent of the detailed structure of the forcing, provided all modes are excited.

It should be recognized that APT can be applied to more general forecast models than linear stochastic models. For instance, APT can be evaluated for regression models with physically different variables for predictors and predictands. Interestingly, if only a single variable is predicted, but more than one predictor is used, then the APT relation (17) still is applicable, but the parameter  $\rho_\tau$  denotes the *multiple correlation*. This result is discussed further in Part II of this paper.

We note that other measures of predictability can be used to define APT, and these alternative measures may be attractive in some cases. Mutual information in particular is appropriate for non-Gaussian or nonlinear systems. Furthermore, mutual information is integrable for linear stochastic systems, even though it is unbounded as lead time approaches zero for perfect initial conditions. For linear stochastic models with Gaussian white noise, the APT derived from mutual information turns out to be proportional to the APT (42) for normal whitened dynamical operators. However, the relation between this alternative form of APT and power spectra is obscure, and the decomposition of this form of APT is not straightforward.

In climate systems, the APT of different components can vary greatly owing to the widely different time scales associated with land, ice, and ocean processes. Thus, characterizing the predictability of the full climate system by a single number might seem to be a gross simplification. However, the APT can be decomposed into uncorrelated components that can be ordered by their fractional contribution to APT, such that the first component maximizes APT, the second maximizes APT subject to being uncorrelated with the first, and so on. This decomposition therefore allows the full spectrum of predictability times to be diagnosed. This decomposition can be used to study predictability on different time scales without time filtering, provided the predictabilities on

different time scales are characterized by different spatial structures. This decomposition and its practical implementation are discussed in Part II.

*Acknowledgments.* We thank three anonymous reviewers for detailed comments that lead to an improved manuscript. This research was supported by the National Science Foundation (ATM0332910), the National Aeronautics and Space Administration (NNG04GG46G), and the National Oceanic and Atmospheric Administration (NA04OAR4310034). MKT is supported by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration (NA05OAR4311004). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies.

APPENDIX

A Cauchy–Schwartz Inequality

The Cauchy–Schwartz inequality states that for any two vectors  $\mathbf{x}$  and  $\mathbf{y}$

$$\|\mathbf{x}^H \mathbf{y}\|^2 \leq (\mathbf{x}^H \mathbf{x}) (\mathbf{y}^H \mathbf{y}) \tag{A1}$$

with equality if and only if  $\mathbf{x} = \alpha \mathbf{y}$  for some scalar  $\alpha$  (we use the notation  $\|\mathbf{x}\|^2 = \mathbf{x}^H \mathbf{x}$ ). The Cauchy–Schwartz inequality can be written equivalently as

$$\left\| \sum_{ij\omega} X_{ij\omega} Y_{ij\omega}^* \right\|^2 \leq \left( \sum_{ij\omega} \|X_{ij\omega}\|^2 \right) \left( \sum_{ij\omega} \|Y_{ij\omega}\|^2 \right), \tag{A2}$$

where each vector is a function of three indices. Now define

$$Y_{ij\omega} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \tag{A3}$$

Substituting this particular function into the Cauchy–Schwartz inequality gives

$$\left\| \sum_{i\omega} X_{i\omega} \right\|^2 \leq \left( \sum_{ij\omega} \|X_{ij\omega}\|^2 \right) \left( \sum_{i\omega} 1 \right). \tag{A4}$$

We now identify  $X$  with the whitened power spectrum matrix

$$X_{ij\omega} = \tilde{P}_{ij\omega}. \tag{A5}$$

Furthermore, let the  $\omega$  index represent equally spaced frequencies between  $-\pi$  and  $\pi$ , where the number of frequencies increases indefinitely. To the extent that the resulting sum can be interpreted as a Riemann integral, the Cauchy–Schwartz inequality becomes

$$\left\| \sum_i \int_{-\pi}^{\pi} \tilde{P}_{ii\omega} d\omega \right\|^2 \leq \left( \int_{-\pi}^{\pi} \sum_{ij} \|\tilde{P}_{ij\omega}\|^2 d\omega \right) \left( \sum_i \int_{-\pi}^{\pi} d\omega \right). \tag{A6}$$

Because whitened variables have unit variance, the integral on the left-hand side equals unity. Therefore, the above inequality becomes

$$K^2 \leq \left( \int_{-\pi}^{\pi} \sum_{ij} \|\tilde{P}_{ij\omega}\|^2 d\omega \right) (2\pi K), \tag{A7}$$

which gives (60). The above inequality becomes equality when  $\tilde{P}_{ij\omega}$  is proportional to (A3), which corresponds to (61).

REFERENCES

Cohen, J. E., 1988: Spectral inequalities for matrix exponentials. *Linear Algebra Appl.*, **111**, 25–28.

DelSole, T., 2001: Optimally persistent patterns in time-varying fields. *J. Atmos. Sci.*, **58**, 1341–1356.

—, 2004: The necessity of instantaneous optimals in stationary turbulence. *J. Atmos. Sci.*, **61**, 1086–1091.

—, and M. K. Tippett, 2007: Predictability: Recent insights from information theory. *Rev. Geophys.*, **45**, RG4002, doi:10.1029/2006RG000202.

—, and —, 2008: Predictable components and singular vectors. *J. Atmos. Sci.*, **65**, 1666–1678.

—, and —, 2009: Average predictability time. Part II: Seamless diagnoses of predictability on multiple time scales. *J. Atmos. Sci.*, **66**, 1188–1204.

Farrell, B. F., and P. J. Ioannou, 1996: Generalized stability theory. Part I: Autonomous operators. *J. Atmos. Sci.*, **53**, 2025–2040.

Fukunaga, K., 1990: *Introduction to Statistical Pattern Recognition*. 2nd ed. Academic Press, 591 pp.

Gelb, A., Ed., 1974: *Applied Optimal Estimation*. MIT Press, 382 pp.

Horn, R. A., and C. R. Johnson, 1999: *Topics in Matrix Analysis*. Cambridge University Press, 607 pp.

Ioannou, P. J., 1995: Nonnormality increases variance. *J. Atmos. Sci.*, **52**, 1155–1158.

Julian, P. R., and R. M. Chervin, 1978: A study of the Southern Oscillation and Walker Circulation phenomenon. *Mon. Wea. Rev.*, **106**, 1433–1451.

Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, **59**, 2057–2072.

Kwon, W. H., Y. S. Moon, and S. C. Ahn, 1996: Bounds in algebraic Riccati and Lyapunov equations: A survey and some new results. *Int. J. Control*, **64**, 377–389.

Lancaster, P., and M. Tismenetsky, 1985: *The Theory of Matrices: With Applications*. Academic Press, 570 pp.

- Leung, L.-Y., and G. R. North, 1990: Information theory and climate prediction. *J. Climate*, **3**, 5–14.
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.
- Majda, A., R. Kleeman, and D. Cai, 2002: A mathematical framework for quantifying predictability through relative entropy. *Methods Appl. Anal.*, **9**, 425–444.
- Priestley, M. B., 1981: *Spectral Analysis and Time Series*. Academic Press, 890 pp.
- Schneider, T., and S. Griffies, 1999: A conceptual framework for predictability studies. *J. Climate*, **12**, 3133–3155.
- Tippett, M. K., and D. Marchesin, 1999: Upper bounds for the solution of the discrete algebraic Lyapunov equation. *Automatica*, **35**, 1485–1489.
- , and P. Chang, 2003: Some theoretical considerations on predictability of linear stochastic dynamics. *Tellus*, **55**, 148–157.
- Weiss, J. B., 2003: Coordinate invariance in stochastic dynamical systems. *Tellus*, **55A**, 208–218.